



Computer Science and Information Systems

Published by ComSIS Consortium

Volume 8, Number 1
January 2011

ComSIS Consortium:

University of Belgrade:

Faculty of Organizational Science, Belgrade, Serbia
Faculty of Mathematics, Belgrade, Serbia
School of Electrical Engineering, Belgrade, Serbia

Serbian Academy of Science and Art:

Mathematical Institute, Belgrade, Serbia

Union University:

School of Computing, Belgrade, Serbia

University of Novi Sad:

Faculty of Sciences, Novi Sad, Serbia
Faculty of Technical Sciences, Novi Sad, Serbia
Faculty of Economics, Subotica, Serbia
Technical Faculty "M. Pupin", Zrenjanin, Serbia

University of Montenegro:

Faculty of Economics, Podgorica, Montenegro
Faculty of Mathematics and Science, Podgorica, Montenegro

EDITORIAL BOARD:

Editor-in-Chief: Mirjana Ivanović, University of Novi Sad

Vice Editor-in-Chief: Ivan Luković, University of Novi Sad

Managing editor: Zoran Putnik, University of Novi Sad

Managing editor: Miloš Radovanović, University of Novi Sad

Editorial Assistant: Vladimir Kurbalija, University of Novi Sad

Editorial Assistant: Jovana Vidaković, University of Novi Sad

Editorial Assistant: Ivan Pribela, University of Novi Sad

Editorial Assistant: Slavica Aleksić, University of Novi Sad

Editorial Assistant: Srđan Škrbić, University of Novi Sad

Associate editors:

P. Andrae, Victoria University, New Zealand
D. Bojić, University of Belgrade, Serbia
D. Bečejski-Vujaklija, University of Belgrade, Serbia
I. Berković, University of Novi Sad, Serbia
G. Bhutkar, Vishwakarma Institute of Technology, India
L. Böszörményi, University of Clagenfurt, Austria
K. Bothe, Humboldt University of Berlin, Germany
Z. Budimac, University of Novi Sad, Serbia
G. Devedžić, University of Kragujevac, Serbia
V. Devedžić, University of Belgrade, Serbia
J. Đurković, University of Novi Sad, Serbia
S. Guttormsen Schar, ETH Zentrum, Switzerland
P. Hansen, University of Montreal, Canada
L. C. Jain, University of South Australia, Australia
V. Jovanović, Georgia Southern University, USA
Lj. Kaščelan, University of Montenegro, Montenegro

P. Mahanti, University of New Brunswick, Canada
M. Maleković, University of Zagreb, Croatia
N. Mitić, University of Belgrade, Serbia
A. Mitrović, University of Canterbury, New Zealand
N. Mladenović, Serbian Academy of Science, Serbia
P. Mogin, Victoria University, New Zealand
S. Mrdalj, Eastern Michigan University, USA
G. Nenadić, University of Manchester, UK
Z. Ognjanović, Serbian Academy of Science, Serbia
A. Pakstas, London Metropolitan University, England
P. Pardalos, University of Florida, USA
M. Racković, University of Novi Sad, Serbia
B. Radulović, University of Novi Sad, Serbia
M. Stanković, University of Niš, Serbia
D. Tošić, University of Belgrade, Serbia
J. Trninić, University of Novi Sad, Serbia
J. Woodcock, University of Kent, England

EDITORIAL COUNCIL:

S. Ambroszkiewicz, Polish Academy of Science, Poland
Z. Arsovski, University of Kragujevac, Serbia
D. Banković, University of Kragujevac, Serbia
T. Bell, University of Canterbury, New Zealand
S. Bošnjak, University of Novi Sad, Serbia
Ž. Branović, University of Novi Sad, Serbia
H. D. Burkhard, Humboldt University of Berlin, Germany
B. Chandrasekaran, Ohio State University, USA
V. Ćirić, University of Belgrade, Serbia
D. Domazet, Faculty of Information Technology, Belgrade, Serbia
S. Đrođević-Kajan, University of Niš, Serbia
P. Hotomski, University of Novi Sad, Serbia
P. Janičić, University of Belgrade, Serbia
Z. Jovanović, University of Belgrade, Serbia
L. Kalinichenko, Russian Academy of Science, Russia
Z. Konjović, University of Novi Sad, Serbia

I. Koskosas, University of Western Macedonia, Greece
W. Lamersdorf, University of Hamburg, Germany
T. C. Lethbridge, University of Ottawa, Canada
A. Lojpur, University of Montenegro, Montenegro
Y. Manolopoulos, Aristotle University, Greece
A. Mishra, Atılım University, Turkey
S. Mishra, Atılım University, Turkey
J. Protić, University of Belgrade, Serbia
D. Simpson, University of Brighton, England
D. Starčević, University of Belgrade, Serbia
S. Stojanov, Faculty for Informatics, Plovdiv, Bulgaria
D. Surla, University of Novi Sad, Serbia
M. Tuba, University of Belgrade, Serbia
P. Tumbas, University of Novi Sad, Serbia
P. Zarate, IRIT-INPT, Toulouse, France
K. Zdravkova, University of Ss. Cyril and Methodius, Skopje, FYROM

ComSIS Editorial office:

**University of Novi Sad, Faculty of Sciences,
Department of Mathematics and Informatics**
Trg Dositeja Obradovica 4, 21000 Novi Sad, Serbia
Phone: +381 21 458 888; **Fax:** +381 21 6350 458
www.comsis.org; Email: comsis@uns.ac.rs

Volume 8, Number 1, 2011
Novi Sad

Computer Science and Information Systems

ISSN: 1820-0214

ComSIS Journal is sponsored by:

Ministry of Science and Technological Development of Republic of Serbia -
<http://www.nauka.gov.rs/lat>



Computer Science and Information Systems

AIMS AND SCOPE

Computer Science and Information Systems (ComSIS) is an international refereed journal, published in Serbia. The objective of ComSIS is to communicate important research and development results in the areas of computer science, software engineering, and information systems.

We publish original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. ComSIS also welcomes surveys papers that contribute to the understanding of emerging and important fields of computer science. Regular columns of the journal cover reviews of newly published books, presentations of selected PhD and master theses, as well as information on forthcoming professional meetings. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

ComSIS publishes invited and regular papers in English. Papers that pass a strict reviewing procedure are accepted for publishing. ComSIS is published semiannually.

Indexing Information

ComSIS is covered or selected for coverage in the following:

- Science Citation Index Expanded (also known as SciSearch) and Journal Citation Reports / Science Edition by Thomson Reuters,
- Computer Science Bibliography, University of Trier (DBLP),
- EMBASE (Elsevier),
- Scopus (Elsevier),
- Summon (Serials Solutions),
- EBSCO bibliographic databases,
- IET bibliographic database Inspec,
- FIZ Karlsruhe bibliographic database io-port,
- Index of Information Systems Journals (Deakin University, Australia),
- Directory of Open Access Journals (DOAJ),
- Google Scholar,
- Center for Evaluation in Education and Science and Ministry of Science of Republic of Serbia (CEON) in cooperation with the National Library of Serbia,
- Serbian Citation Index (SCIIndeks),
- doiSerbia.

Information for Contributors

The Editors will be pleased to receive contributions from all parts of the world. An electronic version (MS Word or LaTeX), or three hard-copies of the manuscript written in English, intended for publication and prepared as described in "Manuscript Requirements" (which may be downloaded from <http://www.comsis.org>), along with a cover letter containing the corresponding author's details should be sent to one of the Editors.

Criteria for Acceptance

Criteria for acceptance will be appropriateness to the field of Journal, as described in the Aims and Scope, taking into account the merit of the content and presentation. The number of pages of submitted articles is limited to 30 (using the appropriate Word or LaTeX template).

Manuscripts will be refereed in the manner customary with scientific journals before being accepted for publication.

Copyright and Use Agreement

All authors are requested to sign the "Transfer of Copyright" agreement before the paper may be published. The copyright transfer covers the exclusive rights to reproduce and distribute the paper, including reprints, photographic reproductions, microform, electronic form, or any other reproductions of similar nature and translations. Authors are responsible for obtaining from the copyright holder permission to reproduce the paper or any part of it, for which copyright exists.

Computer Science and Information Systems

Volume 8, Number 1, January 2011

CONTENTS

Editorial

Papers

- 1** **VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises**
Sebastian Marius Roşu, George Drăgoi
- 27** **Towards the Methodology for Development of Fuzzy Relational Database Applications**
Srđan Škrbić, Miloš Racković, Aleksandar Takači
- 41** **Design Pattern Instantiation Directed by Concretization and Specialization**
Peter Kajsa, Lubomir Majtas, Pavol Navrat
- 73** **A Grammar-based Model for the Semantic Web**
Hyosook Jung, Seongbin Park
- 101** **Ontology-based Multi-label Classification of Economic Articles**
Sergeja Vogrinčič, Zoran Bosnić
- 121** **An Analysis of an Assessment Model for Participation in Online Forums**
Mee-Chin Wee, A. Abrizah
- 141** **A New Method for Constructing Kernel Vectors in Morphological Associative Memories of Binary Patterns**
Yiannis.S. Boutalis

Papers selected from "The 4th International Conference on Information Technology", June 3 – 5, Amman, Jordan 2009

- 167** **Supply Chain Performance Measurement System Based on Scorecards and Web Portals**
Nenad Stefanovic, Dusan Stefanovic
- 193** **Usage of Agents in Document Management**
Dragoslav Pešović, Milan Vidaković, Mirjana Ivanović, Zoran Budimac, Jovana Vidaković
- 211** **Integration of Recommendations and Adaptive Hypermedia into Java Tutoring System**
Aleksandra Klačnja-Milićević, Boban Vesin, Mirjana Ivanović, Zoran Budimac

EDITORIAL

This, the first regular issue of Computer Science and Information Systems in 2011, marks the start of a new publication schedule for the two regular yearly issues of ComSIS: January–June, replacing the old June–December scheme. We hope that the new schedule will assist in bringing more up-to-date information and research results to the community of ComSIS readers. In addition, we are pleased to announce that ComSIS has been included in the Directory of Open Access Journals (DOAJ), contributing to the widening international recognition of our journal.

The content of this issue of ComSIS consists of seven regular papers and three papers selected from ICIT 2009, 4th International Conference on Information Technology, held during June 3–5 in Amman, Jordan. We thank the organizers of ICIT, especially the General Chair of the conference, professor Yahya Abdelfatah A., Al-Zaytoonah University, Jordan, Dean of the Faculty of Science and IT, and Program Chair of the conference, professor Ali Al-Dahoud, Al-Zaytoonah University, for assisting in the contribution of high-quality articles to this issue of ComSIS.

The first of the regular papers, entitled “VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises,” by Sebastian Marius Roşu and George Drăgoi, analyzes the network architecture for geographically dispersed enterprises, which are viewed as virtual enterprises, in the context of supporting virtual project development by virtual teams. The article also presents an enterprise networks monitoring solution using open source software.

The next regular paper “Towards the Methodology for Development of Fuzzy Relational Database Applications,” by Srđan Škrbić, Miloš Racković, and Aleksandar Takači, tackles the problem of extending the classic relational data model with mechanisms that can handle imprecise, uncertain and inconsistent attribute values using fuzzy logic and fuzzy sets. Besides presenting a fuzzy relational data model used for fuzzy knowledge representation in relational databases, the paper also introduces a novel CASE tool for fuzzy database model development.

The article “Design Pattern Instantiation Directed by Concretization and Specialization” by Peter Kajsa, Lubomir Majtas, and Pavol Navrat, describes a method of design pattern instantiation support that allows specification of pattern instance occurrence via semantic extension of UML directly on the context, enabling the use of higher levels of abstraction in the modeling of patterns. The approach also provides the means to determine and control the

results of transformations and can represent a framework for supporting custom model structures.

In “A Grammar-Based Model for the Semantic Web”, Hyosook Jung and Seongbin Park present a two-level model for the Semantic Web from the perspective of formal language theory. The model consists of two grammars where the first-level grammar supports ontology creation, and the second-level grammar is used for specification of ontological instances. Based on the model, the authors implemented a system which enables construction of small-scale Semantic Web environments.

“Ontology-Based Multi-Label Classification of Economic Articles” by Sergeja Vogrinčić and Zoran Bosnić, presents an approach to the task of automatic multi-label document categorization in the field of economics, where labels represent keywords, using supervised machine learning. The article describes the complete cycle of corpus preparation using a tool which integrates ontology construction with text mining methods, and evaluation of three groups of multi-label classification approaches: transformation to single-class problems, specialized multi-label models, and hierarchical/ranking models, observing potential for successful use of all evaluated methods.

Wee Mee Chin and A. Abrizah, in “An Analysis of an Assessment Model for Participation in Online Forums,” analyze an assessment model implemented in online discussion forum software. The model is aimed at automating the assessment of students’ participation in forums, and is formulated based on four different participation indicators and educators’ feedback. Evaluation of the model in a real-world scenario indicates a high level of correlation between performance indicator scores generated by the model and the actual scores given by educators, suggesting that the assessment model is suitable for the task.

“A New Method for Constructing Kernel Vectors in Morphological Associative Memories of Binary Patterns” by Yiannis S. Boutalis examines kernel vectors as a representation for the retrieval of pattern associations, and focuses on the case when patterns are corrupted by a particular kind of erosive noise. Theoretical justification for the failure of kernel vectors under erosive noise is given, and a new method is proposed for the construction of kernel vectors for binary patterns associations that, instead of binary values, use ‘gray’ values in the [0, 1] range. The article shows that the new kernel vectors carry the good properties of conventional kernel vectors, can be easily computed, and do not suffer from the particular noise deficiency of conventional kernel vectors.

The following three articles represent selected and revised versions of papers published in the proceedings of the 4th International Conference on Information Technology (ICIT), June 3–5, 2009, Amman, Jordan.

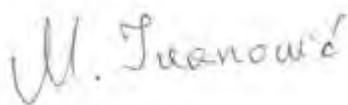
In “Supply Chain Performance Measurement System Based on Scorecards and Web Portals” Nenad Stefanović and Dušan Stefanović introduce an architecture of a pervasive performance measurement (PM) system for supply chain management which aims to unify different business elements, concepts, technologies and tools. The article describes the main system elements: the process model, metrics and data warehouse, and presents a specialized PM Web portal which enables proactive performance monitoring and fosters the improvement and optimization.

Dragoslav Pešović, Milan Vidaković, Mirjana Ivanović, Zoran Budimac, and Jovana Vidaković, in “Usage of Agents in Document Management” consider the application of agents and workflows in the area of document management systems. The employed multi-agent system is the EXtensible Java-based Agent Framework (XJAF) – a pluggable architecture for hierarchical intelligent agent systems with communication based on KQML. Workflows, on the other hand, are handled by “Workers, Inc.” – a workflow management system implemented using mobile agents.

Finally, the article “Integration of Recommendations and Adaptive Hypermedia into Java Tutoring System” by Aleksandra Klačnja-Milićević, Boban Vesin, Mirjana Ivanović, and Zoran Budimac, presents an approach to the integration of recommender system technology into an existing web-based Java tutoring system, in order to provide various adaptive programming courses. The considered methods for making such an e-learning system adaptive include recommending and adapting the appearance of hyperlinks, as well as recommending actions and resources.

On behalf of the Editorial Board and the ComSIS Consortium, we would like to thank the authors for their high-quality contributions, and also the reviewers for the effort and time invested into the preparation of this issue of Computer Science and Information Systems.

Editor-in-Chief
Mirjana Ivanović



Managing Editor
Miloš Radovanović



VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises

Sebastian Marius Roșu^{1,2} and George Drăgoi²

¹ Special Telecommunications Service, Information Technology Department,
323A Splaiul Independetei Street, Bucharest 6, 060032, Romania
sebastianrosu@stsnet.ro

² Polytechnic University of Bucharest, PREMINV Research Center,
313 Splaiul Independetei Street, Bucharest 6, 060042, Romania
gdragoi@mix.mmi.pub.ro

Abstract. In order to be competitive enterprises continuously implement ICT strategies & architectures to improve manufacture, research, products quality, sales, services and costs control. All large enterprises have a local area network, a virtual private network, an Intranet and Internet, servers and workstations for operations, administration and management working together for the same objective: profits. The virtual enterprise and the virtual team's concepts are discussed in this article. This work analyzes the network architecture for geographically dispersed enterprises (seen as virtual enterprises) as support for virtual project development by virtual teams work. In addition, the paper presents an enterprise networks monitoring solution using open source software (OSS).

Keywords: Virtual Enterprise, Virtual Team, Enterprise Network, Virtual Network, Network Management.

1. Introduction

All organizations use today Internet or Internet technologies to attract, retain and cultivate relationships with customers, streamline supply-chain, manufacturing, procurement systems and automate corporate processes to deliver the right products and services to customers quickly and cost-effectively, also to capture, explore, analyze, and automate corporate processes information on customers and company operations in order to provide better business decisions [14, 29]. Development of ICT leaves much more freedom to the designers and consultants to accommodate organizations to other influences, both internal and external [11]. For business, e-service is going to be a new way to save money, to revenue growth, and faster development model. For end-users, e-services increase productivity and simplify life, take advantage of more sophisticated and specialized services on as needed basis. At the level of production-dedicated

enterprises, e-services are [13]: business-to-business (supply-side), intra-business (internal-side), and business-to-customer (customer-side).

New enterprise model architecture uses the Intranet/Internet/Extranet infrastructure [14] and technologies (see figure 1).

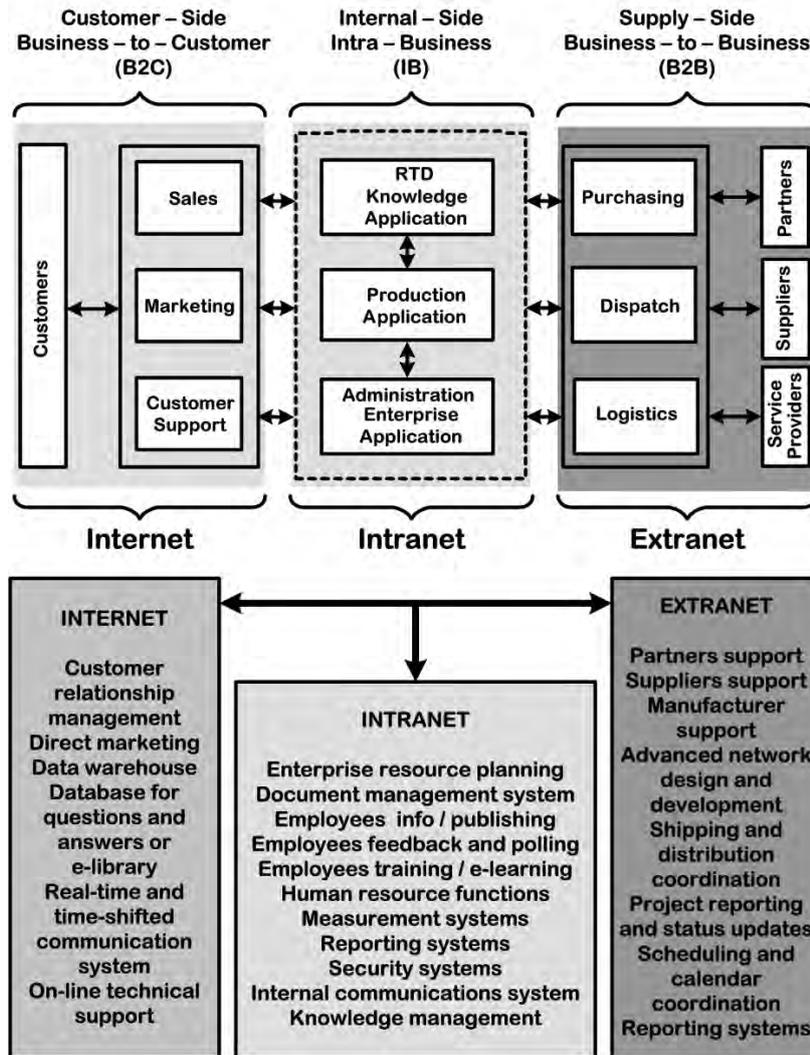


Fig. 1. The exchange of information services and goods through networks (up) and Internet/Intranet/Extranet based tools for the enterprise in the new e-economy (down)

As a general requirement for an infrastructure support the enterprises must be able to inter-operate and exchange information and knowledge in real time so that they can work as a single integrated unit, although keeping their independence/autonomy. For the future, e-services and e-business, as were

defined, require the enterprise re-thinking and re-modeling, with the system and applications design for an efficient use of new network technologies.

The perspectives of this kind of manufacturing and economy, named shortly new digital economy (e-economy), are based on the product perspective (holistic product view, product life-cycle, value-network integration, etc.), business organizational perspective (new organizational form, customers and suppliers integration, collaborating organization etc.), the technology perspective (technological building blocks, infrastructures, interoperability etc.) and the individual perspective (skills, workspaces, collaborating individual, different rolls: worker, consumer, citizen), [28].

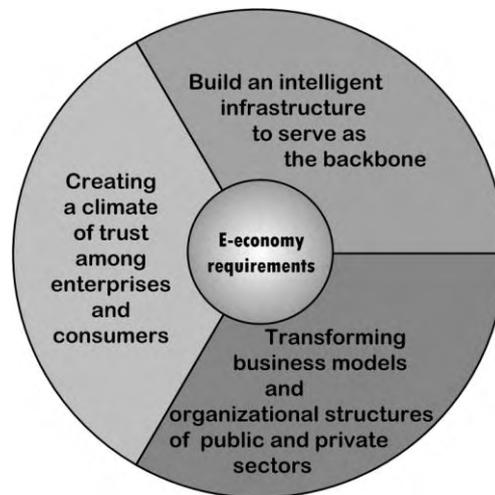


Fig. 2. The new digital economy requirements

Building an e-economy for the 21st century is a complex challenge. It requires (see figure 2):

- To transform business models and organizational structures of public and private sectors to generate continuous streams of productivity gains and product innovations, through the applications and use of ICTs;
- To create a climate of trust among consumers and businesses that fosters the growth of the e-economy in each country and internationally and creates global markets for electronic goods and services;
- To build an intelligent infrastructure to serve as the backbone of the e-economy – by encouraging investment, strengthening research, enhancing commercialization and ensuring that all persons have access to this infrastructure and know how to use it.

Developing and implementing these strategies will require partnership and collaboration among the private, public and academic sectors as well as other agencies and organizations that strive to link these together. It will require the active involvement of consumers and citizens.

Also, today, the critical and strategically questions are:

- What is needed to make the e-economy a priority?
- What overall strategies are needed to catalyze actions that respond successfully to the opportunities and challenges presented by the e-economy?
- Are there other factors in addition to those already identified that are yet to be understood or fully harnessed, and that will enable each enterprise to benefit more fully as we progress towards a mature e-economy?
- What additional measures are needed to address the broader challenges of the e-economy?

Enterprises are now facing growing global competition and the continual success in the marketplace depends very much on how efficient and effective the companies are able to respond to customer demands. The formation of virtual enterprise network is taking up momentum to meet this challenge.

The concepts of a virtual enterprise (VE) and of virtual teams, enabled by a new generation of Internet/Intranet/Extranet - based services are discussed in this article, as a means to stay competitive and to thrive in a turbulent market. This work analyzed the network architecture for an enterprise geographically dispersed (virtual enterprise) as support for virtual private networks (VPNs) possible structures and presents a network monitoring solution using open source software.

2. Virtual Enterprise

The production process does not continue in a single company in a same geographic location. Companies feel the need to focus on their core competence and join together in virtual industrial groups, dispersed geographically to meet requirements of new products/services required in the market [29]. Also, the most classic examples of organizational network can be found in several fields of economy such as automotive – this trend is prevalent in many other areas, including agriculture and food industry. Hereby, the VE concept appears. As a definition, virtual enterprise is a temporary alliance of enterprises that aim to share resources and skills in order to respond better and faster to emerging opportunities in the market, based on a technical infrastructure and information technologies represented by communications/computer networks [6]. This concept is supported by new technologies and globalization beginning to dominate the orientation of industrial development. The functions and activities linked to the products begin to decentralize, distributing on large geographic areas. Design, production planning and marketing, production, services, etc., can be realized in any place in a country, a continent or the world, because the infrastructure facilities allow the exchange of information, goods and services. As a general requirement for infrastructure to support VE, we emphasize that companies must be able to work together, exchanging information among themselves,

goods and services in a very short time (in real time), giving the impression that works as an integrated enterprise, even if each retains its autonomy.

Data, information, knowledge and processing or storing them is distributed temporarily to global scale, using facilities offered by technologies like network developed in the last years.

Virtual enterprise is a *temporary alliance of partners* focus on their core competencies, able to ensure cooperation, the process of innovation in network and to respond rapidly to business requirements [30]. In this case, the essential features of VE are cooperation partners focus on their core competencies and innovation in network, which is characterized by [30]:

- *Innovation capacity* (in the culture of the organization, power of innovation - financial strength, strategy, market knowledge, innovation management, project management, etc.);
- *Processing capacity* (dynamic organization - dynamic structures, work flow management, new information systems, expertise and technological opportunities);
- *Cooperation Ability* (teamwork ability, mentality on removing barriers in attitude/mentality on cooperation, thinking in the network, etc.).

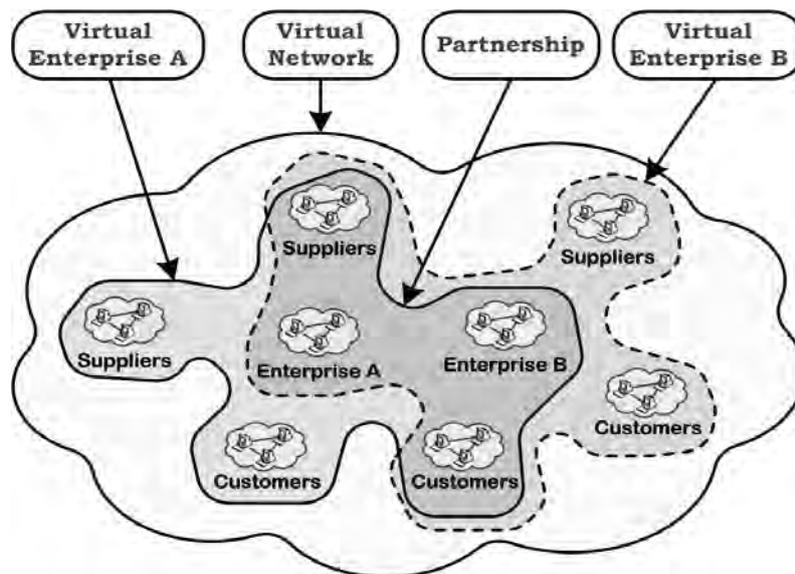


Fig. 3. The virtual enterprise partnership

Choosing partners to partnership creation (see figure 3) is very important when seeking to increase the competitiveness of the enterprise in a VE system [27] and represent a step in the process of VE forming (see figure 4).

The computers network and telecommunications, and other tools of information technology support cooperation between enterprises involved in a virtual enterprise. Modern enterprise with a virtual enterprise oriented

production type is a geographically distributed system with the following functions [18]:

- receiving orders and quick response to them;
- setting the structure on the virtual communication network;
- global planning system;
- local system planning and authorizing;
- control proactively at VE level;
- reactive control at alliance VE partner level.

These functions provide [4]:

- verifying real-time orders, in terms of opportunities for achievement (feasibility) and terms of delivery;
- VE configuration through negotiation and determination/verification of the ability to deliver products on time limits set by contract;
- establishment of order necessary to meet the order, and optimization of consumption and routes to maximize profits;
- acquisition and processing of data for monitoring the status of orders to avoid delays in delivery;
- control at the local and VE level and the manufacturer level, aiming to maintain the virtual alliance in normal operation area;
- exchange of information necessary virtual organization, to support all the functions provided;
- standard interfaces to other applications that VE interconnect.

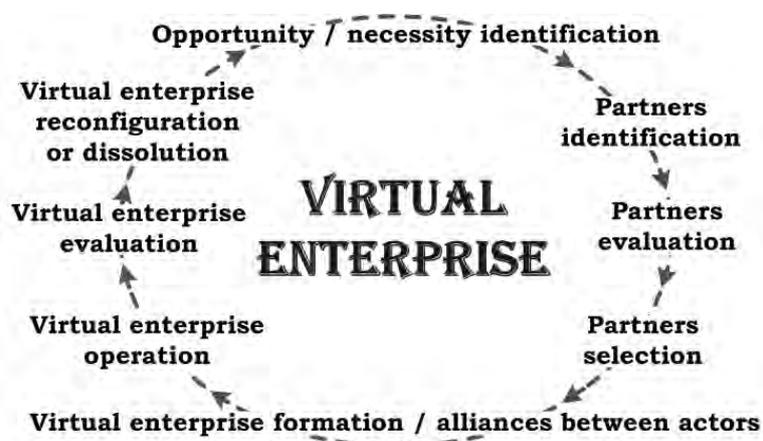


Fig. 4. The virtual enterprise framework

The idea of *virtual enterprise network* (VEN) is meant to establish a dynamic organization by the synergetic combination of dissimilar companies with different core competencies, thereby forming a *best of everything* consortium to perform a given business project to achieve maximum degree

of customer satisfaction [30]. In this context a VEN is a way for businesses to achieve virtual scale enabling them to operate as if they had more resources and capacity than they actually have.

In this emerging business model of VEN, the decision support functionality, which addresses the issues such as selection of business partners, coordination in the distribution of production processes and the prediction of production problems, is an important domain to be studied. In order to achieve collaboration between different actors in the VE, there needs to be common processes supporting the distributed product development process.

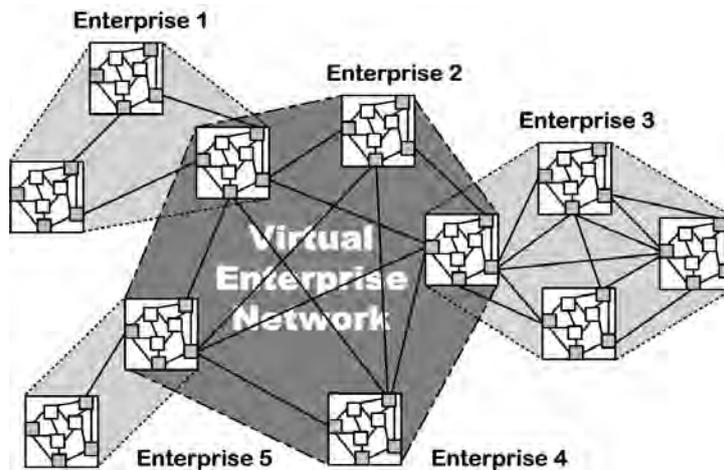


Fig. 5. The cooperative enterprise project in VEN

A virtual enterprise network (see figure 5) needs its own Private Member Collaboration System to communicate and develop its projects and bids. A hierarchical network design model breaks the complex problem of network design into smaller, more manageable problems. Each level, or tier, in the hierarchy addresses a different set of problems so that network hardware and software can be optimized to perform specific roles. Devices at the lowest tier of the hierarchy are designed to accept traffic into a network and then pass traffic up to the higher layers. The core of the network has one purpose: to provide an optimized and reliable transport structure by forwarding traffic at very high speeds. In other words, the core layer should switch packets as fast as possible. It needs its distinctive Network Business Applications such as Capability Aggregation and Tender Matching to enable it to function effectively as a co-operative in both pre-sales and contract operations. Also, it needs a Public Web Site to manage its interactions with potential customers and new members.

In this context the ingredients of VEN are presented in the figure 6. So, what is a virtual enterprise network solution [14, 28]?

- A way for businesses to achieve virtual scale enabling them to operate as if they had more resources and capacity than they actually have.

- A network of small enterprises allowing them to operate with all the resources and reach of a large enterprise but without sacrificing their speed, agility and low overheads.
- A network of enterprises enabling them to compete for bigger contacts with higher innovation and design elements with bigger customers who are prepared to have strategic partnerships with their suppliers [32].



Fig. 6. The elements of VEN

Developing information and communication technologies (ICT) in recent years allowed materialization concept of virtual alliance [30], which will lead to restructuring production, goods and services, based on these new technologies and trends of markets globalization and competition.

3. Virtual Teams

The term *virtual team* is used to cover a wide range of activities and forms of technology-supported activities [3]. A Virtual team is a group of people and sub-teams who interact through interdependent tasks guided by common purpose and work across links strengthened by information, communication and transport technologies [17]. With rare exceptions [22] all organizational teams are virtually to some extent. This era is growing popularity for virtual team structures in organizations [33]. The virtual teams are the teams whose members use technology to varying degrees in working across location, temporal, and relational boundaries to accomplish an interdependent task [22]. Enterprise virtual team's members are located in more than one

physical location. This team trait has fostered extensive use of a variety of forms of computer-mediated communication that enable geographically dispersed members to coordinate their individual efforts and inputs [24].

Enterprise virtual teams work across boundaries of time and space by utilizing modern computer-driven technologies. Although virtual teamwork is a current topic in the literature on global organizations, it has been problematic to define what virtual means across multiple institutional contexts [9]. Virtual teams are groups of individuals collaborating in the execution of a specific project (see figure 7) while geographically and often temporally distributed, possibly anywhere within (and beyond) their parent organization [21]. The organizational context of a virtual team is a conglomeration of pieces related to the life worlds, organizational structures and work practices of the local organizational contexts (local sites), the distributed organizational context (global company) and the professional context (software process improvement) [5].

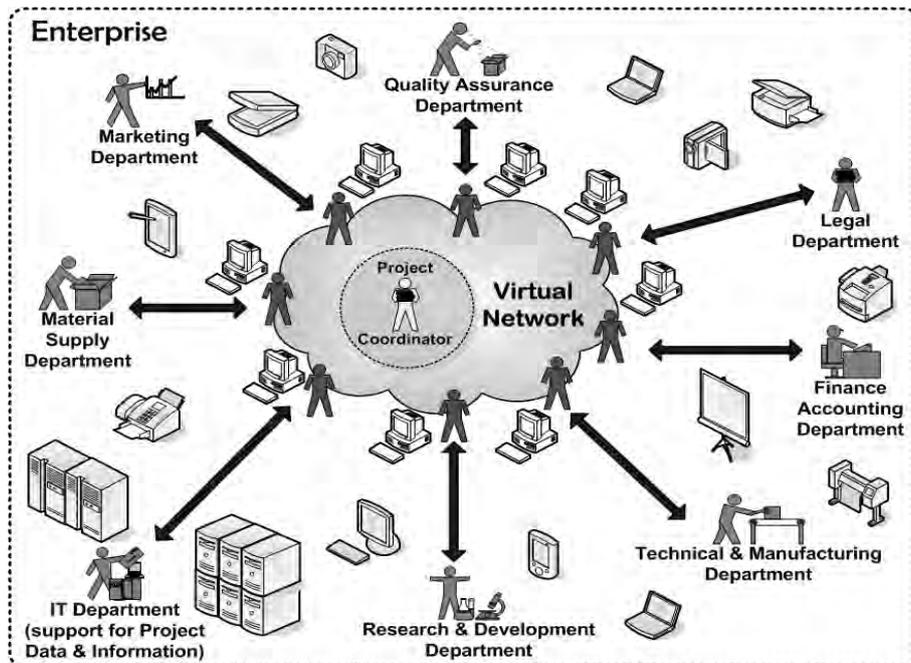


Fig. 7. The enterprise virtual team for project development

Virtual teams can be defined as groups of workers geographically, organizationally and/or time dispersed brought together by information technologies to accomplish one or more organization tasks [26]. The degree of geographic dispersion within a virtual team can vary widely from having one member located in a different location than the rest of the team to having each member located in a different country [31]. The term *virtual team* is a misnomer as although it makes reference to virtual reality and the concept of

creating a virtual space that can be experienced it also suggests that the virtual team isn't actually a team and as such can lead to a loss of performance.

In the initial stages it became clear that whilst a distributed computer based-platform could support distributed teams it could not completely replace face-to-face contact [13].

There are many different types of virtual teams (see figure 8) such as:

- *Management Teams* – comprised of managers who form teams because of their position in the organization. These teams collaborate every day and are responsible for taking decisions at the organization level.



Fig. 8. The virtual teams for project development

- *Project Teams* – are created to perform a specific task. Team members are selected depending on the role that it must fulfill in the project. These teams have a life equal to the project and are interdepartmental.
- *Research & Development Teams* – prepares and lead projects established with the management team. Between research team members should be complementary and a judicious sharing of the concrete tasks of each component to be grafted on to specialized training.
- *Production Teams* – team members regularly conducted activities at work; each member having a function within the team. Membership is defined very clearly.
- *Communities of Practitioners* – are bigger work entities, geographically dispersed (distributed), in which members participate, through Internet, guided by common goals, roles and rules – e.g. open source software project teams. These teams provide support for people working in the

same professional field and benefit from sharing experience. Membership in these teams is voluntary. The teams are focused on learning and / or knowledge development [30].

- *Concurrent Teams* – formed the short term for making recommendations for improvement of a process or system.
- *Service Teams* – created for the organization or customer support in a service – are designed to provide technical support.
- *Intervention Teams* – these teams provide answers or solutions in a very short time. They are usually activated in emergency situations.

The availability of a flexible and configurable base infrastructure is one of the main benefits of virtual teams [2]. Benefits of the virtual team's implementation at the enterprise level are presented in figure 9.

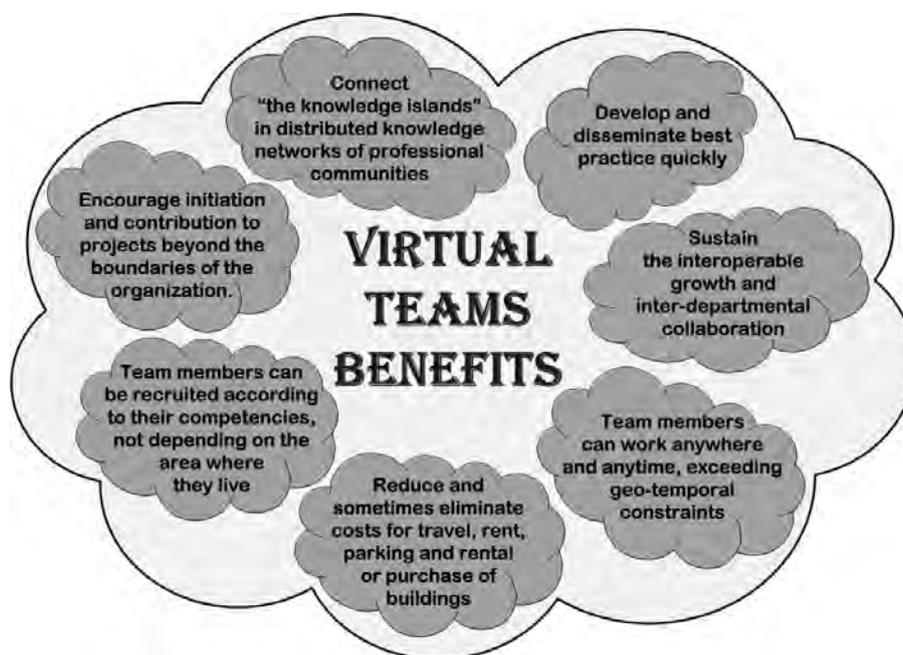


Fig. 9. The main enterprise virtual team's advantages

Virtual teams do not operate like traditional physical teams, as their requirements reflect a whole new way of communicating, working collaboratively, sharing information and mutually supporting other team members. The new technologies and approaches required to achieve this are completely alien to most of our present organizational culture. And this is why they fail. Cooperative processes are not the automatic results of implementing collaborative, real-time communication technologies, but the result of a carefully designed and systematically maintained virtual team development plan [7]. Product development is the process that covers the following: product design, production system design and product introduction

processes and start of production [19]. Product development is widely recognized as a key to corporate prosperity and is vital and needs to be developed both innovatively and steadily [8, 20].

Virtual product development through virtual teams in virtual enterprises is a temporary alliance of teams who share the skills, abilities and resources in pursuit of a project and whose cooperation is supported by computer network and appropriate tools, skills and special applications software. The ultimate objective of all product development teams is superior marketplace success of the new products [1] and services.

4. Virtual Enterprise Network Solutions

Traditional infrastructures type Internet/Intranet/Extranet have now a fast dynamic, marking the transition to new generation networks to provide higher speeds to the user (end to end), for different types of transactions and a reduction in the number of servers by passing information between two nodes.

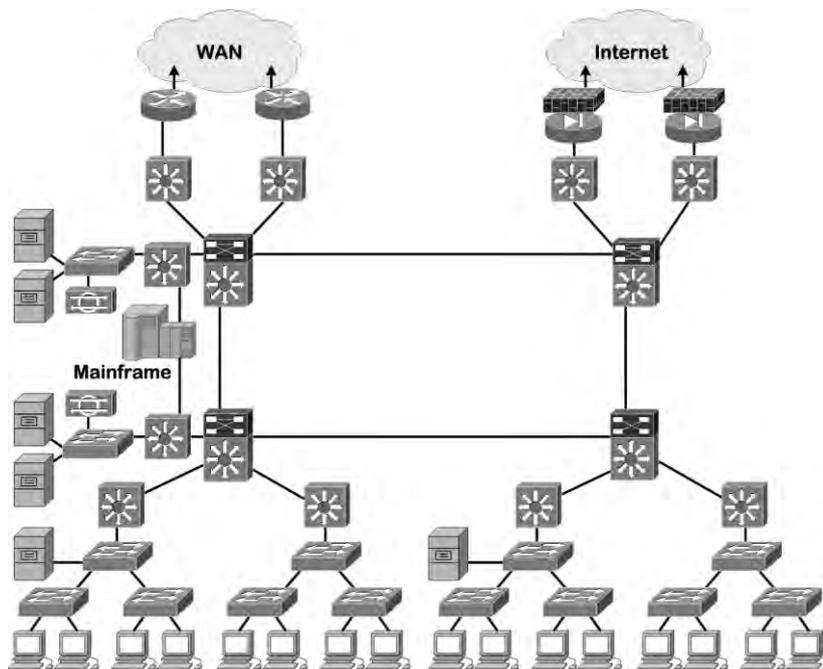


Fig. 10. A hierarchical scalable network design model

A hierarchical network design model (see figure 10) breaks the complex problem of network design into smaller, more manageable problems. A network is a chain where each link must be strong for the network to be

resilient. In this network computers are called host. The hosts are connected through communication subnets. The main task of the subnet is to send messages from one host to another.

Each level, or tier, in the hierarchy addresses a different set of problems so that network hardware and software can be optimized to perform specific roles. Devices at the lowest tier of the hierarchy are designed to accept traffic into a network and then pass traffic up to the higher layers.

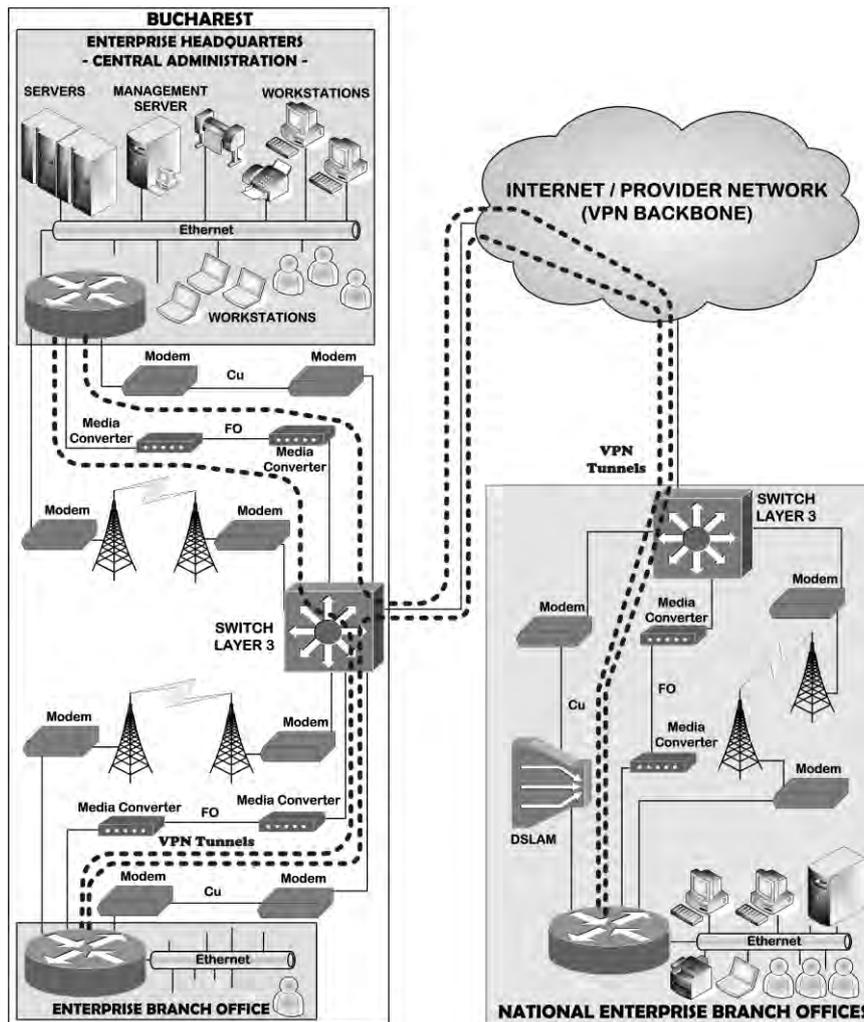


Fig. 11. A network general architecture with 3 loops (copper, fiber optic and radio), using the Internet or a provider network, for a large enterprise geographically dispersed to support transfer of large volumes of data over long distances

We propose in figure 11 a general architecture using Internet or a provider network for a large enterprise or an industrial holding (with headquarters and branches), geographically dispersed, implemented in the PREMINV platform [30].

Large area networks (WAN – Wide Area Network, specific large enterprise or businesses geographically dispersed) were designed to solve connection problems between workstations and local networks, or only the local network where the distances are too large to be able to use a simple cable connection. Large area networks are generally required for the transfer of large volumes of data over long distances. To implement a large area network can use the following transmission media: the public switched telephone network (PSTN), leased circuits of large bandwidth and high speed, high speed fiber optic, satellite links, radio links. A high performance backbone has an intrinsic value for ultra-fast Internet connection only if the points of connection and network users, POP (Point of Presence), providing an equivalent level of performance.

Appearance of virtual networks is related to the evolution switches. A *virtual network* is to combine a group of users regardless of their geographical position but such a manner that it flows together and to provide the best performance. The second advantage of a virtual network consists [14] of administrative solutions which accompany the products, allowing users moving from one group to another through a simple reconfiguration of the equipment. A virtual local network (VLAN) is a logical grouping of local network components without regard to their physical grouping. Common commands to create VLANs (for Cisco equipments) are the following:

```
vlan 75
  name Client_PREMINV_Provider1
vlan 76
  name Client1_PREMINV_Provider2
vlan 77
  name Client2_ PREMINV_Provider2

interface FastEthernet1/0/1
  description PREMINV_Provider1
  switchport access vlan 75
  switchport mode access

interface FastEthernet1/0/2
  description PREMINV_Provider2
  switchport trunk encapsulation dot1q
  switchport trunk allowed vlan 76,77
  switchport mode trunk

interface Vlan75
  description Client_PREMINV_Provider1
  ip address 5.10.34.1 255.255.255.252

interface Vlan76
  description Client1_PREMINV_Provider2
  ip address 5.10.36.1 255.255.255.252
```

```
interface Vlan77
  description Client2_PREMINV_Provider2
  ip address 5.10.36.5 255.255.255.252
```

Benefits of virtual local networks are the following:

- In current applications, physical topologies and/or logical networks are often modified, which in many cases requires changing the structure of wired, acquisition of new data communication equipment, reconfiguration of bridge, router equipment or changing database management. Thus the definition of virtual local networks to reduce costs and efforts dedicated to the topology change.
- Substantially reduce network loading by dividing the areas of broadcast (broadcasting refers to transmitting a packet that will be received (conceptually) by every device on the network).
- Increase the level of network security by restricting the number of users in each network and the separation of broadcast domains.

The *virtual private network* (VPN) is a network emulated (the *virtual*) built on public infrastructure (*shared*), dedicated to a client (the *private*) to connect users in locations and to ensure similar conditions of integrity, confidentiality and quality similar with those of a private network. VPNs allows the provisioning of private network services for an organization or organizations over a public or shared infrastructure such as the Internet or service provider backbone network.

The shared service provider backbone network is known as the VPN backbone and is used to transport traffic for multiple VPNs, as well as possibly non-VPN traffic.

VPNs provisioned using technologies such as Frame Relay and Asynchronous Transfer Mode (ATM) virtual circuits (VC) have been available for a long time, but over the past few years IP and IP/Multiprotocol Label Switching (MPLS) – based VPNs have become more and more popular. VPNs may be service provider or customer provisioned and falls into one of two broad categories [12, 23]:

- *Site-to-site* VPNs connect the geographically dispersed sites of an organization or organizations
- *Remote access* VPNs connect mobile or home-based users to an organization's.

There are three primary models for VPN architectures that can be implemented at the enterprise level [30]:

- *Host-to-host* – used to protect communication between two computers. The model is most used when a small number of users must be online or is given a remote that requires protocols that are normally uncertain.
- *Host-to-gateway* – protects communications between one or more individual hosts belonging to a specific network of an organization. Host-to-gateway is used to allow hosts of unsecured networks, access to internal organization services such as email and web servers.

- *Gateway-to-gateway* – this model protects communications between two specific networks, such as organization’s headquarters networks and organization’s branch offices or two business partners’ networks.

A VPN solution typically requires integration of several services (design, network management services, dial-up or dedicated access). Company that ensures coordination of services included in the solution is called *integrator*. Depending on the case, VPN solution integrator can be:

- customer himself (by the network administrator) ;
- one of equipment providers (e.g. equipments supplier can ensure the services integration against payment) ;
- an independent company, specializing in turnkey solutions.

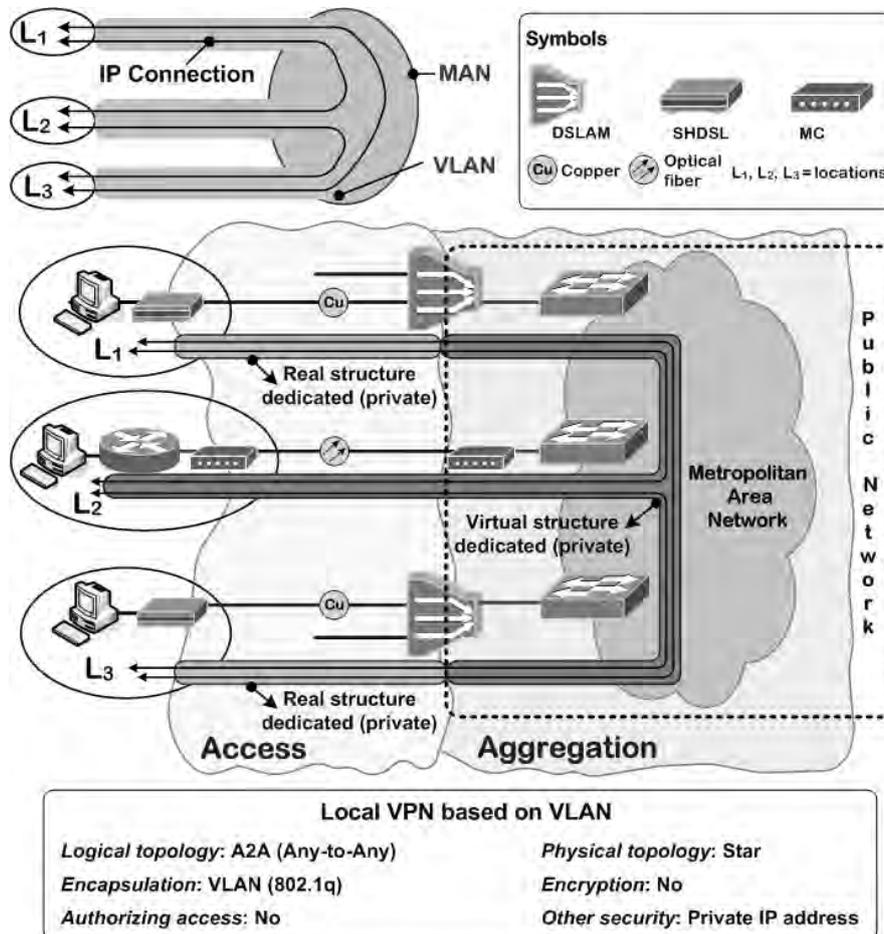


Fig. 12. Local VPN based on VLAN example

Possible solutions to implement a VPN structures for a VE system realization in a geographically dispersed enterprise can be:

- Local VPN based on VLAN (Virtual Local Area Network) – see figure 12;
- Local VPN based on IPsec (Internet Protocol Security) – see figure 13;
- VPN wide area based on IPsec;
- VPN wide area based on MPLS (Multiprotocol Label Switching);
- VPN based on PPPoL2TP (Point-to-Point Protocol over Layer 2 Tunnelling Protocol);
- UMTS (Universal Mobile Telecommunication System), etc.

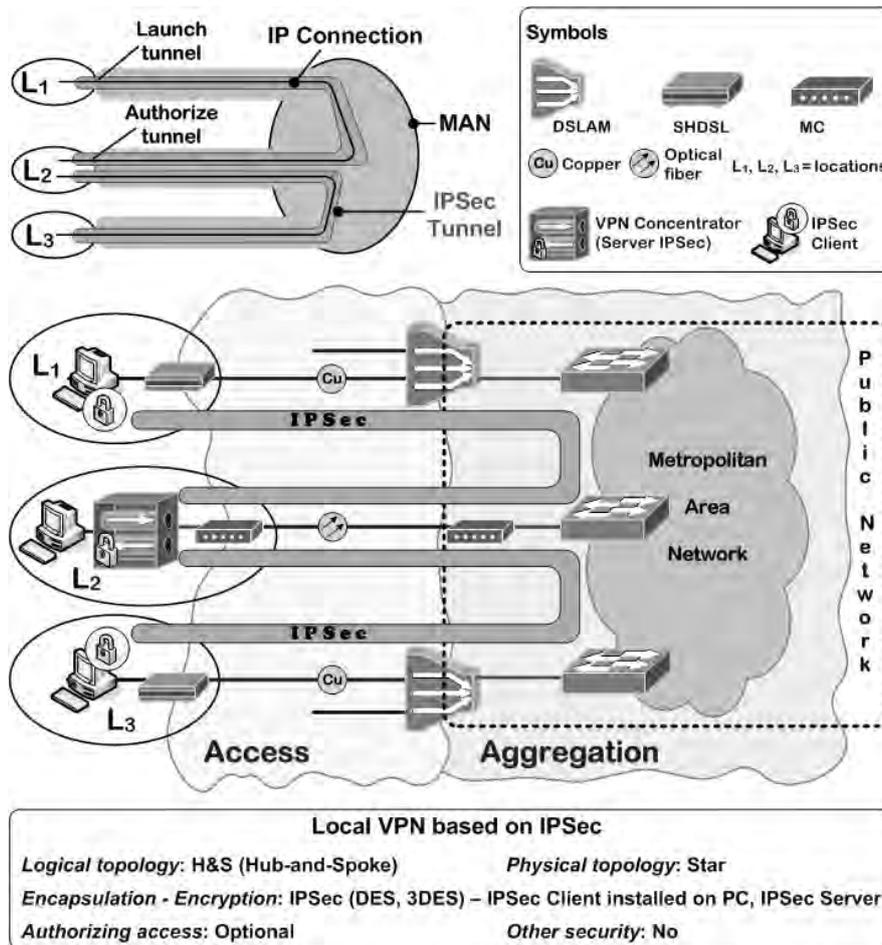


Fig. 13. Local VPN based on IPsec example

The main trend is now evolving to logic defined intranets and extranets, which will lead to the reintegration of the various networks in single logical subdivisions with no physical.

Structures that allow the approximation of this goal are virtual private networks. Newer, VPNs can be used in different ways to support business processes, is the ideal solution if it is not efficient in terms of construction

costs of a particular network for a firm with a workforce highly mobile, or for small firms that can not justify the cost of their telecommunications network.

Benefits services provided by VPN are [30]:

- the voice, video and data services convergence is done with low costs;
- Secure Remote Access to company resources;
- costs predictable and easier to budget, independent of traffic;
- the possibility of transferring any-to-any of data-voice-video applications;
- reliable support for LANs integration;
- security of data transmission;
- constant transfer rate, technological guaranteed;
- smart management solutions.

VPNs can be purchased from a telecommunications company and as an alternative they can create by using existing network infrastructure as the Internet or public switched telephone network, and software through the tunnel crossing.

The VPNs will be done according to enterprise network territorial expansion: local, metropolitan (county) and national (international).

5. Virtual Enterprise Network Monitoring

Network management represents the activities, methods, procedures, and tools (software and hardware) that pertain to the operation, administration, maintenance, and provisioning of networked systems [25]. The corporations can development permanently project with virtual teams support if have a good network management.

Proposed and implemented solution by us is to use a host and service monitor designed to inform as of network problems before your clients, end-users or managers do.

A system and network monitoring application is Nagios®. We found and took this software from the Internet by download. This software is licensed under the terms of the *GNU General Public License Version 2* as published by the *Free Software Foundation* (GNU General Public License is a free, copy left license for software and other kinds of works). This gives you legal permission to copy, distribute and/or modify Nagios® under certain conditions.

We've installed Nagios® on a server with the following technical characteristics: 2 Dual Core Intel Xeon (TM) 3.6 GHz processors (64-bit), 2 GB RAM, 2 x 80 GB Hard Drives and Debian Linux 4.0 Operating System.

Some of the many Nagios® features include [16]:

- Monitoring of network services (SMTP, POP3, HTTP, NNTP, PING) and monitoring of host resources (processor load, disk and memory usage, running processes, log files, etc.)

VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises

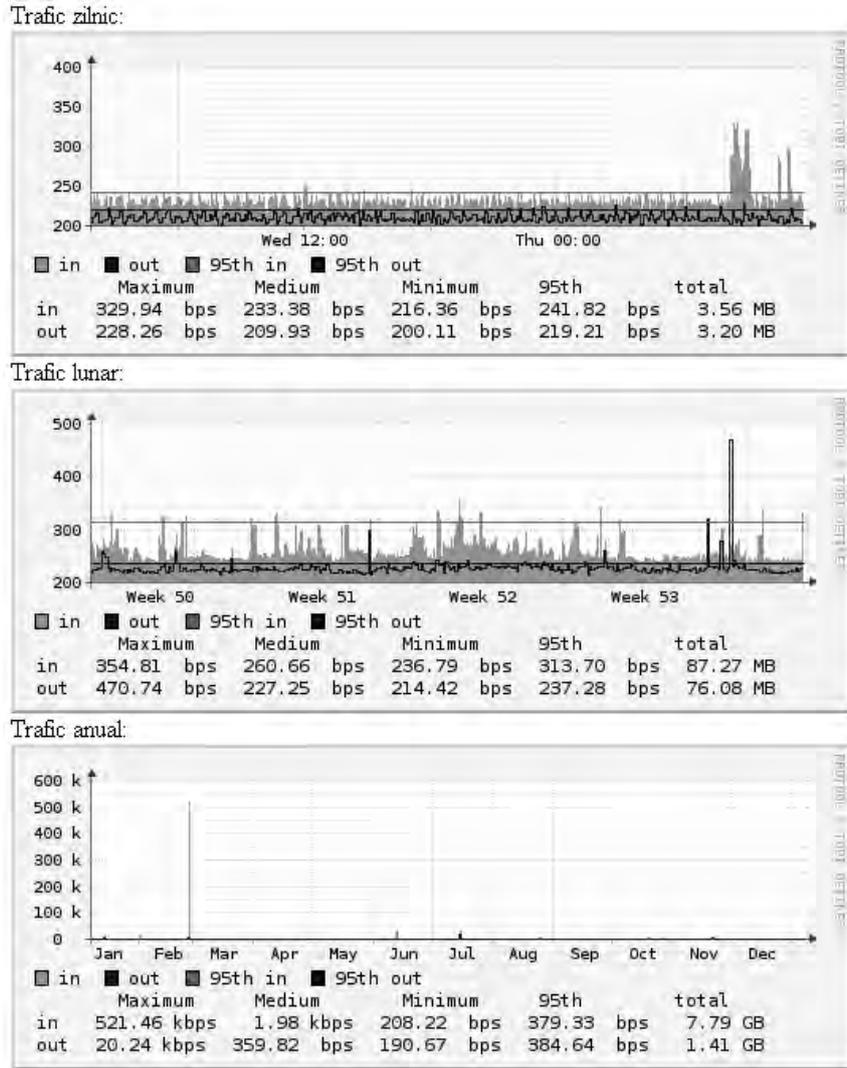


Fig. 14. The Nagios® traffic statistics for an enterprise location (daily, monthly and annual traffic)

- Monitoring of environmental factors such as temperature (see in figure 14 Nagios® traffic statistics);
- Support for implementing redundant and distributed monitoring servers;
- Simple plug-in design that allows users to easily develop their own host and service checks and ability to define network host hierarchy, allowing detection of and distinction between hosts that are down and those that is unreachable;

- Contact notifications when service or host problems occur and get resolved (via email or other user-defined method);
- Optional escalation of host and service notifications to different contact groups;
- Ability to define event handlers to be run during service or host events for proactive problem resolution;
- External command interface that allows on-the-fly modifications to be made to the monitoring and notification behavior through the use of event handlers, the web interface, and third-party applications;
- Retention of host and service status across program restarts;
- Scheduled downtime for suppressing host and service notifications during periods of planned outages and ability to acknowledge problems via the web interface;
- A Web interface – viewing current network status, notification and problem history.
- Simple authorization scheme that allows restricting what users can see and do from the web interface.

Also, Nagios® can do [25]:

- The connections state verification by PING command at monitored equipments;
- The loops state verification by OSPF routing protocol state monitoring;
- For monitored equipments configurations does default intervals saves.

We implemented Nagios® to a large enterprise which has its headquarters in Bucharest and branch offices (agencies) in the country – in big as well as in medium and small cities.

All enterprise locations have a local area network and communicate among themselves through a virtual private network. In each location were made two or three loops – one copper, one optical fiber and/or radio. To implement this application we have used over a hundred locations. In figure 15 are presented the locations monitored for this large enterprise – we eliminated the beneficiary name for advertising reason. We realized more scripts as support for different operations.

An example is *OSPF Verifying*. Generally, specific settings for OSPF (Open Shortest Path First – a dynamic routing protocol developed for Internet Protocol networks) [10], settings associated interfaces (example is for Cisco equipments) are:

```
router ospf 1
 network 5.10.34.0 0.0.0.255 area 34
 network 5.10.36.0 0.0.0.255 area 36
```

VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises



Fig. 15. Status for an enterprise geographically dispersed network - the local Enterprise Agencies of Bihor, Calarasi, Dolj & Salaj Counties are down and the local Agency of Arges County have long response times.

Using *OSPF Verifying* script below (realized in PHP programming language) we can monitor the OSPF Protocol status – generally used in locations that have at least 2 loops and prioritization is done by OSPF (if OSPF Protocol is upended for a provider loop monitoring system with loop UP).

```
<?php
unset($res_ospf);
$ip=$argv[2];
$comm=$argv[4];
$vec=$argv[6];
list($ip_v,$neighbor)=explode(':', $neigh);
$ip_v=trim($ip_v);$neighbor=trim($neighbor);
//echo "$ip\n$ip_v\n$comm\n$neighbor\n\n";
exec(" snmpwalk -v 1 -c $comm $ip
.1.3.6.1.2.1.14.10.1.6.$ip_v.0 ", $res_ospf);
$car=trim(substr($res_ospf[0],strlen($res_ospf[0])-1));
if(!strncmp($res_ospf[0], 'Time', 3)){$car='0';}
if($car=='8'){
    echo "Link to $neighbor is Up: Link OK ! \n";
}else{
    echo "Link to $neighbor is DOWN !!!\n";
}
?>
```

6. Conclusions

Considering future product development as collaboration and communication oriented we implemented in the PREMINV platform [15] a solution based on a virtual network concept using consistent collaboration integrated data sets and tools [13].

As a general requirement for this infrastructure support the companies must be able to inter-operate and exchange data, information and knowledge in real time so that they can work as a single integrated unit, although keeping their independence/autonomy.

A complete redesign of an existing enterprise would represent a not justifiable effort as companies are not replacing easily their running systems. A better strategy is to try to separate the internal functionalities from the network-related ones and develop the necessary mappings to legacy systems, to correspond to the new aggregator model for modern electronic system. In applying this strategy a virtual team could be formed with members located in different geographic locations. A virtual local area network is created for each project in the PREMINV platform. In addition to the team's full-time members, the team also includes contributing members who are recruited for specific components of the project. As such, a core group is responsible for leading the project and a sub-group is involved in specific components of the project while the full time employees from the central core of the team, experts in the different problems of the project (control systems, mechanic systems, electronic systems, programmer's etc.) are also team members. Today, in order to survive, whether small, medium or large organizations are composed of one or many enterprises (holding) it is necessary to learn from the past, supervise the present and plan the future.

The benefits of using Web Service technology as a core IT platform for the PREMINV comprises minimized processing times and costs and the improvement of the following features:

- *functionality* (system supporting and/or fully automating product development);
- *process*;
- *integration* (system to (internal or external) system communication);
- *usability* (effortless communication between the human user and the system);
- *security* (protection of the enterprise knowledge);
- *flexibility* (easily adjustable to a fast-changing business environment).

The decision support system presented in this paper includes two main modules: an information (classification and retrieval) module and a decision inference module. These modules implicitly guide users to follow systematic decision procedures to structure a problem, collect information, develop a model and analyze a decision step by step. The input/output interface provides an interactive mechanism for communication between users at the

client sites and the other components in the system. The database manages the domain knowledge embodied in the system.

The overall developed system infrastructure was based on object-oriented and agent technology. Every agent can communicate with other software agents to smooth out the decision process. When an agent conducts a decision support activity, it consults the knowledge base located in various sectors of a platform.

A solution for a large enterprise geographically dispersed network monitoring using open source software (Nagios®) has been presented in this paper. For an enterprise, network monitoring is a critical and very important function, which can save significant resources, increase network performance, employee productivity and maintenance cost of infrastructure.

This software (Nagios®) can be developed and implemented at a corporate level but also in a company that provides telecommunication services.

This work was realized at the UPB-PREMINV Research Centre. The validation of this solution by a case study in the PROGPROC research project was meant to determine the conceptual model for a new organization type integrating the virtual enterprise medium and outsourcing shared resources from UPB-PREMINV research centre to industrial partners.

References

1. Akgun, A.E., Lynn, G.S., Yilmaz, C.: Learning process in new product development teams and effects on product success: A socio-cognitive perspective. *Industrial Marketing Management*, Vol. 35, 210 – 224. (2006)
2. Ale Ebrahim, N., Ahmed, S., Taha, Z.: Innovation and R&D Activities in Virtual Team [Literature review]. *European Journal of Scientific Research*, Vol. 34, No. 3, 297-307. (2009)
3. Anderson, A.H., McEwan, R., Bal, J., Carletta, J.: Virtual team meetings: An analysis of communication and context. *Computers in Human Behavior*, Vol. 23, No. 5, 2558–2580. (2007)
4. Ash, J.: *Next Generation Knowledge Management*. Ark Group Press, London, United Kingdom. (2007)
5. Bjørn, P., Ngwenyama, O.: Virtual team collaboration: building shared meaning, resolving breakdowns and creating translucence. *Information System Journal*, Vol. 19, 227-253. (2009)
6. Camarinha – Matos, L.M., Lima, C., Osorio, A.: The PRODNET Platform for Production Planning and Management in Virtual Enterprises. In: *Proceedings of the 4th International Conference on Concurrent Engineering*, Nottingham. (1997)
7. Camarinha-Matos, L.M., Afsarmanesh, H., Ollus M.: *Methods and Tools for Collaborative Networked Organizations*, Springer, New York. (2008)
8. Chen, H.H., Kang, Y.K., Xing, X., Lee, A.H.I., Tong, Y.: Developing new products with knowledge management methods and process development management in a network. *Computers in Industry*, Vol. 59, 242-253. (2008)

9. Chudoba, K.M., Wynn, E., Lu, M., Watson-Manheim, M.B.: How virtual are we? Measuring virtuality and understanding its impact in a global organization. *Information Systems Journal*, Vol. 15, 279-306. (2005)
10. Cisco Systems, Inc.: OSPF Design Guide (2008-2009). Available: www.cisco.com (current December 2009)
11. Čudanov, M., Jaško, O., Jevtić, M.: Influence of Information and Communication Technologies on Decentralization of Organizational Structure. *ComSIS*, Vol. 6, No. 1, 93-109. (2009)
12. Deal, R.: *The Complete Cisco VPN Configuration Guide*. Published by Pearson Education, Cisco Press. (2005)
13. Drăgoi, G., Coteș, C., Roșu, L., Roșu, S.M.: Internet/Intranet/Extranet-Based Systems in the CESICED Platform for Virtual Product Development Environment. *Advances in Integrated Design and Manufacturing in Mechanical Engineering II*, Tichkiewitch, S., Tollenaere, M., Ray, P. (Eds.), Springer-Verlag New York, XIV, pp. 293-307. (2007)
14. Drăgoi, G., Drăghici, A., Roșu, S.M., Coteș, C.E.: Virtual Enterprise Network Solutions for Virtual Product Development in SMEs. *Proceedings of the CENTERIS'2009 – Conference on ENTERprise Information Systems*, Ofir, Portugal, 613-628. (2009)
15. Drăgoi, G., Guran, M., Cotet, C.E.: The PREMINV Platform for Training in Management and Engineering for Virtual Enterprise, in *Collaborative Business Ecosystems and Virtual Enterprises*. In: Camarinha - Matos, L.M. (eds): *Collaborative Business Ecosystems and Virtual Enterprise*. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 571-578. (2002)
16. Galstad, E.: *Nagios® Version 2.x Documentation (1999-2006)*. Available: www.nagios.org (current September 2009)
17. Gassmann, O., Von Zedtwitz, M.: Trends and determinants of managing virtual R&D teams. *R&D Management*, Vol. 33, No. 3, 243-262. (2003)
18. Husband, J., Bair, J.: *Making Knowledge Work: Impact of Web*, Ark Group Press, London, United Kingdom. (2007)
19. Johansen, K.: *Collaborative Product Introduction within Extended Enterprises*. Dissertation no. 943, University of Linköpings, Linköping, Sweden. (2005)
20. Lam, P.K., Chin, K.S., Yang, J.B., Liang, W.: Self-assessment of conflict management in client-supplier collaborative new product development. *Industrial Management & Data Systems*, Vol. 107, No. 5, 688 - 714. (2007)
21. Leenders, R.T.A.J., Engelen, J.M.L.V., Kratzer, J.: Virtuality, communication, and new product team creativity: a social network perspective. *Journal of Engineering and Technology Management*, Vol. 20, 69–92. (2003)
22. Martins, L.L., Gilson, L.L., Maynard, M.T.: Virtual teams: What do we know and where do we go from here? *Journal of Management*, Vol. 30, No. 6, 805–835. (2004)
23. Mason, A.G.: *Cisco Secure Virtual Private Networks*. Published by Pearson Education, Cisco Press. (2001)
24. Peters, L.M., Manz, C.C.: Identifying antecedents of virtual team collaboration. *Team Performance Management*, Vol. 13, 117-129. (2007)
25. Popescu, M.M., Roșu, S.M., Șofron, E.: Enterprise Network Monitoring Using Open Source Software. *Annals of DAAAM & Proceedings of the 20th International DAAAM Symposium*, Vienna, Austria, 223-224. (2009)
26. Powell, A., Piccoli, G., Ives, B.: Virtual teams: a review of current literature and directions for future research. *The Data Base for Advances in Information Systems*, Vol. 35, 6–36. (2004)

27. Roșu, S.M., Drăgoi, G.: Knowledge Bases Building to Partners Assessment in Virtual Enterprises Using the Weighted Point Evaluation Method. In: Annals of DAAAM & Proceedings of the 20th International DAAAM Symposium, Vienna, Austria, 89-90. (2009)
28. Roșu, S.M., Drăgoi, G.: Virtual Enterprise Network General Architecture. Proceedings of the 8th International Conference on Communications, Bucharest, Romania, ©2010 IEEE, 313-316. (2010)
29. Roșu, S.M., Drăgoi, G., Guran, M.: A Knowledge Management Scenario to Support Knowledge Applications Development in Small and Medium Enterprises. Advances in Electrical and Computer Engineering, Vol. 9, No. 1, pp. 8-15. (2009)
30. Roșu S.M.: Research on the implementation of Knowledge Management System for an Industrial Holding. Ph.D. Diss., Faculty of Automatic Control and Computers, "Politehnica" University of Bucharest. (2008)
31. Staples, D.S., Zhao, L.: The Effects of Cultural Diversity in Virtual Teams Versus Face-to-Face Teams. Group Decision and Negotiation, Vol. 15, 389-406. (2006)
32. Thuraisingham, B.: Web Data Mining and Applications in Business Intelligence and Counter-Terrorism. CRC Press LLC, Boca Raton, Florida. (2003)
33. Walvoord, A.A.G., Redden, E.R., Elliott, L.R., Coover, M.D.: Empowering followers in virtual teams: Guiding principles from theory and practice. Computers in Human Behavior, Vol. 24, No. 5, 1884-1906. (2008).

Sebastian Marius Roșu received the B.E. in Aerospace Construction and in Informatics from University "Polytechnica" of Bucharest, in 2000 and University „Dunărea de Jos” of Galați in 2002, respectively. In 2009 obtained doctoral degree in Automatics from University "Polytechnica" of Bucharest. From 2001-2004, he was a Researcher in Machines and Thermal Equipment at the GRUP ROMET Buzău. Since 2004, he was Quality Surveillance Receiving Inspector and Pressure Test Supervisor at the Nuclear Power Plant Unit Two Site of Cernavodă. Since as 2006 he is IT engineer at the Romanian Special Telecommunication Service. He has been active in the fields of the thermodynamic, design, nuclear systems, quality control, risk management, industrial informatics, knowledge management, computers, computer integrated enterprise, computer networks, networks design and collaborative design systems. Since 2004, he has been associated with the PREMINV Research Laboratory at the Polytechnic University of Bucharest first as PhD Student and from 2009 as PhD. He has published 5 chapters in international scientific books, 13 articles in international and national journals, 41 articles in the proceedings of international conferences and 9 research reports.

George Drăgoi received the B.S. in Electronics & Telecommunications and doctoral degree in Industrial Informatics from University "Polytechnica" of Bucharest, in 1982 and 1994, respectively. From 1982-1986, he was an Associate Researcher of the Romanian National Research Center in Telecommunication of Bucharest. From 1986 to 1990 he was a Research Assistant, from 1990 to 1993 lecturer, from 1993 to 1998 associate professor and from 2006 to present professor at the University "Polytechnica" of

Sebastian Marius Roşu and George Drăgoi

Bucharest. He has been active in the fields of the industrial robots control, industrial informatics, computer integrated enterprise, computer networks and telecommunications, virtual enterprise and collaborative design systems. Since 1996, he has been associated with the Institute National Polytechnic at Grenoble, France, as Associate and Visiting Professor. He received the Romanian Academy Award “Aurel Vlaicu” in 1993 and it has served as a member of the National Council for Technological Education and Innovation after 1999.

Received: January 27, 2010; Accepted: September 15, 2010.

Towards the Methodology for Development of Fuzzy Relational Database Applications

Srđan Škrbić¹, Miloš Racković¹, and Aleksandar Takači²

¹ Faculty of Science, Trg Dositeja Obradovića 3,
21000 Novi Sad, Serbia
{shkrba, rackovic}@uns.ac.rs

² Faculty of Technology, Bulevar Cara Lazara 1,
21000 Novi Sad, Serbia
stakaci@tf.uns.ac.rs

Abstract. In this paper we examine the possibilities to extend the relational data model with the mechanisms that can handle imprecise, uncertain and inconsistent attribute values using fuzzy logic and fuzzy sets. We present a fuzzy relational data model which we use for fuzzy knowledge representation in relational databases that guarantees the model in 3rd normal form. We also describe the CASE tool for the fuzzy database model development which is apparently the first implementation of such a CASE tool. In this sense, this paper presents a leap forward towards the specification of a methodology for fuzzy relational database applications development.

Keywords: fuzzy database, fuzzy-relational data model, fuzzy-relational CASE tool

1. Introduction

Relational model's disability to model uncertain and incomplete data can be viewed as its disadvantage in some applications. The idea to use fuzzy sets and fuzzy logic to extend existing database models to include these capabilities has been utilized since the 1980s. Although this area has been researched for a long time, concrete implementations are rare. Methodologies for fuzzy-relational database applications development are nonexistent.

However, literature contains references to several models of fuzzy knowledge representation in relational databases, as well as different approaches to fuzzy-relational database querying mechanisms. Some good overviews of these research areas can be found in [1,2,3].

In the next section of this paper we give a detailed description of a number of references that describe research in the use of fuzzy logic in relational databases. This overview stretches from the very beginnings of this idea, to the most recent approaches. Third section contains a description of our

approach to fuzzy-relational data modelling. We give a detailed description of our fuzzy meta model and investigate its theoretical value. Fourth section contains a description of the fuzzy-relational data modelling CASE tool, the first of its kind. We compare our approach to some previous approaches in the fifth section. At the end, we give the conclusion.

2. Related Work

One of the early works, the Buckles-Petry model [4] is the first model that introduces similarity relations in the relational model. This paper gives a structure for representing inexact information in the form of a relational database. The structure differs from ordinary relational databases in two important aspects: components of tuples need not be single values and a similarity relation is required for each domain set of the database. Zvieli and Chen [5] offered a first approach to incorporate fuzzy logic in the ER (Entity-Relationship) model. Their model allows fuzzy attributes in entities and relationships. It defines three levels of fuzziness. At the first level, entity sets, relationships and attribute sets may be fuzzy, i.e. they have a membership degree to the model. The second level is related to the fuzzy occurrences of entities and relationships, and on notion which instances belong to the entity or relationship with different membership degrees. Finally, the third level is concerned with the fuzzy values of attributes of special entities and relationships.

Fuzzy functional dependencies and fuzzy normal forms, as well as algorithms for dependency preserving and lossless join decompositions of fuzzy relations in specific fuzzy extensions of relational model are investigated in [6,7].

Umano and Fukami proposed FREEDOM-O, a fuzzy database system which is an extension of relational model of data [8]. This system supports a fuzzy data model, and querying. It is the first implementation of a fuzzy database system. After that result, other researchers have proposed similar fuzzy extensions to the relational model in such as in [9,10,11,12].

Another serious attempt to implement a fuzzy database system is given in [13,14]. Authors propose fuzzy extensions of the classical SQL and implement a system that allows using fuzzy conditions in place of Boolean ones.

The GEFRED (Generalized Model of Fuzzy Relational Databases) model [15] is a probabilistic model that refers to generalized fuzzy domains and admits the possibility distribution in domains. This is a fuzzy relational database model that has representation capabilities for a wide range of fuzzy information. In addition, it describes a flexible way to handle this information. Also, it contains the notion of unknown, undefined and null values. The GEFRED model experienced subsequent expansions, such as [16,17,18,19].

Chen and Kerre [20,21] introduced the fuzzy extension of several major EER (Extended Entity-Relationship) concepts. Fuzzy logic was applied to

some of the basic EER concepts connected to the notion of subclass and super class. Chaudhry, Moyne and Rundensteiner [22] proposed a method for designing fuzzy relational databases following the extension of the ER model of Zvieli and Chen. They also proposed a design methodology for FRDBs (Fuzzy Relational Databases), which contains extensions for representing the imprecision of data in the ER model, and a set of steps for the derivation of a FRDB from this extended ER model.

Galindo, Urrutia and Piattini [23] describe a way to use the fuzzy EER model to model the database and represent modelled fuzzy knowledge using relational database in detail. This work gives insight into some new semantic aspects and extends EER model with fuzzy capabilities. The model is called FuzzyEER model. Also, a way to translate FuzzyEER model to the FIRST-2, a database schema that allows representation of fuzzy attributes in relational databases is given. The FIRST-2 schema introduces a concept of Fuzzy Meta-knowledge Base (FMB). For each attribute type, it defines how to represent values and what information about them has to be stored in the FMB. In addition, in this work, authors introduce and describe specification and implementation of the FSQL - an SQL language with fuzzy capabilities in great detail. This language is an extension of the SQL language that allows users to write flexible conditions in queries, using all extensions defined by the FuzzyEER model.

We conclude that the current state of the art in this area includes mature fuzzy EER model extensions that describe a wide range of modelling concepts for full flavoured fuzzy database modelling. These conceptual models are supported by robust models for fuzzy data representation in relational databases, such as the FIRST-2. The possibilities to translate conceptual models to the relational-based ones are also studied in detail. In addition, the FSQL is the first implementation of fuzzy database query language that incorporates the majority of fuzzy logic concepts.

In [24,25,26] authors have studied the possibilities to extend the relational model with the fuzzy logic capabilities. The subject was elaborated in [27,28], where a detailed model of Fuzzy Relational Databases (FRDB) was given. One of the main features of the model is that it allows any fuzzy subset of the domain to be the attribute value which was not the case in previous FRDB models.

Moreover, using the concept of the Generalized Priority Constraint Satisfaction Problem (GPF CSP) from [29] and [30], authors have found a way to introduce priority queries into FRDB, which resulted in the PFSQL query language [31,32]. In [33] authors introduce similarity relations on the fuzzy domain which are used to evaluate the FRDB conditions. The PFSQL allows the conditions in the WHERE clause of the query to have different priority i.e. importance degree. It is one of the first languages with such capabilities. The GPF CSP gives the theoretical background for the implementation of priority queries.

In this paper, we focus on an innovative fuzzy relational data model designed to include a more detailed structure and allow better performance. We analyze concordance of this model to theoretical concepts of relational

model, especially normal forms. In addition, we describe the CASE tool that allows development of fuzzy databases using our model. This appears to be the first implementation of such a CASE tool. Proposed fuzzy relational data model and the CASE tool that supports it give a good foundation for development of the database part in fuzzy relational database (FRDB) applications. We discuss the possibilities to define a complete methodology for FRDB applications development and describe steps that need to be made in that direction.

3. Fuzzy Relational Data Model

In this section we describe our relational model extensions that constitute our variant of fuzzy relational data model. Our model stores crisp values in the same way as relational model does, while, for fuzzy values, we define fuzzy meta data model. In addition, here we provide an insight into the process of transformation of an example of the classical relational model with fuzzy attributes to the corresponding fuzzy relational data model.

If we wish to store a fuzzy value, we need to find a way to store data about its characteristic function. Theoretically, in this way, we could store any fuzzy value. But, in practice, only a handful of characteristic functions are in use. Let us name them fuzzy data types from this aspect. That is why we cover only a limited number of fuzzy data types and obtain an efficient and relatively simple data model.

An example relational model shown at Fig. 1 contains tables Worker and Car as well as an intersection table Uses that we use to model this many- to-many relationship. Tables Worker and Car have two fuzzy attributes each.

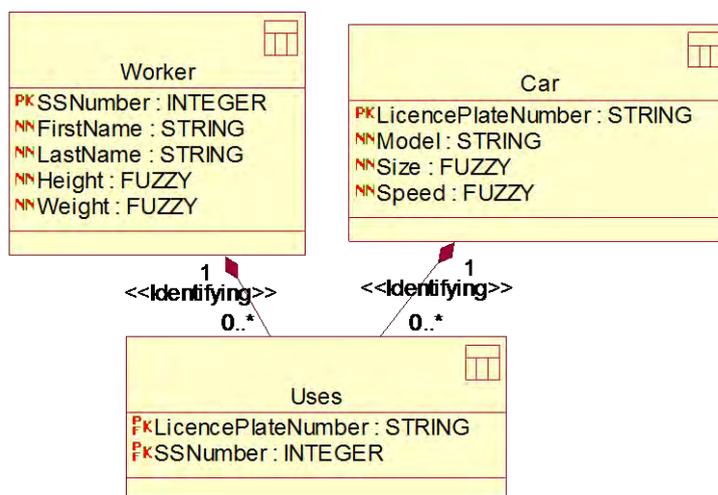


Fig. 1. Example fuzzy data model

The corresponding fuzzy-relational data model is shown at Fig. 2. The tables *Worker*, *Car* and *Uses* are shown at the top of the figure. They are the same as they were before except for the data type of fuzzy columns. In this model, they are of type INTEGER. Moreover, they became foreign keys that originate from the attribute *ValueID* in the table *FuzzyValue*. In order to represent these fuzzy values in the database, we extend this model with some additional tables that make the fuzzy meta data model.

The table *IsFuzzy* simply stores the information whether an attribute is fuzzy or not. All attribute names in the database are stored here, and beside the table and the attribute name (attributes *TableName* and *AttributeName*), the information whether the attribute is fuzzy (value of the attribute *IsFuzzy* is 1) or not (value of the attribute *IsFuzzy* is 0) is present.

The table *FuzzyValue* represents a connection between the fuzzy data model and the fuzzy data meta model. Every fuzzy value in every table is a foreign key that references attribute *ValueID* - the primary key of the table *FuzzyValue*. Thus, we have one record in the table *FuzzyValue* for every record with the fuzzy value in the database. The attribute *Code* is a foreign key from the table *FuzzyType*. This table stores the name of every possible type of fuzzy value allowed in the model.

These types are as follows:

- interval - fuzzy value is an interval,
- triangle - fuzzy value is a triangular fuzzy number,
- trapezoid - fuzzy value is a trapezoidal fuzzy number,
- general - fuzzy value is a general fuzzy number given by points,
- fuzzyShoulder - fuzzy value is a fuzzy shoulder,
- linguisticLabel - fuzzy value is a linguistic label,
- crisp - fuzzy value is actually a crisp value.

For every value in this list, there is a separate table in the meta model that stores data for all fuzzy values of specific fuzzy type. Every one of these tables has the attribute *ValueID*, foreign key from the table *FuzzyValue*. In this way, the value for the specific fuzzy attribute is stored in one of these tables depending on its type.

The attribute *ForValueID* in the table *FuzzyValue* is a foreign key that represents a recursive relationship and references the primary key of the *FuzzyValue* table. This attribute is used to represent linguistic labels. It has a value different than null if the type of the attribute that it represents is linguisticLabel. As mentioned before, linguistic labels only represent names for previously defined fuzzy values. In this fashion, if an attribute is a linguistic label, then its name is stored in the table *LinguisticLabel*. In this case, the attribute *ForValueID* has the value of *ValueID* of a fuzzy value that this linguistic label represents. We conclude that, in order to represent a linguistic label, two records in the table *FuzzyValue* are needed.

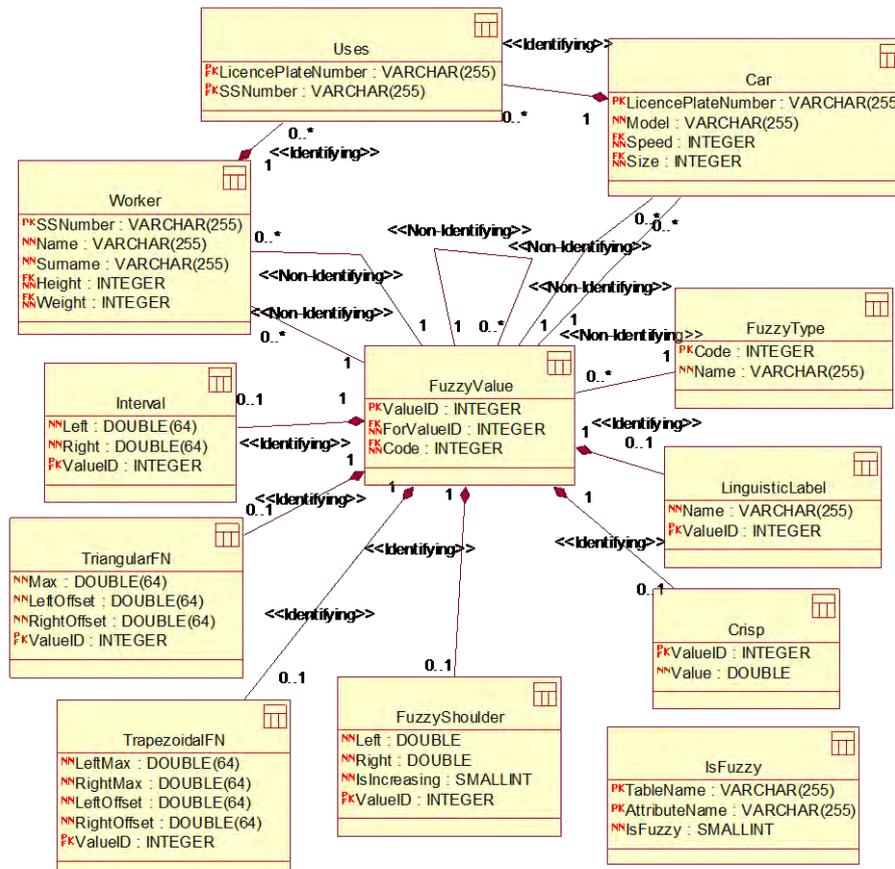


Fig. 2. Fuzzy relational data model.

For example, let us suppose that worker John Doe has a height designated with linguistic label Tall which represents the fuzzy shoulder with the membership function:

$$\begin{cases} 0, & x < 180 \\ \frac{x}{20} - 9, & 180 \leq x < 200 \\ 1, & x \geq 200 \end{cases} \quad (1)$$

and weights 80kg. In the table Worker there is a corresponding record with attribute values *SSNumber*=001, *Name*='John', *Surname*='Doe', *Height*=1, *Weight*=2. In the table FuzzyType there are three records with the following attribute values: Record T1: *Code*=1, *Name*='linguisticLabel'; Record T2: *Code*=2, *Name*='crisp'; Record T3: *Code*=3, *Name*='fuzzyShoulder'. Attributes *Height* and *Weight* in the table Worker are foreign keys that refer to the attribute *ValueID* in the FuzzyValue table. In this case, the table FuzzyValue contains three records with the following attribute values: Record

F1: *ValueID=1, ForValueID=3, Code=1*; Record F2: *ValueID=2, ForValueID=NULL, Code=2*; Record F3: *ValueID=3, ForValueID=NULL, Code=3*. Record F1 represents linguistic label Tall, Record F2 represents crisp value for weight and Record F3 represents fuzzy shoulder which corresponds to the linguistic label Tall. Further descriptions of these values are given in tables corresponding to their types described in the FuzzyType table. According to this, the table LinguisticLabel contains one record with the following attribute values: *Name='Tall', ValueID=1*. It describes the linguistic label Tall. The table FuzzyShoulder contains one record with the following attribute values: *Left=180, Right=200, IsIncreasing=1, ValueID=3*, describing the membership function given in equation (1), while the table Crisp contains a record describing the weight: *ValueID=2, Value=80*.

The rest of the values, for other fuzzy types, are stored in the database in a similar way. The complete description of all values and types that can be stored in the database can be found in [25, 33, 28]. The difference between the data model described there and this improved version is in structures that store fuzzy values. In the previous model, we had only two tables in the fuzzy meta data model - IsFuzzy and FuzzyValue. In that model, fuzzy values were stored in the table FuzzyValue as strings with predefined structure - one type for every type of fuzzy value that can be stored. In this way, the value of a column in a database record was not atomic, so it had to be decomposed in order to be used. That implies that our previous model was not even in the first normal form. Here we use one table for every fuzzy data type and have atomic values in the whole database.

Presented fuzzy meta model has been put through the synthesis algorithm [34] that guarantees that resulting model conforms to the 3rd normal form. Of course, fulfilment of theoretical conditions for the 3rd normal form relational model depends on the ground database model that we are creating too. In any case, this feature guarantees that if a database model is at least in the 3rd normal form, then the addition of the presented fuzzy meta model will result in a complete model at least in the 3rd normal form. In this way, presented fuzzy meta model significantly improves its theoretical and practical performance. The main reason for insisting on 3rd normal form in this model is the efficiency of the complete software system. In the previous case, when the values were not atomic, the system parsed the strings which correspond to the appropriate fuzzy values. Now, all information is obtained by following the primary-foreign key pairs, which results in better overall performance.

4. The CASE Tool

In this section we give an overview of the CASE tool for fuzzy relational model development. The application is implemented using Java programming language and Swing platform for the GUI (Fig. 3). It can be downloaded together with the source code and accompanying UML model from <http://www.is.pmf.uns.ac.rs/fuzzydb>.

4.1. Requirements

Requirements set in the process of modelling of the CASE tool include functions for simplified building of a fuzzy relational data model, as well as functions for its transformation to the SQL script.

Our intention was to implement a CASE tool capable for visual modelling and easy administration of all components of a fuzzy relational data model - tables, attributes, fuzzy data types and relationships. The CASE tool's GUI works in the similar way as in all modern tools of this type that allow modelling of classical relational models. So we do not describe the details here. In the model building process all the automation related to the migration of keys through relationships is included. This feature includes cascade deletion process, migration of keys during the relationship creation process and circular reference detection and prevention.

In addition, the CASE tool is required to allow easy SQL script generation for the specific database management system. In this sense, capabilities to specify the data types used by the DBMS and rules for mapping of the types used in the model (together with fuzzy data types) to these types had to be included.

The complete UML model that includes the use case diagrams, the static system model (class diagrams) and the dynamic model of the main processes, as mentioned before, is available for download.

4.2. Elements

The main window consists of five parts shown at Fig. 3: menu, toolbar, navigation tree, main panel and status bar.

The menu and the toolbar contain all of the commands available in the CASE tool. Using those commands user can manage data model, specify data types in the model and their mapping to SQL data types. The only data type that exists by default is Fuzzy. Crisp data types have to be explicitly defined together with their mapping to SQL data types that are specific to the DBMS used.

The navigation tree is a visual representation of data about the model that exists in the repository. The repository can contain a set of tables, while the tables contain attributes and relationships (Fig. 3). The main panel is located to the left of the navigation tree. Its content depends on the element selected at the navigation tree. For every type of element at the navigation tree, the panel that allows editing of that element is defined. The status bar with some useful information is located at the bottom of the application window.

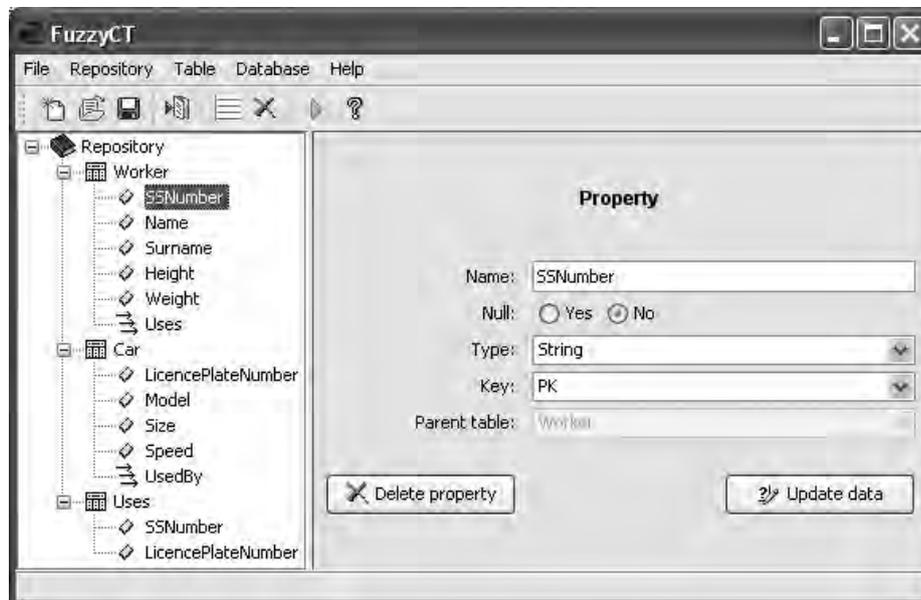


Fig. 3. The main window

4.3. Generating the SQL DDL script

This section provides insight into the process of generation of the SQL DDL (Structured Query Language Data Definition Language) script based on the fuzzy relational data model.

Activity diagram at Fig. 4 models the actions that need to be conducted in order to generate data definition SQL code from the model existing in the CASE tool. The first activity in the code generation is related to the FuzzyValue table definition. As mentioned above, this table is essential in our model, and presents a link between the model and the fuzzy meta model. The rest of the fuzzy meta model is defined in the similar way.

After that, the SQL code for the tables in the model is created. This activity has three sub activities. At first, for every table in the model we open a CREATE TABLE clause. After that, based on the table attributes, SQL code describing those attributes is generated inside the clause. At the end, the primary key constraint is added.

When all tables are created, foreign key constraints that connect tables need to be added. At first, code that links tables in the model to the fuzzy meta model via FuzzyValue table is created. Then the rest of the foreign key constraints are generated. Result of this process is a sequence of SQL clauses written into the file system as a DDL (Data Definition Language) text file.

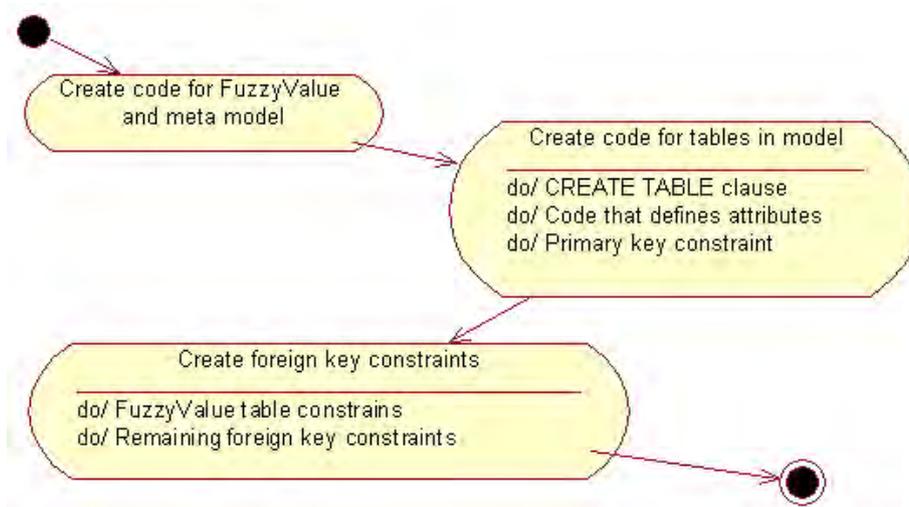


Fig. 4. The SQL DDL code generation process.

5. Motivation and Comparison

Here we give a comparison of our data model to the most advanced fuzzy relational data model available today - the FIRST-2 [6]. Our conclusion is that there are several similarities between them. Although the methods for fuzzy value representation are completely different, functionally, our model is a subset of the FIRST-2 model. Our intention was to define the simplest possible model that supports most widely used fuzzy concepts, and stores values as effectively as possible without too much overhead.

Fuzzy attributes of the type 1 in the FIRST-2 model are crisp values that our model also supports. Fuzzy types that our model covers are a subset of those represented by the fuzzy attributes type 2 and 3. Null values, intervals and trapezoidal fuzzy numbers in the FIRST-2 are represented by the structures that have these same names. A subset of the set of triangular fuzzy numbers, isosceles triangle, is represented by the approximate value with explicit margin in the FIRST-2 model. All other types of triangular fuzzy numbers, as well as fuzzy quantities can be represented by the possibility distributions with 2 and with 4 values in the FIRST-2, although these distribution types are more general.

The general fuzzy number from our model is known as the fuzzy attribute type 3 in the FIRST-2 model. Moreover, the FIRST-2 model describes a wider range of other possibilities for fuzzy values and combines atomic values according to their respective structure. In this paper we described an advanced version of our model that treats fuzzy values similarly. Although, functionally, our model is a subset of the FIRST-2, it gives theoretical

contribution in modelling from the aspect of relational model theory because it conforms to the 3rd normal form. The basic disadvantage of the FIRST-2 model is non conformance even to the 1st normal form.

The fuzzy database query language FSQL is built on top of the FIRST-2 model using Oracle DBMS and PL/SQL stored procedures [23]. Similarly, we used the fuzzy-relational data model described in this paper to build an interpreter for the PFSQL language. We have developed the PFSQL query language from ground up, extending the features of SQL into the fuzzy domain. The PFSQL language is an extension of the SQL language that allows fuzzy logic concepts to be used in queries. Among other features described in [27,31,32] in detail, this query language allows priority statements to be specified for query conditions. For calculating the membership degree of query tuples when priority is assigned to conditions, we use the GPF CSP systems mentioned in introduction. Although the FSQL language has more features than the PFSQL, it does not allow usage of priority statements. The PFSQL is the first query language that does. Moreover, the PFSQL is implemented using Java, outside the database, which makes our implementation database independent.

Following this idea, we implemented the CASE tool described here in order to ease the fuzzy-relational model development and its usage in the real world applications with the PFSQL.

6. Conclusion

In this paper we present an innovative fuzzy-relational data model and a unique CASE tool for FRDB model development. Presented data model extends the relational model with capabilities to store fuzzy values and supports the execution of PFSQL queries. The CASE tool for fuzzy relational model development has been implemented using the Java programming language and the Swing platform for the graphic user interface. To the best of our knowledge, this is the only CASE tool with such capabilities in existence today.

In addition, we give a comparison between this model and the more general FIRST-2 model. It is our conclusion that this model represents a significant improvement compared to our previous model given in [28]. On the other hand, it is a functional subset of the FIRST-2 model, although the methods for the fuzzy value representation are completely different. Its compliance to the 3rd normal form makes it better theoretically founded than other known models of this kind.

In an effort to ease the PFSQL usage further, we implemented a fuzzy JDBC driver [24,27,32] that allows easy PFSQL statement execution within the Java environment. This set of tools supports the idea to specify the complete methodology for fuzzy-relational database applications development. A working version of our model, the CASE tool and the PFSQL is available for download from <http://www.is.pmf.uns.ac.rs/fuzzydb>.

References

1. Bosc, P., Kraft, D., Petry, F.: Fuzzy Sets in Database and Information Systems Status and Opportunities. *Fuzzy Sets and Systems*, Vol. 156, No. 3, 418–426. (2005)
2. Ma, Z.M., Mili, F.: Handling Fuzzy Information in Extended Possibility-Based Fuzzy Relational Databases. *International Journal of Intelligent Systems*, Vol. 17, No. 10, 925–942. (2002)
3. Prade, H., Testemale, C.: Fuzzy Relational Databases: Representational Issues and Reduction Using Similarity Measures. *Journal of the American Society for Information Science*, Vol. 38, No. 2, 118-126. (1987)
4. Buckles, B., Petry, F.: A Fuzzy Representation of Data for Relational Databases. *Fuzzy Sets and Systems*, Vol. 7, No. 3, 213-226. (1982)
5. Zvieli, A., Chen, P.: ER Modelling and Fuzzy Databases. In *Proceedings of the 2nd International Conference on Data Engineering*, Los Angeles, CA, 320-327. (1986)
6. Bahar, O., Yazici, A.: Normalization and Lossless Join Decomposition of Similarity-Based Fuzzy Relational Databases. *International Journal of Intelligent Systems*, Vol. 19, No 10, 885 – 917. (2004)
7. Raju, K., Majumdar, A.: Fuzzy functional Dependencies and Lossless Join Decomposition of Fuzzy Relational Database Systems. *ACM Transactions on Database Systems*, Vol. 13, No. 2, 129-166. (1988)
8. Umamo, M.: Freedom-O: A Fuzzy Database System. *Fuzzy Information and Decision Processes*. In: Gupta, Sanchez (eds.): *Fuzzy Information and Decision Processes*, North-Holland, Amsterdam. (1982)
9. Kacprzyk, J., Zadrozny, S., Ziolkowski, A.: Fquery III+: A “Human-Consistent” Database Querying System Based on Fuzzy Logic with Linguistic Quantifier. *Information Systems*, Vol. 14, No. 6, 443–453. (1989)
10. Sanchez, E.: Importance in Knowledge Systems. *Information Systems*, Vol. 14, No. 6, 455–464. (1989)
11. Vandenberghe, R., Schooten, A.V., Caluwe, R. D., Kerre, E.: Some Practical Aspects of Fuzzy Database Techniques: An Example. *Information Systems*, Vol. 14, No. 6, 465–472. (1989)
12. Wong, M., Leung, K.: A Fuzzy Database-Query Language. *Information Systems*, Vol. 15, No. 5, 583–590. (1990)
13. Bosc, P., Pivert, O., Farquhar, K.: Integrating Fuzzy Queries into an Existing Database Management System: An Example. *International Journal of Intelligent Systems*, Vol. 9, No. 5, 475–492. (1994)
14. Bosc, P., Pivert, O.: SQLf: A Relational Database Language for Fuzzy Querying. *IEEE Transactions on Fuzzy Systems*, Vol. 3, No. 1, 1-17. (1995)
15. Medina, J.M., Pons, O., Vila, M.A.: GEFRED: A Generalized Model of Fuzzy Relational Databases. *Information Sciences*, Vol. 76, No. 1-2, 87-109. (1994)
16. Galindo, J., Medina, J.M., Aranda, M.C.: Querying Fuzzy Relational Databases Through Fuzzy Domain Calculus. *International Journal of Intelligent Systems*, Vol. 14, No. 4, 375-411. (1999)
17. Galindo, J., Medina, J.M., Aranda-Garrido, M.C.: Fuzzy Division in Fuzzy Relational Databases: an Approach. *Fuzzy Sets and Systems*, Vol. 121 No. 3, 471–490. (2001)
18. Galindo, J., Medina, J.M., Cubero, J.C., Garcia, M.C.: Relaxing the Universal Quantifier of the Division in Fuzzy Relational Databases. *International Journal of Intelligent Systems*, Vol. 16, No. 6, 713-742. (2001)

19. Galindo, J., Medina, J.M., Pons, O., Cubero, J.C.: A Server for Fuzzy SQL Queries. *Lecture Notes in Computer Science*, Vol. 1495, 164-174. (1998)
20. Chen, G., Kerre, E.: Extending ER/EER Concepts Towards Fuzzy Conceptual Data Modelling. In *Proceedings of the IEEE International Conference on Fuzzy Systems*. Anchorage, AK, 1320-1325. (1998)
21. Kerre, E., Chen, G.: Fuzzy Data Modelling at a Conceptual Level: Extending ER/EER Concepts. In: O. Pons (ed.): *Knowledge Management in Fuzzy Databases*, Physica Verlag, Heidelberg, 3-11. (2000)
22. Chaudhry, N., Moyne, J., Rundensteiner, E.: A Design Methodolgy for Databases with Uncertain Data. In *Proceedings of the 7th International Working Conference on Scientific and Statistical Database Management*. Charlottesville, VA, 32-41. (1994)
23. Galindo, J., Urrutia, A., Piattini, M.: *Fuzzy Databases: Modeling Design and Implementation*. IDEA Group Publishing, Hershey, USA. (2006)
24. Škrbić, S., Takači, A.: On Development of Fuzzy Relational Database Applications. In *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Malaga, Spain, 268-273. (2008)
25. Takači, A., Škrbić, S.: How to Implement FSQL and Priority Queries. In *Proceedings of the 3rd Serbian-Hungarian Joint Symposium on Intelligent Systems*, Subotica, Serbia, 261-267. (2005)
26. Takači, A.: Towards a Priority Based Logic. In *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty*, Paris, France, 247-252. (2006)
27. Škrbić, S.: *Fuzzy Logic Usage in Relational Databases*. Doctoral dissertation, Faculty of Science, Novi Sad. (2009) (in Serbian)
28. Takači, A., Škrbić, S.: Data Model of FRDB with Different Data Types and PFSQL. In: Galindo, J. (ed.): *Handbook of Research on Fuzzy Information Processing in Databases*, IGI Global, Hershey, PA, 403-430. (2008)
29. Luo, X., Lee, J., Leung, H., Jennings, N.: Prioritized Fuzzy Constraint Satisfaction Problems: Axioms, Instantiation and Validation. *Fuzzy Sets and Systems*, Vol. 136, No. 2, 151-188. (2003)
30. Takači, A.: Schur-concave Triangular Norms: Characterization and Application in PFCSP. *Fuzzy Sets and Systems*, Vol. 155, No. 1, 50-64. (2005)
31. Škrbić, S., Racković, M.: PFSQL: a Fuzzy SQL Language With Priorities. In *Proceedings of the 4th International Conference on Engineering Technologies*, Novi Sad, Serbia, 58-63. (2009)
32. Škrbić, S., Takači, A.: An Interpreter for Priority Fuzzy Logic Enriched SQL. In *Proceedings of the 4th Balkan Conference in Informatics*, Thessaloniki, Greece, 96-100. (2009)
33. Takači, A., Škrbić, S.: Measuring the Similarity of Different Types of Fuzzy Sets in FRDB. In *Proceedings of the 5th Conference of the EUSFLAT*, Ostrava, Czech Republic, 247-252. (2007)
34. Bernstein, P.A.: Synthesizing Third Normal Form Relations From Functional Dependencies. *ACM Transactions on Database Systems*, Vol. 1, No. 4, 277-298. (1976)

Srđan Škrbić, Miloš Racković, and Aleksandar Takači

Srđan Škrbić received the PhD degree in 2009 from the Department of mathematics and informatics, Faculty of Science at University of Novi Sad. He is currently an assistant professor at the same department. His research interests include intelligent database systems and fuzzy database modeling. He has published more than 30 scientific papers.

Miloš Racković is a full professor at the Department of mathematics and informatics, Faculty of Science at University of Novi Sad. He received the PhD from the same department in 1996. He works in the fields of computer science, artificial intelligence and databases. Professor Racković has published 82 scientific papers, two monographs and three books. He mentored more than 200 Bsc theses, 20 Msc theses and two PhD theses.

Aleksandar Takači, received the PhD degree in 2006 from the Department of mathematics and informatics, Faculty of Science at University of Novi Sad. He is currently an assistant professor at the Faculty of Technology, University of Novi Sad. His research interests include fuzzy systems and fuzzy logic. He has published more than 30 scientific papers.

Received: January 02, 2010; Accepted: April 13, 2010.

Design Pattern Instantiation Directed by Concretization and Specialization

Peter Kajsa¹, Lubomir Majtas¹, and Pavol Navrat¹

¹ Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovičova 3, Bratislava, 842 16, Slovakia
{kajsa, majtas, navrat}@fiit.stuba.sk

Abstract. Design patterns provide an especially effective way to improve the quality of a software system design as they provide abstracted, generalized and verified solutions of non-trivial design problems that occur repeatedly. The paper presents a method of design pattern instantiation support based on the key principles of both MDD and MDA. The method allows specification of the pattern instance occurrence via the semantic extension of UML directly on the context. The rest of the pattern instantiation is automated by model transformations of the specified pattern instances to lower levels of abstraction. Such approach enables the use of higher levels of abstraction in the modeling of patterns. Moreover, the model transformations are driven by models of patterns besides the instance specification, and thus the approach provides very useful ways how to determine and control the results of transformations. The method is not limited to design pattern support only, it also provides a framework for the addition of support for custom model structures which are often created in models mechanically.

Keywords: Design patterns, concretization, specialization, MDD.

1. Introduction

There are many efforts to improve the quality of software system development or maintenance based on identification, acquisition and application of some kind of architectural knowledge [20]. In general, patterns are based on abstractions and generalizations of effective, reliable and robust solutions to recurring problems. Patterns provide abstracted, generalized and verified solutions of non-trivial problems. The concept of patterns was first introduced in the work of Alexander [17] dealing with urban solutions, but soon patterns were also defined and used in software engineering. The idea of applying verified pattern solutions to common recurring problems in the software design attracted considerable attention very quickly (cf. [4] and consequently, e.g. [5]), since the quality of software systems depends greatly on the design solutions chosen by developers.

Patterns have been applied in various phases of the software development lifecycle. Patterns were discovered and defined in software analysis, design, integration, testing and other areas. Currently, design patterns represent an important tool for developers in the process of software design construction, and provide particularly effective ways to improve the quality of software systems. It is evident that design patterns are not the solution to all problems related to software development. Some have noted their limitations and propose new approaches to the knowledge representation in the software development domain, even proposing language architectures [21]. However, it is well known that the application of design patterns in software projects assists in the creation of modifiable, recursive and extensible software design [4]. CASE or other modeling tools provide nowadays some kind of support for design pattern instantiation, but it is often based on simple copying of pattern template into the model with minimal possibilities for modification and with minimal support of instance integration into the context – application model. A more systematic approach to pattern instantiation in interaction between software designer and a supporting tool has been presented in [22].

Since patterns provide abstracted and generalized solutions to recurring problems, their application to a specific problem requires to concretize and to specialize the solution described by the pattern [5] (see Fig. 1).

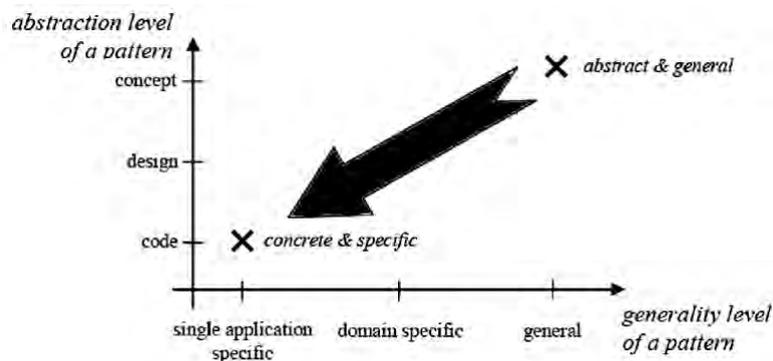


Fig. 1. Concretization and specialization of the solution described by the pattern, when the pattern is applied to a concrete and specific problem [5]

Specialization process of a design pattern typically lies in its integration into the specific context of the problem. The knowledge is mainly available to developers and domain experts involved in the design process, because it requires very specialized and detailed understanding of the domain context and the specific application itself. This is why this process is difficult to automate. Despite this, it is possible to make specializing of a pattern much easier by providing an appropriate mechanism for supporting application of design patterns.

The goal of concretization of a design pattern is to recast its abstract form into a concrete realization with all its parts, methods, attributes and

associations, but only within the scope of the pattern instance and its participants, not the rest of the application model. The more parts the structure of the pattern instance contains, the more concrete it becomes. The most concrete level of a design pattern instance is the source code, because at this level of abstraction the pattern instance contains all parts of its structure. Majority of activities in the concretization process depends on a stable and fixed definition of the design pattern structure so that these activities are fairly routine. This is a good starting point for the automation of this process.

Consequently, we see a fairly good basis for the development of a method that would describe the way how to apply a design pattern based on supporting explicitly its specialization and concretization. We aim at proposing a method that would involve a specially devised tool supporting these two principal lines of design pattern instantiation.

Section 2 introduces several known approaches to tool based design pattern support and section 3 infers the open problems in this area. In the rest of the paper, we focus on the elaborated method of design pattern instantiation. Section 4 presents the theory about the method and it provides the method description. In the following section 5 the article covers particular aspects of method realization. Section 6 contains case study and the method evaluation. The paper is completed by a proposal of future works.

2. State of the Art

There exist several approaches which introduce their own tool-based support for pattern instantiation.

Mapelsden et al. [15] introduce an approach to design pattern application based on the Design Pattern Modelling Language. The authors describe this language which is a notation for the specification of solutions of design patterns and their instantiation into UML models. Design pattern instances are regarded as a part of the object model, providing another construct that can be used in the description of a program. Once all design pattern instance elements are linked to one or more UML design elements, the consistency checks are made. A deficiency of this approach is that the developer needs to model all pattern participants manually and then link these parts into the pattern model. El Boussaidi et al. [11] present model transformations based on the Eclipse EMF and JRule frameworks. Wang et al. [12] provide similar functionality by XSLT-based transformations of models stored in XMI-Light format. Both approaches can be considered as driven by a single template and they focus mostly on the transformation process and do not set space for pattern customization.

Another method was introduced by Ó Cinnéide et al. [13]. They present a method for the creation of behavior-preserving design pattern transformations and apply this method to GoF design patterns. The method involves a refactoring process which provides descriptions of transformations

to modify the spots for pattern instance placement (so called precursors). The placement is achieved by the application of so called 'micropatterns' to the final pattern instances. While Ó Cinnéide's approach is supposed to guide the developers pattern placement in the phase of refactoring (based on source code analysis), Briand et al. [8] try to identify the spots for pattern instances in the design phase (based on UML model analysis). They provide a semi-automatic suggestion mechanism based on a decision tree combining an evaluation of the automatic detection rules with user queries.

All the former approaches focus on the creation of pattern instances. The ones presented by Dong et al. [9, 10] presume the presence of pattern instances in the model. They provide support for evolution of the existing pattern instances resulting from application changes. In the former [9], the implementation employs QVT based model transformations, and in the latter [10] the same is achieved by XSLT transformations over the model stored as XML. However, both work with a single configuration pattern template allowing only changes in the presence of hot spots participants. Other possible variations are omitted.

Debnath et al. [14] propose a level architecture of UML profiles for design patterns. Authors introduce a profile for patterns and analyze the advantages of using profiles to define, document, and visualize the design. Authors provide a guide to the creation of UML Profiles, but they give no concrete way of providing support in any tool. Dong et al. [16] discuss some of the relevant aspects of the UML profile. The paper presents an approach to the creation of UML profiles for design patterns. The approach allows an explicit representation of patterns in software designs and introduces a notation for the names of stereotypes: `Type<name:String [instance:integer], role:String>`; for example: `PatternClass<Observer[1], ConcreteObserver>`. The introduced notation is useful because it visualizes individual instances of design patterns, but the `Type` part of the notation is redundant, because the stereotype definition itself already carries the information.

3. Open Problems

The approaches that focus on the creation of pattern instances are typically based on the strict forward participant generation - participants in all roles are created according to a single template. Similarly, the support of design patterns available in traditional CASE or other modeling tools is usually based on UML templates of each design pattern. They are simply copied into the model with a minimal possibility for modification and integration in the rest of the model when pattern instance is created [1], [2]. However, patterns describe not only the main solution, but also many alternative solutions and variations. However, a developer is not allowed to choose an appropriate variant or a concrete structure of the design pattern. Only one generic form is offered to the developer for use. Any other adjustments need to be performed manually without any tool based support. Further, by the generation of source

code from a model with applied pattern instances, only class structure is generated, and the bodies of the methods of the patterns participants are empty. Consequently, the support of concretization has great deficiencies.

Moreover, the instance of a pattern created by a tool is typically without any connection to the rest of the application model. So the instance of a pattern has not been integrated into the application model, i.e. the context. It lacks associations and the names of pattern participants are general, and so on. All these activities of instance specialization have to be done by the developer manually. Even in the approach presented in [15], the developer needs to model all pattern participants manually, and then to link these parts to the pattern model.

Our intention is to automate these activities. Our vision is that the developer simply specifies a pattern instance occurrence directly in the context, and the rest of the pattern structure is then automatically generated into the application model in an appropriate form.

4. Method Description

Our idea emphasizes collaboration between the developer and the CASE tool. We assume that we do not need to force the developer to explicitly model or mark all the pattern participants. Our aim is to encourage him/her just to suggest the pattern instance occurrence while the rest of the instantiation process is automated.

Patterns are often described as a collection of cooperating roles. Our approach is based on the idea [19] that the pattern roles can be divided into roles dealing with the domain of the created software system and roles performing the pattern's infrastructure. The domain roles can be considered as the "hot spots" while they can be modified, added or deleted according to the requirements of the particular software environment. The roles performing the pattern infrastructure are not changing too much between the pattern instances. Their purpose is to glue the domain roles together to be able to perform desired common functionality. Examples of domain dependent roles are presented in the Table 1.

The employment of patterns into the project allows the developer to think at a higher level of abstraction. When he decides to employ a pattern, the first thing he needs to take care of is how it will be connected to his project, how the solution will be integrated to the rest of his model / code. At this moment the developer does not focus on the entire pattern's inner structure, because it is irrelevant to him at this moment. The way how he integrates the pattern to the project lies in the specification of the domain roles. Their participants can be existing parts of the project or new ones created for this situation. Once the domain roles are specified, the specification of the infrastructure roles takes place. This is quite a routine, when the developer subsequently adds participants of the infrastructure roles according to the sample instance from the pattern catalogue.

Table 1. Examples of domain dependent roles of patterns [19]

Pattern	Domain dependent roles	Description
Composite	Leaf and its Operations	Leaves and their operations provide all domain dependent functionality. Everything else is just the infrastructure allowing the hierarchical access to the leaf instances.
Flyweight	Concrete Flyweight	Concrete Flyweight provides all domain dependent functionality. The rest is infrastructure for storing instances in memory providing access to them.

When we look closer at such instantiation process from the perspective of its division into two more or less independent processes of specialization and concretization (described in the section 1 Introduction) [5], we can see that the user does the specialization process when he is specifying the domain roles. When he is supplementing pattern instance with the infrastructure roles he just completes the concretization process.

In our approach we do not want to replace the developer in the specialization process, but we want to relieve him of the necessity to instantiate the infrastructure roles during the concretization process. We want the developer to make a suggestion by the application of semantics as to where and which design pattern he wishes to be applied in the model and to specify the domain dependent roles. Then he can also specify which variant of the pattern to employ, and in what way he wants it to be generated. Subsequently, the rest of the pattern instance structure will be automatically generated by model transformations to lower levels of abstraction according to the instance specification.

In order to achieve the specified goal, it is necessary to provide an appropriate mechanism of pattern semantics in the application model. It is important to support insertion of semantics directly into the elements of the model, because such approach supports the specialization of pattern instances, and makes the creation of the instance specification effortless. Thanks to the semantics, the model transformations are able to understand the model of the application and recognize its parts.

In case the transformations are driven by an appropriate model of design pattern, and both the model of an application and the model of the pattern contain information on semantics, the transformation is capable to compare these models and to create mappings between them. So in this way the transformation can recognize participants of design patterns that are present in the application model already, and which are not. As a consequence, the transformation is able to generate missing participants in the desired form obtained from the pattern model.

We note that model transformations automate the concretization process. They are driven by pattern instance suggestion and specification and by the pattern model as well. Such transformations have several capabilities. Firstly,

they provide a possibility to choose an appropriate configuration of the pattern by instance specification. Secondly, they enable the modeling of a custom pattern or structure by modification of the pattern model, and this way to achieve its generation into the model.

Moreover, our method assumes that the models and the transformations are split into more levels of abstraction in accord with the ideas of the MDA development process. These levels support work with instances of design patterns at various levels of abstraction. This process is shown in Fig. 2.

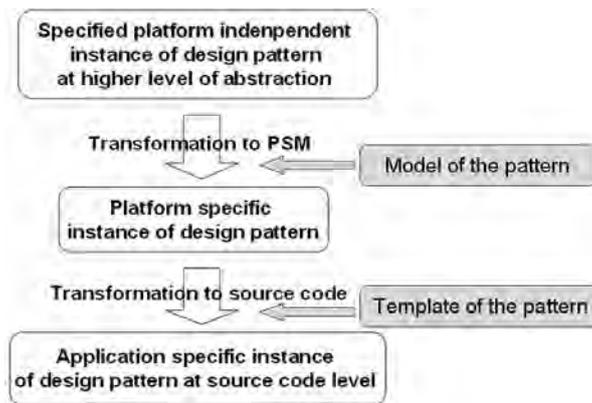


Fig. 2. Proposal of design pattern instantiation process

One of the main objectives of the approach is to consider ideas of model driven, iterative, and incremental development of software systems. It is important to note why the transformation to platform specific models (PSM) is necessary. It is at this level that the first differences in structure between instances of design patterns may occur. For example, some platforms allow multiple inheritance, others provide interfaces, etc.

5. Method Realization

The following subsections explain particular aspects of the method realization.

5.1. Realization of Pattern Instance Suggestion and Specification

The suggestion and the specification of pattern instance are realized by applying information on the semantics into the models provided by semantical extension of UML. We choose the semantical extension of UML in a form of UML profile as a standard extension of UML, since one of our goals

is to remain compliant with the majority of other UML tools. UML profiles provide a standard way to extend the UML semantics in the form of definitions of stereotypes, tagged values - meta-attributes of stereotypes, enumeration and constraints. All these can be applied directly to specific model elements such as Classes, Attributes, and Operations [6]. This way it is possible to specify participants of design patterns and relations between them directly in the context of the elements of the application model (for more details about the UML profile please see the section 5.3).

For example, Fig. 3 shows a suggestion of the Observer pattern instance via applying one stereotype <<Observes>> to a desired element, in this case, an association. From the information the transformation can recognize that the source element of the association represents a Concrete Observer and the destination element is a Concrete Subject. Consequently, on the basis of the information and the available pattern model and semantics, the transformation can recognize the other pattern participants need to be added to the model.

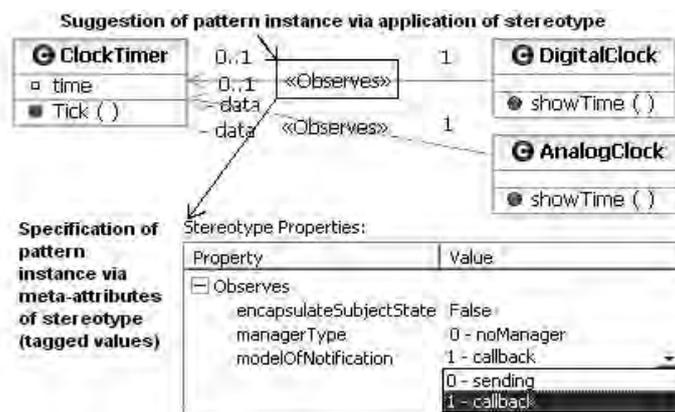


Fig. 3. Example of an application of the Observer pattern to a model. It represents a specified platform independent instance and thus the most abstract form of the Observer pattern instance

The transformation also needs information about how to generate the rest of pattern instance, e.g. variant of pattern, desired adjustments of pattern instance, and so on. The next step is the specification of pattern instance. This goal is achieved by setting up values of meta-attributes of stereotype (Fig. 3). In our approach this step is not mandatory, because default values of meta-attributes of the stereotype are set and are available. Consequently, the application of the desired pattern can consist only of applying one suggestion mark – the stereotype onto the specified model element, when the developer wants the default variant of the pattern. Any other activities will be completed by a tool via model transformations. In this phase, developers do not have to concern themselves with the concrete details of the pattern structure, and

they can comfortably work with the pattern instances at a higher level of abstraction. The Application of the desired pattern is realized on elements of the system model or context, and thus the specialization process is supported.

5.2. Realization of Concretization Process

The concretization process is realized and automated by model transformations to lower levels of abstraction until the source code level is reached. One of the possible results of the transformation of the model from Fig. 3 is shown in Fig. 4. As it can be seen the transformation generates the rest of pattern structure in a desired form in accord with pattern suggestion and specification from Fig. 3. The pattern instance becomes more concrete, so the form of the instance now represents its lower abstraction level. Thanks to the realization of the pattern instance by placing the suggestion and specification directly into the context of elements in the application model, the transformation is also able to integrate the generated participants with participants already present in the model. As a result, the pattern instance is in the application specific form.

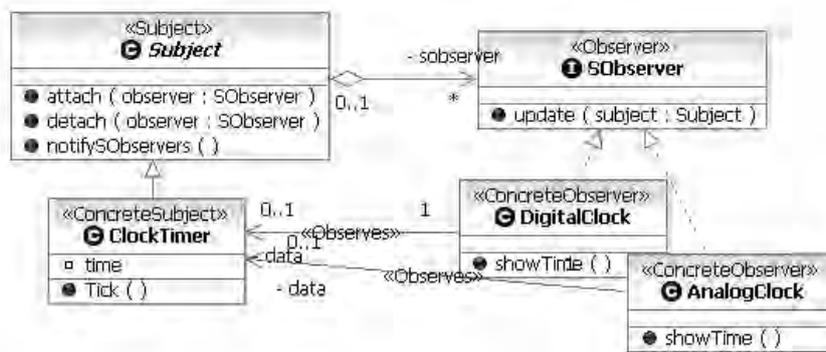


Fig. 4. The result of the transformation to Java target platform of the model from Fig. 3 in accord with the instance suggestion and specification

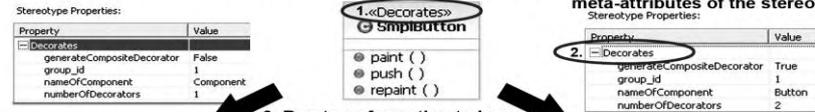
It is important that the transformation is realized and launched with a choice of target platform because, as mentioned earlier, at this point the first differences may occur in the structure of patterns depending on target platform. The choice of a target platform also determines the set of possible choices of data types before subsequent transformation to source code level.

As one can see in Fig. 4, the transformation also adds explicit marks (stereotypes) to all identified and generated pattern participants. The addition of marks and also the whole transformation is performed on the basis of the pattern model (more in Section 5.4). As a consequence, the instance is

clearly visible, and the developer can repeat the instantiation process at a lower level (PSM) directly from the optional second step, i.e. by specifying the instance and choosing a more detailed adjustments of pattern instance (e.g. concrete data types). Again, the default values of the stereotype meta-attributes are set, so the developer can run the transformation to source code directly.

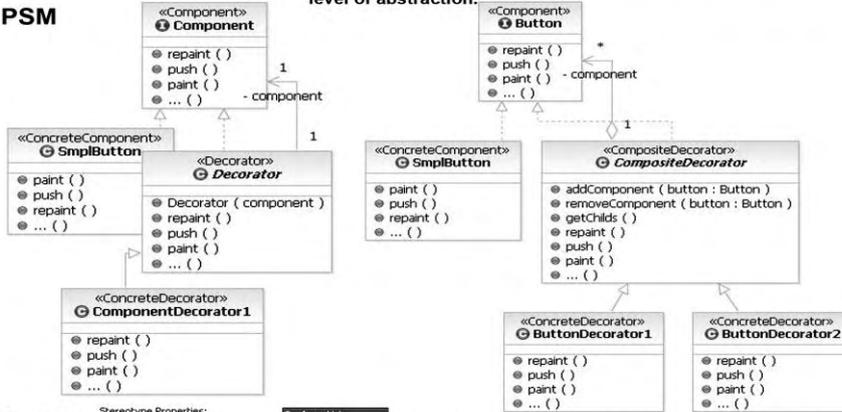
PIM

- 0. Decision to apply the Decorator pattern
- 1. Suggestion of the pattern instance - applying of the stereotype term instance - setting up appropriate meta-attributes of the stereotype
- 2. (Optional) Specification of the pattern instance - setting up appropriate meta-attributes of the stereotype



3. Run transformation to lower level of abstraction.

PSM



Source code

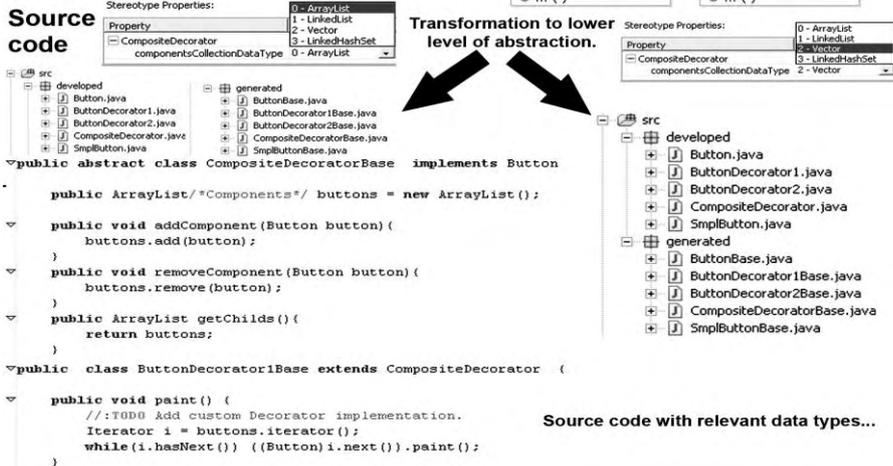


Fig. 5. An overall illustration of the pattern instantiation process

Two separate groups of classes are generated by the initial transformation to source code. The first is the base group which is always overwritten by subsequent source code generation. The second is the development group which is generated only by initial transformation. The developer can write and add a specific implementation here without the threat of it being overwritten.

Overall illustration of the described pattern instantiation process is shown in Fig. 5 using as an example a Decorator pattern application.

This way, our approach has achieved support for working with pattern instances at three different levels of abstraction:

- Pattern suggestion and specification level – PIM
- Design model level – PSM
- Source code level

5.3. Realization of UML Profile for Design Patterns

UML profiles provide a suitable way to define semantics for each design pattern and allow applying of semantics directly onto the elements of model. Consequently, a UML profile allows specification of participants of design patterns, and relations between them, directly on the elements of application model. The snippet of UML profile for Observer pattern is shown in Fig. 6.

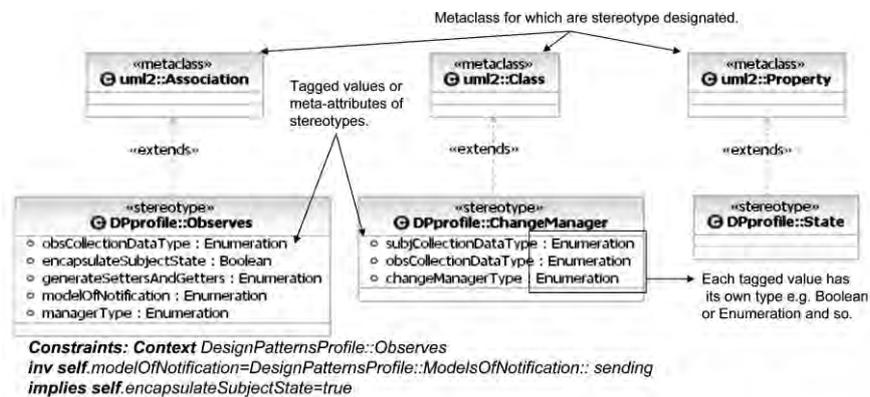


Fig. 6. The snippet of UML profile with some elements for Observer pattern

Authored UML profile provides semantics to various pattern instances adjustments, suggestions and specifications. However, it is not mandatory to apply all the semantics elements (stereotypes). The developer applies and specifies only what he needs to express. On the basis of applied semantics and pattern models with semantics, the transformation generates elements that are missing (more in the next Section 5.4). Because of the default values of meta-attributes of stereotypes, the transformation always has enough information for default behavior. Inconsistent specifications of pattern

instances are handled by OCL constraints which are part of UML profile as well (for example see Fig. 6).

Semantics of patterns is defined in one common UML profile for all supported patterns. However, the semantics of patterns from UML profile is not generalized for all patterns or structures. It contains semantics specific for patterns which are supported and in the consequence, when a developer wants to support new pattern or structure, he needs to add a semantics specific for this new pattern or structure into the profile (for more details see section 5.7 Extending of Support for New Patterns or Structures). It is important to remark, that it is not quite possible to create a profile generalized for all patterns or structures, because each pattern has its own semantics, purpose, variations and so on. Moreover, exactly our goal is to allow the developer to suggest and specify his intentions and design decisions in a specific way via semantics specific for the applied pattern. In case that the semantics applied by developer would be general for all patterns, intentions or decisions, we would not be capable of deducing some required specific information from such general semantics.

We tried to name the stereotypes according to the established names of pattern participants. However, a developer can change these names in the UML profile, but he must, of course, update also the pattern model.

Authored prototype of UML profile with description can be found in [26].

5.4. Realization of Transformations

Transformations performed by the tool are driven by properly specified and marked models of design patterns. These prepared models cover all supported pattern variants and possible modifications. Each element of these models is marked. There are two types of marks in pattern models. The first type of mark expresses the role of the element in the scope of the pattern. On the basis of this type of mark the tool is capable of creating mappings between models. The second type of mark expresses an association of the element with a variant of the pattern. On the basis of this type of mark the tool is capable of deciding which element should be generated into the model, which way and in what form. For the second type of mark the following notation is defined:

```
[~]?StereotypeName::Meta-attributeName::value;
```

An element from the pattern model is generated into the model only if the specified meta-attribute of the specified stereotype has the specified value. These marks can be joined via “;”, while the symbol “~” expresses negation. If an element has no mark, it is always generated into the model. A sample section of the model of the Observer pattern is exposed in the Fig. 7.

Design Pattern Instantiation Directed by Concretization and Specialization

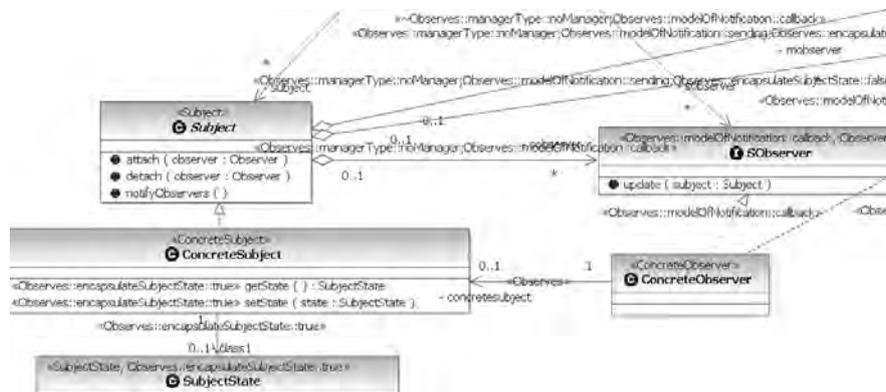


Fig. 7. Sample section of Observer pattern model by which the transformation is driven

The whole algorithm of the transformation is captured in the following Fig. 8.

The first action performed by the tool after the start of the transformation is the comparison of the first type marks in pattern model to the marks in the application model. When an instance of pattern is processed, only the marks with identical value of `group_id` are taken into consideration (for example, see `<<Decorates>>` stereotype in the Fig. 5 or case study in the Fig. 19 - 21). When the mark is without `group_id`, each next occurrence of the mark with the same name is considered as another instance participant. For example, stereotype `<<Observers>>` does not have `group_id` meta-attribute and therefore, when the tool processes one of such marks the others are considered as other instances (for example, see example of Observer instantiation in the section 5.5).

Based on the first type marks comparison the tool is capable of making a mapping between the marked models, and consequently to recognize which parts of the structure of the design pattern instance are in the model of the developing application and which are not. For example, in Fig. 3 in the previous section we have shown the application of the Observer pattern by applying one stereotype `<<Observes>>` on the directed association. From so marked association the tool can recognize that the parts Concrete Observer and Concrete Subject of this Observer pattern instance are present in the model already, and also which elements (in this case classes) in the application model represent these roles or parts.

Decisions about which variant of pattern and which elements from the pattern model need to be generated into the application model are based on the comparison of the second type marks in the pattern model with the values of the meta-attributes of stereotypes. These values are set up by the developer in the second step - specification of the pattern instance (see Sections 5.1 and 5.2 and Fig. 3 and 5).

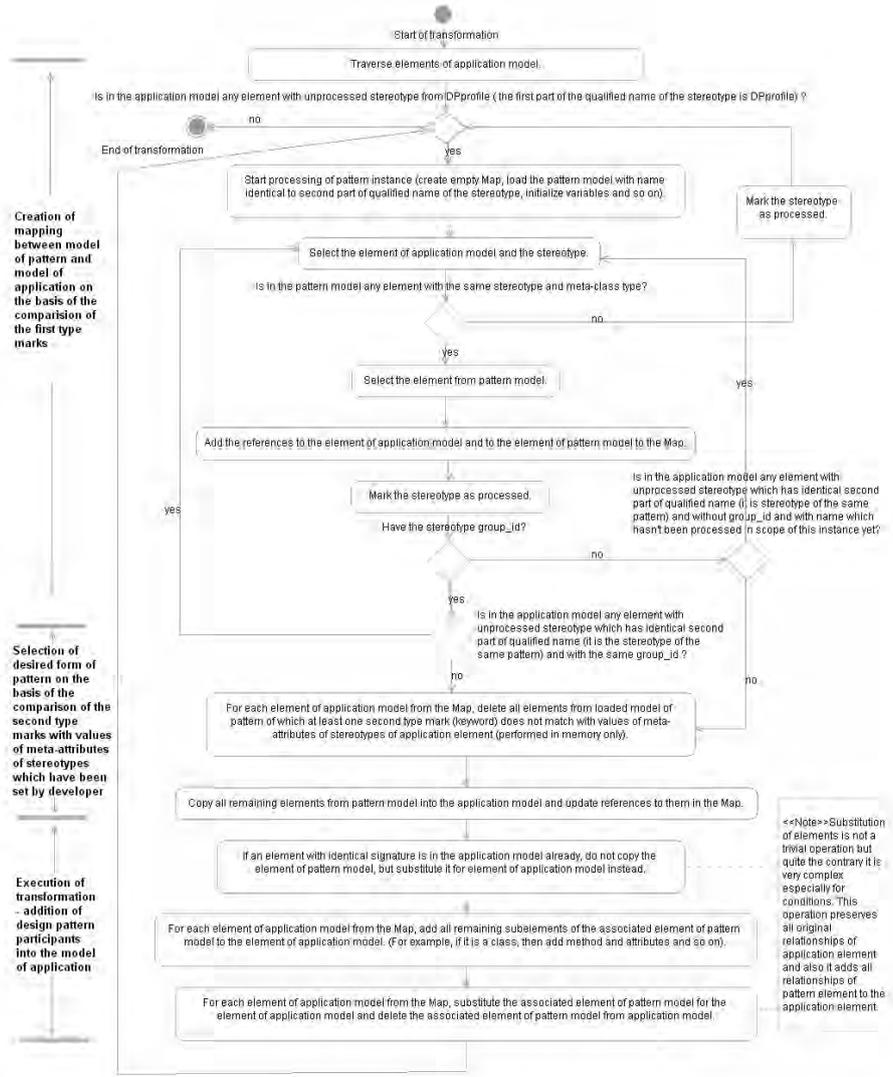


Fig. 8. Principles of tool functioning - tool under the hood.

After decision-making and selection of the desired pattern form, the alone transformation is performed. The results of the transformation are correctly specialized and concrete instances of the patterns created in the desired form, as presented in Fig. 4 and 5 in the previous sub-section.

Driving the model transformations by pattern models allows us to adjust results of transformations by modifying of the pattern models. Marks in the models ensure that the tool is always capable of creating correct mappings between the model of application and the model which drives the

transformation, and consequently decide which element should be generated into the model and in what form. This way it is possible to model any custom structure and achieve support for its application into the model.

The transformation to source code is realized on the basis of the code templates for now. Each pattern participant has own code template. The transformation takes code template with name identical to the stereotype name of the participant and it generates template's content into specified destination. For model elements without any stereotype the common code template is used which generates only signatures of the class, fields and methods with empty body. The inconsistent states, such as duplicity of classes, illegal inheritance and others, are handled by the first transformation of the model of highest level of abstraction to the model of lower level of abstraction. The rules of correcting of such inconsistent states are common for all possible patterns or structures and therefore they are hard coded in the transformation algorithm. The transformation of the model to source code simply generates source code of each element from the model. An example of snippet of `Subject` code template is shown in the following Fig. 9.

The transformation to source code is still under our research. For more details see the section 7 Future Work. We have proposed the improvement of this transformation already.

```

/* ***** These file will be always overwritten!
 */
public <%= if(isAbstract){ %>abstract<%= %> class <%=className%> <%=extend%> <%=implement%> {

    /*generated fields*/

    <%= boolean getsubjectstate=true:
    int pocet = 0;
    for(int i=0; i<attributes.size(); i++){
    Property p=(Property) attributes.get(i);
    if(p.getName().startsWith("soobserver")){
    String typeOfState = subjectStates.size()>pocet ? (String) subjectStates.get(pocet) : "Object";
    pocet++;
    %>

    <%= JavaNameTool.getPropertySignature(p, collType) %>

    public void attach(<%=p.getType().getName()%> observer) {
    <%=p.getName()%>.add(observer);
    }

    public void detach(<%=p.getType().getName()%> observer) {
    <%=p.getName()%>.remove(observer);
    }

    public void notify<%=p.getType().getName()%>() {
    Iterator i = <%=p.getName()%>.iterator();
    while(i.hasNext()){
    <%=p.getType().getName()%> o=(<%=p.getType().getName()%>) i.next();
    o.update(<%=typeOfState%>) this.getSubjectstate();
    }
    }
}

```

Fig. 9. Snippet of code template of Subject participant of Observer pattern

5.5. Detailed View on the Method and the Tool in Action

This section provides illustration of the method and the tool, its functionality and usage by means of an example. The following Fig. 10 shows example of initial form of UML model before application of patterns.

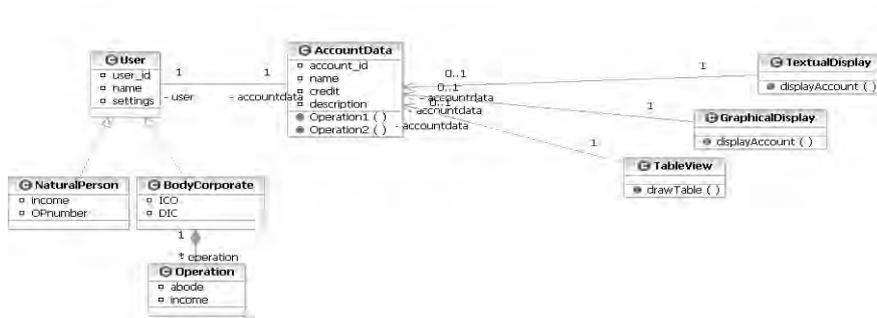


Fig. 10. Example of starting UML model before the application of patterns

The model represents an example of starting point of model into which the developer intends to apply, for example, Observer pattern now. In order to apply the desired pattern (in this case Observer) the developer suggests the instance occurrences via particular semantics marks – stereotypes (in this case stereotype `<<Observes>>`). Notice that the developer performs the suggestion of pattern instance occurrence on existing model elements directly in the context and so, in the consequence, the pattern instance will be integrated in the application model or context and thus there won't be necessary any manual specialization of pattern instance.

The resulting model after pattern instances suggestion is shown in the following Fig 11.

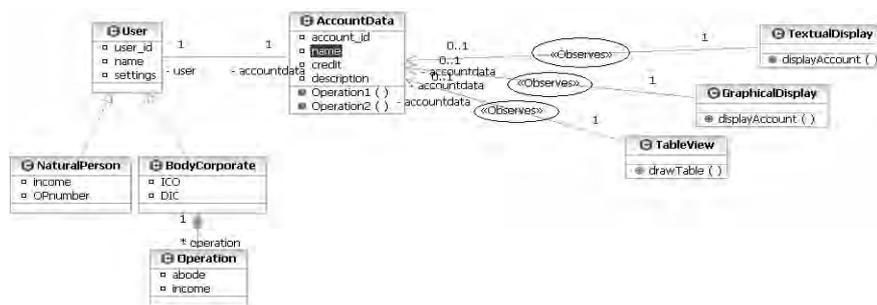


Fig. 11. The resulting model after pattern instances suggestion

It is important to remark, that each stereotype can be applied only on an instance of meta-class onto which is designated. For example, the stereotype

Design Pattern Instantiation Directed by Concretization and Specialization

<<Observes>> extends the meta-class association and the stereotype <<Observer>> extends the meta-class class. Therefore, the tool does not allow to apply the stereotype <<Observer>> to any association or any other model element which is not an instance of meta-class class and also it does not allow to apply the stereotype <<Observes>> to any class or any other model element which is not an instance of meta-class association.

Now the tool knows what design pattern and where the developer wants to apply it. On the basis of comparison of this model to the pattern model by which the tool is driven, the tool also recognizes that the association between classes `TextualDisplay` and `AccountData` corresponds with association between `ConcreteObserver` and `ConcreteSubject` from the pattern model. The recognition is realized on the basis of first type of marks – stereotypes comparison in these models (see Fig. 12) and this way the tool creates mapping between these models.

Because the match of marks occurs on the association, the transformation recognizes that also the source and destination elements of associations (in our case `ConcreteObserver` and `ConcreteSubject`) must be already in the model of the application under development. In consequence, the transformation recognizes which elements of pattern model are in the model of application and which are not.

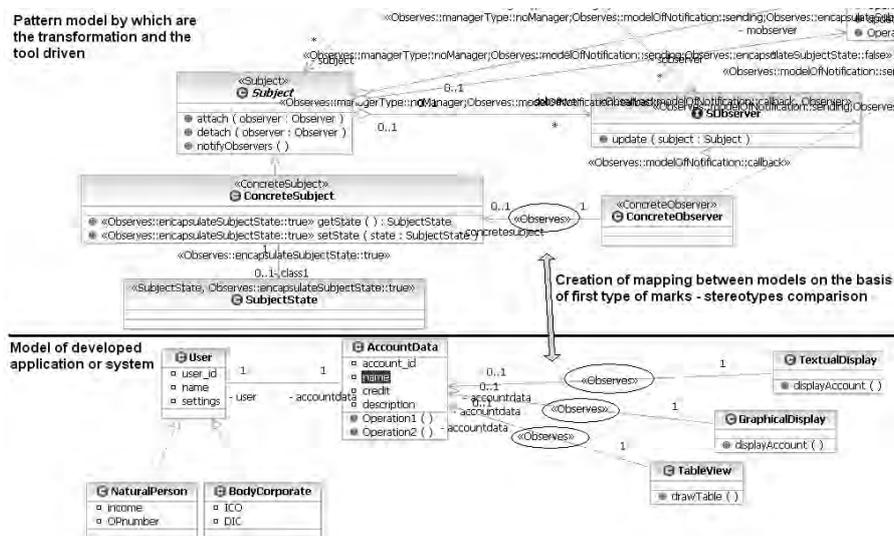


Fig. 12. Creation of mapping between model of developing application or system and pattern model by which the tool and the transformation are driven

Because the pattern model covers all the pattern variants, the tool needs to know which variant of pattern the developer wants to generate. In other words, the tool needs to know which of all identified missing pattern elements from the pattern model and what way it should generate into the model of

application. So the developer chooses the variant or modification of the pattern via setting up the values of particular stereotype meta-attributes in the next step of pattern instantiation (see Fig. 13). It is important to remark that the meta-attributes of stereotypes have set their default values. Therefore, this step is realized only if the developer wants to generate other than default variant of pattern. The possible variants and adjustments of pattern are defined in UML profile via enumerations or elements' primitive type specification such as boolean, integer and so on.

The developer specifies which variant or modification of pattern he desires and so the developer creates the specifications of suggested pattern instances. When the transformation is being executed, the tool processes all identified missing pattern participants from pattern model and it checks the second type of marks – keywords on these missing elements. As it has been introduced in previous section, for the second type of mark the following notation is defined (remind that these marks can be joined via “;”, while the symbol “~” expresses negation):

```
[~]?StereotypeName::Meta-attributeName::value;
```

A missing element from the pattern model is generated into the model only when the specified meta-attribute of the specified stereotype has the specified value.

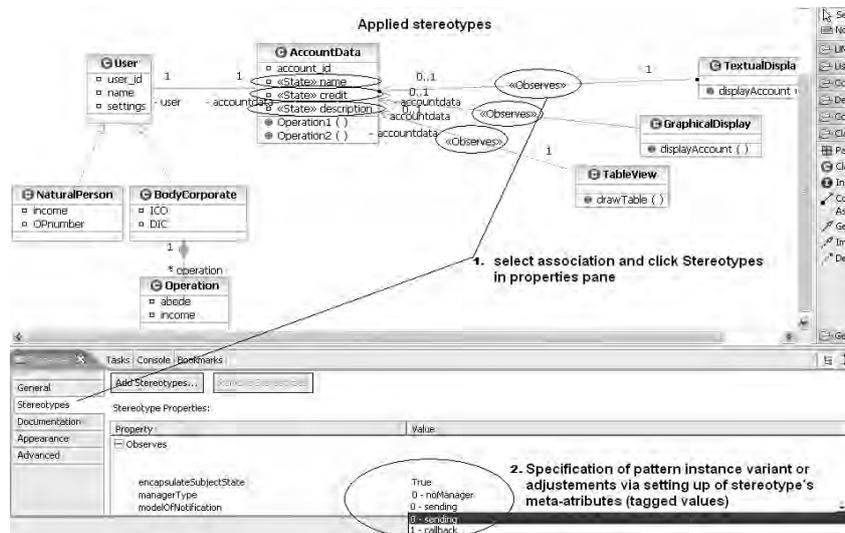


Fig. 13. Setting up of values of stereotype meta-attributes

Elements from pattern model of which at least one second type mark does not match the pattern instance specification are ignored by the tool and so only elements with all positive matches of marks or without any mark are generated into the model. For example, when the element

ConcreteSubject from the pattern model is identified as missing element in the application model, it is always generated into the application model, because it does not have any second type mark. On the other hand, the methods `getState` and `setState` are generated, only if the developer sets the value of meta-attribute `encapsulateSubjectState` of the stereotype `Observes` to `true`, because these methods are marked with the following second type mark `<<Observes::encapsulateSubjectState::true>>` (see Fig. 14, ConcreteSubject class of Observer pattern model).



Fig. 14. Element ConcreteSubject from Observer pattern model

When suggestions and specifications of pattern instances are completed, the transformation can be launched simply from context menu of application model (for more details see user guide on [26]). The resulting model of transformation is shown in the following Fig. 15.

The following sample specification of pattern instances has been set in the second step of pattern instantiation by the developer (i.e. choosing pattern variant and adjustments via setting up the values of stereotype meta-attributes, see Fig. 13).

1. `<<Observes>>` AccountData - TextualDisplay:
 - `modelOfNotification` = `sending` - the interface of Observers which takes reference to the `SubjectState` class as notification parameter has been generated.
 - `managerType` = `noManager` – no manager has been generated
 - `encapsulateSubjectState` = `true` - the state of class `ConcreteSubject` has been encapsulated
2. `<<Observes>>` AccountData - GraphicsDisplay:
 - the same as previous instance `AccountData - TextualDisplay`.
3. `<<Observes>>` AccountData - TableView:
 - `modelOfNotification` = `callBack` - the interface of Observers which takes reference to `Subject` class as notification parameter has been generated.
 - `managerType` = `noManager` - no manager has been generated
 - `encapsulateSubjectState` = `false` – this instance of Observer pattern does not use any encapsulated `SubjectState`, but the `Subject` reference instead.

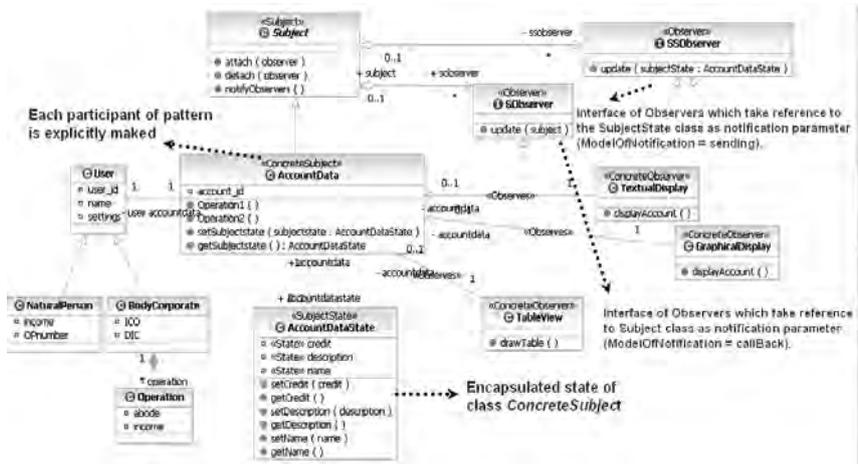


Fig. 15. The resulting model of transformation of model from Fig. 13

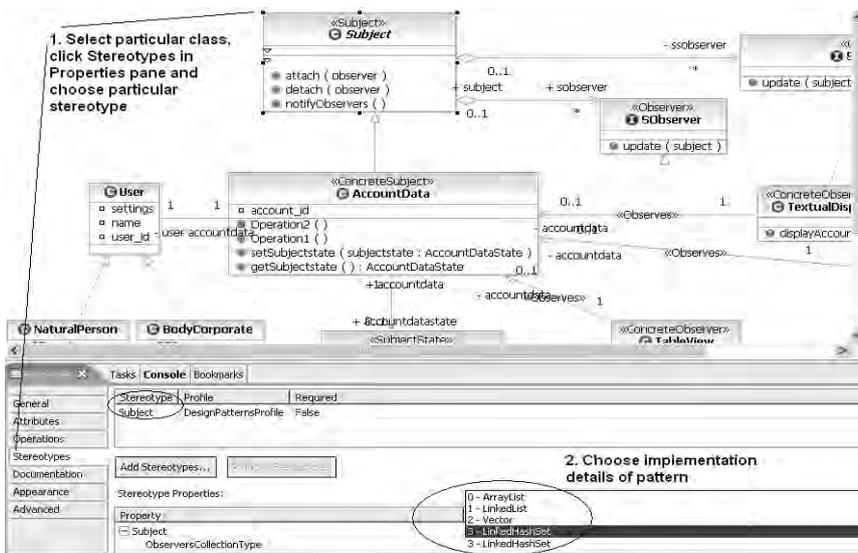


Fig. 16. Choosing of implementation details of pattern instances

The transformation marks explicitly also all the identified and generated participants of pattern instances and in the consequence, it makes the participants clearly visible. Moreover, in the next step of instantiation the developer can repeat the previous instantiation process from second step and can specify implementation details of pattern instances directly without necessity of further stereotype application (see Fig. 16). This step is optional

again, because the default implementations details are set and so the developer can launch the transformation to source code immediately.

The snippet of resulting source code of transformation of model from Fig. 16 to Java source code is shown in the Fig. 17.

The transformation to the source code generates two separate packages (generated and developed). The first is the base package which is always overwritten by subsequent source code generation. The second is the development package which is generated only by the initial transformation. The developer can write and add a specific implementation here without the threat of it being overwritten. Further, the distinct methods of observer notification have been generated for each group of Observers according to their specification (in our case `TextualDisplay` and `GraphicsDisplay` as the first group with `SSObserver` interface and `TableView` as the second group with `SObserver` interface, see Fig. 17). The transformation also uses chosen data types in the code generation. Description of source code generation has been introduced in the section 5.4. The snippet of code template of Subject participant of Observer pattern has been shown in the Fig. 9 as well.

After all, suggested and specified pattern instances from the highest level of abstraction have been transformed to the lowest level of abstraction – source code. The developer can utilize the created model and perform next iteration of development. For more details how the method and the tool work see user guide and video on [26].

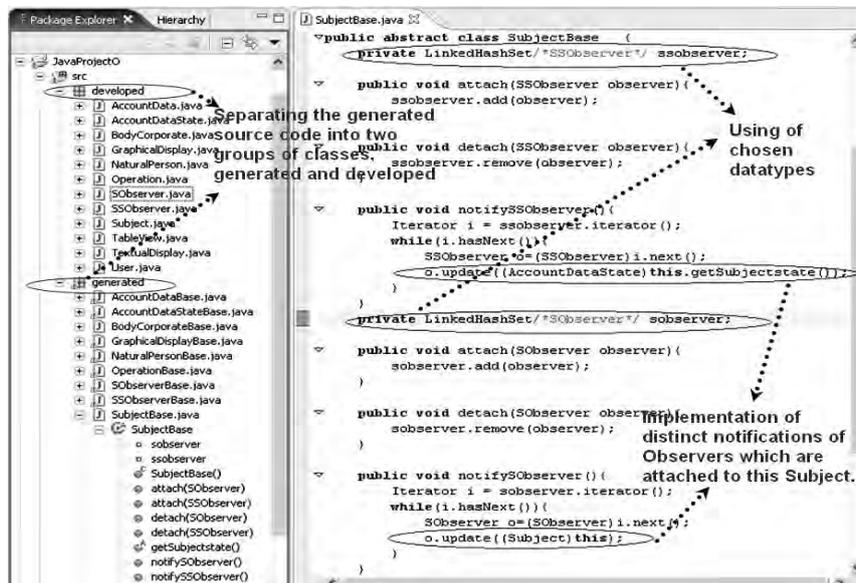


Fig. 17. The snippet of resulting source code of transformation of model from Fig. 16 to Java source code

5.6. Implementation

The presented method and the tool was implemented and verified in the form of an IBM Rational Software Modeler transformation plug-in. The following features have been implemented:

- Semantics in the UML profile for the patterns Factory Method, Decorator, Observer, Chain of Responsibility and Mediator
- Transformation of the highest level of abstraction (PIM) to the lower level (PSM) and transformation of PSM to source code
- Incremental consistency check mechanism
- Visualization of pattern instances and its participants
- Transformation of PIM to the lower level model PSM is driven by pattern models
- Models of design pattern covered all pattern variants and modifications which provide the basis upon which the transformational tool is driven
- Mechanism for adjustments of concrete form or desired variant of pattern instance for the patterns Factory Method, Decorator, Observer and Mediator

The first type of transformation of the highest level of abstraction (PIM) to the lower level (PSM) is implemented by M2M, UML2 and EMF frameworks. These frameworks are subprojects of the top-level Eclipse Modeling Project and they provide ideal infrastructure for model-to-model transformations.

The second type of transformation of model of lower level of abstraction (PSM) to source code is implemented by frameworks JET, UML2 and EMF. The JET is also part of Eclipse Modeling Project in M2T (Model to Text) area. It provides infrastructure for source code generation based on code templates. The architecture of the implemented tool is shown on the following figure 18.

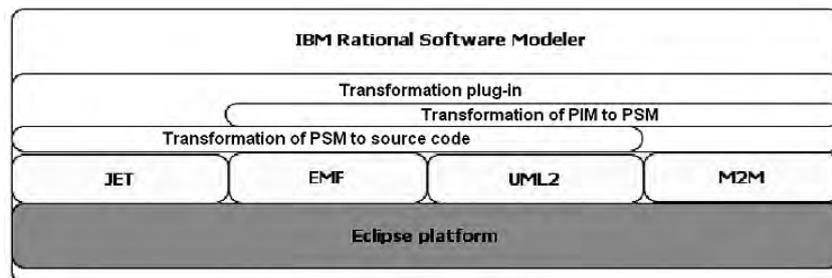


Fig. 18. The architecture of the implemented tool

5.7. Extending of Support for New Patterns or Structures

In order to extend the support for a new pattern or structure it is necessary to add definition of semantics of such new pattern into the existing UML profile. It is necessary to identify participants of a new pattern and to add definition of stereotype for each identified participant into the profile. All defined stereotypes should have the same second part of its qualified name (in RSM the stereotypes should have the same keyword). This part of the name represents the name of the new pattern. It is up to the developer how he names it, but the name should be unique in the set of names of supported patterns. After that it is necessary to identify variants of the new pattern and to create the according meta-attributes of the stereotypes (tagged values) and to create also definition of permissible values of the meta-attributes in form of enumerations or their type definition. If any stereotype can be applied in scope of one instance of a new pattern more than once, then the stereotype should have `group_id` meta-attribute in order to distinguish which stereotype belongs to which instance. In other words if cardinality of any participant of a new pattern is greater than one, then the stereotype of such participant should have defined `group_id` meta-attribute.

In the second step it is necessary to create a class model of the new pattern and to mark the participants with appropriate stereotype defined in the first step. Now the tool would be able to create mapping between models, because the developer places the same marks – stereotypes in the application model. So the tool can compare them simply. The tool still needs to know which participant it should generate and when. So it is necessary to add second type marks - keywords to the elements of class model of new pattern in introduced form:

```
[~]?StereotypeName::Meta-attributeName::value;
```

If the specified meta-attribute of the specified stereotype has the specified value, the element will be generated into the application model. Finally, it is necessary to export created model of a new pattern into XMI structure and place it into the working folder of the tool. The name of the file with the pattern model should be the same as the name of the new pattern (i.e. the second part of qualified name of stereotypes defined in the first step). The refresh or update of original UML profile is also necessary.

How the developer marks the model of the new pattern, thus the tool will generate the pattern into the model of application. So it is up to developer to mark the pattern model in the way that he desires. We do not want to restrict the developer. Our aim is to allow him to model any custom pattern. The tool simply takes a new pattern model, next the tool seeks in it the elements with marks identical to marks from application model placed by developer and then it maps the elements with identical marks. After that on the basis of the comparison of second type marks (keywords) from new pattern model and values of meta-attributes from application model which have been set by developer the tool filters out unwanted elements and it generates desired elements of the pattern. The tool performs all actions according to the

algorithm introduced in Section 5.4 (Fig. 8). That approach allows extension of transformation with new special functionality in form of definition of new rules and notations of marks. In this case the implementation of the new rules and the new notation recognition should be necessary, of course.

6. Evaluation

The presented method and its realization were evaluated in various experiments. In the following case study the aspects of correct pattern instantiation were considered in the evaluation process. The transformation algorithm (in Fig. 8) always checks on the presence of elements with identical definition by adding the pattern elements to the application model. Consequently, the transformation does not duplicate the pattern participants with identical definition when more instances of patterns are applied in the model. In addition, when the transformation of the model is run repeatedly, the incremental consistency of the model is verified. When an element with an identical definition is presented in the model, it is not duplicated. Instead, it is swapped. Illustrations of some case studies are shown in the following Fig. 19, 20 and 21.

The next evaluation was realized through experiments in which we have monitored and focused on the time of carrying out of an assigned task with and without usage of the tool. Also the count of generated and added source code lines has been observed. The tasks consisted of implementing specified instances of design patterns in a specified form. The average results of the experiments on a group of five programmers and five master degree students of software engineering are summarized in the Table 2.

Table 2. Average results of executed experiments

Time with using the tool t1	Time without using the tool t2	Speed up t2/t1	Number of generated code lines Ng	Number of added code lines Nd	Improving coefficient (Ng / Nd) + 1
< 30 min	> 120 min	> 4	478	52	10,2

The quantity of the generated source code has been evaluated for each design pattern via metrics. The results of this evaluation are shown in Table 3.

Table 3. Quantity of generated source code

Design Pattern	LOC	NOA	NOC	NOCON	NOIS	NOM	NOO
<i>Decorator pattern</i>	223	9	6	7	11	103	22
<i>Mediator pattern</i>	212	6	6	7	9	50	15
<i>Observer pattern</i>	193	14	6	1	10	60	14

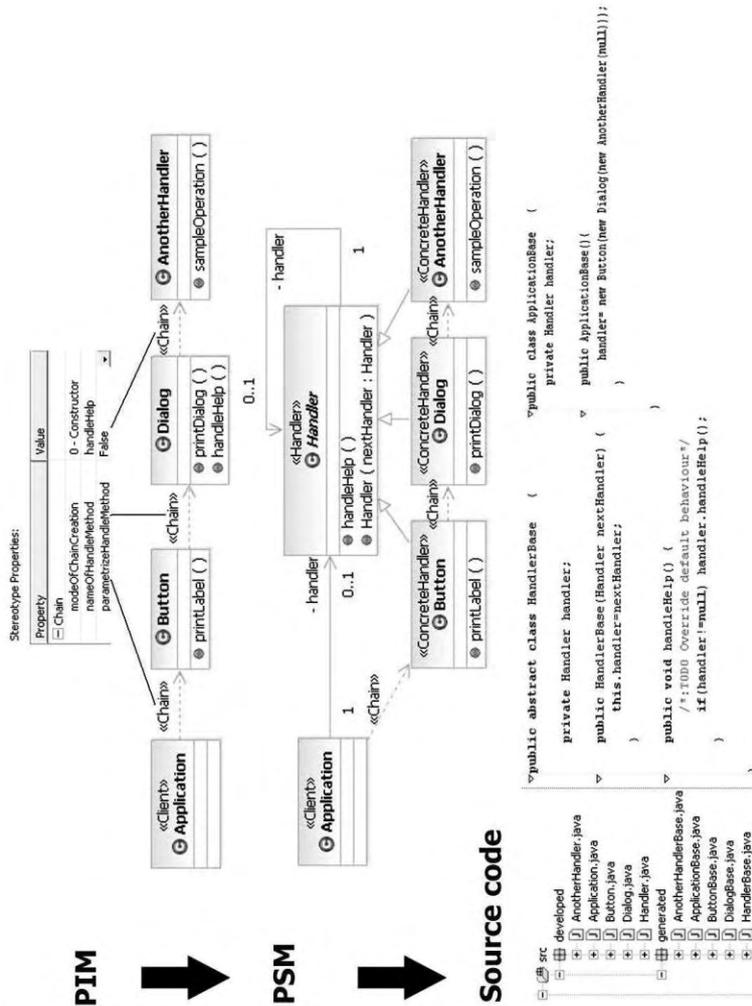


Fig. 19. Case study of simple Chain of Responsibility pattern instantiation

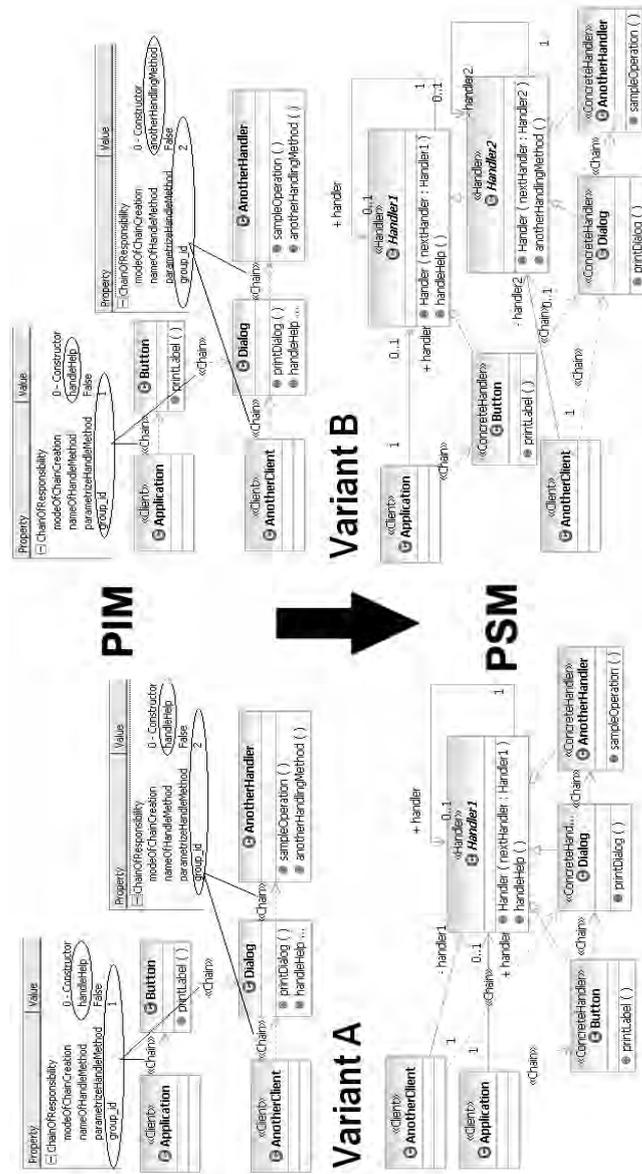


Fig. 20. Case study of advanced instantiation of Chain of Responsibility pattern. In this case, there are two clients. One client uses Button and Dialog as the processing objects and the other client uses Dialog and AnotherHandler as the processing objects. Variant B illustrates the case with different handled method names and Variant A with the same handled method names (in this case, all processing objects have super classes with the same definition, so the tool does not duplicate them, but it substitutes them instead)

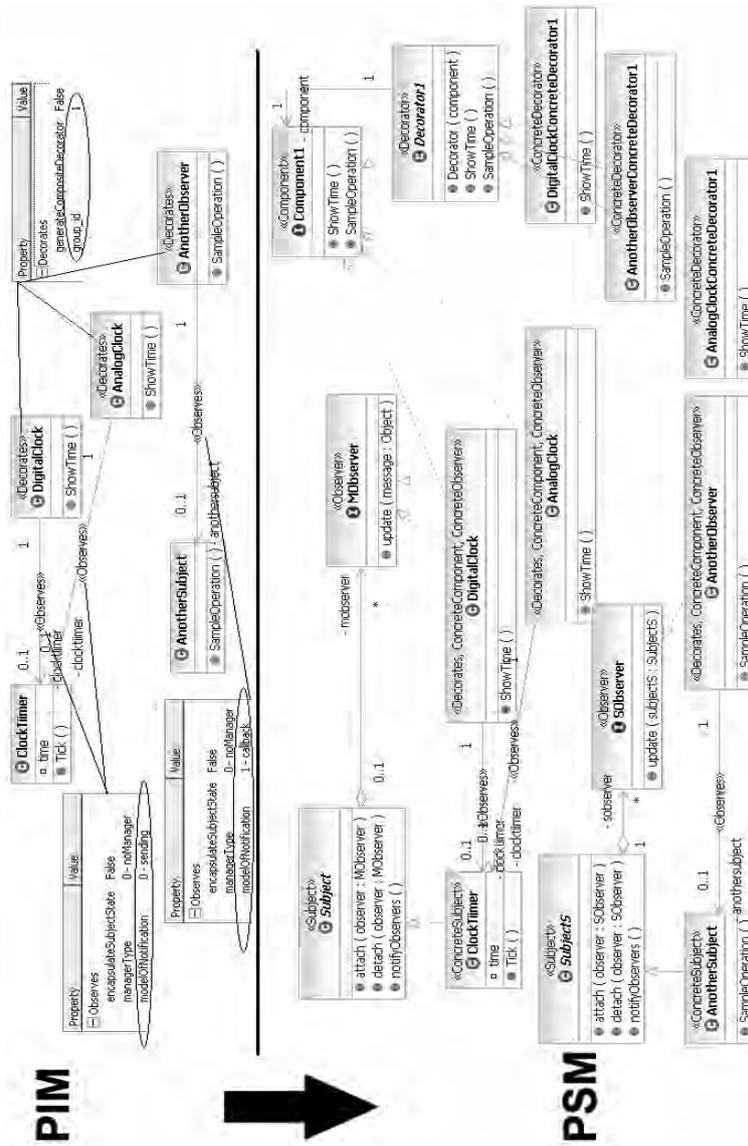


Fig. 21. Case study of sample Observer and Decorator pattern composition. Classes DigitalClock, AnalogClock and AnotherObserver have the same group_id and therefore they are considered as one Decorator instance. Moreover, the tool does not duplicate the elements with an identical definition, but it substitutes them successively as instance by instance are generated

Results of experiments show a significant improvement gained by use of the method and tool in the area.

7. Future Work

In the future, it is important to support also the fourth characteristic of the model driven development – the invertibility of models. The most important problem is to transform the source code to the design level (PSM), because the higher-level semantics cannot be reasoned directly and automatically from the source code in general. The knowledge is mainly available to developers and domain experts involved in the design process. Therefore, our aim is to add the missing semantics into the source code. Our idea is to mark explicitly and make visible higher-level (i.e. design) intentions in the source code via annotations. This way it would be possible to express also the semantics of patterns in the source code and the intention of annotated code as well. Consequently, it would be possible to expand the visibility of pattern instances from model into the source code by annotations. The pattern instances do not become invisible in huge amount of source code lines, quite the contrary, the full visibility of instances and their participants would be achieved by annotations. Consequently, using source code annotations the inverse transformation would be able to recognize pattern instance participants in source code and to transform them into a higher level of abstraction.

Besides this feature, also the traceability of transformations and pattern instances would be enhanced at the source code level. The code annotations make identifying of pattern participants in the source code quite easy. As a result, the tool based support of pattern instantiation or existing instances evolution, validation and identification at the source code level can be achieved in the form of code assists. Thanks to the annotations, the tool would be able to identify the pattern participants already implemented, and subsequently it would be able to offer to the developer the generation of any missing pattern participant or the possible evolution of instance in the given context. The evolution of existing instances of patterns without any tool-based support is quite difficult, because a developer has not a good vision about all concrete participants of pattern instances in the source code. However, this idea would bring significant improvement in pattern instantiation, evolution and validation in the source code.

Nowadays, we have proposed the improvement of the transformation to the source code. The method presented in this paper marks all pattern participants by stereotypes in the model. Our idea is that the transformation to the source code preserves the marking from the model and also extends it via annotations into the generated source code. Therefore no manual annotation of the code would be necessary in the generated source code, in comparison to the other present approaches [23, 24]. For more details about the improvement of the transformation to the source code and the method of

continuous support of the patterns at the source code level see our paper [25].

Currently, the tool does not give any suggestion or guide on what suitable patterns to apply are. In our opinion, this guide is relatively hard to automate by the tool, because the knowledge of what are suitable patterns to apply requires really detailed understanding of the context and the application and, therefore, it is available especially to the developers or designers involved in the design process. But this is also a challenge to the future.

8. Conclusion

The abstraction, semantics and model transformations represent the key aspects of Model Driven Development and Model Driven Architecture. The possible level of the automation of the development process can be improved considerably thanks to them. The semantics applied in the models enables the possibility to understand the model and its elements, and also to recognize which elements play which roles in the model. Consequently, on the basis of the understanding of the model and its elements, it is possible to construct the transformation which transforms the model to a lower level of abstraction.

These principles represent the basis of the elaborated method of the design pattern application support. Thanks to the elaborated semantic extension of UML in form of UML profile, it is possible to specify participants of design patterns and relations between them directly on the elements of the application model. The suggestion and specification of pattern instances in the model allow the transition to higher levels of abstraction in the modeling of pattern instances. The instantiation details are split into more levels of abstraction, so developers do not need to concern themselves with concrete details of pattern structure at higher levels.

The transformations of models to lower levels of abstractions are driven by models of patterns. This aspect provides the key option to the developer to adjust the results of transformations by modification of these pattern models. This way it is possible to model any custom model structure and achieve support of its application to the model. Consequently, the method is not limited to GoF design pattern support only, but it also represents the framework of creation and addition of support for other custom model structures which are often created in models mechanically.

Acknowledgments. This work was partially supported by the Scientific Grant Agency of Republic of Slovakia, grant No. VEGA 1/0508/09 and by Slovak Research and Development Agency, grant No. APVV-0391-06 "Semantic Composition of Web and Grid Services".

References

1. Arlow, J., Neustadt, I.: Enterprise Patterns and MDA: Building Better Software with Archetype Patterns and UML, Addison Wesley, (2003)
2. France, R., Dae-kyoo, K., Ghosh, S.: A UML-Based Pattern Specification Technique, pp. 193-206, IEEE transactions on Software Engineering, (2004)
3. Frankel, D.: Model Driven Architecture: Applying MDA to Enterprise Computing, Wiley Publishing, (2003)
4. Gamma, E. et al.: Design Patterns, Elements of Reusable Object-Oriented Software, Addison-Wesley professional computing series, (1995)
5. Návrát, P. et al.: A technique for modeling design patterns. Knowledge-Based Software Engineering - JCKBSE'98, pp. 89-97, IOS Press, (1998)
6. Object Management Group: MDA, MOF and UML Specifications. (2009) [Online]. Available: <http://www.omg.org/>
7. Borland Software Corporation: Borland Together Architect. (2009) [Online]. Available: <http://www.borland.com/together/>
8. Briand, L., Labiche, Y., Sauve, A.: Guiding the application of design patterns based on uml models. In ICSM '06: Proceedings of the 22nd IEEE International Conference on Software Maintenance, pp 234-243, Washington, DC, USA, (2006). IEEE Computer Society.
9. Dong, J., Yang, S.: Qvt based model transformation for design pattern evolutions. In: J.-N. Hwang (Ed.): Proceedings of the Tenth IASTED International Conference on Internet and Multimedia Systems and Applications (IMSA 2006), Honolulu, Hawaii, USA, August 14-16, 2006. IASTED/ACTA Press 2006, 16-22.
10. Dong, J., Yang, S., Zhang, K.: A model transformation approach for design pattern evolutions. In ECBS '06: Proceedings of the 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems, pp 80-92, Washington, DC, USA, (2006). IEEE Computer Society.
11. Boussaidi, G., Mili, H.: A model-driven framework for representing and applying design patterns. In COMPSAC '07: Proceedings of the 31st Annual International Computer Software and Applications Conference, pp 97-100, Washington, DC, USA, (2007). IEEE Computer Society.
12. Wang, X.-B., Wu, Q.-Y., Wang, H.-M., Shi, D.-X.: Research and implementation of design pattern-oriented model transformation. In ICCGI '07: Proceedings of the International Multi-Conference on Computing in the Global Information Technology, Washington, DC, USA, (2007). IEEE Computer Society.
13. Cinnéide, M., Nixon, P.: Automated software evolution towards design patterns. In IW- PSE '01: Proceedings of the 4th International Workshop on Principles of Software Evolution, pp 162-165, New York, NY, USA, (2001). ACM.
14. Debnath, N.C. et al.: Defining Patterns Using UML Profiles. In IEEE International Conference on Computer Systems and Applications, pp.1147-1150, Washington, DC, USA, (2006). IEEE Computer Society.
15. Mapelsden, D., Hosking, J., and Grundy, J.: Design pattern modelling and instantiation using DPML. In Proceedings of the Fortieth international Conference on Tools Pacific: Objects For internet, Mobile and Embedded Applications, pp 3-11, Darlinghurst, Australia, (2002) ACM International Conference Proceeding Series, vol. 21. Australian Computer Society.
16. Dong, J., Yang, S.: Visualizing design patterns with a UML profile. In Proceedings of the 2003 IEEE Symposium on Human Centric Computing

- Languages and Environments, pp 123-125, Washington, DC, (2003). IEEE Computer Society.
17. Alexander, C. et al.: A pattern language. Towns, buildings, construction. Oxford University Press, New York, USA, ISBN 0-19-501919-9, (1977).
 18. Judson, S. R.: Pattern-based model transformation. In OOPSLA '03: Companion of the 18th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, pp 124-125, Anaheim, CA, USA, (2003). ACM
 19. Majtás, L.: Tool Based Support of the Pattern Instance Creation. In: e-Infomatica Software Engineering Journal, Vol. 3. Iss. 1, 89-102. (2009)
 20. Havlice, Z. et al.: Knowledge Based Software Engineering. In: Computer Science and Technology Research Survey, elfa, Kosice, pp. 1-10, (2009)
 21. Kollár, J., Porubán, J., Václavík, P., Forgác, M., Wassermann, L. et al.: New Generation of Language Architectures. In: Kollar, J. (Edt.): Computer Science and Technology Research Survey, elfa, Kosice, pp. 21-30, (2009)
 22. Marko, V.: Template Based, Designer Driven Design Pattern Instantiation Support. In: LNCS 3255 – SOFSEM 2004, Springer-Verlag, pp. 144-158, (2004)
 23. Sabo, M., Porubán, J.: Preserving Design Patterns using Source Code Annotations. In: Journal of Computer Science and Control Systems. (2009), pp. 53-56.
 24. Meffert, K.: Supporting Design Patterns with Annotations. In: Proceedings of the 13th Annual IEEE international Symposium and Workshop on Engineering of Computer Based System. ECBS'06. IEEE Computer Society, Washington, DC, (2006), pp. 437-445
 25. Kajsa, P., Návrát, P.: Design Pattern Support at Source Code Level Based on Annotations and Feature Models. In: Student Research Conference in Informatics and Information Technologies 2010, Bratislava, (2010), pp. 233-240
 26. Kajsa, P., Majtás M., Návrát, P.: Web page of *Design Pattern Instantiation Directed by Concretization and Specialization*. (2010). [Online]. Available: <http://www.fiit.stuba.sk/~kajsa/tool-based-design-patterns-support/>

Peter Kajsa is a PhD student and a teaching assistant at the Faculty of Informatics and Information Technologies of the Slovak University of Technology in Bratislava. Peter received his Master degree in Software Engineering in the year 2009. His main research interests include design and architecture of software systems, design and architectural patterns, Model Driven Development, Model Driven Architecture and other Object Management Group specifications. He has published several works in the area.

Lubomír Majtás is a PhD candidate and a teaching assistant at the Faculty of Informatics and Information Technologies of the Slovak University of Technology in Bratislava. He received his Master degree in Software Engineering in the year 2006. His main research activities focus on design and architecture of software systems, Model Driven Development and Model-Driven Architecture where he specializes on support automation for design

Peter Kajsa, Lubomir Majtas, and Pavol Navrat

pattern instances creation and their detection in the existing software. He has published several papers in the area.

Pavol Návrat received his Ing. (Master) cum laude in 1975, and his PhD. degree in computing machinery in 1984 both from Slovak University of Technology. He is currently a professor of Informatics at the Slovak University of Technology and serves as the director of the Institute of Informatics and Software Engineering. During his career, he was also with other universities abroad. His research interests include related areas from software engineering, artificial intelligence, and information systems. He published numerous research articles, several books and co-edited and co-authored several monographs. Prof. Návrat is a Fellow of the IET and a Senior Member of the IEEE and its Computer Society. He is also a Senior Member of the ACM and a member of the Association for Advancement of Artificial Intelligence, Slovak Society for Computer Science and Slovak Artificial Intelligence Society. He serves on the Technical Committee 12 Artificial Intelligence of IFIP as the representative of Slovakia.

Received: December 12, 2009; Accepted: October 11, 2010.

A Grammar-based model for the Semantic web

Hyosook Jung¹ and Seongbin Park^{1*}

¹ Department of Computer Science Education,
Korea University, Seoul, Korea
{est0718, hyperspace}@korea.ac.kr

Abstract. The Semantic Web is an extension of the Web where information is represented in a machine processable way. In this paper, we present a two-level model for the Semantic Web from the perspective of formal language theory. The model consists of two grammars where the first level grammar is for creating ontologies and the second level grammar is for creating ontological instances. Based on the model, we implemented a system by which one can easily construct a small-scale Semantic Web environment.

Keywords: Semantic web, ontology, grammar.

1. Introduction

The Semantic Web is a vision for the future of the Web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web [5].

In this paper, we present a grammar-based model for the Semantic Web. As in [3], we view the Semantic Web as the set of ontologies and ontological instances, where an ontology is a document or file that defines the relations among concepts. The proposed model consists of two grammars. The first level intends to represent an ontology about a domain of interests. Ontologies are strings generated by the first level grammar. The second level intends to represent ontological instances which are resources described using concepts and relationships based on the ontology defined at the first level.

While there are approaches to model the Semantic Web [3,9,10], the advantage of the proposed model is that users can easily create a small-scale Semantic Web environment where various experimentations such as whether a current Web browser needs a new functionality or not can be done. To construct the environment, one can define a grammar for an ontology and generate ontological instances.

Our system can serve as an education tool for teaching the Semantic web. For example, non-experts learn about the conceptualization and formalization of ontologies during lectures. Although there are different ontology language

* To whom correspondence should be addressed.

standards, in teaching level the general understanding about the ontology language might be more important than the specific understanding about a certain ontology language standards. Using our system, they can practice how to create ontologies by defining their simple ontology languages. While the languages that they define are not full-fledged ontology languages, they can understand the roles of ontologies and how ontologies are used.

Our system is different from ontology development tools such as Protégé-OWL editor [13], OntoEdit [15], OntoKick [16], WebODE [17] etc. in that it allows non-expert users to generate ontologies by using simple languages defined by themselves without the knowledge of OWL and RDF which are not simple concepts to understand [14]. They can understand the essential elements of ontologies and how they are used.

Figure 1 shows the idea behind our approach.

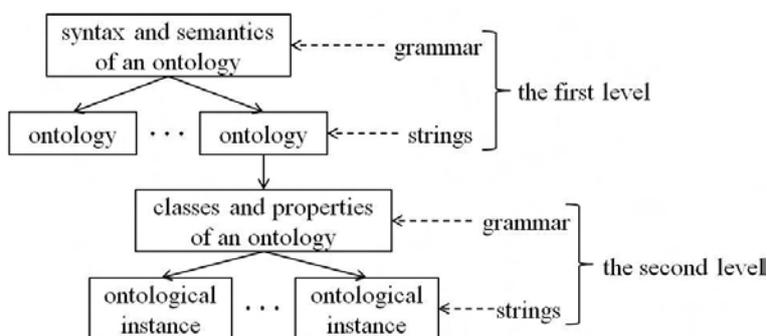


Fig. 1. A grammar-based model for the Semantic Web

This paper is structured as follows. Section 2 describes related works. In section 3, a two-level model is explained. Illustrative examples are given in section 4 and section 5 describes how a small-scale Semantic Web can be constructed using the proposed model. Finally, section 6 concludes the paper.

2. Related Works

The Semantic Web is an environment where Web contents are represented in a form that is machine processable [4]. There are several languages to represent machine interpretable content on the Web. XML offers a surface syntax for structured documents and XML Schema is a language for restricting the structure of XML documents. RDF is a data model for objects and their relations and supports a simple semantics for the data model. RDF Schema is a vocabulary for representing properties and classes of RDF resources. OWL adds more vocabulary for describing properties and classes:

among others, relations between classes, cardinality, equality, richer typing of properties, characteristics of properties, and enumerated classes [1].

OWL ontology represents a domain by defining classes and properties of those classes and defines individuals and asserts properties about them. Ontologies contain computer-usable definitions of basic concepts in the domain and the relationships among them. They encode knowledge in a domain and also knowledge that extends domains [6]. An OWL instance is a description about a resource created by using properties and classes defined in the OWL ontology [2,8].

In Ontobroker [7], ontologies are defined in a representation language based on Frame-Logic which supports queries by using instances of an ontology. The representation language used to define ontologies enables elementary expressions such as classes, attributes, relationships, and axioms. It also allows complex expressions such as facts, rules, double rules and queries. The defined ontology is composed of concept hierarchy which defines the subclass relationship between different classes, attribute definitions given for classes and a set of rules which defines relationships between different concepts and attributes.

The Semantic Web has been modeled in various ways. [9] describes a Semantic Web space as two-tuple $\langle O, R \rangle$, where O is a set of ontologies and R is a set of resources such as web pages, databases, and sensors.

[10] describes the semantic web as a Notebook + Memex where, the Memex emphasizes on engaging with information, developing it, and working with it, the notebook focuses on both the more writerly and the more personal side of engaging with information. It can perform the automatic and logical processing of repetitive thought tasks and the creation of associative links across different resources by connecting into either the similar tasks or creative thought processes.

[3] describes a semantic network as a directed labeled graph. For the Semantic Web, a semantic network substrate is represented by the constraints of the RDF which describes a semantic network as a set of triples where a subject resource points to an object resource according a predicate resource. Subject and predicate resources are identified by URI (Uniform Resource Identifier) and object resources are a literal or URI. The Semantic Web can be defined as $G \subseteq (U \times U \times (U \times L))$, where U is the set of all URIs and L is the set of all literals.

Linked Data is about using the Web to create typed links between data from different sources. It basically uses the RDF data model to publish structured data on the Web and RDF links to interlink data from different data sources [18]. It is associated with the semantic web because the semantic Web isn't just about putting data on the web, but about making links, so that a person or machine can explore the web of data [19]. Our tool lets people represent data based on ontologies, which is a basic process to make semantic links between data.

Cloud computing is a term to describe both a platform and a type of application. A cloud is a pool of virtualized computer resources. A cloud computing platform dynamically provisions, configures, reconfigures, and

deprovisions servers as needed. Cloud applications are extended applications to be accessible through the Internet. These cloud applications use large data centers and powerful servers that host Web applications and Web services [20]. Although the cloud computing providers are publishing various clouds over the Internet, there are no standard, open protocols and discover mechanisms for different kinds of clouds [21]. So, the Cloud Computing Interoperability Forum (CCIF) focus on being placed on the creation of a common agreed upon framework or ontology that enables the ability of two or more cloud platforms to exchange information [22]. A common cloud ontology can support the expression of cloud computing and its related parts by using a common data model. Our tool allows people to define ontologies to represent data semantically. They can experience the way of creating a data model for cloud computing.

Social semantic web is related to the creation of explicit and semantically rich knowledge representations. It can be seen as a Web of collective knowledge systems that which can provide useful information based on human contributions and get better as more people participate. Instead of relying entirely on automated semantics with formal ontology processing and inferencing, humans are collaboratively building semantics aided by socio-semantic information systems [23]. Our tool enables users to create ontologies and represent data based on the ontology by using their own description languages instead of RDF/OWL. The users can also create a small-scale social semantic web that supports semantic browsing by using user-defined ontologies.

A reasoner is a service that takes the statements encoded in an ontology as input and infers new statements from them. In particular, OWL reasoners such as FaCT++ and Pallet can be used to reveal subclass or superclass relationships among classes, determine the most specific types of individuals, and detect inconsistent class definitions [24]. Our tool checks whether the ontology is defined without syntactic and semantic errors and whether the instances are defined by using the classes and properties of the ontology.

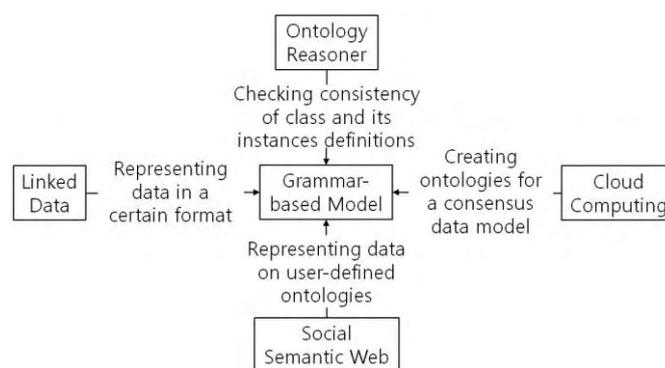


Fig. 2. Related research areas to our work

Figure 2 shows how our research is related to Linked Data, Cloud computing, Social Semantic Web, and Ontology reasoners.

3. A Grammar-based Model

In this section, we describe a grammar-based model for the Semantic Web. The proposed model consists of two grammars. The first grammar is for generating ontologies and the second grammar is for generating ontological instances. More specifically, the first level grammar in our model is used to generate ontologies that describe information about classes or properties. A class has certain restrictions, where a restriction is a data type of a class or a condition about data value. Datatype properties define relations between instances of classes and RDF literals and XML Schema datatypes. Object property defines relations between instances of two classes by connecting instances in a domain class into instances in a range class. A data range is used as the range of a data-valued property such as string, integer, Boolean, and float.

The syntax of the first level grammar¹ is as follows.

```
ontology ::= 'Ontology' ontologyID directive*;
directive ::= import | class | property;
import ::= 'NS:' namespaceID=referenceID
class ::= 'Class' classID description*;
description ::= 'SubClassOf' classID | restriction*;
restriction ::= 'Restriction On
Property' (datatypePropertyID datatype |
objectPropertyID objecttype);
datatype ::= dataRange | cardinality;
objecttype ::= classID | cardinality;
cardinality ::= 'min' digit+ | 'max' digit+ | 'equals'
digit+;
property ::= datatypeProperty | objectProperty;
datatypeProperty ::= 'DatatypeProperty'
datatypePropertyID
('domain' classID)* ('range' dataRange)*;
objectProperty ::= 'ObjectProperty' objectPropertyID
('domain' classID)* ('range' classID)*;
dataRange ::= 'string' | 'integer' | 'boolean' |
'float';
ontologyID ::= identifier;
```

¹ Terminals are quoted (i.e., 'Ontology') and non-terminals are not quoted (i.e., ontologyID). Alternatives are either separated by vertical bars(|) or are given in different productions. Components that can occur at most once are followed by '+' and components that can occur any number of times including zero are followed by '*'.

Using this grammar, an ontology about a movie can be defined. For example, a movie ontology can have Class Film and Genre, and property genreOf which creates a relation between Film's and Genre's instances. Class Film has at least one instance of Class Genre as the value of ObjectProperty genreOf. Using the grammar, a Movie ontology can be derived as follows.

```
'Ontology' ontologyID directive*
→'Ontology' Movie directive*
→'Ontology' Movie class directive*
→'Ontology' Movie 'Class' Film description* directive*
→'Ontology' Movie 'Class' Film restriction directive*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' objectPropertyID objecttype directive*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf cardinality directive*
→'Ontology' Movie Class Film 'Restriction On Property'
genreOf 'min' 1 directive*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 class directive*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 'Class' Genre directive*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 'Class' Genre property
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 'Class' Genre objectProperty
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 'Class' Genre
'ObjectProperty' objectPropertyID ('domain' classID)*
('range' classID)*
→'Ontology' Movie 'Class' Film 'Restriction On
Property' genreOf 'min' 1 'Class' Genre
'ObjectProperty' genreOf 'domain' Movie 'range' Genre
```

Now, this ontology serves as the second level grammar. According to the Movie Ontology, Class Genre has Instance Fantasy and Adventure and Class Film has Harry_Potter_and_the_Sorcerers_Stone whose genre is Fantasy and Adventure. The classes and properties of the ontology become terminals and strings are variables. An instance of the movie ontology can be derived as follows.

```

('Genre' genreInstance)+ ('Film' filmInstance 'genreOf'
(genreInstance)+)+
→'Genre' Fantasy 'Genre' Adventure (Film filmInstance
('genreOf' genreInstance)+)+
→'Genre' Fantasy 'Genre' Adventure 'Film'
Harry_Potter_and_the_Sorcerers_Stone ('genreOf'
genreInstance)+
→'Genre' Fantasy 'Genre' Adventure 'Film'
Harry_Potter_and_the_Sorcerers_Stone 'genreOf' Fantasy
'genreOf' Adventure
    
```

Figure 3 shows Movie ontology and its instances.

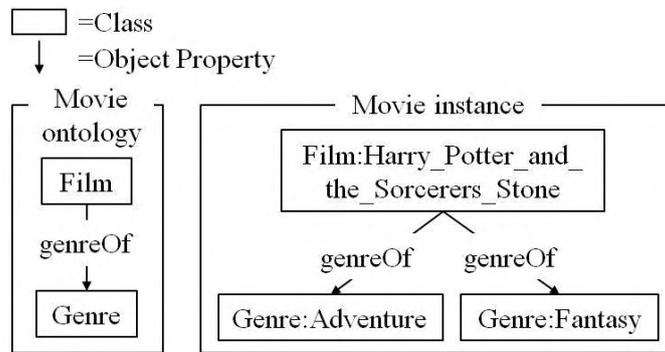


Fig. 3. Movie ontology and its instances

4. Illustrative Examples

In this section, we show how the proposed model can be used to describe various aspects of the Semantic Web environment.

4.1. Scenario 1

In the Semantic Web, an instance of an ontology can be semantically related to an instance of another ontology. For example, figure 4 shows how two domains (Movie and Travel) are connected.

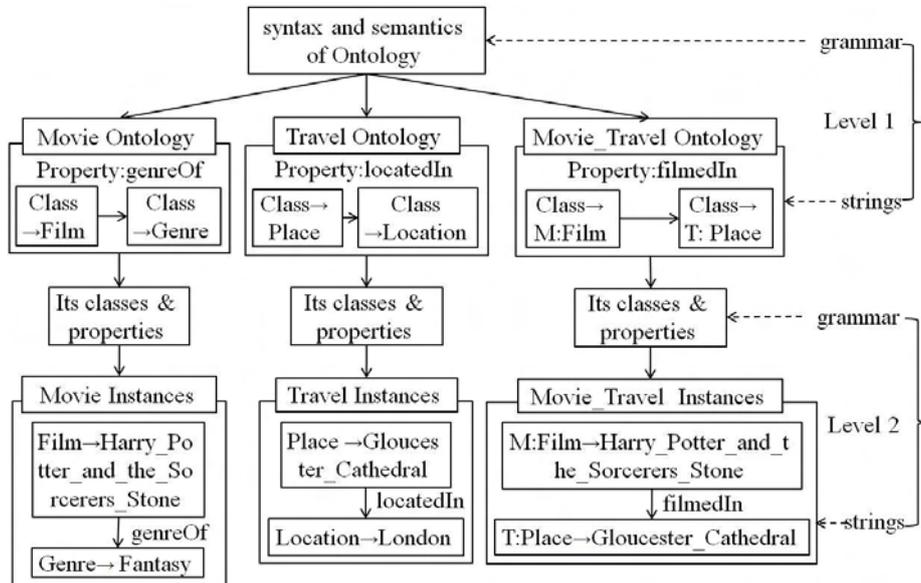


Fig. 4. Movie, Travel, Movie_Travel ontology and their instances

This situation can be easily described using the proposed model. Movie ontology has Film and Genre class, and genreOf property which creates a relation between Film's and Genre's instances.

```

Ontology Movie
Class Genre
Class Film
Restriction on Property genreOf equals 1
ObjectProperty genreOf domain Film range Genre
    
```

Based on Movie Ontology, Film class has Harry_Potter_and_the_Sorcerers_Stone in-instance and Genre class has Fantasy.

```

Genre
Fantasy
Film
Harry_Potter_and_the_Sorcerers_Stone genreOf Fantasy
    
```

Travel Ontology has Location and Spot class, and locatedIn property which creates a relation between Location's and Spot's instances.

```

Ontology Travel
Class Location
Class Place
Restriction on Property locatedIn equals 1
ObjectProperty locatedIn domain Place range Location
    
```

Based on Travel Ontology, Location class has London instance and Place class has Gloucester_Cathedral.

```
Location
  London
Place
  Gloucester_Cathedral locatedIn London
```

We create `Movie_Travel` ontology by importing `Movie` and `Travel` ontology. It uses `Film` class of `Movie` ontology and `Spot` class of `Travel` ontology. It has `filmedIn` property which creates a relation between `Film`'s and `Spot`'s instances.

```
Ontology Movie_Travel
NS:M=Movie
NS:T=Travel
Class T:Place
Class M:Film
Restriction on Property filmedIn equals 1
ObjectProperty filmedIn domain M:Film range T:Place
```

Based on `Movie_Travel` Ontology, `Film` class has `Harry_Potter_and_the_Sorcerers_Stone` instance and `Place` class has `Gloucester_Cathedral`.

```
T: Place
  Gloucester_Cathedral
M: Film
  Harry_Potter_and_the_Sorcerers_Stone filmedIn
  loucester_Cathedral
```

People can get information for traveling a place where a famous movie is filmed by combining `Film` and `Travel` ontology.

4.2. Scenario 2

Historical study often focuses on events and developments that occur in particular blocks of time. Therefore, the events and developments might be organized based on historical periods such as Ancient history, Middle Ages, Early modern period, Modern era and Post-Modern. Assume that there are three types of ontologies which define different classes and properties as follows.

```
Ontology Ancient_History
Class Nation
Class Machine
  Restriction on Property inventedBy equals 1
ObjectProperty inventedBy domain Machine range Nation
```

```
Ontology Middle_Ages
Class Area
Class Invention
  Restriction on Property introducedFrom equals 1
```

Hyosook Jung and Seongbin Park

```
ObjectProperty introducedFrom domain Invention range
Area
```

```
Ontology Modern_Era
Class Country
Class Technology
  Restriction on Property developedIn equals 1
ObjectProperty developedIn domain Technology range
Country
```

Assume that a person wants to organize information on technology in history by using an ontology, but historians already organized the information on technology as well as war, religion, or science based on different ontologies by the historical periods. The person tries to extract the classes or properties related to science and technology from different ontologies and to integrate them in an ontology. That can be done by importing the three ontologies. So, the person can define the namespace for each ontology.

```
Ontology History_Of_Technology
NS:AH = Ancient_History
NS:MA = Middle_Ages
NS:ME = Modern_Era
AH:Nation
  Egypt
AH:Machine
  Ramp AH:inventedBy Egypt
  Lever AH:inventedBy Egypt
MA:Area
  East
MA:Invention
  Compass MA:introducedFrom East
  Gunpower MA:introducedFrom East
  Silk MA:introducedFrom East
  Astrolabe MA:introducedFrom East
ME:Country
  Britain
ME:Techology
  StreamEngine ME:developedIn Britain
Program Code
```

4.3. Scenario 3

People organize resources based on their interests or needs. One resource can be classified differently because their interests or needs are different. The proposed model allows users to create a specification file which contains the lexical definitions and the grammar of their own ontology language and define ontology which represents the meaning of terms and the relationships between those terms by using the ontology language. If the web resources

are reorganized based on their own ontologies, the users can conveniently navigate the web resources according to their interests or needs without wasting a lot of time.

As an example, assume that a user wants to search web resources in an Internet art museum. An Internet art museum has lots of art works and users navigate them based on their interests. Some search the art works of the artists who they like such as van Gogh, Picasso, Millet, etc. Others search the art works according to the trend of art such as realism, impressionism, cubism, etc. The others search the art works of art forms which they are interested in such as drawing, painting, sculpture, etc. If the artworks re-organized by their interests are displayed, the users navigate them conveniently and find out desired resources easily.

If user A wants to browse the art works based on painting styles, the user can use the following ontology.

```
Ontology Painting_Style
Class Work
  Restriction on Property belongTo equals 1
Class Art_Movement
ObjectProperty belongTo domain Work range Art_Movement

Art_Movement
  Realism
  Impressionism
  Cubism
Work
  Work_1 belongTo Realism
```

If user B wants to browse the art works based on artists, the user can use the following ontology.

```
Ontology Painting_Artist
Class Work
  Restriction on Property belongTo equals 1
Class Artist
ObjectProperty paintedBy domain Work range Artist

Artist
  Picasso
  Van_Gogh
  Millet
Work
  Work_1 paintedBy Millet
```

If user C wants to browse the art works based on art forms, the user can use the following ontology.

```
Ontology Painting_Medium
Class Work
  Restriction on Property madeOf equals 1
```

```

Class Medium
ObjectProperty madeOf domain Work range Medium

Medium
  Fresco
  Oil
  Watercolor
Work
  Work_1 madeOf Oil
    
```

Figure 5 shows how the resources can be organized according to users' interests. .

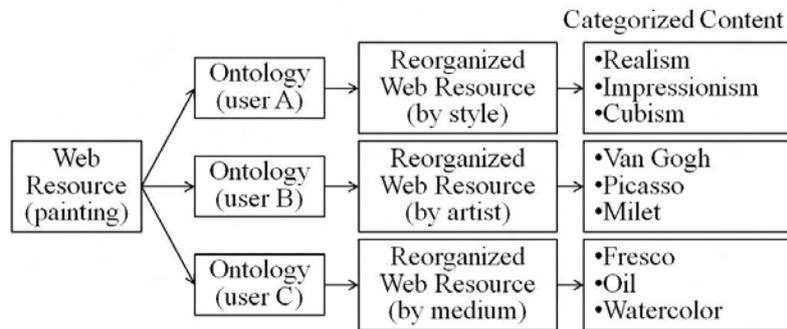


Fig. 5. Organization of resources based on users' interests

4.4. Scenario 4

The proposed model can be used to create Linked Data. Each user can build an ontology about a certain domain by using our model and create instances based on the ontology. Each instance can be regarded as the description about a raw data that each user has and so they are similar to the descriptions of data in Linked Data by using the standards like RDF.

Linked Data is an approach to expose, share, and connect pieces of data, information, and knowledge on the Semantic Web using URIs and RDF. If some users can define a common ontology together, create instances to describe their raw data, and share the instances by using our model, it is possible to construct Linked data. For example, there are people who are interested in art. They open a community in a social network for sharing their data. They first build a general ontology about art domain as follows;

```

Ontology Art
Class Painting
Class Artist
ObjectProperty artist domain Painting range Artist
    
```

```
DatatypeProperty title domain Painting range String
DatatypeProperty year domain Painting range String
```

They also create the instances which describe their data based on the Art ontology as follows;

Table 1. Instances created by each user

User A	User B	User C	User D
Ontology ART	Ontology ART	Ontology ART	Ontology ART
Artist Van_Gogh	Artist Van_Gogh	Artist Picasso	Artist Picasso
Painting work_1	Painting work_2	Painting work_3	Painting work_4
artist Van_Gogh	artist Van_Gogh	artist Picasso	artist Picasso
Painting work_1	Painting work_2	Painting work_3	Painting work_4 title
title Sunflower	title Self_Portrait	title Guernica	Massacre_in_Korea
Painting work_1	Painting work_2	Painting work_3	Painting work_4
year 1889	year 1886	year 1937	year 1951

Then, they share the instances in their community. It is possible kind of linked data services. For example, they can find all data linked to a certain artist such as Van_Gogh or Picasso. Even though each user has small data, they can get an amount of linked data and also create new information from the linked data.

5. Constructing a small scale Semantic Web environment

In this section, we show how a user can construct a small-scale Semantic Web environment using the system we implemented. In order to create a small-scale Semantic Web environment, a user defines a grammar for the ontology language and creates a parser by using SableCC [11] that is a parser generator which creates object-oriented frameworks for building compilers, interpreters, and other text parsers. For describing the ontology, the user needs to create a SableCC specification file which contains the lexical definitions and the grammar productions of an ontology language.

Figure 6 shows how a user can construct a small-scale Semantic Web environment using the system. First, a user defines an ontology language grammar. The grammar file is written in a SableCC specification file format. Then, the user creates an ontology compiler by launching SableCC on the grammar file. The user writes and compiles Java sources for an ontology compiler that checks grammatical errors of the ontology defined by the ontology language and is aware of its classes and their relationships. If there are no errors, an ontology in XML format is written. The user writes and saves the instances in a XML document if it is defined based on the classes and properties of the defined ontology.

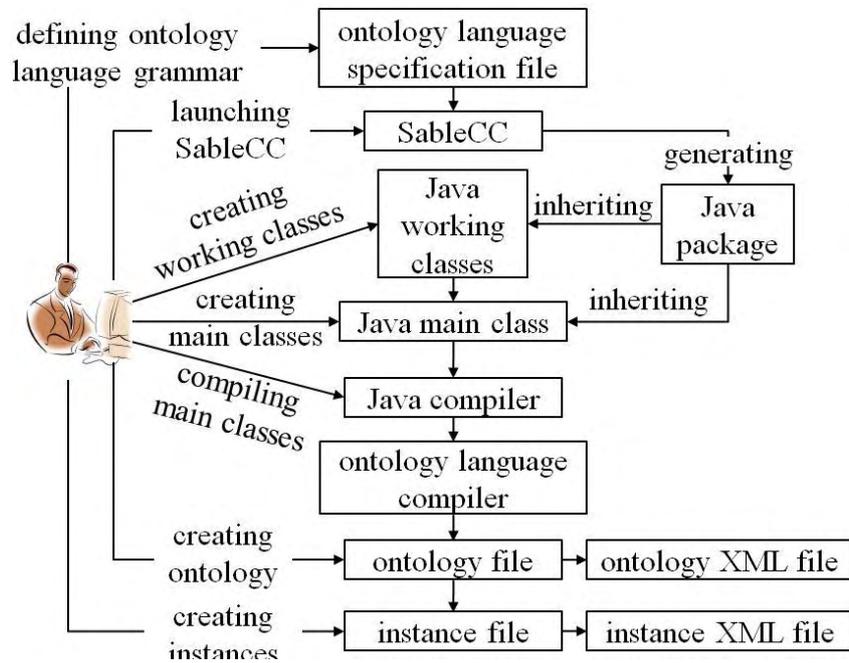


Fig. 6. Steps for constructing a small-scale Semantic Web environment.

Figure 7 shows the screenshot of the user interface captured when a user creates an ontology.

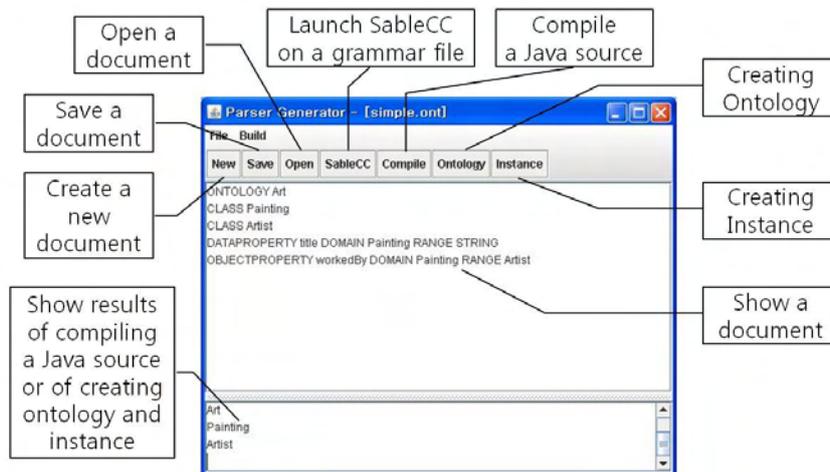


Fig. 7. Art ontology and its instances

Figure 8 shows the situation used in the example that follows.

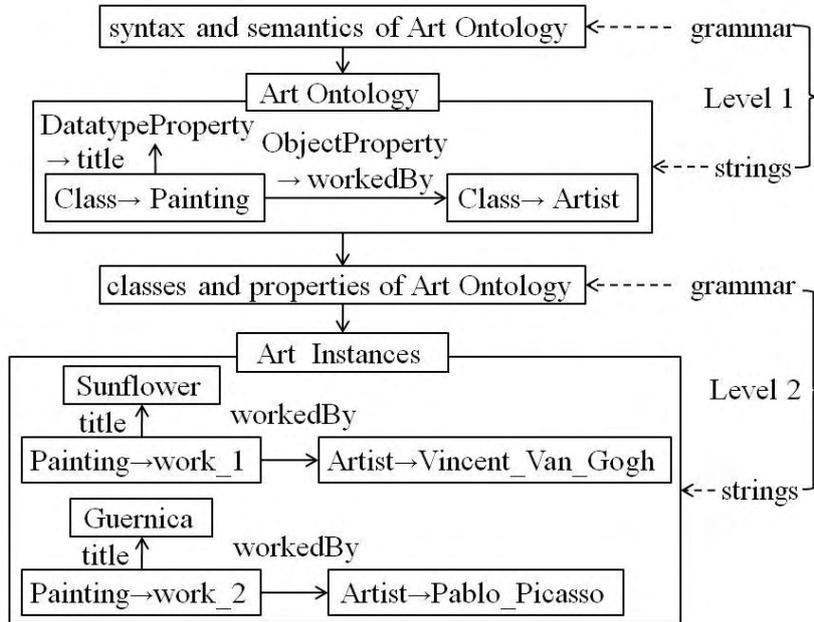


Fig. 8. Art ontology and its instances

We assume that the environment consists of an Art ontology and its instances. More specifically, in the Art ontology, there are a Painting class and an Artist class. The title datatype property defines the Painting class as its domain and a string type as its range. An individual of the Painting class has a title value. The workedBy object property defines the Painting class as its domain and the Artist class as its range. An individual of the Painting class has an individual of the Artist class as its workedBy value. Instances of of Art ontology are created by creating two individuals and assigning their properties. We define a Painting with an ID of work_1 and specify that it is worked by (workedBy) Vincent_Van_Gogh and its title is Sunflower. We also define a Painting is an ID of work_2 and specify that it is worked by (workedBy) Pablo_Picasso and its title is Guernica.

The steps for constructing a small-scale Semantic Web environment are as follows.

1. A user defines the grammar of an ontology language to be compiled and saves it as a specification file. Figure 9 shows the screenshot of the the specification file.

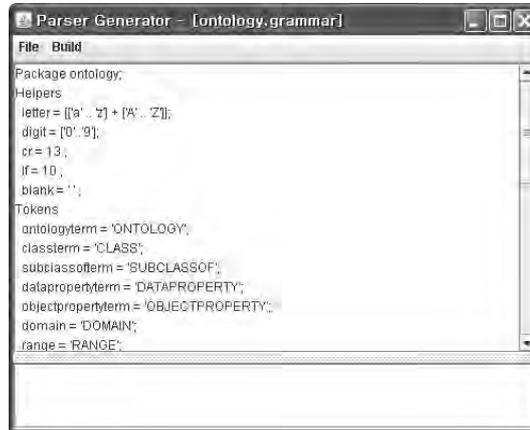


Fig. 9. Screenshot of the specification file

2. The user launches SableCC on the specification file by clicking [Build]-[Launch] (Figure 10). It generates a framework which consists of four packages such as lexer, parser, node and analysis.

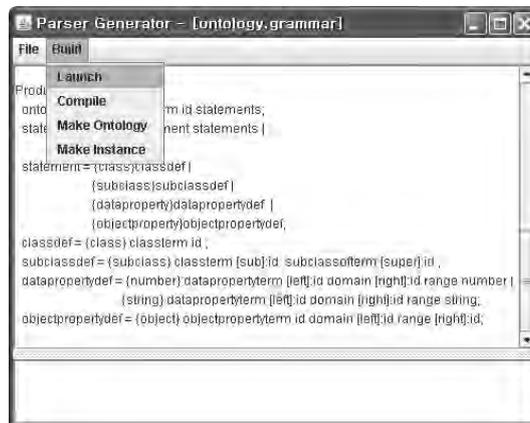


Fig. 10. Launching SableCC on the specification file

3. The user creates working classes which inherit fields and methods from the classes of the Java packages. The working classes contain the core compiler functionalities. If an input is an ontology file, it finds classes and properties of the ontology and saves them and their relationships as an XML file. If an input is an instance file, it finds individuals, their properties and values, and saves them as an XML file.
4. The user also creates a main compiler class which activates lexer, parser, and working classes. The main class reads an ontology file which is

defined by the user. If an input is an ontology file, it checks whether the ontology is defined according to the grammar of the ontology language. If an input is an instance file, it checks whether the instances are defined based on the vocabulary of the ontology.

5. Then, the user compiles the main compiler with a Java compiler and the application generates an ontology language compiler. If the main compiler has any error, the user can debug it. In this example, the user saves a compiler program as "Main.java" and compiles it by clicking [Build]-[Compile].
6. The user creates an ontology which contains classes and properties about a domain according to the ontology language grammar and compiles it with the ontology language compiler. The application generates an ontology XML file. For example, the user defines an Art ontology and saves it as "simple.ont". The ontology can be created if the user clicks [Build]-[Make Ontology]. If there is no syntax and semantic error, the system produces an ontology.
7. The user also creates its instances which are defined by the classes and properties of the ontology and compiles it with the ontology language compiler. The application generates an instance XML file. In this example, the user defines instances based on Art ontology and saves it as "simple.ins". The user creates instances based on the ontology by clicking [Build]-[Make Instance]. If the instances are defined by using the classes and properties of the ontology, the system produces an ontology (Figure 11).

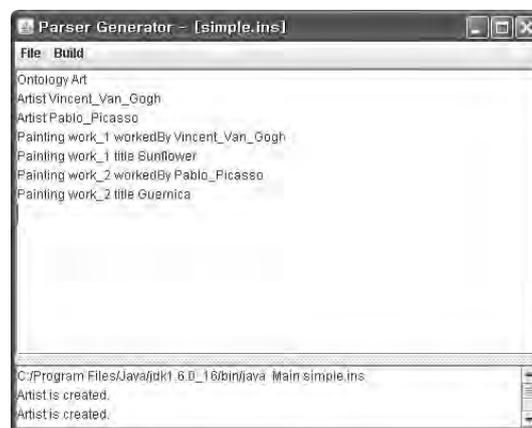


Fig. 11. Creation of instances

6. Comparison with other ontology development tools

In this section, we compare the proposed system with two well-known ontology development tools, Protégé [13] and Apollo [27]. Specifically, we show how the scenario given in section 5 can be realized using Protégé and Apollo in section 6.1 and 6.2, respectively. In addition, we show parts of an ontology and an instance from three systems to illustrate the differences in section 6.3. The following table summarizes briefly the differences.

Table 2. Comparison between Protégé-OWL, Apollo, and our system

System	Language	Prerequisite	Usage
Protégé-OWL	OWL/RDF	Understanding OWL/RDF(S) vocabularies	Building general ontologies & their instances
Apollo	OKBC model	Understanding OKBC Knowledge Model	Building general ontologies & their instances
Our system	User-defined language	Understanding formal languages & parsing	Constructing a semantic web environment easily

Protégé-OWL and Apollo are developed for implementing metadata of ontology using the languages used to encode the ontology. They generally require users to be trained for the languages, knowledge representation, and predicate logic. For example, Protégé-OWL supports the Web Ontology Language (OWL) and exports ontologies to OWL/RDF (Resource Description Framework). It requires users understand the vocabularies of RDF(S) and OWL and their functions. Apollo is a knowledge modeling application based on the internal model of the OKBC (Open Knowledge Base Connectivity) protocol and export ontologies to CLOS (Common LISP Object System) and OCML (Options Configuration Modeling Language). It also requires users understand the meaning of each concept, the operations, and the naming and argument conventions provided in OKBC specifications. Our system allows users to use languages that they define. It requires the users to have a basic understanding of formal languages and parsing which undergraduate students generally learn from a compiler course.

Protégé-OWL and Apollo are developed for all stages of the ontology lifecycle such as creation, population, validation, deployment, maintenance and evolution. However, our system is developed for undergraduate students to construct an environment that is structurally similar to the Semantic web which consists of ontologies and their instances so that they can understand the structural properties of the Semantic Web while studying the Semantic Web. The combination of theory and practice can help them understand the Semantic Web clearly. On top of this, they can also conduct

experimentations on Semantic Web applications that run in the environment so constructed.

The ontologies in Protégé-OWL or Apollo are represented by using general languages or knowledge representation such as RDF(S), OWL, OKBC model, etc. They can be reused and shared with other applications using the same language. However, users should understand the technical terminologies of the language or its specification and it can be difficult [28,29]. In our system, users can simply build ontologies and their instances which do not use technical terminologies and logic. Even though the ontologies are not represented by the general ontology language, the system can be helpful for the undergraduate students who have basic knowledge about computer science to understand the Semantic web. In addition, although our system now represents the ontology by XML, it can be easily extensible to represent the ontology by OWL/RDF(S).

6.1. Protégé-OWL

The steps to create an Art ontology given in section 5 are as follows.

1. We start Protégé-OWL and create a new OWL project by clicking “Create New Project”. When “Create New Project” wizard appears, we select a project type, “OWL/RDF Files” and specify a unique URI that will become the identifier for the ontology. Then, we select an OWL/RDF dialect such as OWL DL.
2. We create classes for concepts in the ontology. We select the “OWL Classes” tab. It shows the hierarchy of classes. All the classes will be created subordinate to owl:Thing. We click the Create subclass button. A class is created with a generic name such as “Class_1”. We rename the class using the “class name widget” to “Artist”.

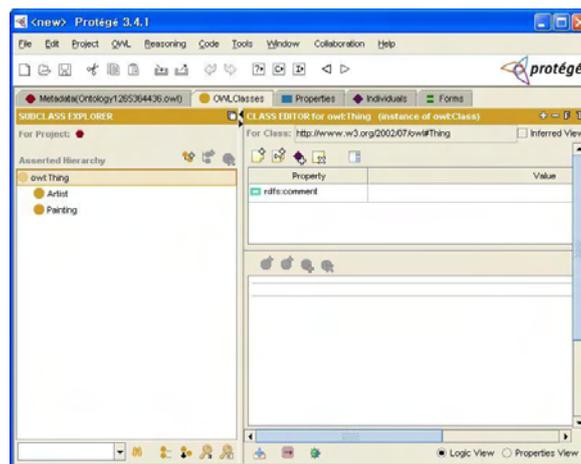


Fig. 12. Creation of classes

We repeat the previous step to add the class “Painting” (Figure 12).

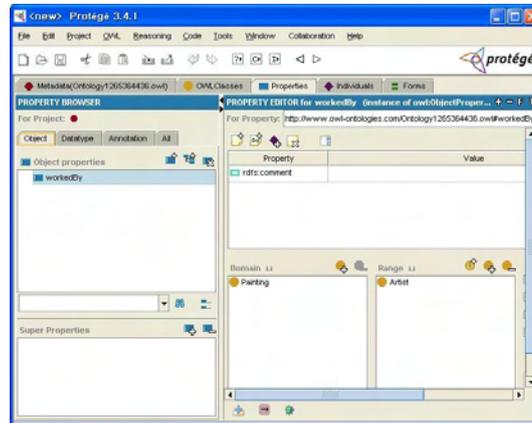


Fig. 13. Creation of an Object property

3. We create properties of these classes, for example, the title of the painting and the artist that painted it. We switch to the “Properties” tab. We click the “Create Object Property” button to create a new Object property. An Object property is created with a generic name. We rename the property to “workedBy”. Then, we specify a domain and a range of the Object property. We press the “Add named class” button on the “Domain Widget” and select the class “Painting”. We also press the “Add named class” button on the “Range Widget” and select the class “Artist” (Figure 13).

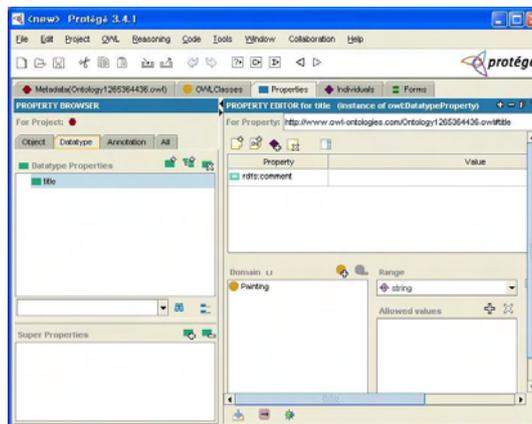


Fig. 14. Creation of a Datatype property

4. We click the “Create Datatype Property” button to create a new Datatype property. A Dataproperty is created with a generic name. We rename the

property to “title”. Then, we specify a domain and a range of the Datatype property. We press the “Add named class” button on the “Domain Widget” and select the class “Painting”. We select the item “string” on the “Range Widget” (Figure 14).

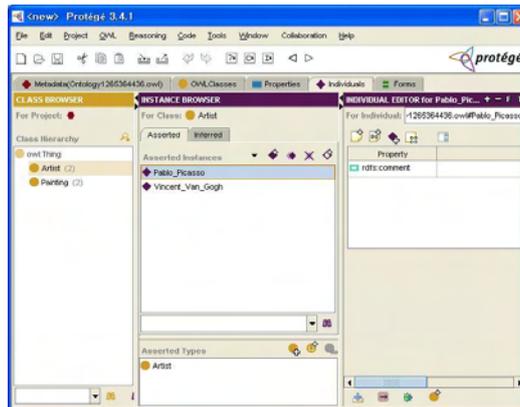


Fig. 15. Creation of instances of the class Artist

5. We create some instances of the classes. We switch to the “Individuals” tab. We select the class “Artist”. We press the “Create Instance” button. An instance is created with a generic name. We rename the instance to Vincent_Van_Gogh. We also create another instance called “Pablo_Picasso” (Figure 15).

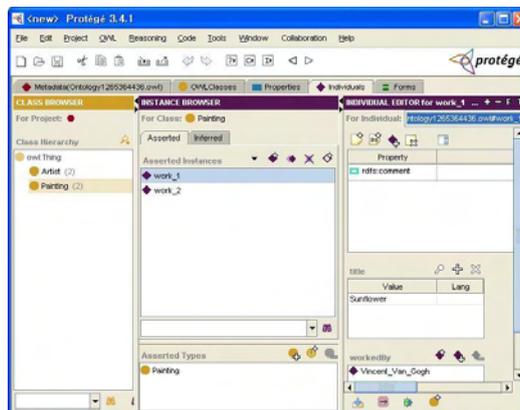


Fig. 16. Creation of instances of the class Painting

6. We select the class “Painting” and press the “Create Instance” button. An instance is created with a generic name. We rename the instance to “work_1”. We press the “Add new value” button in the Datatype property

“title” and type “Sunflower” as its value. We also press the “Add new value” button in the Object property “workedBy” and select the instance “Vincent_Van_Gogh” as its value.

7. We repeat step 6 to create another instance called “work_2”. We press the “Add new value” button in the Datatype property “title” and type “Guernica” as its value. We also press the “Add new value” button in the Object property “workedBy” and select the instance “Pablo_Picasso” as its value (Figure 16).

6.2. Apollo

The steps to create an Art ontology given in section 5 are as follows.

1. We start Apollo and a new project. We click “Create New Project” and open the “Create new ontology” dialog and enter its name, “Art”.
2. We create classes for concepts in the ontology. We open the “New class” dialog in the focused ontology and type the class name “Artist”. We repeat the previous step to add the class “Painting” (Figure 17).

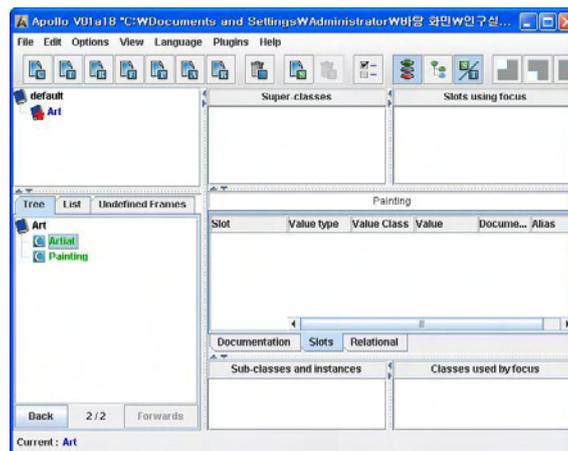


Fig. 17. Creation of classes

3. We create property slots of the classes. A slot contains a number of facets such as value, value type, and value class, etc. We open the “New slot” dialog in the class “Artist”. We type its name “workedBy” and set its value type “instance” and its value class “Painting”. We also open the “New slot” dialog in the class “Painting”. We type its name “title” and set its type “string” (Figure 18).

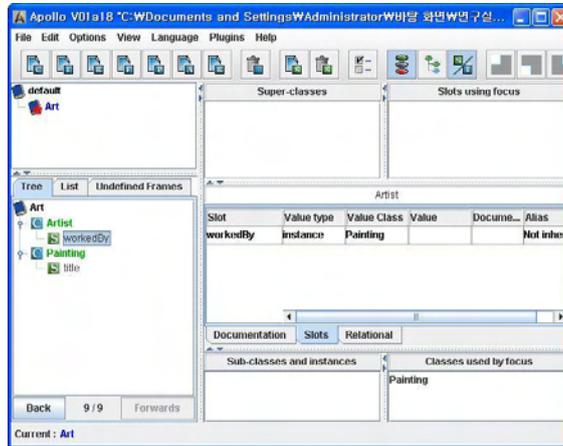


Fig. 18. Creation of property slots

4. We create some instances of the classes. We select the class “Artist” and open the “New instance” dialog. We type the instance name “Vincent_Van_Gogh” and its type “Artist”. We also open the “New instance” dialog again. We type the instance name “Pablo_Picasso” and set its type “Artist” (Figure 19).

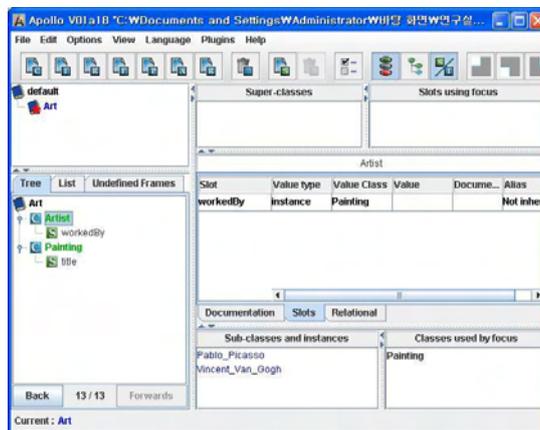


Fig. 19. Creation of instances of the class Artists

5. Similarly, we select the class “Painting” and open the “New instance” dialog. We type the instance name “work_1” and set its type “Painting”. We also open the “New instance” dialog again. We type the instance name “work_2” and set its type “Painting”.
6. We specify the facets of the instances of the class “Artist”. We select the class “Artist” and the instance “Vincent_Van_Gogh” in sub-classes panel.

We double-click the value of the slot “workedBy” and select the instance “work_1”. We select the instance “Pablo_Picasso” in sub-classes panel. We double-click the value of the slot “workedBy” and select the instance “work_2” (Figure 20).

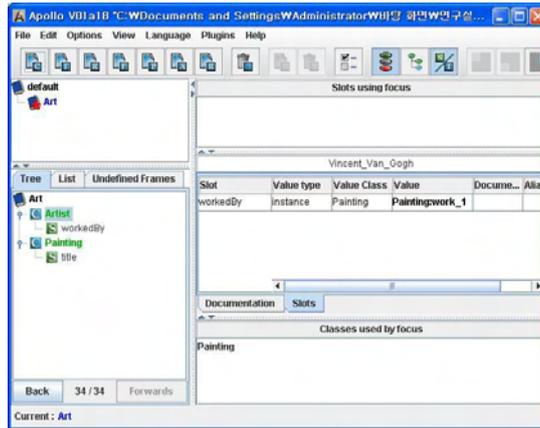


Fig. 20. Specifying the facets of the instances of the class Artist

7. Similarly, we specify the facets of the instances of the class “Painting”. We select the class “Painting” and the instance “work_1” in sub-classes panel. We edit the string value of the slot “title” to “Sunflower”. We select the instance “work_2” in sub-classes panel. We edit the string value of the slot “title” to “Guernica”.

6.3. Comparison of the results

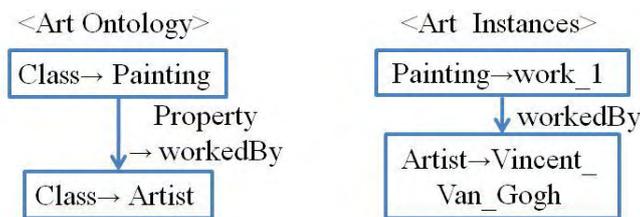


Fig. 21. Two classes and instances

To illustrate the differences and similarities among our system, Apollo, and Protégé-OWL, we extract parts of the documents where they save the ontology and its instances. Protégé-OWL saves them in OWL format, and Apollo and our system do in XML format. There are two classes “Artist” and “Painting” and the property “workedBy”. There are also instances “work_1”

and “Vicent_Van_Gogh” which are connected with the property “workedBy” (Figure 21).

The following table shows the ontology generated from the three systems.

Table 3. Ontologies generated by Protégé-OWL, Apollo, and our system

Task	Apollo (XML)	Protégé-OWL (OWL)	Our system (XML)
Creating Ontology	<pre><classes> <class name="Artist"/> <class name="Painting"/> <slots> <slot name="workedBy"> <type value="instance"/> <is_own value="false"/> <value_class value="Artist"/> <value_type value="instance"/> </slot> </slots> </classes></pre>	<pre><owl:Class rdf:ID="Painting"/> <owl:Class rdf:ID="Artist"/> <owl:ObjectProperty rdf:ID="workedBy"> <rdfs:range rdf:resource="#Artist"/> <rdfs:domain rdf:resource="#Painting"/> </owl:ObjectProperty></pre>	<pre><CLASS name='Painting'/> <CLASS name='Artist'/> <OBJECTPROPERTY Name='workedBy'> <DOMAIN name='Painting'/> <RANGE name='Artist'/> </OBJECTPROPERTY></pre>

The ontological instances generated from the systems are as follows.

Table 4. Ontological instances generated by Protégé-OWL, Apollo, and our system

Task	Apollo (XML)	Protégé-OWL (OWL)	Our system (XML)
Creating Instance	<pre><instance name = "Vincent_Van_Gogh"Class = "Artist"/> <instance name ="work_1" class="Painting"> <slots> <slot name="title"> <value value="Sunflower"/> </slot> <slot name="workedBy"> <value value="Artist: Vincent_Van_Gogh"/> </slot> </slots> </instance></pre>	<pre><Painting rdf:ID="work_1"> <workedBy> <Artist rdf:ID="Vincent_Van_Gogh"/> </workedBy> </Painting></pre>	<pre><Artist name="Vincent_Van_Gogh"> <Painting name="work_1"> <workedBy value="Vincent_Van_Gogh"> </workedBy> </Painting></pre>

7. Conclusions and Future Works

In this paper, we presented a two-level model for the Semantic Web. The model consists of two grammars, where one grammar is used to model ontologies and the other grammar is used to model ontological instances. We implemented a system by which a user can easily construct a small-scale Semantic Web environment.

Our model can be utilized as follows.

First, a personalized semantic web can be easily constructed. The Semantic Web is a linked information space where data is being enriched and added based on the standards to formalize the syntactic and semantics of web contents. It encourages users to create, share, and reuse resources related to their needs and interests. Especially, the rapid increase of communities promotes the interaction with each other and development of a shared repository of resources. However, it is not easy for average users to handle the languages to construct the semantic web such as RDF or OWL. Our model enables the users to design an ontology language for a domain of their interests and represent web contents by using the language. The users can represent their own contents and connect to others in a shared domain of interests easily because they use languages that are easier than RDF or OWL. They can construct their personalized semantic webs.

Second, constructing knowledge is easily done. People construct their knowledge by connecting existing knowledge into new knowledge, but their knowledge construction is different each other. Although they are given the same resources, they organize them in different ways because they have various views about the resources. The proposed model enables users to define and utilize resources according to their views easily. They define ontologies and describe resources based on the ontologies to organize the resources. They can reuse resources made by others in new and exciting contexts as well. It can help them build knowledge.

Third, a semantic social network can be easily created. In the Semantic Web, users create online communities where they can create, collect and share resources. Especially, a social network is a community where members with a shared interests interact and develop shared contents. If they can construct a small semantic web suited to their own community, they can represent resources semantically and share meaningful information. The proposed model enables them to construct a semantic social network according to their interests and needs.

We are currently investigating ways by which a logical inference mechanism can be supported in the proposed model. We are also working on a tool that can exploit the structural properties of the Semantic Web such as Magpie [25], Piggy Bank [12], Potluck [26], etc. using a structurally similar environment to the Semantic Web created by our system.

References

1. McGuinness, D. L. and van Harmelen, F.: OWL Web Ontology Language Overview, W3C Recommendation, <http://www.w3.org/TR/owl-features>. (2004)
2. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F. and Stein, L. A.: OWL Web Ontology Language Reference, W3C Recommendation, <http://www.w3.org/TR/owl-ref>. (2004)
3. Rodriguez, M. A. and Bollen, J.: Modeling Computations in a Semantic Network, Computing Research Repository (CoRR), ACM, abs/0706.0022. (2007)
4. Antoniou, G., and van Harmelen, F.: A Semantic Web Primer, The MIT Press (2004)
5. Berners-Lee, T., Hendler, J., and Lassila, O.: The Semantic Web, Scientific American Special online Issue. (2001)
6. Heflin, J., Volz, R. and Dale, J.: Web Ontology Requirements, Proposed W3C Working Draft, <http://km.aifb.uni-karlsruhe.de/projects/owl/index.html>. (2002)
7. Decker, S., Erdmann, M., Fensel, D. and Studer, R.: Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information, Kluwer Academic Publishers. (1998)
8. Patel-Schneider, P. F., Hayes, P. and Horrocks, I.: OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation, <http://www.w3.org/TR/owl-semantics>. (2004)
9. Pan, Z., Qasem, A. and Heflin, J.: An Investigation into the Feasibility of the Semantic Web, In Proceedings of the Twenty First National Conference on Artificial Intelligence (AAAI 2006). (2006)
10. Schraefel, M. C.: What is an analogue for the semantic web and why is having one important?, In Proceedings of the eighteenth conference on Hypertext and hypermedia, 123-132. (2007)
11. Gagnon, E. M. and Hendren, L. J.: SableCC, an Object-Oriented Compiler Framework, In Proceedings of the Technology of Object-Oriented Languages and Systems, IEEE Computer Society. (1998)
12. Huynh, D., Mazzocchi, S., and Karger, D.: Piggy Bank: Experience the Semantic Web Inside Your Web Browser, International Semantic Web Conference. (2005)
13. <http://protege.stanford.edu>
14. Martin, B., Mitrovic, A., Suraweera, P.: ITS Domain Modelling with Ontology. Journal of Universal Computer Science, Vol. 14, No. 17, 2758-2776. (2008)
15. Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., Wenke, D.: OntoEdit: Collaborative Ontology Development for the Semantic Web. Proceedings of ISWC 2002, Springer, LNCS 2342, 221-235. (2002)
16. Mizoguchi, R.: Tutorial on ontological engineering - Part 2: Ontology development, tools and languages, New Generation Computing, OhmSha&Springer, Vol.22, No.1, 61-96. (2004)
17. Corcho, O., Fernandez-Lopez, M., Gomez-Perez, A., Vicente, O.: WebODE: An Integrated Workbench for Ontology Representation, Reasoning and Exchange, Proceedings of EKAW2002, Springer, LNAI 2473, 138-153. (2002)
18. Bizer, C., Cyganiak, R., Heath, T.: How to Publish Linked Data on the Web. <http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial>. (2007)
19. Berners-Lee, T.: Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>. (2009)
20. Boss, G., Malladi, P., Quan, D., Legregni, L., Hall, H.: Cloud Computing. IBM Corporation (2007)

Hyosook Jung and Seongbin Park

21. SHEU, P. C-Y, Wang, S., Wang, Q., Hao, K., Paul, R.: Semantic Computing, Cloud Computing, and Semantic Search Engine. IEEE International Conference on Semantic Computing, 654-657. (2009)
22. <http://www.cloudforum.org>
23. http://en.wikipedia.org/wiki/Social_Semantic_Web
24. http://syntheticbiology.org/Semantic_web_ontology/Software.html
25. <http://projects.kmi.open.ac.uk/magpie/main.html>
26. <http://simile.mit.edu/potluck>
27. <http://apollo.open.ac.uk>
28. Wang, H., Horridge, M., Rector, A.: Debugging OWL-DL Ontologies: A Heuristic Approach, Proceedings of ISWC 2005, LNCS 3729 (2005)
29. Chung, M., Oh, S., Kim, K., Cho, H., Cho, H.: Visualizing and Authoring OWL in ezOWL, International Conference on Advanced Communication Technology, 528-531. (2005)

Hyosook Jung received bachelor's degree in the department of elementary education from Seoul National University of Education in Seoul, Korea, master's degree in the department of elementary computer education from Seoul National University Graduate School of Education in Seoul, Korea and doctoral degree in the department of computer science education from Korea University in Seoul, Korea. Her research interests include Semantic Web, adaptive hypermedia and computer science education.

Seongbin Park received bachelor's degree in the Department of Computer Science from Korea university in Seoul, Korea, and both master's degree and doctoral degree in the department of computer science from the University of Southern California. He is currently an associate professor at the department of computer science education of Korea University in Seoul, Korea. His research interests include Semantic Web, adaptive hypermedia and computer science education.

Received: October 11, 2009; Accepted: February 23, 2010.

Ontology-based multi-label classification of economic articles

Sergeja Vogrinčič¹ and Zoran Bosnić²

¹ Jožef Stefan International Postgraduate School,
Jamova 39, 1000 Ljubljana, Slovenia
sergeja.sabo@mps.si

² University of Ljubljana, Faculty of Computer and Information Science
Tržaška cesta 25, 1000 Ljubljana, Slovenia
zoran.bosnic@fri.uni-lj.si

Abstract. The paper presents an approach to the task of automatic document categorization in the field of economics. Since the documents can be annotated with multiple keywords (labels), we approach this task by applying and evaluating multi-label classification methods of supervised machine learning. We describe forming a test corpus of 1015 economic documents that we automatically classify using a tool which integrates ontology construction with text mining methods. In our experimental work, we evaluate three groups of multi-label classification approaches: transformation to single-class problems, specialized multi-label models, and hierarchical/ranking models. The classification accuracies of all tested classification models indicate that there is a potential for using all of the evaluated methods to solve this task. The results show the benefits of using complex groups of approaches which benefit from exploiting dependence between the labels. A good alternative to these approaches is also single-class naive Bayes classifiers coupled with the binary relevance transformation approach.

Keywords: ontology, multi-label classification, machine learning, text categorization, economics, document classification.

1. Introduction

Classification of textual data has become increasingly important during the last decade, along with its many applications on the World Wide Web. People are using intelligent agents to find content of their interest as well as the articles pertaining to their research fields. Traditionally, the librarians, authors and field experts were in charge of categorizing documents with keywords, numerical classifications and other metadata which is used to summarize the documents' contents as well as to enable more efficient document retrieval through keyword search.

To ease the task of the document classification and retrieval in various domains, the usage of automatic approaches is welcome. By replacing the tedious work of manually categorizing documents, automatic document

classification can utilize computer resources to perform the task more efficiently. In addition to more efficient execution of the task, the automatic approach can be in practice used to categorize large sets of documents which have not been annotated in the past, hence enabling their automatic retrieval which was not possible so far.

In this paper, we address a problem of automatic document classification, focusing on the economic domain in particular. Given a corpus of economic scientific papers' abstracts, we aim at finding the answers to the following questions:

1. How to define the possible classes for the documents in the corpus, having no prior knowledge about their contents? We approach this challenge by utilizing a tool for semi-automatic ontology generation, using which we detect and define the most notable concepts in the domain, represented by the corpus. In the following, we use these concepts as possible document classes.

2. What are the appropriate classification algorithms, suitable for this task? In real-life scenarios, documents are frequently annotated with more than a single category. With our aim of classifying such document, multi-label classification approaches, which can predict many classes simultaneously, are required. In addition to relevance of multiple document categories, a hierarchical dependence of the categories may exist in a domain. Based on these facts, in our experimental work we comparatively evaluate: (i) adapted single-class classifiers, coupled with a data transformation approach which enables predicting multiple classes simultaneously [1], (ii) specialized models for multi-label prediction and (iii) a specialized hierarchical multi-label classification model and a ranking model.

We expect that the performance of the machine learning approach should justify the usage of the automatic document classification approach in this domain. Additionally, we also expect the specialized models to perform better than the adapted single-class classifiers.

The paper is organized as follows. In Section 2, we summarize the related work on ontologies (tools, economic ontologies), multi-label classification and text classification approaches. In Section 3, we describe our construction of the ontology of economics using a semi-automatic ontology tool and other text mining software. In Section 4, we empirically evaluate the performance of various types of the multi-label prediction approaches and present their results. We conclude the paper in Section 5, where we present the ideas for further work as well.

2. Related work

As noted in the Introduction, in our work we combine the field of semi-automatic ontology generation with the field of supervised multi-label classification in machine learning. In the following we present the related work in both fields.

2.1. Ontology tools and economic ontologies

Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain. Ontologies have become increasingly important research topics in many areas, dealing with a particular domain structure, categories, entities and their inter-relations. Ontology consists of *concepts*, their *hierarchical relations*, their *additional arbitrary relations*, and *axioms*. Additionally, it may also contain other *constraints* and *functions* [2]. The connection between nodes within ontology can be represented by a graph, namely ontology chart.

Ontology creation and management tools. Ontologies can be constructed, structured and managed either manually or using a certain degree of automatism [3]. Among the latter techniques, we can distinguish between semi-automatic and fully automatic methods for construction, maintenance and evolution strategy of the ontology, depending on the interaction they require from the user. The most of the methods are semi-automatic, being implemented as specialized ontology tools.

Currently, there are many such tools for managing an ontology, such as *Protégé* [4], *OntoEdit* [5], *Ontolingua ontology development environment* [6], *Chimaera* [7] and *OntoGen* [8, 9]. *OntoGen* is a semi-automatic and data-driven ontology editor focusing on editing of topic ontologies. The system combines the text mining techniques with an efficient user interface to bridge the gap between the complex ontology editing tools and the domain experts who are constructing the ontology. Due to efficient integration of text mining techniques with the ontology authoring, we chose *OntoGen* as a tool for creation of the economic ontology, which we describe in Section 3.

Ontological representations of economics. Although some works on economic ontologies exist, none of them systematically covers the whole field of economics, but focuses mainly on the ontological representation of some economic sub-area pertaining to the article topic. Zuniga [10] deals with the ontology of economic objects. The author describes economic categories and laws that provide the conditions for settling objectively whether individuals' views about an instance of any category indeed correspond to that category. Blomqvist [3] deals with a fully automatic construction of enterprise ontologies using design patterns and deals with the creating an enterprise ontology for an automotive supplier. Siricharoen [11] uses the economic domain to illustrate to the economists how ontologies work, as well as to guide the computer software developers to understand the basic concepts of economy. There are also additional works in this field, approaching the economics through the views of a philosophical science, social science and politics [12, 13].

As shown in the following section, in our work we construct a new representation of the economic domain, based on the corpus of economic documents. Since the ontology is built using the underlying economic

documents, it will include the concepts from those documents which are suitable for their further categorization.

2.2. Multi-label classification in machine learning

A large body of research in supervised learning deals with the analysis of a single label data, where training examples are associated with a single label l from a set of disjoint labels L (i.e. a single-label classification). However, nowadays training examples in most of application domains are associated with a set of labels $Y \subseteq L$, being therefore multi-labeled. Learning from such examples and predicting their labels therefore calls for multi-label prediction approaches [1,14].

The two major tasks in supervised learning from multi-label data are: multi-label classification (MLC) and label ranking (LR). MLC [14] is concerned with learning a model that outputs a bipartition of set of labels into relevant and irrelevant with respect to a query instance. LR [15], on the other hand, is concerned with learning a model that outputs an ordering of the class labels, according to their relevance to a query instance. In certain classification problems, the labels belong to a hierarchical structure, then we call the task hierarchical classification. If each example is labeled with more than one node of the hierarchical structure, then the task is called the hierarchical multi-label classification [14].

Many real-world problems in the areas of text mining, semantic annotation of images and videos [16, 17], web page categorization [18], music categorization [19], bioinformatics (gene functional analysis, functional genomics) [18, 20, 21, 22], and many others, are multi-label problems. This has attracted attention from many researchers who were motivated to find a number of new applications to solve these problems.

2.3. Text classification

Text classification (also known as text categorization), where each document may belong to several topics (or labels, keywords, categories, classes), is the task of building learning systems capable of classifying text documents into one or more predefined categories or subject codes. Textual data, such as documents and web pages, are frequently annotated with more than a single label. The categorization of textual data is perhaps the most dominant multi-label application. One of the well-known approaches to solve the problem of text classification is BoosTexter proposed by Schapire and Singer [23], which is extended from the ensemble learning method AdaBoost. A Bayesian approach to multi-label document classification proposed by McCallum [24] combines a mixture probabilistic model and the EM algorithm. Ueda and Saito [25] proposed two parametric mixture models (PMMs) for multi-label text classification, where basic assumption under PMMs is that multi-labeled text

has a mixture of characteristic words, appearing in single-labeled text that belong to each category of the multi-categories.

The classification of the textual data from the domain of economics, which is the focus of our paper, is certainly a case of a multi-label classification which has not received much systematic attention so far.

3. Ontology-based assignment of document classes

To construct the ontology of the economic domain, we created a corpus of economic documents. With the aim to include such texts which contain representative phrases and words for this domain, we decided the corpus to include the abstracts of the published papers in the distinguished economic journals (such as *Quarterly Journal of Economics*, *Journal of Economic Literature*, *Journal of Economic Perspective*, *Econometrica* etc.). We used the JSTOR (<http://www.jstor.org/>) online service, which includes contents of over thousand academic journals and other scholarly content, to collect 1015 such abstracts. Although short in size, we expect the abstract to hold enough of condensed information required to categorize the paper into the appropriate categories.

3.1. Construction of the economic ontology

As mentioned in Section 2.1, we used OntoGen to partially make the construction of the ontology automatic. OntoGen [8,9,26] is a semi-automatic and data-driven ontology editor. The system combines text-mining techniques with a user interface to ease and integrate automatic analysis of texts and ontology construction based on the found important keywords and concepts. The authors [26] define the system to be *semi-automatic* (it suggests concepts, relations between them, visualizes instances within a concept and provides a good overview of the ontology through concept browsing, while the user is always in full control of the system's actions and can accept or reject the system's suggestions) and *data-driven* (most of the aid provided by the system is based on the underlying data provided by the user typically at the beginning of the ontology construction; the data provided as a document corpus serves for an automatic extraction of instances for the concept and relation learning).

After creating a document corpus for OntoGen, we utilized the following functionalities of OntoGen to construct the economic ontology using the:

1. k-means clustering (unsupervised learning): we automatically generated a list of possible sub-concepts for concepts of interest by using k-means clustering. We performed the clustering many times for different possible numbers of clusters (sub-concepts) and selected the most

reasonable grouping¹. The sub-concept generation was repeated recursively to create the sub-sub-concepts etc., until the desired granularity of the concepts was reached.

2. querying for a particular concept and active learning: by providing a set of keywords describing the sought concept, the system followed by asking a series of yes/no questions, if particular documents belonged to the concept in question. The system automatically identified the documents that corresponded to the topic, and the selection was further refined by the user-computer interaction through an active learning loop, using a machine learning technique for a semi-automatic acquisition of the user knowledge. The questions were chosen from the instances on the border between being relevant to the query or not and were therefore most informative to the system. The system then refined the suggested concept after each our reply, and we decided when to stop the process based on how satisfied we were with the suggestions. After the concept was constructed, it was added to the ontology as a sub-concept of the selected concept.

3. visualization-based assignment to concepts: OntoGen has a functionality to visualize a document corpus. In the document space the similar documents are visualized as the neighboring points, while the less similar are visualized more apart from each other. By interactively selecting very dense subgroups of the documents in the visualized space and analyzing their keywords, we were able to assign those document groups to a concept in the ontology.

4. manual assignment of documents to concepts: in the last step of ontology construction we manually analyzed which document corresponded to each concept in the ontology, and fine-tuned the document categorizations by: categorizing documents to some other concepts, categorizing documents to additional (more than one) concepts or categorizing documents which were not categorized so far.

After performing the semi-automatic construction steps (1 and 2 above), we therefore additionally improved the ontology by performing manual steps (3 and 4 above). As a result, the hierarchy of concepts (ontology) shown in Figures 1 and 2 was obtained. As shown in the both figures, we developed the hierarchy to the maximum depth of three levels. The lowest level contains 16 different concepts, corresponding to 16 different classes (not considering their parent concepts). Note that a particular document could had been assigned to different categories (using concept querying, visualization-based selection and manual assignment), thus the same document belongs to more than one class which is a scenario for the multi-label classification.

¹ according to the first author's judgment, who is a bachelor of economics

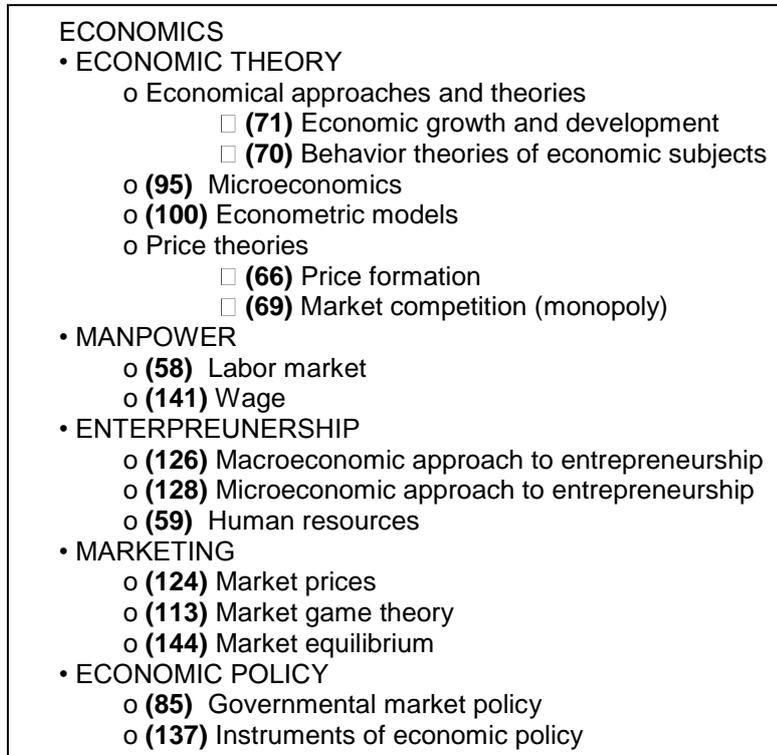


Fig. 1. The hierarchy of the concepts in the ontology of the economic domain. The numbers of the bottom-level concepts denote the number of documents (of total 1015), associated with that category (since each document can be associated to many concepts, the sum of these numbers is higher than 1015).

3.2. Dataset for multi-label learning

After creating the ontology of economics, OntoGen allowed us to export the constructed hierarchy, along with the information which document belongs to which concepts as a set of Prolog² clauses. In parallel, we used the R Project for Statistical Computing software [27] to transform the document corpus into the example-attribute relational dataset, suitable for machine learning.

² Prolog is a general purpose declarative logic programming language associated with artificial intelligence and computational linguistics. Prolog has its roots in formal logic; the program logic is expressed in terms of relations, represented as facts and rules.

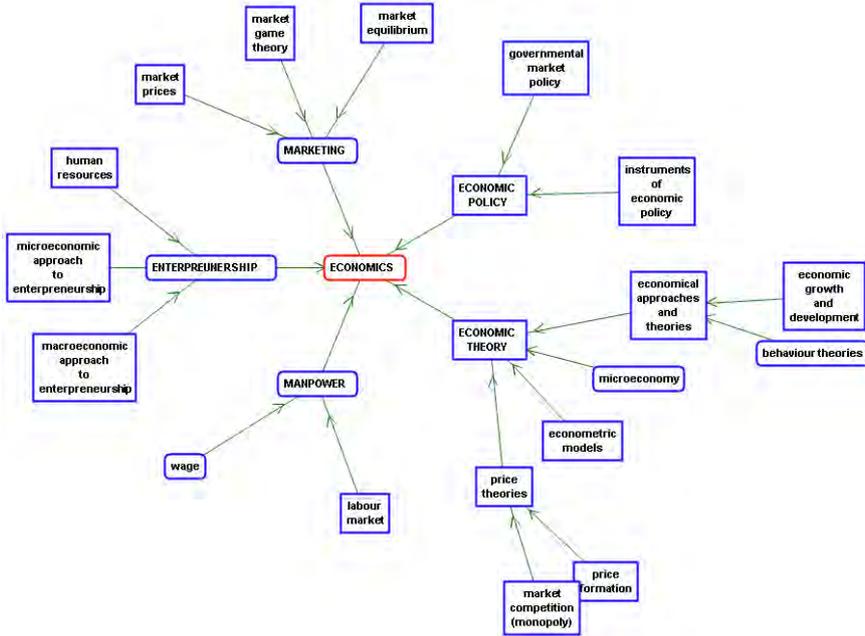


Fig. 2. Visualization of the constructed ontology

For our purpose, we used the R text mining library *tm* [28]. For each document we performed the typical preprocessing operations: converting to lower case, removal of punctuation and numbers, removal of stopwords (common words such as *and*, *the*, *of*, etc.), stemming (removal of different suffixes to keep only the root of the word).

For each document we computed the attributes, representing each appearing word by using the popular weighting schema, the normalized term (word) frequency TFIDF [29] which is based on the logic that the word is more important if it appears several times in a target document, and if it appears in less documents in the corpus:

$$tfidf(w) = tf(w) \cdot \log\left(\frac{N}{df(w)}\right) \quad (1)$$

where $tf(w)$ denotes the term frequency (number of word occurrences in a document), $df(w)$ denotes the document frequency (number of documents containing the word), and N denotes the number of all documents.

The computed attributes were combined with the corresponding document classes which were exported from OntoGen as described in the first paragraph. After performing the feature selection using the information gain measure, we selected the 100 best evaluated attributes out of 4991 total, yielding the final dataset ready for the multi-label classification.

In the following, we tested the performance of various multi-label classification models on the obtained dataset. Note that the ontology generation tool was only used to automatically label the documents in a corpus, and that the further testing scenario will independently assure unbiasedness of the testing procedure by cross-validating the dataset. In addition, the authors have taken care that the developed ontology covers all general fields of economics, independently of the underlying documents' contents.

4. Multi-label classification

In this Section we test the prediction performance of various multi-label classification approaches, including hierarchical multi-label classifier and classification ranking algorithm. We present and compare the empirical results of the evaluated methods, achieved using the tenfold cross-validation evaluation of the models. We implemented the experiment in the Weka [30] environment, using the additional library Mulan [1] which is an open source Java library for multi-label learning.

In the following, we will use $L = \{\lambda_j, j = 1 \dots q\}$ to denote a finite set of labels in a multi-label learning task, and $D = \{(x_i, Y_i), i = 1 \dots m\}$ to denote a set of multi-label examples, where x_i denotes a feature vector and $Y_i \subseteq L$ the set of labels of the i -th example. To evaluate the performance of the tested methods, we observed measures of their classification accuracy (CA) and the average precision (AP). The CA measure is in the context of multi-label classification a very strict measure, as it requires the predicted set of labels to be an exact match to the true label set:

$$ClassificationAccuracy = \frac{1}{m} \sum_{i=1}^m \#(Z_i = Y_i) \quad (2)$$

where $\#(condition)$ returns the number of occurrences for which the condition holds, Y_i denotes a set of true labels and Z_i denotes a set of predicted labels for the i -th example.

The AP is a ranking measure which evaluates the average fraction of labels that were ranked above a particular label $\lambda \in Y_i$ which actually are in Y_i :

$$AveragePrecision = \frac{1}{m} \sum_{i=1}^m \frac{1}{|Y_i|} \sum_{\lambda \in Y_i} \frac{|\{\lambda' \in Y_i: r_i(\lambda') \leq r_i(\lambda)\}|}{r_i(\lambda)} \quad (3)$$

where m denotes the number of examples, Y_i a set of labels of the i -th example and $r_i(\lambda)$ a ranking of the label λ which was predicted for the i -th example.

While the classification accuracy measures the number of exact matches between the sets of the true and the predicted labels, the average precision is a softer measure. High average precision can be interpreted as an indicator

that among the predicted labels, the true labels were given more priority (in terms of the predicted higher rank or class probability) than the irrelevant labels.

4.1. Transformation to a single-label classification problem

The group of methods which transform the learning set to traditional single-label classification task can be applied to any multi-label classification problem. They transform the learning task into one or more single-label classification tasks in combination with which one can use an arbitrary single-label classifier.

Among possible dataset transformation practices we used the following three approaches [1]:

- instance copy transformation (CO) which replaces every multi-label example (x_i, Y_i) with $|Y_i|$ examples (x_i, λ_i) , for every $\lambda_i \in Y_i$. A single-label classifier that outputs class probability distributions can afterwards be used to learn the ranking and predict the relevant labels for a query instance. To output a bipartition of the relevant and irrelevant labels and thus solve a multi-label classification problem, a threshold needs to be applied to the predicted label probability scores. In our experimental work, the default selected probability threshold was 0.5. In cases where probabilities of all labels were less than 0.5, the most probable class was output as relevant, regardless of its probability score.
- label powerset (LP) which considers each unique set of labels that exists in a multi-label training set as one of the classes of a new single-label classification task. Given a new instance, the single-label classifier of LP outputs the most probable class, which is actually a set of labels.
- binary relevance (BR) which learns q binary classifiers, one for each different label in L . It transforms the original dataset into q datasets $D_{\lambda_j}, j = 1 \dots q$ that are intended to perform binary classification tasks for each label in $L = \{\lambda_j, j = 1 \dots q\}$. The transformed datasets contain all examples of the original dataset labeled positively if the label set of the original example contained λ_i and negatively otherwise. For the classification of a new instance BR outputs the union of the labels λ_i that are positively predicted by the q classifiers.

Figure 3 provides an example which illustrates the effect of the above transformations (CO, LP and BR) on an example dataset. Note that the LP transformation results in a transformed dataset, where each example is labeled with a *single* label which represents a combination of all possible labelings in the label powerset. Additionally, note that the BR transformation results in a set of binary classification datasets on which a separate binary classifier is afterwards trained.

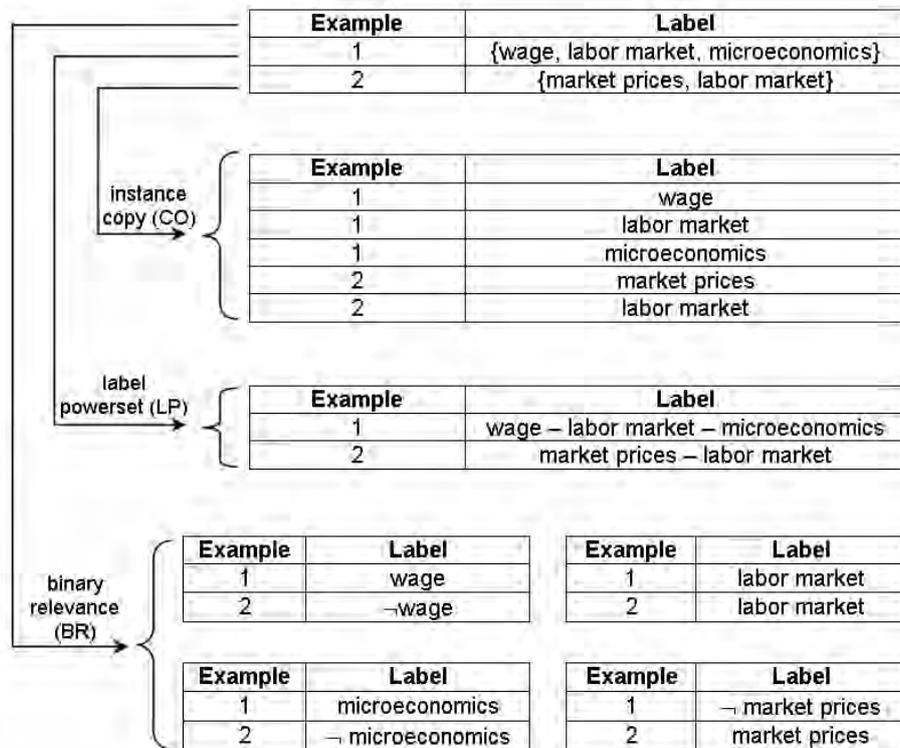


Fig. 3. Illustration of applying data transformation techniques (instance copy, label powerset and binary relevance) on an example multi-label dataset which contains two examples (in our case paper abstracts where each paper is labeled with a set of labels).

Since the above problem transformation approaches can be used with an arbitrary classifier that can output a probability distribution over classes, we coupled each of these approaches with one of the four following classification models:

- support vector machines (SVM) [31]: SVM implementation in Weka which uses a third-degree radial basis function kernel, $\gamma=1/(\text{number of attributes})$, parameter $C=1$ and sequential minimal optimization training,
- decision tree (DT) [32]: implementation of a recursive partitioning tree in Weka (J48) which splits nodes by examining the normalized information gain (difference in entropy) of attributes. Each leaf is assigned a majority class of the examples in the leaf,
- k-nearest neighbors (kNN): an implementation of the instance-based lazy learning algorithm IBk in Weka which uses 5 nearest neighbors for the

classification of a new example. Similarity of two examples is computed based on the Euclidean distance (all attributes are continuous),

- naive Bayes (NB) classifier: multinomial variant of the common naive Bayes probabilistic classifier.

By combining each of the three transformation approaches with each of the four classifiers we therefore tested 12 different scenarios. Since the tested single-class models return the probabilities over class distributions, the labels can be ranked according to their predicted probability. In addition to the classification accuracy (Eqn. 2), their order was therefore evaluated using the average precision (Eqn. 3) as well. The experimental results are shown in Table 1.

Table 1: Performance of four different classifiers (decision tree – DT, support vector machine – SVM, naive Bayes – NB, k-nearest neighbors – kNN) coupled with three problem transformation approaches (binary relevance – BR, label powerset – LP, copy transformation – CO). The best achieved individual and average results are denoted by underlining. The table shows the average values of the performance measures and standard deviations.

CLASSIFICATION ACCURACY				
	BR	LP	CO	average
DT	0.282±0.024	0.311±0.040	0.318±0.048	0.304
SVM	0.053±0.010	0.225±0.034	0.289±0.046	0.189
NB	0.346±0.014	0.352±0.023	<u>0.391±0.035</u>	<u>0.363</u>
kNN	0.195±0.042	0.288±0.039	0.282±0.047	0.255
average	0.219	0.294	0.320	
AVERAGE PRECISION				
	BR	LP	CO	average
DT	0.531±0.022	0.397±0.040	0.450±0.040	0.459
SVM	0.255±0.019	0.374±0.035	0.434±0.035	0.354
NB	<u>0.618±0.024</u>	0.475±0.019	0.527±0.027	<u>0.540</u>
kNN	0.525±0.038	0.275±0.023	0.431±0.033	0.410
average	0.482	0.380	0.461	

We can see from the table that, on the average, the highest CA was achieved using the copy transformation approach (0.320) and the highest AP using the binary relevance transformation (0.482). Among the average performance of the classifiers, the naive Bayes performed the best with the average CA of 0.363 and the average AP of 0.540. By analyzing the results of each combination individually, we can see that the naive Bayes classifier achieved the best CA (0.391, coupled with CO transformation) and AP (0.618, coupled with BR transformation).

Note that classifying into the majority class (the most documents are labeled with a single category {*econometric_models*}) would give a default classification accuracy of 0.080. This means that the usage of all but one tested model-transformation combinations (namely, SVM-BR), outperforms

this default accuracy and shows a potential of using the machine learning algorithms to solve this task.

In the following Section 4.2, we move on to specialized multi-label classification models that can learn from and predict the data in their original multi-label form.

4.2. Multi-label models

The second group of methods extends the specific learning algorithms in order to handle multi-label data directly. In the previous section, we tried to solve multi-label problem by transforming the dataset to enable usage of the single-label models. However, these kinds of methods do not consider the correlations between the different labels. In the field, several approaches especially designed for multi-label learning tasks have been proposed, among which we evaluate the following two:

- back-propagation multi-label neural network (BP-MLL) [33]: This model is derived from the popular back-propagation algorithm through replacing its error function with a new function defined to capture the characteristics of multi-label learning, that is, the labels belonging to an instance should be ranked higher than those not belonging to that instance. The neural network was trained by epoch and contained one hidden layer of 20 neurons,
- multi-label k-nearest neighbors algorithm (ML-kNN) [18]: an adapted *k*-nearest neighbors lazy learning algorithm which retrieves the *k* nearest examples and aggregates the label sets of these examples by combining the statistical information gained from the neighboring instances. The algorithm uses normalized Euclidean distance as a distance function and parameter $k=10$.

The accuracy and the average precision of these two approaches are shown in Table 2.

Table 2: Results of multi-label models BP-MLL and ML-kNN, hierarchical model HOMER and ranking model CLR. The cells contain the average CA and AP values and their standard deviations.

	BP-MLL	ML-kNN	HOMER (hierarchical)	CLR (ranking)
Classification Accuracy	0.207±0.079	0.202±0.031	0.466±0.028	0.365±0.021
Average Precision	0.433±0.107	0.576±0.032	0.627±0.023	0.683±0.012

The results show that the ML-kNN achieves comparable CA to BP-MLL classifier, but outperforms it with better AP (0.576 compared to 0.433). If we compare these results to the ones in the previous subsection, we can see that the CA of the single-label naive Bayes classifier is on the average still better

than CA of any tested multi-label model. However, both multi-label models achieve slightly higher CA than the SVM on the average.

As for the AP, the multi-label classification approach seems to benefit from exploiting the dependence of the labels. Namely, ML-kNN achieves higher AP (0.576) than any of the single-class models on the average (the highest AP was on the average achieved using the naive Bayes – 0.540). However, note that the combination of the naive Bayes and the binary relevance transformation approach still achieves better AP of 0.618.

4.3. Hierarchical and ranking models

In our domain, there is a hierarchical dependence of the categories. This calls for evaluation of the additional models which are able to consider the relations between various labels (classes) in the hierarchy. As only 16 bottom-level ontological concepts were used so far (see Figures 1 and 2), to apply such classification method we need to expand the dataset we used so far with the additional classes, corresponding to the parent concepts in the ontological hierarchy (first-level and second-level concepts). Doing this, we introduce additional 7 classes, giving altogether 23 classes which are given to the learning algorithm along with the information about their inter-relations.

In this section, we evaluate a couple of methods that focus on dealing with problems with large number of labels by decomposing the original multi-label classification problem into a series of simpler problems [34]:

- HOMER (Hierarchy Of Multi-label classifiERs) [35,36] approach decomposes the problem into a hierarchy of simpler problems, where each problem uses a reduced number of possible labels. The hierarchical structure of the labels is obtained by applying recursive clustering to the initial set of labels. The main idea is the transformation of a multi-label classification task with a large set of labels L into a tree-shaped hierarchy of simpler multi-label classification tasks by recursively partitioning the set of labels into a number of nodes using a balance clustering algorithm. Then it builds one multi-label classifier at each node apart from leaves, following the hierarchical binary relevance approach [1]. A calibrated label ranking algorithm was used as an underlying multi-label classifier, and the number of selected clusters was 3.
- The Calibrated Label Ranking approach (CLR) [37] interprets a multi-label problem as a special case of a preference learning problem. Besides using an underlying classifier to rank the labels according to their prediction score, it also uses an additional neutral label which represents a breaking point of the ranking into relevant and irrelevant sets of labels. The binary models that learn to discriminate between the virtual label and each of the other labels correspond to the models of binary relevance. This way CLR can be used to perform the multi-label ranking task. In our experiments, the underlying binary models were built using the J48 decision trees.

The CA and the AP of HOMER and CLR are shown in Table 2. The results show that HOMER outperforms CLR in terms of its CA, but performs worse than CLR in terms of its AP. Compared to the results of the multi-label models (BP-MLL and ML-kNN) we can see that both, HOMER and CLR, achieve better CA and AP. Based on these results we can conclude that the learners from this section benefit from the information on hierarchical class structure in our problem domain.

Additionally, compared to the results of the single-class learners, applied to the transformed datasets (see Table 1), we can see that the CA and AP of HOMER are higher than the average CAs and APs of all models. However, the individual AP denoting the combination of naive Bayes and binary relevance transformation method seems to be only slightly worse than AP of HOMER and CLR. The visual comparison of performance of all tested models is shown in Fig. 4.

5. Conclusion

In the paper, we focused on the task of document classification in the field of economics, proposed an approach to this task and empirically evaluated how the selected machine learning methods are successful performing it.

We approached to the problem of **assigning the classes to documents** in the corpus by collecting a database of 1015 economic paper abstracts and semi-automatically constructing an ontology. We used an ontology tool which, during the construction, automatically distributed the documents among the concepts being recognized from the underlying text documents. Noting a scenario for the **multi-label classification approaches** we evaluated three groups of approaches to classify the documents: transformation to single-class problems, specialized multi-label models, and hierarchical/ranking models.

Classification accuracies of all tested classification models compared to the default majority classifier indicate that there is a potential for using the majority of the evaluated methods to solve this task. An exception to this rule is the SVM-BR model/transformation combination, which performed the worst classification accuracy of 0.053. The hierarchical multi-label classifier HOMER achieved the highest accuracy of 0.466. Given that classification accuracy in the context of multi-label classification is a very strict measure, which measures only the ratio of the exact matches between the true and the predicted labels, we can interpret this as a good result. By analyzing the performance using the average precision which is a less strict measure in terms of requiring the equivalence of the predicted and true label sets, the ranking approach CLR performed the best (average precision of 0.683).

The results show the benefits of using more complex (multi-label models, hierarchical and ranking) approaches, since they benefit from exploiting dependence between the labels. However, a good comparably-performing

alternative to these approaches is a single-class naive Bayes classifier which performed comparably to the best more complex approaches.

Our ideas for the further work include:

1. it shall be considered, how the described approach of ontology-based document classification can be further automatized, not requiring the interaction of the user,
2. alternative feature selection approaches shall be tested, along with the effect of using different numbers of selected attributes,
3. the usefulness of this approach shall be tested in other domains and using alternative document corpora (including larger documents instead of abstracts).

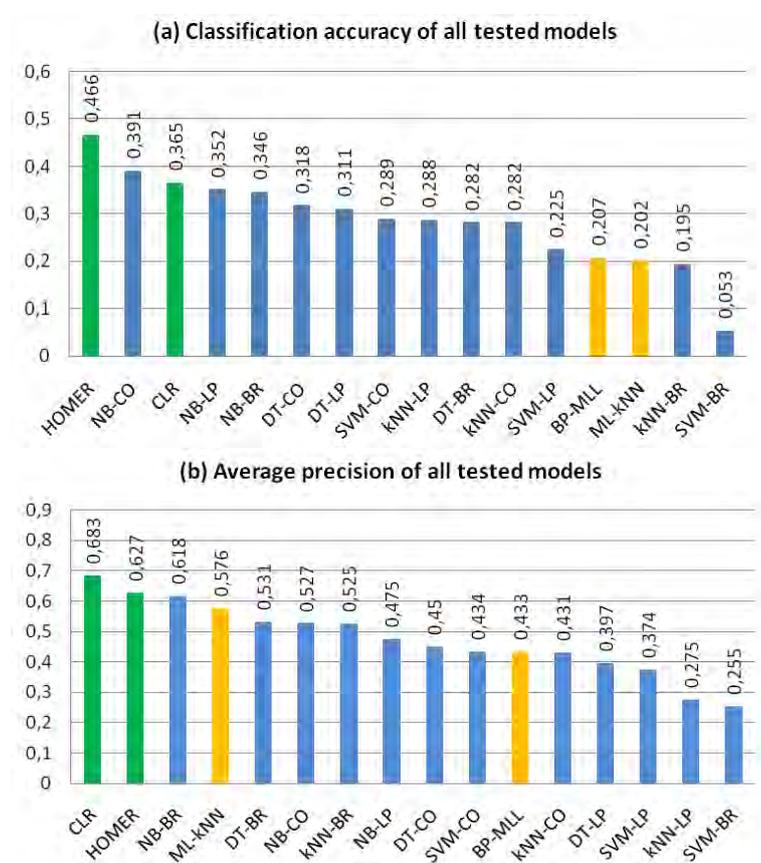


Fig. 4. Graphical representation of the obtained results: classification accuracies (above) and average precisions (below) of all tested models. The results are ranked in the decreasing order of the performance measure.

References

1. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining Multi-label Data. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook*, 2nd edition, Springer, Heidelberg. (2010)
2. Noy, N. F., McGuinness, D. L.: *Ontology Development 101: A Guide to Creating Your First Ontology*. Knowledge Systems, AI Laboratory, Stanford University, No. KSL-01-05. (2001)
3. Blomqvist, E.: *Fully Automatic Construction of Enterprise Ontologies Using Design Patterns: Initial Method and First Experiences*. Lecture Notes in Computer Science, Vol. 3761. Springer-Verlag, Berlin-Heidelberg, 1314-1329. (2005)
4. Stanford Center for Biomedical Informatics Research: Protégé [Online]. Available: <http://protege.stanford.edu/> (current March 2010)
5. Sure, Y., Angele J., Staab, S.: *OntoEdit: Guiding Ontology Development by Methodology and Inferencing*. In: *Proceedings of the Confederated International Conferences on the Move to Meaningful Internet Systems CoopIS DOA and ODBASE 2002*, Lecture Notes in Computer Science, Vol. 2519. Springer-Verlag, 1205-1222. (2002)
6. Farquhar, A., Fikes, R., Rice, J.: *Tools For Assembling Modular Ontologies in Ontolingua*. In: *Proceedings of 14th American Association for Artificial Intelligence Conference (AAAI-97)*. AAAI/MIT Press., Menlo Park, CA, 436-461. (1997)
7. Stanford University: Chimaera [Online]. Available: <http://www-ksl.stanford.edu/software/chimaera/> (current March 2010)
8. Fortuna, B., Grobelnik, M., Mladenić, D.: *Semi-automatic Data-driven Ontology Construction System*. In: *Proceedings of the 9th International multi-conference Information Society IS-2006*, Bohanec, M. et al. (eds.), Ljubljana, 223-226. (2006)
9. Fortuna, B., Grobelnik, M., Mladenić, D.: *OntoGen: Semi-automatic Ontology Editor* [Online]. Available: <http://ontogen.ijs.si/> (current March 2010)
10. Zuniga, G. L.: *An Ontology of Economic Objects*. *American Journal of Economics and Sociology*, Vol. 2, No. 58, 299-312. (1999)
11. Siricharoen, W. V., Puttitanun, T.: *Creating Ontology Chart Using Economy Domain Ontologies*. *International Journal of Digital Content Technology and its Applications*, Vol. 3, No. 3, 74-80. (2009)
12. Mäki, U.: *The Economic World View: Studies in the Ontology of Economics*. Cambridge University Press, UK. (2001)
13. Fullbrook, E.: *Ontology and Economics: Tony Lawson and His Critics*. Routledge, Abingdon, UK. (2009)
14. Tsoumakas, G., Katakis, I.: *Multi-label Classification: An Overview*. *International Journal of Data Warehousing and Mining*, Vol. 3, No. 3, 1-13. (2007)
15. Vembu, S., Gärtner, T.: *Label Ranking Algorithms: A Survey*. In Fürnkranz, J., Hüllermeier, E. (eds.): *Preference Learning*. Springer-Verlag. (to appear 2010)
16. Yang, S., Kim, S.-K., Yong M. R.: *Semantic Home Photo Categorization*. In: *Circuits and Systems for Video Technology*, *IEEE Transactions on*, Vol. 17, No. 3, 324-335. (2007)
17. Qi, G. J., Hua, X. S., Rui, Y., Tang, J., Mei, T., Zhang, H. J.: *Correlative Multi-label Video Annotation*. In: *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, New York, USA, 17-26. (2007)
18. Zhang M.-L., Zhou Z.-H.: *ML-kNN: A Lazy Learning Approach to Multi-Label Learning*. *Pattern Recognition*, Vol. 40, No. 7, 2038-2048. (2007)
19. Li, T., Ogihara, M.: *Detecting Emotion in Music*. In: *Proceedings of the International Symposium on Music Information Retrieval*, Washington D.C., USA, 239-240. (2003)

20. Schietgat, L., Blockeel, H., Struyf, J., Džeroski, S., Clare, A.: Decision Trees for Hierarchical Multilabel Classification: A Case Study in Functional Genomics. *Lecture Notes in Computer Science, LNAI*, Vol. 4213, 18-29. (2006)
21. Elisseeff, A., Weston, J.: A Kernel Method for Multi-labeled Classification. In: *Advances in Neural Information Processing Systems 14*, 681-687. (2002)
22. Clare, A., King, R. D.: Knowledge Discovery in Multi-label Phenotype Data. In: *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001)*, Freiburg, Germany, 42-53. (2001)
23. Schapire, R. E., Singer, Y.: Boostexter: A Boosting-based System for Text Categorization. *Machine Learning*, Vol. 29, No. 2/3, 135-168. (2000)
24. McCallum, A.: Multi-label Text Classification with a Mixture Model Trained by EM. In: *Working Notes of the AAAI'99 Workshop on Text Learning*, Orlando, Florida. (1999)
25. Ueda, N., Saito, K.: Parametric Mixture Models for Multi-labeled Text. In: Becker, S., Thrun, S., Obermayer, K. (Eds.): *Advances in Neural Information Processing Systems*, Vol. 15, MIT Press, Cambridge, MA, 721-728. (2003)
26. Fortuna, B., Grobelnik, M., Mladenić, D.: OntoGen: Semi-automatic Ontology Editor. In: *HCI International 2007*, Beijing, China. *Lecture Notes in Computer Science*, Vol. 4558. Springer-Verlag, Berlin-Heidelberg, 309-318. (2007)
27. R Development Core Team: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (2006).
28. Feinerer, I., Hornik, K., Meyer, D.: Text Mining Infrastructure in R. *Journal of Statistical Software*, Vol. 25, No. 5, 1-54. (2008)
29. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, Vol. 24, No. 5, 513-523. (1988)
30. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, Vol. 11, Issue 1, 10-18. (2009)
31. Christiannini, N., Shawe-Taylor, J.: *Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press. (2000)
32. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: *Classification and Regression Trees*. Wadsworth International Group, Belmont CA. (1984)
33. Zhang M.-L., Zhou Z.-H.: Multi-label Neural Network with Applications to Functional Genomics and Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No.10, 1338-1351. (2006)
34. Tsoumakas, G., Mencia, L. E., Katakis, I., Park, S.-H., Fürnkranz J.: On the Combination of two Decompositive Multi-label Classification Methods. *Workshop on Preference Learning, ECML PKDD 09*, Hullermeir, E., Fürnkranz, J. (Ed.), Bled, Slovenia, 114-133. (2009)
35. Tsoumakas, G., Katakis, I., Vlahavas, I.: Effective and Efficient Multilabel Classification in Domains with Large Number of Labels. In: *Proceedings ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08)*, Antwerp, Belgium. (2008)
36. Koller, D., Sahami, M.: Hierarchically classifying documents using very few words. In: *Proceedings of the 14th International Conference on Machine Learning (ICML-97)*, Nashville, 170-178. (1997)
37. Brinker, K., Hüllermeier, E.: Case-based Label Ranking. In: *Machine Learning: ECML 2006*, Vol. 4212, 566-573. (2006)

Sergeja Vogrinčič obtained her M.Sc. degree from Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, in 2010. She has been employed at the same establishment, since 2006. Her research interests include machine learning, classification and ontology construction.

Zoran Bosnić obtained his M.Sc. and Ph.D. degrees from the University of Ljubljana, Faculty of Computer and Information Science (Slovenia) in 2003 and 2007, respectively. Since 2006 he has been employed at Faculty of Computer and Information Science and currently works as an assistant professor in the Laboratory of Cognitive Modelling. His research interests include artificial intelligence, machine learning, regression, and reliability estimation for individual predictions, as well as applications in these areas.

Received: April 20, 2010; Accepted: September 15, 2010.

An Analysis of an Assessment Model for Participation in Online Forums

Mee-Chin Wee¹ and Abdullah Abrizah¹

¹Faculty of Computer Science and Information Technology, University of Malaya,
50603 Kuala Lumpur, Malaysia
{weemc, abrizah}@um.edu.my

Abstract. Existing online forum software support limited assessment features. This paper presents an analysis of an assessment model which has been implemented in online discussion forum software. The assessment model is aimed to automate the assessment of students' participation in online discussion forums. The model is formulated based on four different participation indicators and educators' feedback. The model was tested by a group of students who used the online forum to complete a project. Pearson product-moment correlations were calculated using the scores (performance indicator scores) generated by the model and the actual scores given by five educators. The performance indicator scores generated using the assessment formula was highly correlated with the actual grades assigned by the educators. The results suggest that the assessment model is reliable and can be used to evaluate students' participation in online discussion forums.

Keywords: information systems, online discussion forums, students assessment, online participation, performance indicator.

1. Introduction

Students' discussion on learning-related issues in online discussion forums is common. Pendergast noted that a successful online discussion forum requires the following mechanisms: a collection of stimulating discussion topics, a sound technique to implement them, and a consistent way to assess student participation [1]. Assessing students' participation in online discussion forums is important as students will contribute more readily and meaningfully to online discussion when they knew that the discussion is assessed [2]. Assessment criteria for online forums can serve as a clear guideline to students for the expected quality of thinking and discussion, and as means of aligning teaching and learning behaviors and goals [3, 4]. This is because there are more opportunities for students to engage in online discussions that utilise the higher level cognitive skills such as analysis, synthesis and evaluation – rather than face-to-face, especially when the learners read and

interpret postings, as well as formulate and articulate their own opinions. However, high levels of participation without focus and coherence could create confusion and information overload for other students [5].

The assessment criteria found in the literature include students performance based on total number of students' postings, total message length, keyword contribution [6], content analysis using Henri's Analytical Model [7], and Garrison and Anderson's Practical Inquiry Model of Cognitive Presence [8]. Existing online discussion forums do not have feature to automate the assessment of students' participation in online forums. Most of the assessments of students' participation in online discussion forums are performed manually by educators. To ease the work of educators, an assessment formula has been developed and incorporated into online forum software. This paper presents the results and reliability of the assessment formula. The formula has been tested in a real life learning context, and the results show that the assessment formula is reliable in measuring students' contributions to the forum software.

2. Current Online Assessment Criteria

There are two (2) approaches for assessing constructed response materials such as online discussions. They are holistic scoring and analytic scoring [9]. In the holistic scoring approach, students' response is scored as a whole. Discussions posted by a student are collected and educators evaluate and assign a single score. Initial criteria are established prior to scoring, and these criteria are taken as a whole. An example of the criteria used (such as research depth, feedback to other students, regular input, citing examples from professional practice, and demonstrating leadership qualities) for holistic scoring is presented in Table 1 [10]. Some learning management systems support viewing of individual student's postings in each forum. This is reflected for example, in myLMS where instructors can click on a student's name to display all messages posted by the student in online learning participation interface; and in Blackboard Learning Management System, where instructors can view online messages sorted by the creator of the message. Assessment of participation in online forums using holistic approach can only be performed once the discussion is over. Students have no knowledge about their online discussion performance while the discussion is going on.

On the other hand, analytic scoring is more detailed where the assessment is done across multi dimensions of performance rather than based on overall impression (holistic scoring). Individual criterion is established using a point allocation method. Points are allocated to each student discussion based on the criteria. The sum of the score gives the student an overall score. The advantage of the analytic scoring method is that it helps to identify the students' strengths and weaknesses. However, using this approach for assessment is time consuming and it focuses on different aspects or

dimensions which may result in overlooking the overall quality of discussion and communication of ideas.

Table 1. Sample criteria for Holistic scoring (source:[10])

1.0 – 2.0 points
<ul style="list-style-type: none">• Posted main topic information.• Replied to one other student posting.• No depth of presentation, no research base, opinion only.• Information posted only one time or several posts at one time.• Comments were barely related to main discussion question and/or other student posting.• No constructive comments to help class discussion.• All posts made within 24 hours of assignment due date.
3.0 – 4.0 points
<ul style="list-style-type: none">• Posted main topic information and one response on same day.• Several posts, but all on same day.• Time between posting indicated student had read and considered substantial number of student postings before responding.• Replied to other student postings and provided relevant responses and constructive feedback to the students.• Enhance quality of discussion (that is illustrated a point with examples, suggested new perspectives on issues, asked questions that helped further discussion, cited current news events, etcetera.).• Referred to other research, gave examples, and evoked follow-up responses from other students.
5.0 Points
<ul style="list-style-type: none">• Demonstrated leadership in discussions.• Posted regularly during the week.• Replied to main topic. Substantially enhanced quality of discussion (that is illustrated a point with examples, suggested new perspectives on issues, asked questions that helped further discussion, cited current news events, etcetera.).• Replied to several other student postings on a regular basis and provided relevant responses and constructive feedback to student postings.• Time between posting indicated student had read and considered substantial number of student postings before responding.• Referred to other research, gave examples, and evoked follow-up responses from other students.

The analysis of students' contributions to online discussion forums in analytic scoring approach includes looking at the frequency/patterns of interaction (for example, counting messages and participation rates), message length and message content analysis. The number of messages posted by a student could be considered as a participation indicator since it shows the number of times the student reads and responds in writing to others' postings [11, 12]. For instance, Cheung and Hew associated the degree of learners' participation in online discussions to the quantity of message posted by the learners [13]. The messages of a particular minimum length are used as one assessment criterion by some instructors [12]. Knowlton emphasized on the importance of giving a minimum length for messages since it takes some length to construct perspectives that can become the basis of knowledge [14]. The minimum length requirement assures that students are spending time to become engaged in the discussion. However, a message should not be too long as it may be overlooked by the student participants because of the length [15]. Previous studies have found a positive relationship between the amount of time students spend reading messages and engaged in virtual dialogue with their classmates and their achievement of course objectives [16]. Therefore, students' effort in the online discussion forums could be reflected by the amount of words they post to the system. The message length is measured by counting all the words in the student's messages. The total number of messages and message length do not give insight into the quality of the contributions made by each participant to the online discussion forum. Some learning management systems support online learning participation where the total number of messages for individual student in each forum is displayed and updated instantly. However, display of the total message length posted by an individual student in online forums is rather rare.

Wu and Chen adopted the idea of term weighting in information retrieval to assign weights to keywords [6]. Usage of keywords was taken as an indicator for the learning quality. A keyword is defined as a simple, non-recursive noun phrase or a base noun phrase. The unique noun phrases extracted from all class messages are defined as the class concept base. The importance of a keyword is measured by its frequency. The more frequently a keyword is used, the more important it is. However, if a keyword is used by more students, it becomes less important in terms of differentiating one student's contributions to the class concept base compared to others. The usage of more specific keywords does not necessarily result in high quality work. However, this method may be used to indicate if a student is bringing in new concepts, and not merely repeating the existing ones. Wu and Chen explored the use of message length, message count and keyword contribution to evaluate students' performance in online forums. The results of the experiment showed that the combination of all three measures performed better than any of the individual three measures.

Another way to measure effective participation is utilizing Bloom's Taxonomy of educational objectives to interpret discourse contributed by students [17]. The taxonomy identifies six (6) educational objectives, listed in

order of cognitive complexity, namely, knowledge, comprehension, application, analysis, synthesis and evaluation. Knowledge in the taxonomy involves the psychological process of remembering information. Comprehension involves the lowest level of understanding where the students know what is being communicated and can use it in its immediate context. Application consists of the application of abstractions (ideas, principles, generalized methods and theories) to particular concrete situations. Analysis is demonstrated through the students' discrimination of information and ability to compare and differentiate. Synthesis requires the combination of information to find solutions to unfamiliar problems, or in the production of an original work. Evaluation is evidenced through the students' ability to formulate value judgments about theories and methods for a given purpose. Bloom's six major categories were changed from noun to verb forms to add relevance for 21st century [17]. The lowest level of the original, knowledge was renamed and became remembering. Comprehension and synthesis were re-titled to understanding and creating. To determine the effectiveness of a student's participation in online discussions, an educator manually categorizes messages according to Bloom's taxonomy of cognitive objectives by reading the text messages. Effective discussions address higher order thinking skills in Bloom's hierarchy by engaging students in applying theories, distinguishing between facts and opinions, evaluating responses of other students, providing opposing viewpoints as well as feedback on other discussion posts. Knowlton quoted that for a durable discussion, students are required to interact with one another at higher levels of learning, that is, 'analysis', 'synthesis', and 'evaluation' [14]. However, Bloom's taxonomy is not designed for online assessment. Message categories such as 'questioning', 'suggesting' and 'acknowledging' are not listed in the Bloom's taxonomy. As such, Bloom's taxonomy may not be comprehensive enough to be used as criteria for online assessment. Educators who apply Bloom's taxonomy in online assessment may have to list down other message categories that are posted by students, and then decide on the suitable assessment criteria to use before grading.

Henri proposed a model for analyzing the process of learning. Messages are broken into thematic units of meaning which are then analyzed along the model [7]. Henri's Analytical Model has five (5) dimensions. The five (5) dimensions, participative, social, interactive, cognitive and meta-cognitive, are detailed out in Table 2[7]. The model allows analysis of a range of aspects of an online discussion: the level of participation in the form of usage statistics, the nature of the interaction between contributors, and an indication of the learning process through an analysis of the cognitive activity evident in the message content. For the participative dimension, data can be further categorized as overall data, learner and educator participation data. Interactivity in the interactive dimension can be further broken into explicit interaction, direct response, direct commentary, implicit interaction, indirect response, indirect commentary, and independent statement. Elementary clarification, in-depth clarification, inference, judgment and strategies are the categories under cognitive skills. In the elaboration of the metacognitive

model, Henri made a theoretical distinction between metacognitive knowledge and metacognitive skills. Metacognitive knowledge is declarative knowledge concerning the person, the task, and the strategies. Metacognitive skills refer to procedural knowledge relating to evaluation, planning, regulation and self-awareness.

McLoughlin and Luca found Henri’s content analysis model applicable to a teacher-centered discussion model, however they felt that it is unsuitable for a constructivist student-centered discussion model [18]. This suggests that the model is difficult to be implemented in a less structured online discussions environment. Aspects of this model have been taken up and expanded upon by others interested in comparing the level of critical thinking between face to face seminars and computer conferences [19]. Newman et al. developed ten (10) paired indicators of critical versus uncritical thinking in their model [19]. The ten indicators are: relevance, importance, novelty, bringing outside knowledge or experience, clarity, linking ideas or interpretation, justification, critical assessment, practical utility, and width of understanding. Each of the ten indicators has its own list of paired opposites. “Relevant statements” versus “Irrelevant statements or diversions” is a case in point.

Table 2. Henri’s Analytical Framework (Source:[7])

Dimension	Definition	Indicators
Participative	Compilation of the number of messages or statements transmitted by one person or group	Number of messages Number of statements
Social	Statement or part of statement not related to formal content of subject matter	Self-introduction Verbal support “I’m feeling great....”
Interactive	Chain of connected messages	“In response to
Cognitive	Statement exhibiting knowledge and skills related to learning process	“As we said earlier...” Ask questions Making inferences Formulating hypotheses
Metacognitive	Statement related to general knowledge and skills and showing awareness, self-control, and self-regulation of learning	“I understand....” “I wonder.....”

Another approach for analyzing the process of learning is the Practical Inquiry Model of Cognitive Presence [8]. The goal of this model is to judge

the nature and quality of critical reflection and disclosure in a collaborative community of inquiry. The practical inquiry model includes four phases, namely trigger, exploration, integration, and resolution, in describing cognitive presence in an educational context generally. Table 3 provides the descriptors and indicators that correspond to each phase of the practical inquiry process. In this model, a message is the unit which is analyzed along the four phases. Two heuristics, “code up” and “code down”, were developed to resolve contradictory categorization. “Code down” (i.e. select the earlier phase of the model) is used if the phase the message is reflected is not clear, while “code up” (i.e. select the later phase) is used if there is clear evidence of multiple phases in the message. This is justified as higher levels of critical thinking such as integration and resolution borrow characteristics and processes from the two previous phases. Corich et al. used this model to assess the quality of forum contribution for students participating in a first year undergraduate degree course [20].

Table 3. Practical Inquiry Descriptors and Indicators (Source:[8])

Phase	Descriptor	Indicator
Trigger event	Evocative	Recognize problem Puzzlement
Exploration	Inquisitive	Divergence Information Exchange Suggestions Brainstorming Intuitive leaps
Integration	Tentative	Convergence Synthesis Solutions
Resolution	Committed	Apply Test Defend

Dringus and Ellis developed a meaningful and usable schema for categorizing and describing contributions in online discussion forums [21]. The tool, named Scale for Forums / Online Discussion Assessment (SCAFFOLD), is built upon a range of participation indicators identified in the literature. Dringus and Ellis summarized all the participation indicators identified in a table format and developed a list of 19 participation indicators as depicted in Table 4 [21]. SCAFFOLD was used by faculty and students to rate the 13 postings contained in a discrete segment of a masters-level discussion forum in a multimedia systems course [22]. The results of the evaluation showed that SCAFFOLD could be used for developing and conveying feedback on Internet forums. A principal component factor analysis was conducted and three components were identified in the data reduction analysis that tracked closely to the Mezirow’s classification of learning across three dimensions of reflection [23].The dimensions are

content, process, and premise level. Content level reflection entails acquiring facts and building skills. Participation indicators grouped under content level are ‘broadened’, ‘originality’, ‘problem’, ‘questioning’, ‘comprehensive’, ‘evaluative’ and ‘synthesis’. Process level reflection entails developing problem solving ability. Participation indicators grouped under this level are ‘closure’, ‘resolution’, ‘solutions’, ‘summarizing’, and ‘analysis’. Premise level reflection, which is the most cognitively demanding learning dimension, entails an analysis and evaluation of the value and relevance of the subject matter. Participation indicators grouped under the premise level are ‘acknowledging’, ‘clarification’, ‘reflective’, ‘social’, and ‘resources’. Participation indicators such as “error free” and ‘topical’ did not correlate with other indicators. Some of the indicators listed in the instrument could be combined as suggested in the study. SCAFFOLD can be used to categorize and describe messages but it does not evaluate the content of a message. Evaluating the quality of contents posted for discussions is difficult since it may vary from one educator to another. In this study, the authors explored the possibility of using SCAFFOLD as one of the participation indicators for students’ participation in online forums.

Table 4. SCAFFOLD (Source:[21])

No.	The Contribution
1	Acknowledging: responded to another contribution
2	Analysis: provided analysis of the problem being discussed
3	Broadened: increased the scope of the discussion
4	Clarification: supplied or sought clarification as needed in responses
5	Closure: helped lead to a conclusion on a topic
6	Comprehensive: was complete, but not overly lengthy
7	Error Free: contained accurate information
8	Evaluative: was evaluative, assessing the meaningfulness or validity of ideas being shared
9	Originality: contained new ideas or approaches to the topic
10	Problem: identified a worthy problem related to the topic
11	Questioning: raised thoughtful questions about the topic
12	Reflective: interjected personal commentary or experiences
13	Resolution: promoted cooperation to resolve issues of debate or disagreement
14	Resources: exchanged useful resources with others such as links or citations
15	Social: conversational or social in nature
16	Solutions: suggested meaningful solutions
17	Summarizing: summarized the topic discussion overall
18	Synthesis: contained well formed, clear, connected, and synthesized ideas
19	Topical: was on topic

All the models identified earlier (Bloom's model, Henri's Analytical model and Practical Inquiry Model of Cognitive Presence) involve reading and classifying messages of the online discussion forums in an electronic copy or in a print copy by educators. The limitation of the models is the difficulty for instructors to make consistent judgments. Inconsistent judgments happen due to the complexity of the instrument (too few categories for Bloom's model; too many categories or codes for Henri's model and Practical Inquiry Model of Cognitive Presence; and lack of mutual exclusiveness among these models), and the use of an inappropriate unit of analysis (anything other than the sentence in the online discourse) [24]. Although the models can provide useful data for exploring the way in which participants are contributing to an online discussion forum, the large volume of postings makes it difficult for educators to process and analyze data into meaningful information.

3. The Gap

Existing online discussion forums do not automate the assessment of students' participation in these forums. Most of the assessments on students' participation in online discussion forums are performed manually. Some online discussion forums support view feature for total number of messages of individual student in each online forum and the total messages of each forum. There has been no implementation of total messages length for individual student in online forums. Some forum software support viewing of individual student's messages for each forum. With this feature, educators read and grade students' messages using assessment criteria stated in different framework or model such as Henri's framework, Bloom's Taxonomy, or the Practical Inquiry Model of Cognitive Presence. For example, a proprietary discussion forum software system named ANGEL allows the educator to view and grade individual entries to the forum or grade a student's aggregate contributions [1]. ANGEL produces an output in HTML format and can be used for further analysis. To accurately assess students' work using the ANGEL discussion forum, an instructor has to read all the discussions, and grading the online discussions could take more than 50% of the instructors' time for each online class [25]. Wu and Chen investigated the use of keyword, message count and message length to evaluate students' contribution to online forums. They found the significant of combining the three different variables in computing students' online participation [6]. Assessment of students' contributions using keyword approach is performed when the discussion is over. As such, no ongoing feedback is given to the students on their online participation. Students have no knowledge related to their online discussion performance for ongoing improvement. Another issue is that abbreviated words are often found in messages, which can make the keyword contribution difficult to be applied in online discussion forums.

4. An Assessment Formula

To facilitate assessment of students' participation and ease the tasks of educators, an assessment formula was developed and incorporated into online forum software. The formula is used to calculate the score for students' participation based on four (4) different participation indicators and educators' feedback gathered from a survey [26]. The objective of the survey is to elicit the requirements for assessing students' participation in Internet forums. The survey respondents comprised Information Systems (IS) educators in institutions of higher education in Malaysia. IS discipline was selected as the context in this study for the following reasons which leads to its uniqueness for requirements and research in how ICT is used in instructions: (a) it has a good blend of Information Technology and non Information Technology courses, and (b) the programme is either administratively positioned in business or computing schools. The survey results showed that timeliness of message, message category, number of messages posted and message length were opted 67.1%, 58.9%, 56.2% and 24.7% respectively by the respondents (n=73). The survey findings showed that a criterion in a performance indicator is and will remain as a subjective option of the respondents. However, consistent with Winkler and Clemen's [27] study, the combination of assessment criteria from multiple aspects has been proven useful for increasing the forecast accuracy.

The authors assume that the quality of learning in online forums is revealed by the category or description of the messages. The category of a message is analyzed along SCAFFOLD to reflect the depth of student's knowledge; therefore the message category is taken as an indicator for the learning quality. Messages that are categorized as 'analysis', 'synthesis', and 'evaluation' reflect students interaction with one another at higher levels of learning. These types of messages could be considered as indicators for high learning quality. SCAFFOLD was adopted since it is comprehensive and contains elements of higher order thinking skills (i.e. 'analysis', 'synthesis', and 'evaluation') and was build upon the theoretical foundation established in the literature.

Knowlton opined that "for the benefits of online discussion to be realized, students must have formal opportunities for self evaluation" [14]. As such students must practise evaluating their own contributions to an online discussion against a clearly articulated set of criteria. An approach to resolve the challenges of collecting and coding large data sets might be to directly involve students in a process of categorizing their own discussion in such context. As such, students are required to categorize their messages using SCAFFOLD when posting (as shown in Figure 1). Educators are able to change students' message categories from Control Panel if necessary.

Due dates are best for stimulating the discussion online. It is important that messages are posted or replied on time [1, 14, 28]. Some students will not participate until other students have posted their work or they will wait until just before the discussion ends [1]. The timeframe for discussion is important to be considered. If the timeframe is too long the discussion may be slow to

take off and early contributors may have no responses to further their discussion. On the other hand, a short timeframe may not provide the flexibility in time to allow participants to undertake the associated coursework requirements and reflection. The best interaction comes when postings are made in a quick sequence. Therefore, educators have to observe and decide based on their experience about timeliness of posting. For instance, the timeliness of posting could be three (3) days if the discussion topic is relatively new to the students. As such students' contributions in the virtual dialogue could be measured by setting a date where only messages posted before the stated date are included for assessment.

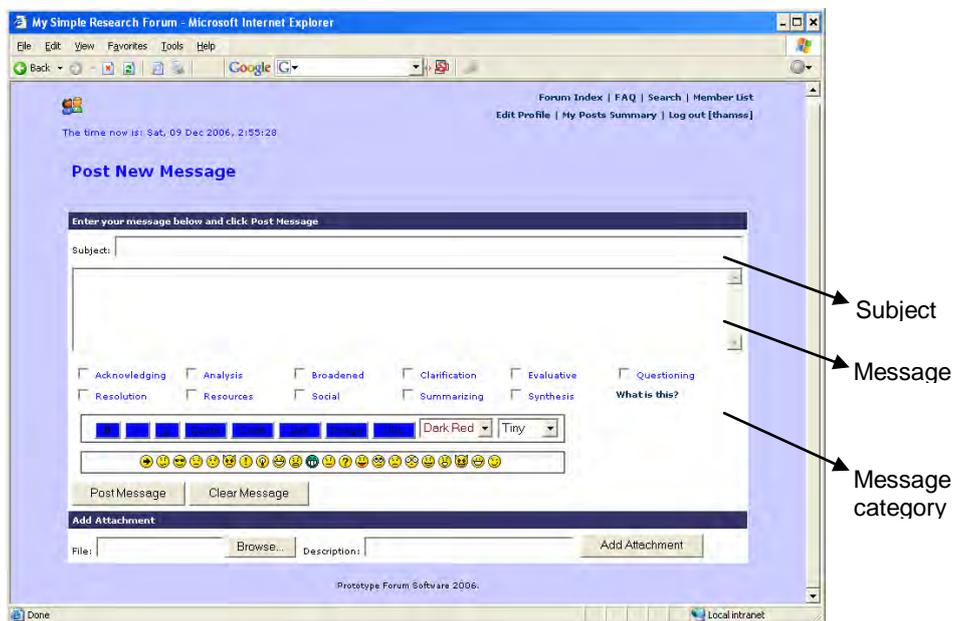


Fig. 1. Screen capture of “post new message” interface

The authors apply the idea of weighting to assign weights to the assessment criteria. The four measures are combined to compute the following Performance Indicator (PI) score:

$$PI \text{ score} = a * Tot_Mess + b * Tot_Length + c * Count_Acknowledging + d * Count_Analysis + e * Count_Broadened + f * Count_Evaluative + g * Count_Clarification + h * Count_Questioning + j * Count_Resolution + k * Count_Resources + m * Count_Social + n * Count_Summarizing + p * Count_Synthesis .$$

Where

a, b, c, d, e, f, g, h, j, k, m, n, p are coefficients.

Tot_Mess – total messages posted by a member.

Tot_Length – total message lengths posted by a member.

Count_Acknowledging – total count of ‘acknowledging’ messages posted by a member.

Count_Analysis - total count of ‘analysis’ messages posted by a member.

Count_Broadened - total count of ‘broadened’ messages posted by a member.

Count_Evaluative - total count of ‘evaluative’ messages posted by a member.

Count_Clarification - total count of ‘clarification’ messages posted by a member.

Count_Questioning - total count of ‘questioning’ messages posted by a member.

Count_Resolution - total count of ‘resolution’ messages posted by a member.

Count_Resources - total count of ‘resources’ messages posted by a member.

Count_Social - total count of ‘social’ messages posted by a member.

Count_Summarizing - total count of ‘summarizing’ messages posted by a member.

Count_Synthesis - total count of ‘synthesis’ messages posted by a member.

For timeliness of posting, a date is set so that the system includes only messages that fall before the input date for assessment purpose. All posts after the specified input date will not be updated to the posts statistics.

5. Evaluation of the Assessment Formula

The main objective of the evaluation is to determine the accuracy of the assessment formula in predicting students’ score for their contributions to online discussion forum. The evaluation of the forum software was conducted in the second semester of an academic year at the Faculty of Computer Science and Information Technology at a research-intensive university in Malaysia, on two different IS courses (referred to in this study as W1 and W2). Two (2) IS educators and a total of sixty four (64) students enrolled in W1 and W2 participated in the evaluation of the forum software. The educators (referred to as Educator A and Educator B) were the lecturers teaching the two (2) courses respectively. The students were required to complete a group project in the two courses, of which they were instructed to discuss learning issues in the online discussion forum. At the end of the project, all messages in the forum were tabulated into 39 tables. A total of 26 tables were stored in a file (named courseW1) for course W1; the other 13 tables were stored in another file (named courseW2) for course W2. Each table contained posts detail of a student such as posts’ subjects, time and date of posts, posts category(s), and aggregate contributions of a student in the online forum (Figure 2). A total of 61% (39) students contributed to the discussion. Almost 39% (25) of the students were either silent or invisible in the online forum and these students would very likely to get zero mark for the

An Analysis of an Assessment Model for Participation in Online Forums

online discussion. The authors decided not to include these invisible students in the comparison of the PI scores with the students' actual grades assigned by the IS educators for the two courses involved. This is simply because it may increase the correlation between the PI scores and the actual grades, and subsequently may not reflect the accuracy of the assessment formula. The compiled data files were sent to the IS educators involved for assessment purpose. This is a common approach for manual grading. Three (3) other IS educators (referred to as independent educator X, Y, and Z) were contacted independently to assess the same set of compiled data files (students' messages). These three IS educators admitted that they did not know who the students were. All five (5) educators involved have had more than six (6) years of teaching experience. The educators were given the freedom to decide on their assessment criteria. They admitted feeling comfortable reviewing the discussions. The online forum software was made available to all five of them for viewing the learning context even though the discussion was over. The project titles and their description were sent to the educators as well. The project for course W1 required the students to discuss the use of intelligent agents in information retrieval whereas in course W2, students were required to describe the decision rules approaches and explore the application of data mining technology in practical approaches within Malaysia or international context.

Username: Fahimy Kamaruddin, total post: 1

Under subject: Assignment 1	marks
<p>Posted on: Thu, 11 Jan 2007, 3:28:33 Post subject: What's your approach to solve this assignment? Post category: Synthesis</p> <p>Lets see... 2 questions Whats an IA & How IA is used to support the IR. Some people browse through the internet to find some uncertain answer... But, should we retrieve the answer instead of browse?</p> <p>Well, you can google for Intelligent Agent or IA + IR for browsing, but how about searching for the exact answer such as googling for "intelligent agent properties" or "IA's implementation".</p> <p>I mean look for and answer not anything related to IA and IR.</p> <p>What do you think? Act speaks louder than words...</p>	

Username: jwee_ching, total post: 3

Under subject: Assignment 1	marks
<p>Posted on: Mon, 15 Jan 2007, 3:58:09 Post subject: Just focuses on only one type of IA? Post category: Questioning</p> <p>Hi,Pooi Yee. After reading your message, I got another question. Can we do our research for this project by focuses on one type of intelligent agent such</p>	

<p>as intelligent multimedia agent that is developed for efficient retrieval and processing of information stored in multimedia databases , or in general? 😊</p> <p>Posted on: Mon, 15 Jan 2007, 7:39:55 Post subject: Just an example... Post category: Questioning</p> <p>Hi, Pooi Yee 😊 The Intelligent Multimedia Agent that i mentioned is just one of the example of IA. It is almost same with the Information Agent that mentioned by Soon Ee, the only different thing is that IMA more for retrieving documents that contains very large data objects such as images and video. Now, my problem is that i not sure whether we need to do our research in general or just focuses on only one type of IA. Can anyone give me some opinions. 😊 Thnx 🙏🙏</p> <p>Posted on: Thu, 18 Jan 2007, 1:34:28 Post subject: Methods used by IA to retrieve information Post category: Analysis, Resolution</p> <p>Dear mwai, In my opinion, i don't think that we need to focus on the importance of searching process because the question states "How IA support IR" and not "How importance IA to IR". From the articles that i found, i think if we focus on the method used by IA in searching for a relevant information will be better. Like a IA may use keyword searching method, agent-based complex query method and so on. Hope to get some opinions from all of you if i get any misunderstanding about mwai ideas. Thnx... Have a nice day 🙏😊:</p>	
---	--

Fig. 2. Two tables in a compiled data file (courseW1 file)

The authors set the coefficients (c, d, e, f, g, h, j, k, m, n, p) of SCAFFOLD to 1, a = 1 and b = 0.01. All coefficients, except b, are set to 1, which means all criteria used in the formula are equally important. The coefficient for message length (b) is set to 0.01. The authors assumed that the message length takes about 100 words to construct perspectives that can become the basis of knowledge. The same coefficients were set for the two IS courses throughout the evaluation. This is because the IS educators' grading criteria were not known. However, when the grading criteria are known, it is easy to adjust the coefficients to reflect the actual grading. Table 5 and Table 6 present a summary of actual marks given by the educators and PI score generated using the assessment formula for course W1 and W2. The assessment criteria used by the educators is unknown. Twenty six (26) and thirteen (13) students from course W1 and W2 respectively were assessed for their online participation. To measure the accuracy of the assessment model, Pearson product-moment correlations between the PI scores and the actual grades were calculated. The correlations between individual measures and the actual grades were also calculated as shown in

Table 7 for course W1 and Table 8 for course W2. The results in the second row of Table 7 and Table 8 demonstrate that there is a high correlation between the PI and the actual grades (0.827 – 0.996). The results in the third, fourth, fifth and sixth column of Table 7 and Table 8 demonstrate the correlation between individual assessment variables and the actual grades given by the independent IS educators. According to a report by Williams [29] on essay grading, the agreement between computer graders and human judges varies from 0.4 to 0.9 approximately, and that is comparable to or even better than the agreement between two human graders. The results of this study also show that, in most cases, PI performs slightly better than any of the three measures i.e. total message count, total message length and SCAFFOLD (highlighted in Table 7 and Table 8).

Table 5. A Summary of Actual Grades given by educators and PI scores for course W1(N=26)

Students	Educator A	Independent educator X	Independent educator Y	Independent educator Z	PI scores
Student 1	2.0	3.0	1.0	1.0	2.08
Student 2	8.0	17.0	3.0	4.0	16.82
Student 3	5.0	8.0	2.0	2.0	7.24
Student 4	4.0	9.0	2.0	3.0	10.19
Student 5	3.0	7.0	2.0	3.0	7.10
Student 6	3.0	5.0	1.0	2.0	5.13
Student 7	9.0	18.0	3.0	4.0	16.88
Student 8	4.0	6.0	1.0	2.0	5.18
Student 9	3.0	5.0	1.0	2.0	4.12
Student 10	5.0	9.0	1.0	3.0	7.24
Student 11	4.0	6.0	1.0	2.0	5.27
Student 12	3.0	3.0	1.0	1.0	3.11
Student 13	6.0	9.0	1.0	3.0	8.29
Student 14	6.0	6.0	1.0	2.0	6.24
Student 15	1.0	2.0	1.0	1.0	2.03
Student 16	1.0	3.0	1.0	1.0	2.03
Student 17	3.0	3.0	1.0	1.0	2.15
Student 18	6.0	9.0	2.0	3.0	7.47
Student 19	4.0	3.0	1.0	1.0	2.21
Student 20	1.0	4.0	1.0	1.0	4.10
Student 21	3.0	3.0	1.0	1.0	2.17
Student 22	4.0	3.0	1.0	1.0	2.38
Student 23	2.0	3.0	1.0	1.0	2.02
Student 24	1.0	3.0	1.0	1.0	2.01
Student 25	1.0	3.0	1.0	1.0	2.02
Student 26	1.0	3.0	1.0	1.0	2.02

Table 6. A Summary of Actual Grades given by educators and PI scores for course W2(N=13)

Students	Educator B	Independent educator X	Independent educator Y	Independent educator Z	PI scores
Student 27	5.0	5.0	4.5	2.0	5.5
Student 28	8.0	17.0	4.5	4.0	17.2
Student 29	2.0	2.0	4.0	1.0	2.0
Student 30	5.0	6.0	4.0	1.0	6.1
Student 31	2.0	2.0	4.0	1.0	2.0
Student 32	5.0	5.0	4.0	2.0	4.1
Student 33	8.0	14.0	4.5	3.0	14.2
Student 34	3.0	2.0	4.0	1.0	2.0
Student 35	3.0	3.0	4.0	1.0	2.0
Student 36	2.0	3.0	4.0	1.0	3.1
Student 37	2.0	2.0	4.0	1.0	2.0
Student 38	2.0	2.0	4.0	1.0	2.0
Student 39	2.0	3.0	4.0	1.0	2.0

Table 7 Correlations for IS course W1 (N=26)

		Educator A	Independent educator X	Independent educator Y	Independent educator Z
PI generated	Pearson Correlation	0.885**	0.852**	0.988**	0.936**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000
Total message count	Pearson Correlation	0.900**	0.806**	0.979**	0.928**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000
Total message length	Pearson Correlation	0.797**	0.905**	0.887**	0.758**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000
SCAFFOLD	Pearson Correlation	0.863**	0.857**	0.980**	0.935**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000

** . Correlation is significant at the 0.01 level (2-tailed).

Table 8 Correlations for IS course W2 (N=13)

		Educator B	Independent educator X	Independent educator Y	Independent educator Z
PI generated	Pearson Correlation	0.827**	0.930**	0.996**	0.942**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000
Total message count	Pearson Correlation	0.789**	0.923**	0.993**	0.937**
	Sig. (2-tailed)	0.001	0.000	0.000	0.000
Total message length	Pearson Correlation	0.901**	0.905**	0.853**	0.835**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000
SCAFFOLD	Pearson Correlation	0.848**	0.926**	0.991**	0.901**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000

** . Correlation is significant at the 0.01 level (2-tailed).

6. Conclusion

The authors have come up with a solution to automate the assessment of student online participation using an assessment model. The model consists of four (4) different participation indicators namely, total number of messages posted, total message lengths, total counts for each SCAFFOLD category, and timeliness of messages. The evaluation results showed that the PI score generated by the assessment model was highly correlated with the actual grades assigned by the educators. It is reasonable to assume that such correlation is comparable to what has been reported in the automatic essay grading literature [29]. Thus, the evaluation results suggest that the assessment model is reliable to be used. The assessment model can be implemented in online discussion forum to help educators obtain a reference to students' online performance without reading through the huge amount of class messages. The strength of using this assessment model is that the PI score for individual student is updated instantly and it reflects the students' ongoing performance in online forums; as such students have the opportunities to improve their online performance before the discussion is over. However, the delimitation of the evaluation was that it was conducted on two (2) different IS courses for a final year degree programme. As such it

would not be prudent to generalize the evaluation results. Future research includes broadening the investigation of the research context and scope of the users to other academic disciplines.

References

1. Pendergast, M.: An Analysis Tool for the Assessment of Student Participation and Implementation Dynamics in Online Discussion Forums. ACM SIGITE Newsletter, Vol 3, No 2. (2006)
2. Reushle, S., Dorman, M., Evans, P., Kirkwood, J., Mcdonald, J. and Worden, J.: Critical Elements: Designing for online teaching. Proceedings of the 16th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education, Brisbane, Queensland University of Technology. (1999)
3. Ho, S. : Evaluating students' participation in on-line discussions. The Eighth Australian World Wide Web Conference (AusWeb 2002). Sunshine Coast, Queensland. (2002)
4. Jones, A., Scanlon, E. and Blake, C.: Conferencing in communities of learners: examples from social history and science communication. Educational Technology & Society Vol 3, No 3. (2002)
5. Harasim, L. M. : Teaching and learning on-line: Issues in computer-mediated graduate courses. Canadian Journal of Educational Communication Vol 16, No 2, 117-35. (1987)
6. Wu, Y-f, B. and Chen, X.: Assessing Student Learning with automated text processing techniques. Journal of Asynchronous Learning Networks. Vol 9, No 3. (2005) Available: <http://www.sloan-c.org/publications/jaln/v9n3/index.asp> [Cited August 2006]
7. Henri, F.: Computer Conferencing and Content Analysis. In A. R. Kaye (Ed), Collaborative learning through computer conferencing: The Najaden Papers 117-136, Berlin: Springer-Verlag. (1992)
8. Garrison, D.R., Anderson, T.: E-Learning in the 21st Century. London, RoutledgeFalmer. (2003)
9. Popham, W. James.: Classroom Assessment: What teachers need to know (3rd ed.) Allyn & Bacon, Massachusetts. (2002)
10. Hazari, S. I.: Strategy for assessment of online course discussions. Journal of Information Systems Education, Vol 15, No 4, 349-355. (2004)
11. Picciano, A.: Beyond Student Perceptions: Issues of Interaction, Presence, and Performance In An Online Course, Journal of Asynchronous Learning Networks Vol 6, No 1. (2002)
12. Dennen V. P.: Looking for evidence of learning: Assessment and analysis methods for online discourse. Computers in Human Behavior, Vol 24, Issue 2, 205-219 (2008)
13. Cheung, W.S. and Hew, K.F.: Examining facilitators' habits of mind and learners' participation. Proceedings ascilite Melbourne 2008. Deakin University, Australia. (2008)
14. Knowlton, D. S.: Promoting durable knowledge construction through online discussion. Sixth Annual Instructional Technology Conference :“Today's successes/tomorrow's horizons”. Middle Tennessee State University. (2001)

15. Murphy, E. and Loveless, J.: Students' self analysis of contributions to online asynchronous discussions. *Australasian Journal of Educational Technology*, Vol 21, No 2, 155-172. (2005)
16. Levenburg, N. M. and Major, H. T.: Motivating the Online Learner: The Effect of Frequency of Online Postings and Time Spent Online on Achievement of Learning Goals and Objectives. *Proceedings of the International Online Conference on Teaching Online in Higher Education*. Indiana University-Purdue University: Fort Wayne. (2000)
17. Anderson, L. W., & Krathwohl, D. R. (Eds.): *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition*, New York : Longman. (2001)
18. McLoughlin, C. and Luca, J.: Lonely Outpourings Or Reasoned Dialogue? An Analysis Of Text-based Conferencing As a Tool To Support Learning. The 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education. (Ascilite 99). Brisbane, Australia. (1999) Available <http://www.ascilite.org.au/conferences/brisbane99/papers/mcloughlinluca.pdf>
19. Newman, D. R., Johnson, C., Webb, B. and Cochrane, C.: Evaluating the quality of learning in computer supported cooperative learning. *Journal of the American Society of Information Science*, Vol 48, 484-495. (1997)
20. Corich, S., Kinshuk & Hunt, L. M.: Assessing discussion forum participation: In search of quality [Electronic Version]. *International Journal of Instructional Technology and Distance Learning*, Vol 1, No 12. (2004). [viewed 02 June 2010] Available : http://www.itdl.org/Journal/Dec_04/article01.htm
21. Dringus L. P. and Ellis T. J.: Building the SCAFFOLD for Evaluating Threaded Discussion Forum Activity: Describing and Categorizing Contributions. *ASEE/IEEE Frontiers in Education Conference*. Savannah, GA. (2004)
22. Dringus L. P. and Ellis T. J.: Evaluating Threaded Discussion Forum Activity: Faculty and Student Perspectives on Categories of Activity. *ASEE/IEEE Frontiers in Education Conference*. Indianapolis. (2005)
23. Mezirow, J.: *Transformative Dimensions of Adult Learning*. San Francisco: Jossey-Bass. (1991).
24. Fahy P. J. : Addressing some common problems in transcript analysis. *The International Review of Research in Open and Distance Learning*, Vol 1, No 2 (2001).
25. Lazarus, B. D.: Teaching Courses Online: How Much Time Does it Take? *Journal of Asynchronous Learning Networks* Vol 7, No 3. (2003)
26. Wee, M. C.: *A Computer Managed Instructional (CMI)-based Tool to Support Information Systems Instructional Processes* (Doctoral dissertation). University of Malaya, Malaysia. (2009)
27. Winkler, R. L. and Clemen, R. T.: Multiple Experts vs. Multiple Methods: Combining Correlation Assessments. *Decision Analysis*. Vol 1, No 3. 167-176 (2004)
28. Chinmoy Sahu,: An evaluation of selected pedagogical attributes of online discussion boards. *Proceedings ascilite Melbourne 2008*. Deakin University, Australia. (2008)
29. Williams, R.: Automated essay grading: An evaluation of four conceptual models. In A. Herrmann and M.M. Kulski(Eds), *Expanding Horizons in Teaching and Learning*. *Proceedings of the 10th Annual Teaching Learning Forum*. (2001)

Mee-Chin Wee and Abdullah Abrizah

Wee Mee Chin is a lecturer in Information Systems at the University of Malaya, Kuala Lumpur, Malaysia. Her research interests include computer in education, computer managed instruction, computer assisted assessment, and data visualization techniques.

Abrizah Abdullah is an Associate Professor in Library and Information Science (LIS) at the University of Malaya, Kuala Lumpur, Malaysia. She teaches and researches on non-conventional technologies in LIS focusing on digital libraries, open access repositories and personal information management. She is the executive editor of the Malaysian Journal of Library & Information Science and has co-edited three proceedings of the International Conference of Libraries, Information and Society. She is an active author and has published widely in international journals and 30 assorted conferences proceedings. Abrizah can be reached at abrizah@um.edu.my.

Received: January 13, 2010; Accepted: November 22, 2010.

A new method for constructing kernel vectors in morphological associative memories of binary patterns

Yiannis.S. Boutalis¹

¹ Department of Electrical and Computer Engineering,
Democritus University of Thrace,
GR-67100 Xanthi, Greece
ybout@ee.duth.gr

Abstract. : Kernel vectors represent an elegant representation for the retrieval of pattern associations, where the input patterns are corrupted by both erosive and dilative noise. However, their action completely fails when a particular kind of erosive noise, even of very low percentage, corrupts the input pattern. In this paper, a theoretical justification of this fact is given and a new method is proposed for the construction of kernel vectors for binary patterns associations. The new kernels are not binary but 'gray', because they contain elements with values in the interval $[0, 1]$. It is shown, both theoretically and experimentally that the new kernel vectors carry the good properties of conventional kernel vectors and, at the same time, they can be easily computed. Moreover, they do not suffer from the particular noise deficiency of the conventional kernel vectors. The recalling result is in general a gray pattern, which in the sequel undergoes a simple thresholding action and passes through a simple Hamming network to produce high recall rates, even in heavily corrupted patterns. Retrieval of pattern associations is very significant for a variety of scientific disciplines including data analysis, signal and image understanding and intelligent control.

Keywords: Neural networks, Associative memory, Kernel vectors, Noise Robustness.

1. Introduction

Morphological Neural Networks (MNNs) represent artificial neural networks whose neurons perform an elementary operation of mathematical morphology [1], [2]. Unlike Hopfield network [3,4], MNNs provide the result in one pass through the network, without any significant amount of training. A number of researchers devised MNNs for a range of applications like those appearing, for example, in [5-13].

Artificial neural network models are specified by the network topology, unit characteristics, and training or learning rules. The underlying algebraic system used in these models is the set of real numbers R together with the operations of addition and multiplication and the laws governing these operations. This algebraic system, known as ring, is commonly denoted by $(R, +, \cdot)$. The basic computations occurring in morphological networks are based on the algebraic lattice structure $(R, +, \wedge, \vee)$, where \wedge, \vee denote the binary operations for minimum and maximum, respectively. The basic axiomatic operations of lattice algebra [14] are:

$$\begin{aligned} (\alpha \vee -\infty) &= (-\infty \vee \alpha) = \alpha \quad \forall \alpha \in R_{-\infty} \\ \alpha \wedge \infty &= \infty \wedge \alpha = \alpha \quad \forall \alpha \in R_{\infty} \\ \alpha + 0 &= \alpha \quad \forall \alpha \in R \\ 0 + \alpha &= \alpha \quad \forall \alpha \in R \end{aligned} \tag{1.1}$$

An *associative memory* (AM) is an input-output system that describes a relation $R \subseteq R^m \times R^n$. If $(\mathbf{x}, \mathbf{y}) \in R$, i.e. if the input \mathbf{x} produces the output \mathbf{y} , then the associative memory is said to *store* or *record* the *memory association* (\mathbf{x}, \mathbf{y}) . NN models serving as associative memories are generally capable of retrieving a complete output pattern \mathbf{y} even if the input pattern \mathbf{x} is corrupted or incomplete. The purpose of *auto-associative* memories is the retrieval of \mathbf{x} from corrupted or incomplete versions $\tilde{\mathbf{x}}$. If an artificial associative memory stores associations (\mathbf{x}, \mathbf{y}) , where \mathbf{x} cannot be viewed as a corrupted or incomplete version of \mathbf{y} , then we speak about *hetero-associative* memory.

Research on neural associative memories goes back in the 1950s [15-17]. *Linear associative memory* or *correlation memory* [17-19] is one well known neural associative memory. Association of patterns $\mathbf{x} \in R^n$ and $\mathbf{y} \in R^m$ is achieved by means of a matrix-vector product $\mathbf{y} = \mathbf{W} \cdot \mathbf{x}$. If we suppose that the goal is to store k vector pairs $(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^k, \mathbf{y}^k)$, where $\mathbf{x}^r \in R^n$ and $\mathbf{y}^r \in R^m$ for all $r = 1, \dots, k$, then \mathbf{W} is an $m \times n$ matrix given by the following outer product rule:

$$\mathbf{W} = \sum_{r=1}^k \mathbf{y}^r \cdot (\mathbf{x}^r)' \tag{1.2}$$

If \mathbf{X} denotes the matrix whose column are the input patterns $\mathbf{x}^1, \dots, \mathbf{x}^k$ and \mathbf{Y} denotes the matrix whose columns are the output patterns $\mathbf{y}^1, \dots, \mathbf{y}^k$, then this equation can be written in the simple form $\mathbf{Y} \cdot \mathbf{X}'$. If the input patterns $\mathbf{x}^1, \dots, \mathbf{x}^k$ are orthonormal, then

$$\mathbf{W} \cdot \mathbf{x}^r = ((\mathbf{y}^1 \cdot \mathbf{x}^1)' + \dots + (\mathbf{y}^k \cdot \mathbf{x}^k)') \cdot \mathbf{x}^r = \mathbf{y}^r \tag{1.3}$$

Thus, we have perfect recall of the output patterns $\mathbf{y}^1, \dots, \mathbf{y}^k$. In case the patterns are not orthonormal the capacity of the memory is extremely limited and its ability to retrieve associations is further reduced when input patterns are noisy [20].

Morphological associative memories (MAMs) are based on the algebraic lattice structure $(R, +, \wedge, \vee)$. They have been initially proposed in [14], [21], [22] for associating binary pattern vectors and are far more robust to noise

than the conventional linear associative memories. Moreover, their memory capacity in stored associations is significantly larger than the capacity of conventional linear associative memories. These properties and the fact that a number of notable features such as optimal absolute storage capacity and one-step convergence have been shown to hold in the general case for real-valued patterns [14], motivated for their extension to gray-coded vector associations [23] or recently in general gray-scale MAMs and fuzzy oriented treatments [24], [25],[26 - 28]. Recent works report also on recall and storage performance of various types of MAMs in color patterns under various types and percentages of noise [29], [30], while other works present new dynamic MAMs and test their performance in recalling gray and color pattern associations [31], [32], [33]. According to [14], there are two alternative approaches to construct MAMs. In the first approach the constructed memory is very robust in the presence of input patterns containing erosive noise. In the dual approach the constructed memory is very robust in the presence of input patterns containing dilative noise. In case the input pattern $\tilde{\mathbf{x}}^r$ is corrupted by both erosive and dilative noise none of the memories is able to recall \mathbf{y}^r .

To overcome this problem Ritter and his coworkers [34-36] proposed the idea of two-step MAMs and the production of the so-called kernel vectors, as an elegant representation of the associations $(\mathbf{x}^r, \mathbf{y}^r)$. These vectors, in conjunction with appropriate morphological operations, are suitable to recall \mathbf{y}^r when an arbitrarily corrupted (both eroded and dilated) input pattern $\tilde{\mathbf{x}}^r$ appears at the input of the MAM. A binary vector \mathbf{z} is said to be a kernel of the association (\mathbf{x}, \mathbf{y}) between the binary vectors \mathbf{x} and \mathbf{y} if it is a subset of \mathbf{x} and satisfies some conditions given in [34], [35]. More recent developments on reconstructing patterns from noisy inputs extend the definition of kernels and associate them with the property of strong morphological independence which should be present in the stored patterns [37-40].

There are two major issues related to the kernel method. The first is the very selection of representative kernel vectors. The conditions for a vector \mathbf{z} to be a kernel vector, given in [34], [35], they do not provide with a fast method for selecting such vectors. As it will be pointed out in section 3 the procedure of selecting kernel vectors is quite time consuming, especially when the pattern vectors are of large dimension. Additional conditions for kernels and selection procedures, which alleviate in some extent this problem, are proposed in [35],[41-43] while in [37-40] the procedure of selecting the kernels emerges through the steps of the proof of the relevant theorems, provided that the patterns are initially modified to meet the property of strong morphological independence.

The second major issue, which is of this paper concern, is related to the behavior of kernel method in the presence of noisy input patterns. Although this issue has received only minor attention in the relevant literature [43], the kernel method is not entirely robust to noise. It can be experimentally verified that, although kernel vectors are quite satisfactory in retrieving associations based on noisy (both dilated and eroded) data, in some cases, they entirely fail even in the presence of very small noise percentages. This

happens when one or more nonzero element of the input pattern, which correspond to respective nonzero elements of the kernel vector are "hit" by erosive noise, which converts them to zero-valued elements. In this case, the retrieval procedure recalls nothing, even if the amount of noise is quite limited.

In this paper, a theoretical proof of this observation is established and the mechanism of producing this failure is explained. Next, a new kernel definition method is proposed which overcomes this problem and is more robust to such kind of noise than the traditional kernel vectors. The new kernels are not anymore binary but they contain elements of variable values in the range [0, 1]. The values of the elements of each selected kernel are related to the frequency of appearance of the corresponding elements of the input training patterns. Different alternatives of selecting kernels based on frequencies are proposed and are compared in respect to their robustness in noise. All frequency based kernel vectors no longer suffer from the particular noise deficiency of conventional kernels. Moreover, the procedure of selecting the kernel vectors is quite simple and well defined; therefore it is not time consuming. The development is given for an auto-associative scheme, but hetero-association can be tackled with the same procedure. The pattern association recall is performed in three steps. The first step is the (*input pattern* -> *kernel*) and (*kernel* -> *output pattern*) recall. The result is an output pattern, which in general is not binary but gray and keeps in a large extent the shape of the output pattern to be recalled. Therefore a simple thresholding action is the second step and may produce the corresponding binary pattern. The third step increases the performance of the proposed method by passing the recalled binary pattern through a simple Hamming network equipped with index selection. The performance of the new kernel in binary pattern recall is demonstrated by using extensive character recognition experiments under various types and percentages of noise.

The paper is organized as follows. First, in section 2, a brief but complete introduction to the morphological associative memories is given. The kernel method is also reported in the same section and its weaknesses are demonstrated. In the sequence, in section 3, the theoretical justification of the failure of the kernel method in the presence of the particular kind of erosive noise is given. In section 4, the new kernel definition method is introduced and its validity and noise robustness is theoretically established. Section 5 provides experimental results regarding the capacity and noise robustness of the proposed kernels, presenting also comparisons between the alternative versions of the frequency based kernels. Conclusions and discussion are finally given in section 6.

2. Morphological associative memories and the kernel method

There are two basic approaches to record k vector pairs $(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^k, \mathbf{y}^k)$ using a *morphological associative memory* (MAM) [14]. The first approach consists of structuring an $m \times n$ matrix \mathbf{W}_{XY} with elements computed by

$$w_{ij} = \bigwedge_{r=1}^k (y_i^r - x_j^r), \quad i = 1, \dots, m \quad j = 1, \dots, n \quad (2.1)$$

The dual approach consists of constructing an $m \times n$ matrix \mathbf{M}_{XY} with elements computed by

$$m_{ij} = \bigvee_{r=1}^k (y_i^r - x_j^r), \quad i = 1, \dots, m \quad j = 1, \dots, n \quad (2.2)$$

If matrix \mathbf{W}_{XY} receives a vector \mathbf{x}^r as input, the product $\mathbf{y}^r = \mathbf{W}_{XY} \otimes \mathbf{x}^r$ is formed. The product is called *max product* and each element of the resulting vector \mathbf{y}^r is computed by the formula

$$y_i^r = \bigvee_{j=1}^n (w_{ij} + x_j^r), \quad i = 1, \dots, m \quad (2.3)$$

Likewise, if matrix \mathbf{M}_{XY} receives \mathbf{x}^r as input the so-called *min product* $\mathbf{y}^r = \mathbf{M}_{XY} \oplus \mathbf{x}^r$ is formed, where each element of the resulting vector \mathbf{y}^r is computed by the formula.

$$y_i^r = \bigwedge_{j=1}^n (m_{ij} + x_j^r), \quad i = 1, \dots, m \quad (2.4)$$

Matrices \mathbf{W}_{XY} and \mathbf{M}_{XY} , computed by (2.1) and (2.2) respectively, constitute the memory of the MAM. The only required training of the network is simply the computation of either of the two matrices \mathbf{W} or \mathbf{M} . The difference between the two memories arises when noisy patterns appear at their input. Memory \mathbf{M}_{XY} is able to retrieve \mathbf{y}^r in case a noisy vector $\tilde{\mathbf{x}}^r$ corrupted by dilative noise appears at its input, while memory \mathbf{W}_{XY} is able to retrieve \mathbf{y}^r in case an eroded version of the input pattern appears at its input. The nature of dilative and erosive noise in binary patterns, as well as the retrieval results of \mathbf{M}_{XY} and \mathbf{W}_{XY} respectively is depicted in Fig. 2.1 and 2.2. In case the input pattern $\tilde{\mathbf{x}}^r$ is corrupted by both erosive and dilative noise none of the memories is able to recall \mathbf{y}^r .

Other types of memories have also been proposed in the recent years. The *median* operator (instead of min or max) is initially proposed in [31] to produce a memory that is robust in strictly mixed type of noise, which presents perfect recall if this noise is of median zero and if the fundamental (uncorrupted) pattern set fulfills some strict conditions. To overcome these restrictions the authors in [31] propose the construction of the memory using specially constructed surrogates of the fundamental pattern set and an algorithm for the training and recall phase using these surrogates. The idea of using surrogates appears also in [32], [33], where a dynamic type of memory is presented and the *mid* operator is employed. Although successful in storing and recalling general pattern associations, dynamic associative memories are

not appropriate for binary pattern associations, because an essential part of their algorithm, namely the determination of active region [32], fails when binary patterns are involved.

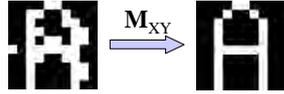


Fig. 2.1. Pattern recall of a dilated input using the min product and memory matrix M .

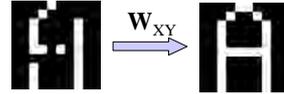


Fig. 2.2. Pattern recall of an eroded input using the max product and memory matrix W .

Another type of memories for binary pattern associations follows from fuzzy oriented treatments [25 - 28]. A new set theoretic interpretation of recording and recall of binary AMM is given in [25-27] and a generalization is provided using fuzzy set theory. This results in the definition of the so-called fuzzy morphological associative memories and the use of the *fuzzy max product* of W and x and the *fuzzy min product* of M and x . When a corrupted binary pattern appears as input to memory W or M the outcome is a gray-level pattern. The respective binary recall is obtained by using appropriate normalized threshold, which however has to be different depending on the type of memory used. An iterative algorithm is also proposed in [25] for the determination of the appropriate threshold. The recall scores are further improved in [28], where an enhanced fuzzy autoassociative morphological memory is presented and is combined with a discrete Hamming neural network that increases the accuracy of the recall. Experiments with patterns corrupted with various noise levels are presented, however the pattern set used is very limited and is not sufficient for drawing clear conclusions.

A well established approach to overcome the problem of the complete failure of memories W (eq. 2.1) and M (eq. 2.2) in dilative and erosive noise respectively is the *kernel* method. To overcome the problem Sussner, Ritter and coworkers [34], [35] proposed the idea of two-step MAMs and the production of the so-called *kernel vectors*, as an elegant representation of the associations (x', y') . These vectors, in conjunction with matrices M and W , are suitable to recall y' when an arbitrarily corrupted (both eroded and dilated) input pattern \tilde{x}' appears at the input of the MAM. A binary vector z is said to be a kernel of the association (x, y) between the vectors x and y if it is a subset of x and satisfies the following conditions.

$$M_{zz} \oplus x = z \tag{2.5}$$

$$W_{zy} \otimes z = y, \tag{2.6}$$

where matrices M_{zz} and W_{zy} are computed according to (2.1), (2.2). In case of autoassociation (2.5) and (2.6) are written as

$$M_{zz} \oplus x = z \tag{2.7}$$

and

$$W_{zx} \otimes z = x \tag{2.8}$$

respectively. Fig. 2.3 shows a sample of binary patterns (letters) and their corresponding kernel patterns, which satisfy equations (2.7) and (2.8). When an arbitrarily corrupted input pattern \tilde{x} appears at the input of a morphological autoassociative network, the original uncorrupted input pattern x is recalled by the following equation:

$$W_{zx} \otimes (M_{zz} \oplus \tilde{x}) = x \quad (2.9)$$

More recent developments on reconstructing patterns from noisy inputs extend the definition of kernels and associate them with the property of strong morphological independence which should be present in the stored patterns [37-40]. A procedure for selecting such kernels is proposed in [37], which also involves a mechanism of altering the initial patterns so that they fulfill the property of strong morphological independence.



Fig. 2.3. Input patterns and underneath their respective kernels

There are two major issues related to the traditional kernel method. The first is the very selection of representative kernel vectors. Equations (2.5), (2.6) (or (2.7), (2.8) for autoassociation) provide the necessary and sufficient conditions for a vector z to be a kernel vector but they do not provide with a fast method for selecting such vectors. An apparent procedure is the random selection of kernel vectors and their acceptance or rejection on the basis of (2.5) and (2.6). This procedure is quite time consuming, especially when the pattern vectors are of large dimension. Additional conditions for kernels and selection procedures are proposed in [35],[41-43].

The second major issue, which is of this paper concern, is related to the behavior of kernel method in the presence of noisy input patterns. Although this issue has received only minor attention in the relevant literature [43], the kernel method is not entirely robust to noise. It can be experimentally verified that, although kernel vectors are quite satisfactory in retrieving binary patterns associations based on noisy (both dilated and eroded) data, in some cases of erosive noise, they entirely fail even in the presence of very small noise percentages. This happens when one or more nonzero elements of the input pattern, which correspond to respective nonzero elements of the kernel vector are "hit" by noise, which converts them to zero-valued elements. In this case, the retrieval equation (2.9) recalls nothing. A theoretical explanation of this fact is given in the following section.

Closing this section, it is worth noting that even in the dynamic MAM presented in [32], [33], there is a part of the memory, namely a one column vector of the respective matrix of the memory, that the authors call *kernel* of the associative memory. The elements of such a kernel play an important

role in recalling patterns altered by some kind of noise. The recall procedure will fail if *kernel* changes, leading to “loss of memory” [32], [33]. However, in this paper, we will not further refer to this type of memories, because dynamic MAM is not appropriate for storing and recalling binary pattern associations. This is due to the failure of determining the so-called “active region” [32] when binary patterns are involved. Another reason is that, with binary patterns, the low accuracy condition (see [32], section 4.2) is very often activated.

3. Noise deficiency of the kernel method

The basic idea of kernel method is to construct a vector (the kernel vector), which is a subset, or otherwise a sparse version, of the binary input vector it represents. Actually, the stricter the subsethood it is the better the performance of the method is. Two very useful conditions for binary kernels are proposed in [34] and in [35]. They are described by the following equations

$$\mathbf{z}^r \sqsubseteq \mathbf{x}^r \tag{3.1}$$

$$\mathbf{z}^r \wedge \mathbf{z}^s = \mathbf{0} \tag{3.2}$$

Equation (3.1) expresses the condition of subsethood described above, while equation (3.2) expresses the demand that different kernels should not have common nonzero points. A pictorial interpretation of the subsethood condition can be drawn from Fig. 2.3. Each binary letter of dimension (10x10) can be scanned row by row and be represented by a (100x1) vector containing only the values 0 and 1. Here, 1 represents the black pixel and 0 represents the white pixel. Similarly, the corresponding kernel patterns can be represented by the (100x1) kernel vectors. The 1s in the kernel vectors are much fewer than the 1s of the corresponding patterns and this is a direct interpretation of (3.1). Moreover, the kernel patterns do not have common non-zero elements and therefore condition (3.2) is fulfilled.

Since each kernel vector satisfies (3.1), each input vector \mathbf{x}^r can be considered as a dilated version of its corresponding kernel vector \mathbf{z}^r . Therefore, since memory \mathbf{M}_{ZZ} is very capable in retrieving patterns corrupted by dilative noise it can recall \mathbf{z}^r by using equation (2.5) (or (2.7)) and the vector \mathbf{x}^r as its input. In the sequel, each recalled kernel vector \mathbf{z}^r can be used as the input to the memory \mathbf{W}_{ZX} or \mathbf{W}_{ZY} to retrieve \mathbf{x}^r or \mathbf{y}^r by using equation (2.6) or (2.8) respectively. The combination of these two steps is expressed in equation (2.9). In case an arbitrarily corrupted version $\tilde{\mathbf{x}}^r$ of the input pattern is considered, equation (2.9) can still provide reasonably good results, provided that $\tilde{\mathbf{x}}^r$ can be considered as a dilated version of \mathbf{z}^r . If, however, $\tilde{\mathbf{x}}^r$ can be considered as a corrupted version \mathbf{z}^r , which contains even the slightest amount of erosive noise, the recall procedure entirely fails and equation (2.5) and consequently (2.9) recalls nothing. The theorems that follow theoretically explain this situation. A pictorial representation of the

situation is also given in Fig. 3.1. It can be observed that if the input pattern is hit by erosive noise, which affects even one pixel that corresponds to a nonzero element of the corresponding kernel pattern, then the recall procedure completely fails.

For notational convenience, we call as kernel point any element of the kernel vector which has nonzero value (value 1). We also recall that, kernel vectors satisfy (3.1) and (3.2).

Theorem 3.1. The rows of \mathbf{M}_{zz} are vectors with values

1. All zeros ($m_{ij} = 0, \forall j$) if the row index (i) does not correspond to any kernel point of any kernel vector.
2. All ones ($m_{ij} = 1, \forall j$) if the row index (i) corresponds to a kernel point belonging to any kernel vector z^r , except when the column index (j) corresponds to another kernel point of the same kernel vector z^r . In this case $m_{ij} = 0$

The proof of Theorem 3.1 is given in the Appendix. A pictorial representation of \mathbf{M}_{zz} for the kernels of Fig. 2.3 is given in Fig. 3.2, where the values of \mathbf{M}_{zz} are displayed in the form of a binary image (0 = black, 1= white). In this example, \mathbf{M}_{zz} is of dimension (100x100).

Theorem 3.2. If there is at least one index i , such that $\tilde{x}_i^r = 0$ and $z_i^r = 1$, then equation (2.5) recalls nothing. That is

$$\mathbf{M}_{zz} \oplus \tilde{\mathbf{x}} = \mathbf{0} \quad (3.3)$$

and consequently

$$\mathbf{W}_{zx} \otimes (\mathbf{M}_{zz} \oplus \tilde{\mathbf{x}}) = \mathbf{0} \quad (3.4)$$

The proof of Theorem 3.2 is given in the Appendix. It explains why there is such a complete failure in the presence of this particular kind of erosive noise, even if the noise percentage is very low.

4. The new kernel definition method

In the sequel we propose a new method for recalling associations between binary patterns, by constructing kernel vectors which does not present the noise deficiency of the conventional kernel vectors described in the previous section. Moreover, the procedure of constructing the kernels requires only moderate processing time and therefore is computationally much more efficient than the conventional kernel method. In our approach kernel vectors are not binary but their elements are allowed to take values in the interval [0, 1]. Each kernel vector \mathbf{z} is constructed to be a subset of an input vector \mathbf{x} (see equation (3.1)), but the subsethood is defined in a fuzzy-like manner.

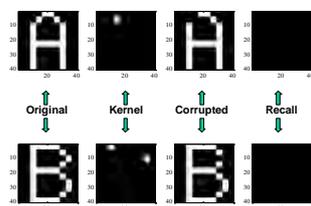


Fig. 3.1. Failure of the recall procedure when even a slight amount of erosive noise affects pixels, which correspond to nonzero kernel elements.

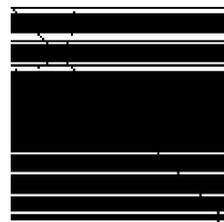


Fig. 3.2. Pictorial representation of the values of \mathbf{M}_{zz} . A black pixel corresponds to $m_{ij} = 0$, while a white pixel corresponds to $m_{ij} = 1$.

Definition 4.1. The new kernel vectors are formed based on the following conditions.

1. Kernel vectors contain values that are in the interval $[0,1]$.
2. The values of the elements of each kernel vector are zero when the corresponding values of the input vector are zero and nonzero when the corresponding input vector values are nonzero.
3. The nonzero values of each kernel vector are formed based on the frequency, f_j , of appearance of the corresponding nonzero input vector element in all the input vectors according to the following **method 1**:
4. $z_j^r = 1$ if $f_j = 1$, $z_j^r = 0.8$ if $f_j = 2$, $z_j^r = 0.7$ if $f_j = 3$, $z_j^r = 0.4$ if $f_j \geq 4$

We call the non-zero elements of a kernel vector **kernel points**. Each kernel point is associated with its strength. The kernel points with $z_j^r = 1$ are the **strongest kernel points**, while kernel points with $z_j^r = 0.8, 0.7, 0.4$ are weaker kernel points. In using the frequencies to construct kernels there are more than one alternative. Two such alternatives, termed **method 2** and **method 3**, are presented later on in this section. However, **method 1** is enough to demonstrate the idea and draw the relevant conclusions.

Fig. 4.1 shows examples of kernel vectors constructed according to **method 1** and associated with the input patterns (letters) of Fig. 2.3. Different gray levels represent the kernel points of different strength with the stronger kernel points corresponding to the brighter intensity. The algorithm for constructing kernel vectors is quite simple and obvious and therefore its computational burden is quite low in comparison with the conventional kernel method. Indeed, to construct a kernel vector \mathbf{z} that corresponds to an input vector \mathbf{x} we proceed as follows: If the value of the element of the input vector is 0 the corresponding element of the kernel vector is set to 0. For each nonzero element of the input vector we count the nonzero occurrences of the same element in the other input vector. We call this frequency of

appearance. Next we assign a nonzero value to the corresponding element of the kernel vector according to the counted frequency and according to *method 1* of definition 4.1. The same procedure is applied for the construction of all the kernel vectors.

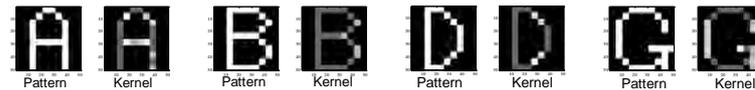


Fig. 4.1. Pictorial representation of patterns and their corresponding kernel patterns constructed using the proposed new method. The brighter pixels correspond to stronger kernel points.

Each new kernel vector is a subset of its corresponding input vector according to equation (3.1), but the subsethood is now defined in a fuzzy like manner since for each element of the kernel vector the following relation holds

$$\mathbf{z}_i^r \leq \mathbf{x}_i^r, \forall i \quad (4.1)$$

The new kernel definition comes as an intuitive extension of conventional binary kernels. In meeting (3.1) and (3.2), denoting subsethood and uniqueness, the resulting binary kernels are sparse vectors, carrying all information in the few nonzero values. All these nonzero values are crucial and, as pointed out in section 3, even if one of them is missing (has zero value) in the input pattern the whole recalling procedure collapses. In the new definition, the “gray” kernel vectors are allowed to include also other supporting non-zero values, associated with non-zero pattern elements appearing in more than one pattern. Due to the definition of the new kernel vectors, which is based on the frequencies of the elements of the corresponding input vectors, the form of memory matrix \mathbf{Mzz} is now completely different than the corresponding matrix produced by conventional kernels. The new form of \mathbf{Mzz} allows the recalling procedure to be successful even if the input vector is corrupted by the special erosive noise described in section 3. The particular *method 1* for selecting the gray values $\{1, 0.8, 0.7, 0.4\}$ to be attributed in kernel points associated with the respective pattern non-zero element frequencies $\{1, 2, 3, \geq 4\}$, came out of author’s observations and is sufficient for demonstrating the idea and drawing the conclusions. As it is proved in the next theorems, the proposed *method 1* serves the purpose of removing the particular noise deficiency of traditional kernels. Other frequency based methods may also be valid. Two such alternative methods (*method 2* and *method 3*) are presented at the end of this section, which carry the good property of *method 1*, and perform better in large pattern sets and are more robust in mixed noise.

Theorem 4.1. The rows of \mathbf{Mzz} are vectors with values

- All zeros ($m_{ij} = 0, \forall j$) if the row index (i) does not correspond to any kernel point of any kernel vector (frequency 0).

- $m_{ij} \in I_1 = \{1, 0.6, 0.3, 0.2, 0\}$ if the row index i corresponds to kernel value $z_i^r = 1$ (frequency 1).
- $m_{ij} \in I_2 = \{0.8, 0.4, 0.1, 0\}$ if the row index i corresponds to kernel value $z_i^r = 0.8$ (frequency 2).
- $m_{ij} \in I_3 = \{0.7, 0.3, 0\}$ if the row index i corresponds to kernel value $z_i^r = 0.7$ (frequency 3)
- $m_{ij} \in I_4 = \{0.4, 0\}$ if the row index i corresponds to kernel value $z_i^r = 0.4$ (frequency ≥ 4).

The proof of Theorem 4.1 is given in the Appendix.

Theorem 4.2. If there are one or more indices i , such that $\tilde{x}_i^r = 0$ and $z_i^r \neq 0$, then equation (2.5) recalls the kernel vector or an eroded version of it. That is

$$\mathbf{M}_{ZZ} \oplus \tilde{\mathbf{x}} = \tilde{\mathbf{z}} \tag{4.2}$$

and consequently

$$\mathbf{W}_{ZX} \otimes (\mathbf{M}_{ZZ} \oplus \tilde{\mathbf{x}}) = \mathbf{x} \tag{4.3}$$

The proof of Theorem 4.2 is given in the Appendix. Fig. 4.2 shows examples of perfect recall using the new kernel vectors constructed by *method 1* and as input vectors uncorrupted patterns. Figure 4.3 shows examples of perfect recall using the new kernel vectors and as input vectors patterns corrupted by the erosive noise described in section 3.

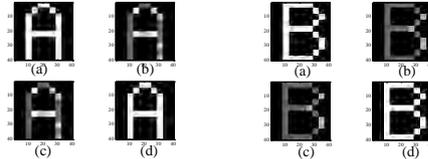


Fig. 4.2. Examples of perfect recall using the new kernel vectors and input vectors the uncorrupted patterns: (a) the input, uncorrupted vector (b) the original kernel vector (c) perfect recall of the kernel vector by (2.5) (d) perfect recall of the pattern by (2.9)

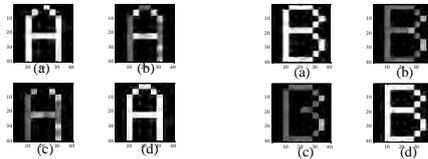


Fig. 4.3. Examples of perfect recall using the new kernel vectors and input vectors corrupted by the erosive noise described in section 3. (a) the corrupted vector used as input (b) the original kernel vector (c) non-perfect recall of the kernel vector by (2.5) (d) perfect recall of the pattern by (2.9)

It has to be noted that term $(\mathbf{M}_{ZZ} \oplus \tilde{\mathbf{x}})$ in (4.3) will produce $\tilde{\mathbf{z}}$, which can be considered as an eroded version of \mathbf{x} . Memory \mathbf{W}_{ZX} is robust in erosive noise but its robustness depends also on the amount of the erosion. Therefore, with the exception of experiments involving small pattern sets with probably small percentages of mixed noise, eq. (4.3) will not produce a binary pattern but a gray surrogate of it, having a shape quite close to it. A binary version of the recall can be obtained after a simple thresholding action. Similar to [28], the recalling results can be further improved if the binary recall passes through a simple Hamming network to produce an index pointing to a pattern of the initial uncorrupted pattern base. The overall recalling scheme is shown in Fig. 4.4 and its performance is demonstrated by the experiments appearing in the next section.

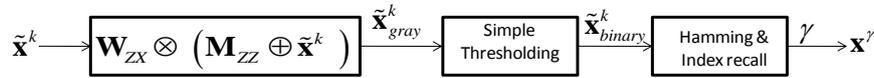


Fig. 4.4. The three step recall procedure. The corrupted binary pattern $\tilde{\mathbf{x}}^k$ passes through eq. (4.3) to produce the gray recall $\tilde{\mathbf{x}}_{gray}^k$, which after simple thresholding will produce a probably corrupted binary pattern recall. The third step will produce an index γ to be used for retrieving the uncorrupted binary pattern \mathbf{x}^γ . If $\gamma = k$ we have perfect recall.

In definition 4.1 the values of the elements of kernel vectors are determined by their frequency of appearance of the corresponding element in the patterns. However, in **method 1** only 4 different values are employed, as all elements having frequency greater than 4 receive the same value $z_j^r = 0.4$. Two alternative frequency based methods are presented below.

Method 2: A plausible extension of method 1 is to allow for a finer determination of kernel point values taking into account all available frequencies of appearance. Since the maximum frequency of a non-zero pixel cannot be greater than the number of available patterns (*nop*) the kernel elements receive the following grading

$$z_j^r = 1 \text{ if } f_j = 1, \quad z_j^r = 1 - \frac{f_j}{nop} \text{ if } f_j \geq 2$$

Regarding the special noise deficiency of kernels, **method 2** carries the same good properties of **method 1**, because theorems 4.1 and 4.2 can be easily extended to cover this method by simply increasing the number of different cases examined to at most *nop* cases. In this approach the sets I_j in Theorem 4.1 now contain more and different values than those of **method 1**, but the rationale of their development and their usage in Theorem 4.2 remain the same. Regarding noise robustness, in the experiments carried out in the next section, **method 2** proves to be more robust than **method 1**.

Method 3: A slightly different approach for constructing the gray kernels is the frequency dependent two-step kernel construction method, which is described here in detail. In the first step frequency patterns are created. Then the kernel vectors are determined by giving values to the kernel points according to the strength of common frequencies in the frequency patterns. More specifically,

Step 1. For each initial pattern \mathbf{x}^r a corresponding frequency pattern p^r is constructed. The elements of the frequency pattern are constructed according to the following rule

$$p_j^r = 0 \text{ if } x_j^r = 0, \quad p_j^r = f_j \text{ if } x_j^r \neq 0$$

Where, as usual, f_j is the frequency of appearance of the non-zero j^{th} element in all patterns. Apparently the minimum value f_j can take is 1.

Step 2. Based on the set of “frequency” patterns, the set of common frequencies F is determined. That is, F contains all the distinct frequencies that appear in **all** the “frequency” patterns. Apparently, the cardinality c_F of F is at most equal to the number of patterns nop ; usually it is smaller. Let also $f_{\min} = \min f_i \in F$ and $f_{\max} = \max f_i \in F$ denote the minimum and maximum of the frequencies participating in F . Next, for each pattern, its kernel vector z^r is constructed according to the following rule.

$$z_j^r = 0 \text{ if } p_j^r = 0, \text{ or } p_j^r \notin F$$

$$z_j^r = 1 \text{ if } p_j^r = f_{\min}, \quad z_j^r = 1 - \frac{p_j^r - f_{\min}}{f_{\max} - f_{\min}} \text{ if } p_j^r > f_{\min}$$

That is, the nonzero elements of each kernel vector correspond to elements of the “frequency” patterns having frequencies that appear in all other “frequency” patterns. The concept of strong and weaker kernel points remain, because the values of the kernel points are determined by the corresponding frequencies, with the largest value (=1) taken by elements that correspond to the minimum common frequency and the other values (<1) are gradually reducing according to the increase of the corresponding common frequency.

Regarding the special noise deficiency of kernels, *method 3* carries the same good properties of *method 1* and *method 2*, because the construction of **Mzz** is analogous to that of the other methods. Theorems 4.1 and 4.2 can be easily extended to cover this method by simply considering different number of cases examined (the number of cases is $c_F + 1$, counting also the $m_{ij} = 0$ case). The sets l_i in Theorem 4.1 now contain different values than those of the other methods, but the rationale of their development and their usage in Theorem 4.2 remain the same. Regarding noise robustness, in the experiments carried out in the next section, *method 3* proves to be much more robust than the other two methods.

5. Experimental results

To test the performance of the proposed method, a number of experiments were carried out. In order to have a meaningfully sized data set, 26 uncorrupted binary patterns were initially created. The set contains the 35x34 binary images of the capital letters of the Latin alphabet and is shown in Fig. 5.1. Some examples of gray kernels produced by *method 1*, *2* and *3* respectively are shown in Fig. 5.2.



Fig. 5.1. The complete pattern set consisting of 35x34 binary images of the 26 capital letters of English alphabet.



Fig. 5.2. Sample of kernel vectors produced by the three frequency based methods. (a) Kernels of *method 1* (b) Kernels of *method 2* (c) Kernels of *method 3*.

All methods (1, 2 &3) of constructing kernels were initially tested on their recall performance of *uncorrupted* patterns, when the three step procedure of Fig. 4.4 was applied. All methods had a 100% success in recalling the uncorrupted pattern.

Fig. 5.3 shows three examples (one for each method) of the overall procedure. The pattern interacts with memory \mathbf{M} according to (4.2) to produce \mathbf{z} . Then, \mathbf{z} interacts with memory \mathbf{W} according to (4.3) to produce a gray pattern $\tilde{\mathbf{x}}$ having the shape of \mathbf{x} . Next, the gray outcome passes from a simple thresholding scheme to produce a binary pattern. In this simple thresholding scheme every gray value greater than zero receives value 1. In the sequel, the binary recall may pass through a Hamming network to recall an index pointing to a member of the initial pattern set. Finally, the uncorrupted pattern associated with this index appears as the final recall of this procedure. It has to be noted that, in the experiments carried out, we did not use a Hamming network. Instead the Hamming distances

$$H_r = \sum_{i=1}^n |x_i^r - \tilde{x}_i| \quad r = 1, \dots, k \quad (5.1)$$

between the binary recall and each one of the patterns of the initial pattern space were computed and the index of the pattern giving the smallest distance was selected. The pattern recalled is \mathbf{x}^γ , where

$$\gamma = \arg \min_r \{H_r\}, \quad r = 1, \dots, k \quad (5.2)$$

In (5.1), n denotes the size (in pixels) of a pattern, i denotes the i^{th} element of the pattern vector and k is the number of patterns in the pattern space.

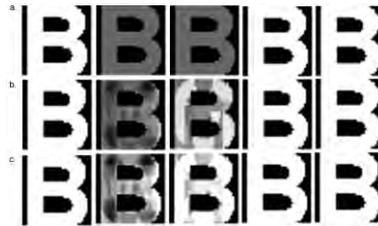


Fig. 5.3. Steps of uncorrupted pattern recall according to the procedure of Fig. 4.4 using the three frequency based kernel methods. **(a)** recall using *method 1*, **(b)** recall using *method 2*, **(c)** recall using *method 3*. From left to right: **1st column** - input pattern, **2nd column** - kernel recall using (4.2), **3rd column** - pattern recall using (4.3), **4th column** – binarization of column 3 recall using simple thresholding, **5th column** – final recall after computing the minimum Hamming distance.

The use of (5.1) and (5.2) provides also a means for indirectly evaluating the storage capacity of the overall scheme. Starting from a small sized pattern space ($k = 5$) and gradually increasing the size by 2 until $k = 26$, the average minimum distance of all pattern recalls at each pattern space for each method is recorded. Figure 5.4 shows the plot of the average minimum distance (AMD) of recalls in respect to the size of the pattern space. It is evident that the AMD increases in the beginning but it approaches an upper bound as the number of patterns increases. That is, after a certain number of patterns the AMD and indirectly the storage capacity of the scheme is not affected by the size of the pattern space. Moreover, Fig. 5.4 demonstrates the superiority of method 3, since it provides always more confident results (smaller AMD).

The next set of experiments is concerned with the robustness of all methods under the presence of various percentages of mixed type (both dilative and erosive) noise. Mixed noise was randomly applied on each binary pattern. The noise percentage is computed by counting the number of the altered (noisy) pixels and divide them with the total number of image pixels. Figure 5.5 shows three examples (one for each method) of the overall recall procedure, when a corrupted by 30% mixed noise pattern appears in the input of the proposed 3-step recall scheme.

A new method for constructing kernel vectors in morphological associative memories of binary

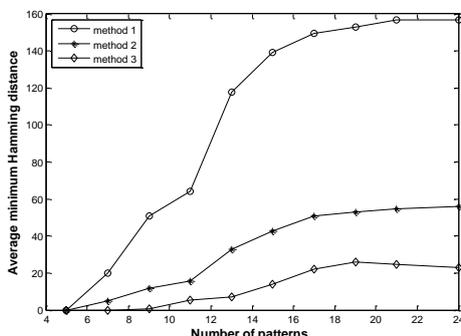


Fig. 5.4. Average minimum Hamming distance for the three frequency based kernel methods in respect to the size of the pattern space.



Fig. 5.5. Steps of successful pattern recall of a corrupted pattern according to the procedure of Fig. 4.4 using the three frequency based kernel methods. **(a)** recall using *method 1*, **(b)** recall using *method 2*, **(c)** recall using *method 3*. From left to right: **1st column** – corrupted input pattern, **2nd column** - kernel recall using (4.2), **3rd column** - pattern recall using (4.3), **4th column** – binarization of column 3 recall using simple thresholding, **5th column** – final recall after computing the minimum Hamming distance.

The various recall stages are similar with those appearing in Fig. 5.3, however the effect of noise is now evident in the intermediate results. It is also clear that *method 2* is more robust to noise than *method 1*, because the intermediate results are closer to the actual ones. Similarly, *method 3* performs much better than the other two.

Finally, Fig. 5.6 depicts the robustness of each method in mixed noise. The failure recall rate of each method is displayed in respect to the percentage of the mixed noise. Since the noise is randomly applied, in order to have statistically reliable results, each experiment corresponding to a different noise percentage was carried out 50 times and the average failure recall rate is actually recorded and displayed. It is evident that *method 3* is far more robust than the other two methods and *method 2* is significantly more robust than *method 1*. Therefore, although all frequency based kernel methods do not present the particular noise deficiency of traditional kernel methods, they present significant differences regarding their noise robustness. This in turn may prompt for future seeking of alternative

Yiannis.S. Boutalis

frequency based kernel construction methods with better performance in recall of noisy patterns.

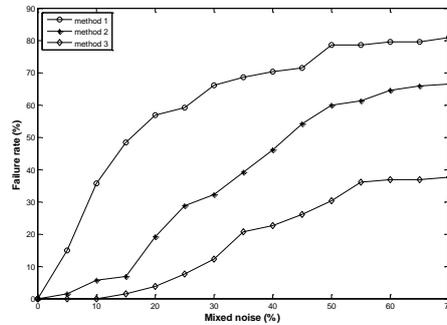


Fig. 5.6. Failure recall rates of the three frequency based kernel vector methods in respect to mixed noise percentage.

6. Conclusion

A new method for constructing kernel vectors was proposed in this paper to be used for recalling associations between binary patterns. The need for a new kernel definition arose from a special noise deficiency of the conventional binary kernel vectors of the relevant literature. The new kernels are not binary but 'gray', because they contain elements with values in the interval $[0, 1]$. These values are determined by the frequency of appearance of nonzero elements of the input pattern vectors. Alternative schemes of producing kernels based on frequency were presented and it was shown, both theoretically and by character recall examples that the new kernel vectors carry the good properties of conventional kernel vectors and, at the same time, they can be easily computed. Moreover, they do not suffer from the particular noise deficiency of the conventional kernel vectors. Experiments, performed on large set of binary patterns corrupted by various degrees of mixed noise demonstrate the robustness and the limits of the proposed approach. Future extension of this work might extend the proposed approach to cover more frequency dependent kernel alternatives as well as an investigation regarding the applicability of the proposed concepts in gray and possibly color pattern associations.

Acknowledgement. The author wants to acknowledge the contribution of Mr. S. Chatzichristofis in performing the experiments carried out.

References

1. Peter Sussner and Marcos Eduardo Valle, Morphological and certain fuzzy morphological associative memories for classification and prediction. In Vassilis Kaburlasos and Gerhard Ritter, (eds), Computational Intelligence Based on Lattice Theory, chapter 8, Springer-Verlag, Heidelberg, Germany, 149–172, (2007).
2. J.L.Davidson, G.X.Ritter, A theory of morphological neural networks, in: Digital Optical Computing II, Vol. 1215, Proc. SPIE, 378–388, (July 1990).
3. J.J.Hopfield, Neural networks and physical systems with emergent collective computational abilities, Vol. 79, Proc. of the National Academy of Sciences, U.S.A., 2554–2558, (April 1982).
4. J.J.Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Vol. 81, Proc. of the National Academy of Sciences, U.S.A., 3088–3092, (May 1984).
5. J.L.Davidson, Simulated annealing and morphological neural networks, in: Image Algebra and Morphological Image Processing III, Vol. 1769, Proc. SPIE, San Diego, 119–127, (1992).
6. J.L.Davidson, F.Hummer, Morphology neural networks: an introduction with applications, Circuits, Systems and Signal Processing, Vol. 12, No. 2, 177–210, (1993).
7. C.P.Suarez–Araujo, Novel neural network models for computing homothetic invariances: an image algebra notation, Journal of Mathematical Imaging and Vision, Vol. 7, No. 1, 69–83, (1997).
8. Y. Won, P. D. Gader, Morphological shared weight neural network for pattern classification and automatic target detection, In Proc. of the IEEE International Conference on Neural Networks, Perth, Australia, 2134–2138, (November 1995).
9. P. D. Gader, M. Khabou, and A. Koldobsky, Morphological regularization neural networks, Pattern Recognition., Special Issue on Mathematical Morphology and Its Applications, Vol. 33, No. 6, 935–945, (2000).
10. L. Pessoa and P. Maragos, Neural networks with hybrid morphological/rank/linear nodes: A unifying framework with applications to handwritten character recognition, Pattern Recognition, Vol. 33, 945–960, (2000).
11. Handbook of Neural Computation, E. Fiesler and R. Beale, (Eds), IOP Publishing and Oxford University Press, pp. C1.8:1–C1.8:14. W. Armstrong, M. Thomas, “Adaptive logic networks”, (1997).
12. B. Raducanu, M. Graña, and X. F. Albizuri, Morphological scale spaces and associative morphological memories: Results on robustness and practical applications, J. Math. Imaging and Vision, Vol. 19, No. 2, 113–131, (2003).
13. M. Graña, J. Gallego, F. J. Torrealdea, and A. D’Anjou, On the application of associative morphological memories to hyperspectral image analysis, Lecture Notes in Computer Science, Vol. 2687, 567–574, (2003).
14. G.X.Ritter, P.Sussner, J.L.Diaz de Leon, Morphological associative memories, IEEE Trans. Neural Networks, Vol. 9, No. 2, 281–293, (1998).
15. W.K. Taylor. Electrical simulation of some nervous system functional activities, Information Theory, Vol. 3, 314–328, (1956).
16. K. Steinbruch. Die lernmatrix. Kybernetick, Vol. 1, 36–45, (1961).
17. J. A. Anderson and E. Rosenfeld, (Eds), Neurocomputing: Foundations of Research, Vol. 1, MIT Press, Cambridge, MA, (1989).
18. T.Kohonen, Correlation matrix memory, IEEE Trans. Comput. Vol. 21, 353–359, (1972).

19. James A. Anderson, Andras Pellionisz, and Edward Rosenfeld, (Eds). *Neurocomputing: Directions for Research*, Vol. 2, MIT Press, Cambridge, MA, (1990).
20. Toshiyuki Tanaka, Shinsuke Kakiya, and Yoshiyuki Kabashima, Capacity Analysis of Bidirectional Associative Memory, In Proc. Seventh Int. Conf. Neural Information Processing, Taejon, Korea, Vol. 2., 779-784, (Nov. 2000).
21. G. X. Ritter and P. Sussner. An introduction to morphological neural networks, In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 709–717, (1996).
22. G. X. Ritter and P. Sussner, Morphological neural networks In *Intelligent Systems: A Semiotic Perspective*, In Proceedings of the 1996 International Multidisciplinary Conference, Gaithersburg, Maryland, 221–226, 1996.
23. G. Costantini, D. Casali, and R. Perfetti, Neural associative memory storing gray-coded gray-scale images, *IEEE Trans. Neural Networks*, Vol. 14, No. 3, 703–707, (2003).
24. P. Sussner and M. E. Valle, Gray-Scale Morphological Associative Memories, *IEEE Trans. Neural Networks*, Vol. 17, No. 3, 559-570, (2006).
25. P. Sussner, Generalizing operations of binary morphological autoassociative memories using fuzzy set theory, *J. Math. Imaging and Vision*, Special Issue on Morphological Neural Networks, Vol. 9, No. 2, 81–93, (2003).
26. P. Sussner, A relationship between binary autoassociative morphological memories and fuzzy set theory, In Proceedings of the Int. Joint Conf. on Neural Networks, Washington DC, (2001).
27. P. Sussner, A Fuzzy autoassociative morphological memory, In Proceedings of the Int. Joint Conf. on Neural Networks, Portland OR, (2003).
28. G. Urcid, J.A. Nieves-V., and C.A. Reyes-G, Enhanced fuzzy autoassociative morphological memory for binary pattern recall, In Proceedings 4th IASTED Int. Conf. on Signal Processing, Pattern Recognition, and Applications, Innsbruck, Austria, (2007).
29. Roberto A. Vazquez and H. Sossa, Behavior of morphological associative memories with true-color image patterns, *Neurocomputing*, Vol. 73, NO. 1-3, 225-244, (2009).
30. M.E. Valle: A Class of Sparsely Connected Autoassociative Morphological Memories for Large Color Images, *IEEE Transactions on Neural Networks*, Vol. 20, No. 6, 1045 – 1050, (2009).
31. H. Sossa, R. Barron and Roberto A. Vazquez, New associative memories to recall real-valued patterns, *Lecture Notes in Computer Science*, Springer, Vol. 3287, 195-202, (2004).
32. Roberto A. Vazquez and H. Sossa, A new associative model with dynamical synapses, *Neural Processing Letters*, Vol. 28, No. 3, 189-207, (2008).
33. Roberto A. Vazquez and H. Sossa, A bidirectional heteroassociative memory for true-color patterns, *Neural Processing Letters*, Vol. 28, No. 3, 131-153, (2008).
34. G.X.Ritter, J.L.Diaz de Leon and P.Sussner, Morphological bidirectional associative memories, *Neural Networks*, Vol. 12, 851-867, (1999).
35. P. Sussner, Observations on morphological associative memories and the kernel method, *Neurocomputing*, Vol. 31, 167-183, (2000).
36. P. Sussner, Associative morphological memories based on variations of the kernel and dual kernel methods, *Neural Networks*, Vol. 16, No. 5, 625–632, (2003).
37. G.X. Ritter, G.Urcid, and L. Iancu, Reconstruction of noisy patterns using morphological associative memories, *Journal of Mathematical Imaging and Vision*, Vol. 19, 95-111, (2003).

38. [G. Urcid and G.X. Ritter, Kernel computation in morphological associative memories for grayscale image recollection, In Proceedings of 5th IASTED Int. Conf. on Signal and Image Processing, Honolulu, Hawaii, (2003).
39. G.X. Ritter and P. Gader, Fixed Points of Lattice Transforms and Lattice Associative Memories, in P. Hawkes (Ed.) Advances in Imaging and Electron Physics Elsevier, Vol. 144, San Diego, California, 2006.
40. G. Urcid, G.X. Ritter, L. Iancu, Kernel computation in morphological bidirectional associative memories, Progress in Pattern Recognition, Speech, and Image Analysis, Springer LNCS 2905 Proceedings, 563-570, (2003).
41. T. Ida, S. Ueda, M. Kashima, T. Fuchida and S. Murashima, On a method to decide kernel patterns of morphological associative memory, Systems and Computers in Japan, Vol. 33, No. 7, 85-94, (2002).
42. M. Hattori, A. Fukui and H. Ito, A fast method of constructing kernel patterns for morphological associative memory, In Proceedings of 9th International Conference on Neural Information Processing, 1058-1063, (Singapore 2002).
43. H. Hashiguchi and M. Hattori, Robust morphological associative memory using multiple kernel patterns, In Proceedings 10th International Conference on Neural Information Processing, (2003).

7. Appendix (Proof of Theorems)

Notation: In the following theorems an element of a conventional binary kernel vector is said to be a kernel point if its value is 1. An element of a new, "gray" kernel vector is said to be a kernel point if its value is > 0 .

Proof of Theorem 3.1. To prove the theorem we start from the formula of computing the elements of \mathbf{Mzz} . According to equation (2.2) the elements m_{ij} of \mathbf{Mzz} are computed by

$$m_{ij} = \bigvee_{r=1}^p (z_i^r - z_j^r), \quad i = 1, \dots, n \quad j = 1, \dots, n \quad (\text{A.1})$$

where n is the length of kernel vector z .

In general the internal term $(z_i^r - z_j^r)$ can take the following values

- (a) $z_i^r - z_j^r = 0$. This happens when for a specific r both z_i^r and z_j^r are kernel points or both are not kernel points
- (b) $z_i^r - z_j^r = 1$ This happens when for a specific r z_i^r is a kernel points and z_j^r is not a kernel points
- (c) $z_i^r - z_j^r = -1$ This happens when for a specific r z_i^r is not a kernel point and z_j^r is a kernel point

Taking now into account that (A.1) is computed by taking the maximum of all the internal terms $(z_i^r - z_j^r)$ for $r = 1 \dots p$, we conclude the following.

(1) If index i does not correspond to any kernel point of any kernel vector then $z_i^r = 0, \forall r$. Consequently, for each j ($z_i^r - z_j^r$), $r = 1 \dots p$ takes the values 0 and -1 . Therefore, the maximum of these values is 0 and therefore $m_{ij} = 0, \forall j$.

(2) In case index i equals the index of a kernel point in any kernel vector z^γ , then taking into account situation (b) above, $m_{ij} = 1, \forall j$ except when $z_j^\gamma = 1$ (that is situation (a) above holds) and simultaneously $z_i^r = z_j^r = 0, \forall r \neq \gamma$, where $m_{ij} = 0$. This means that when the kernel points of a kernel vector do not coincide with the kernel points of another kernel vector (or otherwise kernel points appear only once) (which holds true due to condition (3.2)), then $m_{ij} = 0$ when both i and j correspond to kernel points.

Proof of Theorem 3.2. We take into account that in computing equation (2.5), that is,

$M_{zz} \oplus x = z$, the i^{th} element of z is computed using the elements of the i^{th} row of M_{zz} according to equation (2.4) as follows

$$z_i^r = \bigwedge_{j=1}^n (m_{ij} + x_j^r), i = 1, \dots, n$$

We distinguish the following cases

1. Index of line, i , of M_{zz} does not correspond to any kernel point of any kernel vector. In this case according to Theorem 3.1 $m_{ij} = 0, \forall j$, and

therefore $\bigwedge_{j=1}^n (m_{ij} + x_j^r) = \bigwedge_{j=1}^n x_j^r = 0$. The last equation holds because there is the plausible assumption that at least one element of the input vector x^r is of zero value.

2. Index of line, i , of M_{zz} corresponds to a kernel point of kernel vector z^r . In this case, according to Theorem 3.1, ($m_{ij} = 1, \forall j$), except when the column index (j) corresponds to a kernel point of the same kernel vector z^r . In this case $m_{ij} = 0$. Therefore, $\bigwedge_{j=1}^n (m_{ij} + x_j^r) =$

$$\bigwedge_{j \notin \text{kernel points}} (1 + x_j^r), \bigwedge_{j \in \text{kernel points}} (0 + x_j^r) = \begin{cases} 0 & \text{if } \exists x_j^r = 0 : j \in \text{kernel points} \\ 1 & \text{otherwise} \end{cases}$$

In the above notation an index, j , is said to belong to the kernel points if the element of the kernel vector having the same index, j , is a kernel point. The above equation proves that a kernel vector z^r can be recalled by equation (2.5) only if all the elements of the input vector x^r , which correspond to kernel points, have the value 1. In other words, if even one element of the input vector, which corresponds to a kernel point, is hit by erosive noise then equation (2.5) recalls nothing and so does equation (2.9).

Proof of Theorem 4.1. To prove the theorem we start from the formula (A.1), which computes the elements of \mathbf{Mzz} . Now, the new kernel vectors are used.

In general the internal term $(z_i^r - z_j^r)$ can take values from the set $\{-1, -0.8, -0.7, -0.6, -0.4, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1\}$ with the extreme values to appear as follows

(a) $z_i^r - z_j^r = 0$. This happens when for a specific r both z_i^r and z_j^r are kernel points of the same value (strength).

(b) $z_i^r - z_j^r = 1$ This happens when for a specific r z_i^r is one of the strongest kernel points (value 1) and z_j^r is not a kernel point (value 0)

(c) $z_i^r - z_j^r = -1$ This happens when for a specific r z_i^r is not a kernel points and z_j^r is one of the strongest kernel points

Taking now into account that (A.1) is computed by taking the maximum of all the internal terms $(z_i^r - z_j^r)$ for $r = 1 \dots p$ we conclude the following.

(1) If index i does not correspond to any kernel point of any kernel vector then $z_i^r = 0, \forall r$. Consequently, for each j $(z_i^r - z_j^r), r = 1 \dots p$ takes the values 0, -0.4, -0.7, -0.8, -1. Therefore, the maximum of these values is 0 and therefore $m_{ij} = 0, \forall j$.

(2) If index i corresponds to a kernel point with value $z_i^r = 1$ (frequency 1), then $m_{ij} \in I_1 = \{1, 0.6, 0.3, 0.2, 0\}$ with the value $m_{ij} = 1$ being the most frequent. It has to be noted that the values of m_{ij} are formed by $(z_i^r - z_j^r)$ of kernel r only. The values $(z_i^\gamma - z_j^\gamma), \gamma \neq r$ are always smaller since $z_i^\gamma = 0, \gamma \neq r$ (Frequency 1 means only $z_i^r = 1$) The values of m_{ij} are resolved as follows: $m_{ij} = 1, \text{ if } z_j^r = 0, m_{ij} = 0.6, \text{ if } z_j^r = 0.4, m_{ij} = 0.3, \text{ if } z_j^r = 0.7, m_{ij} = 0.2, \text{ if } z_j^r = 0.8, m_{ij} = 0, \text{ if } z_j^r = 1$

(3) If index i corresponds to a kernel point with value $z_i^r = 0.8$ (frequency 2), then $m_{ij} \in I_2 = \{0.8, 0.4, 0.1, 0\}$ with the value $m_{ij} = 0.8$ being the most frequent. For the formation of m_{ij} only two kernel vectors are actually responsible, which have $z_i^r = 0.8$. We call this set of (two) kernels as forming set and denote it as F . The values of $(z_i^\gamma - z_j^\gamma), \gamma \notin F$ are always smaller since $z_i^\gamma = 0, \gamma \notin F$ (Frequency 2 means only two $z_i^r = 0.8, r \in F$) The values of m_{ij} are resolved as follows:

$$m_{ij} = 0.8 \quad \text{if } z_j^\gamma = 0, \text{ at least for one } \gamma \in F$$

$$m_{ij} = 0.4 \quad \text{if } z_j^\gamma = 0.4, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.4$$

Yiannis.S. Boutalis

$$m_{ij} = 0.1 \text{ if } z_j^\gamma = 0.7, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.7$$

$$m_{ij} = 0 \text{ if } i = j \text{ or}$$

$$m_{ij} = 0 \text{ if } z_j^\gamma = 0.8, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.8$$

(4) If index i corresponds to a kernel point with value $z_i^r = 0.7$ (frequency 3), then $m_{ij} \in I_3 = \{0.7, 0.3, 0\}$ with the value $m_{ij} = 0.7$ being the most frequent. For the formation of m_{ij} only three kernel vectors are actually responsible, which have $z_i^r = 0.7$. We call this set of (three) kernels as forming set and denote it as F . The values of $(z_i^\gamma - z_j^\gamma), \gamma \notin F$ are always smaller since $z_i^\gamma = 0, \gamma \notin F$ (Frequency 3 means only three $z_i^r = 0.7, r \in F$). The values of m_{ij} are resolved as follows:

$$m_{ij} = 0.7 \text{ if } z_j^\gamma = 0, \text{ at least for one } \gamma \in F$$

$$m_{ij} = 0.3 \text{ if } z_j^\gamma = 0.4, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.4$$

$$m_{ij} = 0 \text{ if } i = j$$

$$m_{ij} = 0 \text{ if } z_j^\gamma = 0.7, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.7$$

(5) If index i corresponds to a kernel point with value $z_i^r = 0.4$ (frequency ≥ 4), then $m_{ij} \in I_4 = \{0.4, 0\}$ with the value $m_{ij} = 0.4$ being the most frequent. For the formation of m_{ij} only the kernel vectors, which have $z_i^r = 0.4$ are actually responsible. We call this set of kernels as forming set and denote it as F . The values of $(z_i^\gamma - z_j^\gamma), \gamma \notin F$ are always smaller since $z_i^\gamma = 0, \gamma \notin F$. The values of m_{ij} are resolved as follows:

$$m_{ij} = 0.4 \text{ if } z_j^\gamma = 0, \text{ at least for one } \gamma \in F$$

$$m_{ij} = 0 \text{ if } i = j$$

$$m_{ij} = 0 \text{ if } z_j^\gamma = 0.4, \text{ at least for one } \gamma \in F \text{ and for the remaining } z_j^\gamma \geq 0.4$$

Proof of Theorem 4.2. Similarly to the proof of theorem 3.1, we take into account that in computing equation (2.5), that is, $\mathbf{M}_{zz} \oplus \mathbf{x} = \mathbf{z}$, the i^{th} element of \mathbf{z} is computed using the elements of the i^{th} row of \mathbf{M}_{zz} according to equation (2.4) as follows

$$z_i^r = \bigwedge_{j=1}^n (m_{ij} + x_j^r), i = 1, \dots, n$$

We distinguish the following cases

1. Index of line, i , of \mathbf{M}_{zz} does not correspond to any kernel point of any kernel vector. In this case $m_{ij} = 0, \forall j$, therefore $\bigwedge_{j=1}^n (m_{ij} + x_j^r) = \bigwedge_{j=1}^n x_j^r = 0$.

The last equation holds because there is the plausible assumption that at least one element of the input vector \mathbf{x}^r is of zero value.

A new method for constructing kernel vectors in morphological associative memories of binary

2. Index of line, i , of \mathbf{M}_{zz} corresponds to a kernel point of any strength of kernel vector \mathbf{z}^r . In this case ($m_{ij} \in [0, m_{str}]$), where $m_{str} = \{1, 0.8, 0.7, 0.4\}$ depending on the strength of the kernel point with index i . Therefore, if $x_j^r = 1, \forall j$ being an index corresponding to a kernel point of \mathbf{z}^r of any strength then $\bigwedge_{j=1}^n (m_{ij} + x_j^r) = m_{str} = z_i^r$. On the other hand if $\exists j \in \{\text{kernel points}\}^r : x_j^r = 0$ then $\bigwedge_{j=1}^n (m_{ij} + x_j^r) \in [0, m_{str}) \leq z_i^r$

In the above notation an index, j , is said to belong to the kernel points if the element of the kernel vector having the same index, j , is a kernel point. The above equation proves that a kernel vector \mathbf{z}^r can be exactly recalled by equation (2.5) if all the elements of the input vector \mathbf{x}^r , which correspond to kernel points, have the value 1. In case, however, one or more points of the input vector, which correspond to kernel points, are hit by erosive noise then (2.5) does not entirely fail. In this case (2.5) recalls an eroded version of \mathbf{z}^r . Here however erosion of the value of one pixel means the replacement of its value by another smaller value. This fact is expressed by equation (4.2), which is repeated here

$$\mathbf{M}_{zz} \oplus \tilde{\mathbf{x}} = \tilde{\mathbf{z}}$$

Where $\tilde{\mathbf{z}}$ denotes the eroded version of \mathbf{z} . Since memory \mathbf{W}_{zx} is robust in erosive noise, equation (2.9) results in recalling the original, uncorrupted pattern according to equation (4.3) as follows

$$\mathbf{W}_{zx} \otimes (\mathbf{M}_{zz} \oplus \tilde{\mathbf{x}}) = \mathbf{x}$$

Yiannis.S. Boutalis

Yiannis Boutalis received the diploma of Electrical Engineer in 1983 from Democritus University of Thrace (DUTH), Greece and the PhD degree in Electrical and Computer Engineering (topic Image Processing) in 1988 from the Computer Science Division of National Technical University of Athens, Greece. Since 1996, he serves as a faculty member, at the Department of Electrical and Computer Engineering, DUTH, Greece, where he is currently an Associate Professor and director of the Automatic Control Systems lab. Currently, he is also a Visiting Professor for research cooperation at Friedrich-Alexander University of Erlangen-Nuremberg, Germany, chair of Automatic Control. He served as an assistant visiting professor at University of Thessaly, Greece, and as a visiting professor in Air Defence Academy of General Staff of airforces of Greece. He also served as a researcher in the Institute of Language and Speech Processing (ILSP), Greece, and as a managing director of the R&D SME Ideatech S.A, Greece, specializing in pattern recognition and signal processing applications. His current research interests are focused in the development of Computational Intelligence techniques with applications in Control, Pattern Recognition, Signal and Image Processing Problems.

Received: November 14, 2009; Accepted: July 27, 2010.

Supply Chain Performance Measurement System Based on Scorecards and Web Portals

Nenad Stefanović and Dušan Stefanović

Faculty of Science, University of Kragujevac, Radoja Domanovica 12,
34000 Kragujevac, Serbia
{nenads, dusans}@kg.ac.rs

Abstract. Successful supply chain management becomes essential for the ultimate success of corporations. Companies today seek an effective performance measurement (PM) system to maximize the bottom line. Unfortunately, performance measurement in the supply-chain field has not kept pace with today's world of interdependent business relationships. What companies need is a new PM system that unifies different business elements, concepts, technologies and tools. In this paper, the architecture of such a pervasive PM system is introduced. The main system elements such as process model, metrics and data warehouse are described. Finally, a specialized PM web portal which enables proactive performance monitoring and fosters the improvement and optimization is presented.

Keywords: supply chain, performance, scorecard, web portal.

1. Introduction

Supply chains are growing increasingly complex, from linear arrangements to synchronized, multi-echelon, outward-facing networks of distributed servers. There is much more information that needs to be monitored than there was just a few years ago. Most companies lack the tools that can quickly shift through and present data coming from supply chain partners and systems.

The overall performance of the supply chain significantly affects the financial health of all member companies. Therefore, an effective supply chain performance measurement process should be able to directly address performance areas that create sustainable profitability and financial strength.

In order to accomplish this requirement, the performance measurement process will need to provide a reliable indication of the contribution of supply chain operations to the areas like growth, cost minimization, working capital efficiency and fixed asset utilization

A robust and scalable performance management system is the platform for improvement. It must be exception-based and allow users to prevent problems, resolve issues, capture knowledge, and sustain improvements. The system must be able to handle an increasing number of users and amounts of

information (due to expanded products, members of the supply chain, geography, and time). While it must be personalized and easy to use, it must also ensure high levels of security and privacy.

Supply chain PM cycle is not just for the supply chain, but for all aspects of the enterprise as well as for the extended supply chain. Ultimately, by managing the performance of myriad processes across enterprise boundaries, companies will have achieved the vision of Enterprise Performance Management (EPM) [1].

In supply chain, large volumes of raw transactional data are generated by each process and stored. The challenge for many companies lies in determining what information is necessary to drive improvements and efficiencies at each process in the supply chain, and designing an information management environment to turn the raw data into meaningful metrics and key performance indicators (KPI). Key performance indicators are measurements that directly relate to key business requirements. KPI come in various forms from simple reporting measurements to very complex, cross correlated analytic results.

Information from supply chain management (SCM) processes must be collected, measured, analyzed and continuously monitored. This requires integration of data coming out of ERP (Enterprise Resource Planning), SCM and all other systems supporting these business processes. Data from transactional systems should be summarized into the Data Warehouse (DW), which should be able to scale to large sizes and be continually updated.

A well designed and integrated PM framework increases the capability of business intelligence (BI) systems to provide accurate insights for effective supply chain decision making. BI is evolving from traditional BI to *pervasive* BI (PBI), which empowers everyone in the organization, at all levels, with analytics, alerts and feedback mechanisms. On the benefits side, PBI promises to [2]:

- More effectively leverage the strengths of the whole supply chain by giving every employee the power to contribute to and enhance key performance indicators that have been set by management.
- Increase sustainable competitive advantage by helping every employee to make the right decisions at the right time in step with company and customer objectives.
- Improve operational efficiency by uncovering new best practices and driving those practices from the bottom up as well as the top down.

2. Background Research and Literature Review

Business performance management [3] describes the methods, metrics, processes and systems used in organizations to translate strategies into plans, monitor execution, and provide insight to improve financial and operational performance. It represents the strategic, integrated evolution of business intelligence to support the management process.

The importance of performance measurement in the context of SCM cannot be overstated. Timely and accurate assessment of overall system and individual system component performance is paramount. An effective performance measurement system provides the basis to understand the system, influences behavior throughout the system, and provides information regarding the results of system efforts to supply chain members and outside stakeholders. In effect, performance measurement is the glue that holds the complex value-creating system together, directing strategic formulation as well as playing a major role in monitoring the implementation of that strategy. In addition, research findings suggest that measuring supply chain performance in and of itself leads to improvements in overall performance [4]. Despite its importance, supply chain performance often was measured in oversimplified and sometimes counterproductive (cost-reduction-based) terms [5]. Lack of an appropriate performance measurement system has been cited as a major obstacle to effective supply chain management [6].

Traditionally, companies have tracked performance based largely on financial accounting principles. Financial accounting measures are certainly important in assessing whether or not operational changes are improving the financial health of an enterprise, but insufficient to measure supply chain performance for the following reasons [7]:

- The measures tend to be historically oriented and not focused on providing a forward-looking perspective.
- The measures do not relate to important strategic, non-financial performance.
- The measures do not directly tie to operational effectiveness and efficiency.
- Most performance measurement systems are functionally focused.

Until few years ago, there were several reasons why most companies did not implement supply chain performance measurement systems [8]:

1. No clear established approach or set of measures was available.
2. Software vendor products offered only a limited range of supply chain metrics.
3. Companies were too busy with other more important initiatives.

The traditional approaches to monitoring performance had been metrics projects and balanced scorecards. In metrics projects, functional organizations and workgroups established and tracked metrics that were considered most relevant for measuring performance. Unfortunately, there were a number of limitations with metrics projects:

- By focusing on functional metrics, they ended up driving locally optimized “silo” behavior at the expense of the overall company.
- It was time consuming to compile and analyze information, so visibility often came too late to make a difference. In addition, they only provided information on limited history, not insight into the future.
- Metric tracking was manual, so numbers were often calculated incorrectly or inconsistently over time.
- Many times, workers didn’t know what to do with the data. It wasn’t always clear what constituted poor performance, when to act, or how to act. Or

else, people were so distracted and confused by the measuring process itself that they didn't act.

- Although selected metrics were called key performance indicators, there was no feedback or validation to ensure that organizations were actually measuring the most relevant business drivers.
- Experienced managers learned how to “game” or “tinker with” the metrics to make themselves look good.

In an attempt to overcome some of these limitations, many companies have initiated balanced scorecard (BSC) projects. Based on the methodology of Robert Kaplan and David Norton [9], these organizations created a balanced set of metrics representing financials, customers, internal business processes and innovation. The goal was to enable better decision-making by providing managers with a broader perspective of both tangible and intangible assets. Although conceptually compelling, most balanced scorecards were implemented as static management dashboards, unable to drive action or performance improvement because [10]:

- These dashboards are usually driven out of finance organizations, therefore are typically highly weighted by financial information. Much of the important non-financial data and qualitative information is not captured or synthesized.
- Information is often manually aggregated from operational data sources and is prone to errors and significant delays.
- Infrequent sourcing of information allows people to play tricks operationally to improve the numbers. Who hasn't heard of the manager who shipped orders early or incomplete to reduce inventory levels?
- Where there is data integration, it is often “hard-wired” and difficult to modify over time as strategies and objectives change. Static systems – which encourage the improvement of specific metrics, not necessarily overall business performance – become self-perpetuating due to the fact that those managers successful under the old systems do not want to introduce new ones.
- Executive-level systems are often disconnected from tactics and operations. Because the metrics are high level and presented without regard to their implicit interdependencies, managers are uncertain what action to take to improve overall performance.
- Dashboards do not track decisions and their effectiveness over time so it is difficult for organizations to improve by learning from experience. Moreover, there is no mechanism to embed business rules to help improve the decision-making and problem resolution process itself.
- There is little or no support for collaborative processes across organizations, up and down the chain of command.

The Balanced Scorecard has been successfully implemented at hundreds of companies, however, many companies still need a practical measurement system that will enable them to improve profitability. As Kaplan and Norton state in [11], the execution of the measurement system is more important than the measurement system itself. Accordingly, fewer than 10 percent of the strategies outlined on the Business Scorecard were successfully

implemented. This implies that the measurement strategy must be simplified for a successful execution. 80 percent of enterprises that fail to integrate the balanced scorecard into PM methods and tools will drop the balanced scorecard and return to a less organized and less effective set of metrics [12].

There is a need to establish dynamic supply chain performance measurement systems to effectively manage supply chain operations and meet financial and nonfinancial business objectives. In the following sections, we present a comprehensive supply chain PM model and its realization through a specialized PM web portal that is an integral part of the overall BI model.

3. Architecture of PM System

A robust infrastructure is crucial to realize the benefits of various PM initiatives. Figure 1 shows how main components fit together to create an environment to support PM solution.

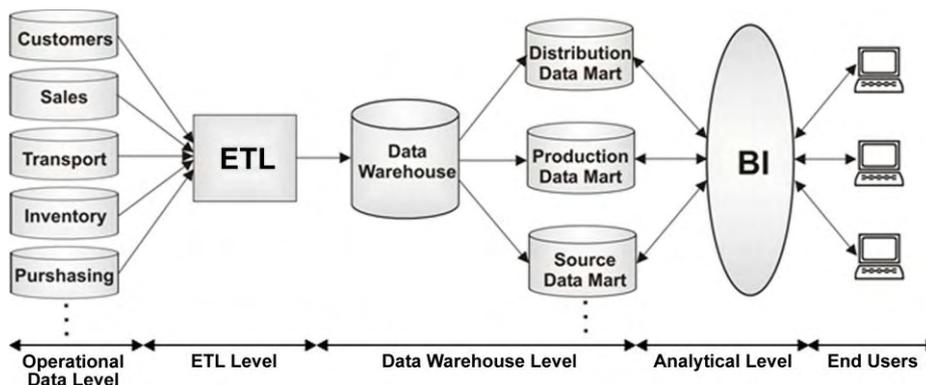


Fig. 1. PM Platform Architecture

The data sources are the basis of all the business intelligence solution extracted from IT systems. They can be anything from ERP backends, CRM systems, HR databases, billing systems, operational databases or any other database that forms a critical part of the SCM business processes. Specific data needs to be extracted from these databases and loaded into a data warehouse to support the kind of heavy analysis and querying that a BI system allows. This layer also referred as ETL (Extraction, Transformation and Loading), is probably the most critical phase in the BI chain. The quality of the data in the warehouse depends on the configuration and capabilities of the ETL engine. A data warehouse is not always the ideal place to fetch data for complex queries and analysis. By extracting data specific to the type of analysis required, data marts improve the speed of queries and processing.

The proposed PM model unifies people, processes, methodologies and tools into a single business solution. PM model is developed in such a way to seamlessly integrate within overall BI and collaboration platform [13].

The architecture of the PM model is presented in Figure 2. The starting point is the SCM process model which provides a library of the supply chain specific set of processes, relationships, metrics and best practices. The developed SCM Metamodel enables flexible modelling and creation of different supply chain configurations (models). These models are the basis for the construction of DW metadata (measures, dimensions, KPI).

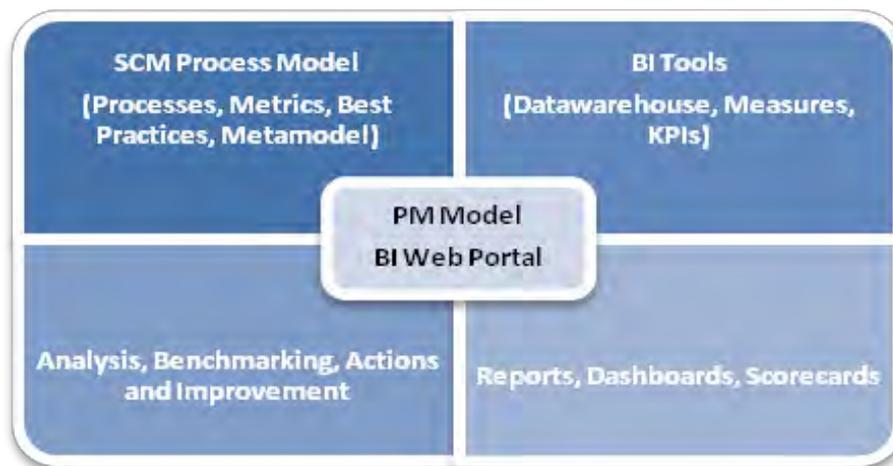


Fig. 2. Architecture of the PM Model

The data warehouse approach, allows the single version of the truth because of the integrated and consolidated data repository. Server-based definition of KPIs offers centralized metadata management and enables creation of various front-end PM and analytical applications. Data warehouse acts as an integrated data source upon which different BI elements are developed such as planning models, performance monitoring systems, reporting and data mining models.

PM web portal holds the central place in the model, as it comprises a portfolio of analytic, collaboration and communication services that connect people, information, processes, and systems both inside and outside the corporate firewall.

We have designed a specialized PM web portal with an adequate set of elements (reports, dashboards and scorecards). The web portal also integrates collaboration and analytical services, and enables users to take actions.

3.1. Supply Chain Metrics

A standardized supply chain process model provides significant benefits. Standardized models provide companies' maps toward business process engineering, establish benchmarking for performance comparison and uncover best business practices for gaining competitive advantage. By standardizing supply chain operations and metrics for managing such operations, companies can not only compare their results against others, but they are also able to gain visibility of operations over a supply chain that may cross corporate borders. Partners in a supply chain can communicate more unambiguously and can collaboratively measure, manage, and control their processes.

The SCM process model contains the standard name for each process element, the notation for the process element, standard definition for the process element, performance attributes that are associated with the process element, metrics that are associated with the performance attributes, and best practices that are associated with the process.

Table 1. SCM Scorecard with Top Level Metrics

Top Level Metrics	Performance Attributes				
	Customer-Facing			Internal-Facing	
	Reliability	Responsiveness	Flexibility	Cost	Assets
Perfect Order Fulfillment	x				
Order Fulfillment Cycle Time		x			
Upside Supply Chain Flexibility			x		
Upside Supply Chain Adaptability			x		
Downside Supply Chain Adaptability			x		
Supply Chain Management Cost				x	
Cost of Goods Sold				x	
Cash-To-Cash Cycle Time					x
Return on Supply Chain Fixed Assets					x
Return on Working Capital					x

All process metrics are an aspect of a performance attribute. The performance attributes for any given process are characterized as either customer-facing (reliability, responsiveness and flexibility) or internal-facing (cost and assets) metrics. A SCM scorecard with all the top-level attributes defined by the SCM process model is provided in Table 1.

These top level metrics are the calculations by which an implementing organization can measure how successful they are in achieving their desired positioning within the competitive market space. Lower level calculations (level 2 metrics) are generally associated with a narrower subset of processes. For example, Delivery Performance is calculated as the total number of products delivered on time and in full based on a commit date. Additionally, even lower level metrics (diagnostics) are used to diagnose variations in performance against plan. For example, an organization may wish to examine the correlation between the request date and commit date.

Each process from the process model has its related metrics, best practices and inputs and outputs. All the metrics follow the same template which consists of the following elements:

- Name
- Definition
- Hierarchical metric structure
- Qualitative relationship description
- Quantitative relationship (optional, if calculable)
- Calculation
- Data collection

For example, the Table 2 contains all the data for the top level metric - Cash-to-Cash Cycle Time.

Based on the SCM process model, we have created the SCM Metamodel [14] (Fig. 3), which enables creation of any supply chain configuration and is the basis for further modelling. The Metamodel is normalized and contains all SCM elements such as processes, metrics, best practices, inputs and outputs. It also incorporates business logic through relationships, cardinality, and constrains.

The Metamodel is extended with additional entities to support supply network modelling. That way, processes, metrics and best practices can be related to the specific node and tier in the supply network. With this Metamodel, processes at different levels can be modelled thus providing more detailed view of supply chain processes and metrics.

This method offers several advantages:

- Better functionality and flexibility of the model
- Metamodel contains SCM knowledge which enables domain-specific modeling.
- The usage of relational database enables integrity of data and models, data importing and exporting, as well as option to use the standard language (SQL) for querying.
- Security and user access control.

Supply Chain Performance Measurement System Based on Scorecards and Web Portals

- It is possible to design front-end web application that can serve as the interface for collaborative supply network modeling.
- Possibility to add or change both the library data (processes, metrics, best practices, etc.) and the data related to models (supply network configurations).

Table 2. Structure of the Cash-to-Cash Cycle Time metric

<i>Metric Definition</i>	
The time it takes for an investment made to flow back into a company after it has been spent for raw materials.	
<i>Hierarchical Metric Structure</i>	
Level 1	
Level 2	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 5px; text-align: center;">Days Sales Outstanding</div> <div style="border: 1px solid black; padding: 5px; text-align: center;">Inventory Days of Supply</div> <div style="border: 1px solid black; padding: 5px; text-align: center;">Days Payable</div> </div>
<i>Qualitative Relationship Description</i>	
The longer the cash-to-cash cycle, the more current assets needed (relative to current liabilities) since it takes longer to convert inventories and receivables into cash.	
<i>Calculation</i>	
<p>[Cash-to-Cash Cycle Time = Inventory Days of Supply + Days Sales Outstanding – Days Payable Outstanding]</p> <p>Level 2 Metrics:</p> <p>Inventory Days of Supply = the amount of inventory (stock) expressed in days of sales. The [5 point rolling average of gross value of inventory at standard cost] / [annual cost of goods sold (COGS) / 365]</p> <p>Days Sales Outstanding = the length of time from when a sale is made until cash for it is received from customers. The [5 point rolling average of gross accounts receivable (AR)] / [total gross annual sales / 365].</p> <p>Days Payable Outstanding = the length of time from purchasing materials, labour and/or conversion resources until cash payments must be made expressed in days. The [5 point rolling average of gross accounts payable (AP)] / [total gross annual material purchases / 365].</p>	
<i>Data Collection</i>	
general ledger system, accounts receivable system, accounts payable system, purchasing system, production reporting system, customer relationship management system	

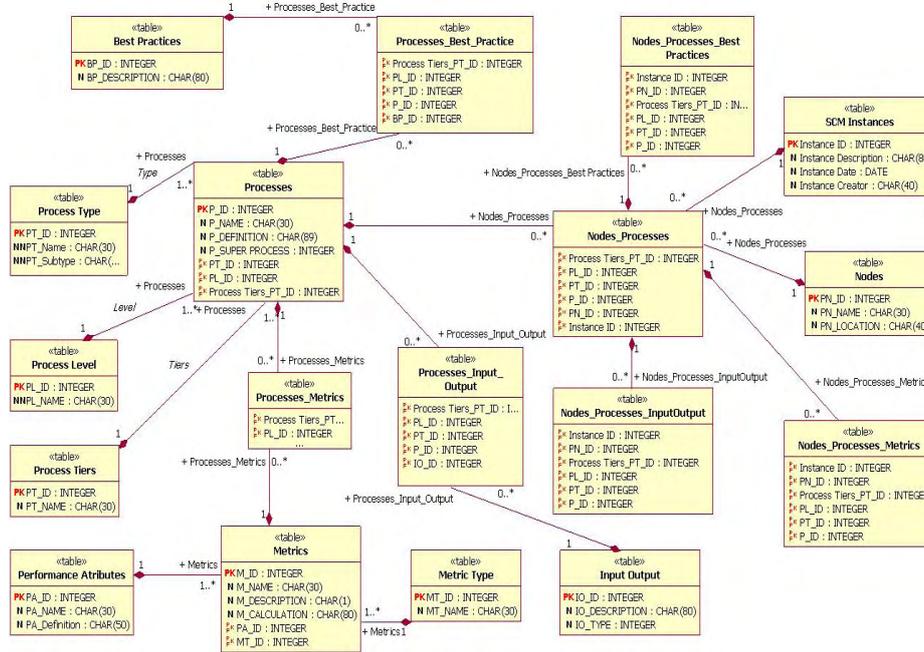


Fig. 3. SCM Metamodel

The presented supply network modeling method and the Metamodel can also be the basis for design of simulation models and scenario analysis.

3.2. PM Data Warehouse

Design of the data warehouse as one of the main components of the supply network BI system is more complex and dynamic than the standard IT projects. The main reasons are: broader business context, heterogeneous and distributed data sources, different software tools and technologies (databases, ETL, analytics, etc.), compound definition of business requirements along the supply chain, the need for specific skilled workforce, cultural and organizational barriers.

A user who wants to retrieve information directly from a data source, such as an ERP database, faces several significant challenges:

- The contents of such data sources are frequently very hard to understand, being designed with systems and developers instead of users in mind.
- Information of interest to the user is typically distributed among multiple heterogeneous data sources.
- Whereas many data sources are oriented toward holding large quantities of transaction level detail, frequently the queries that support business decision-making involve summary, aggregated information.

- Business rules are generally not encapsulated in the data sources. Users are left to make their own interpretation of the data.

In order to overcome these problems, we have constructed the Unified Dimensional Model (UDM) (Figure 3) [15]. The role of a UDM is to provide a bridge between the user and the data sources. A UDM is constructed over one or more physical data sources. The user issues queries against the UDM using a variety of client tools.

Construction of the UDM as an additional layer over the data sources offers more clearly data model, isolation from the heterogeneous data platforms and formats, and improved performance for aggregated queries. UDM also allows business rules to be embedded in the model, as well as option to define actions in relation to query results.

Another advantage of this approach is that UDM does not require data warehouse or data mart. It is possible to construct UDM directly on top of OLTP (On-Line Transactional Processing) systems, and to combine OLTP and DW systems within a single UDM.

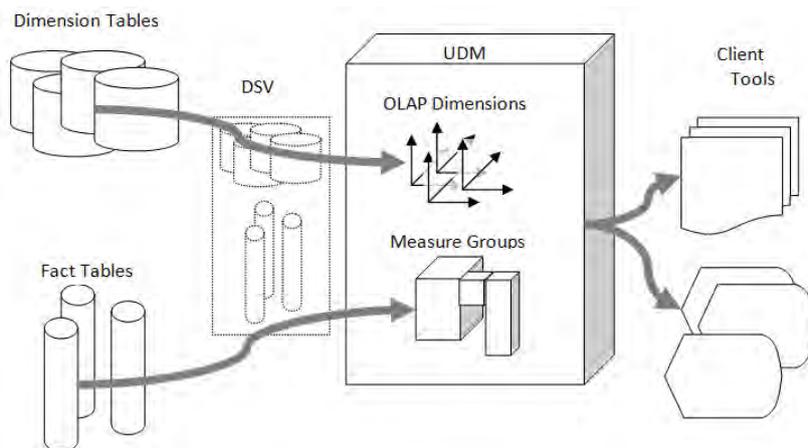


Fig. 3. Unified Dimensional Model

In the UDM we can define cubes, measures, dimensions, hierarchies, and other OLAP (On-Line Analytical Processing) elements, from the DW schemas or directly from the relational database. This enables providing the BI information to the business users even without previously built DW, which can be very useful having in mind facts that within the supply network there can be many nonintegrated data sources which require time to connect, integrate, and design the data warehouse.

Flexibility of UDM also manifests in the fact that tables and fields can be given names and descriptions that are understandable to the end-user and hide unnecessary system fields. These metadata is further used in throughout

the UDM, so all the measures, dimensions and hierarchies that are created based on these table fields will use these new names.

Definitions of all UDM elements are stored as XML (eXtensible Markup Language) files. Each data source, view, dimension or cube definition is stored in separate XML file. For dimensions, these files contain data about tables and fields which store dimension members. OLAP cube definition files also contain information on how the preprocessed aggregates will be managed. This metadata approach enables centralized management of the dimensional model for the entire supply network and provides option for model integration and metadata exchange.

UDM allow creation of one data source view (DSV) for use by the system. The data source view is an abstraction layer that is used to extend the objects (relational tables and views) that are exposed by the data source to a collection of objects from which OLAP server objects are created.

Within the data source view we included all of the relational views that were used to create dimensions, hierarchies, and attributes. Data source views have two important roles:

First, they are a layer of abstraction between the objects that are used by OLAP Server and the data source. This allows creation of objects such as named queries and computed columns, which could be created in the data source itself (for example, in relational views). This is important because OLAP designers might not have the rights necessary to make metadata changes in the supply chain partner's source system.

Data source views allow you to create relationships between tables and views that are not physically located in the database or that are impossible because they are between databases.

Additionally data source views allow you to cache the metadata about the data sources so that you can develop cubes without having a connection to the source data systems.

UDM also enables the creation of hierarchies. Although the consolidation of all the attributes of an entity into a dimension greatly simplifies the model for the user, there are additional relationships between the attributes that a simple list cannot express. For example, product category, subcategory and stock keeping unit, define one of the hierarchies in which products can be organized.

The model can also provide translations of data. An attribute can map to different elements in the data source, and provide the translations for those elements in different languages, which is very beneficial in the supply chain context.

Real-world supply chain models might include tens of measures and dimensions, with each dimension including tens or hundreds of attributes. The UDM also provides special views, called perspectives. A UDM can have many perspectives, each one presenting only a specific subset of the model (measures, dimensions, attributes, and so on) that is relevant to a particular group of users.

Companies often define key performance indicators, which are important metrics used to measure the health of the business. The UDM allows such

KPIs to be defined, enabling a much more understandable grouping and presentation of data. Key performance indicator is a collection of calculations that are associated with a measure group in a cube that are used to evaluate business success. Typically, these calculations are a combination of Multidimensional Expressions (MDX) expressions or calculated members. KPIs also have additional metadata that provides information about how client applications should display the results of the KPI's calculations.

Table 3 lists common KPI elements and their definitions.

Table 3. OLAP KPI Structure

Term	Definition
Goal	An MDX numeric expression that returns the target value of the KPI.
Value	An MDX numeric expression that returns the actual value of the KPI.
Status	An MDX expression that represents the state of the KPI at a specified point in time. The status MDX expression should return a normalized value between -1 and 1.
Trend	An MDX expression that evaluates the value of the KPI over time. The trend can be any time-based criterion that makes sense in a specific business context.
Status indicator	A visual element that provides a quick indication of the status for a KPI. The display of the element is determined by the value of the status MDX expression.
Trend indicator	A visual element that provides a quick indication of the trend for a KPI. The display of the element is determined by the value of the trend MDX expression.
Display Folder	The folder in which the KPI will appear to the user when browsing the cube.
Parent KPI	A reference to an existing KPI that uses the value of the child KPI as part of the KPI's computation.
Current time member	An MDX expression that returns the member that identifies the temporal context of the KPI.
Weight	An MDX numeric expression that assigns a relative importance to a KPI. If the KPI is assigned to a parent KPI, the weight is used to proportionally adjust the results of the KPI value during the calculation.

The use of OLAP-based KPIs allows client tools to present related measures in a way that is much more readily understood by the user. The Figure 4 shows an example of how three KPIs, organized into display folders, might be displayed by a client tool.

Display Structure	Value	Goal	Status	Trend
Customer Scorecard				
Customer Retention Rate	0.92	1		
Financial Scorecard				
Increase Profitability				
Revenue Per Employee	1100	1000		
Sales	1105253	1100000		

Fig. 4. OLAP KPI

4. PM Portal

As the business intelligence market continues to gain momentum, it underscores the increasing need to integrate the growing amounts of information found in disparate locations throughout the supply chain into an integrated framework.

For many organizations, this framework is the enterprise portal. With its ability to integrate the foundations of a strategic BI offering - from collaboration tools to third-party applications to exponentially growing content - enterprise portals remain at the top of the IT agenda as part of a successful PM strategy.

4.1. PM Web Portal

In order to overcome the shortcomings of the existing BI and PM client tools we have designed a specialized web portal that enables supply chain users to monitor business processes, collaborate and take actions [16].

Portal represents the single point of access to all relevant information in a personalized and secured manner. Its composite and service-oriented architecture enables inclusion of different PM components and tools (KPIs, dashboards, scorecards, reports, etc.). PM elements can be personalized and adjusted, and information can be filtered just by using a web browser. PM elements can be defined within the portal and also embedded from the external source (OLAP, another application, spreadsheets) via web services. This information is presented through different special web parts. The portal itself can be a provider (via web services or RSS-Really Simple Syndication) to other applications.

With the PM portal capabilities, supply network partners and teams can:

Supply Chain Performance Measurement System Based on Scorecards and Web Portals

- Use predefined PM portal template with out of the box modules optimized for access and management of reports, data connections, spreadsheets and dashboards. Dashboard pages can contain several web parts, each of them showing information from different data sources.
- Communicate strategy and monitor its execution at different levels of the supply network. KPIs status and trend can be tracked using the special KPI and scorecard web parts. KPIs can display information from different data sources (for example OLAP cubes or Excel spreadsheets).
- Customize and personalize sites, pages or modules by adding or removing certain web parts and by applying filter web parts. Filters allow dashboards to be personalized by communicating shared parameters amongst web parts on a dashboard. For example, the current user filter web part automatically filters information based on who is logged on to the computer. This is useful when you want to display only information, such as customer accounts or tasks that is currently assigned to that user.

Figure 5 shows the specialized SCM Scorecard for the global supply chain performance management.

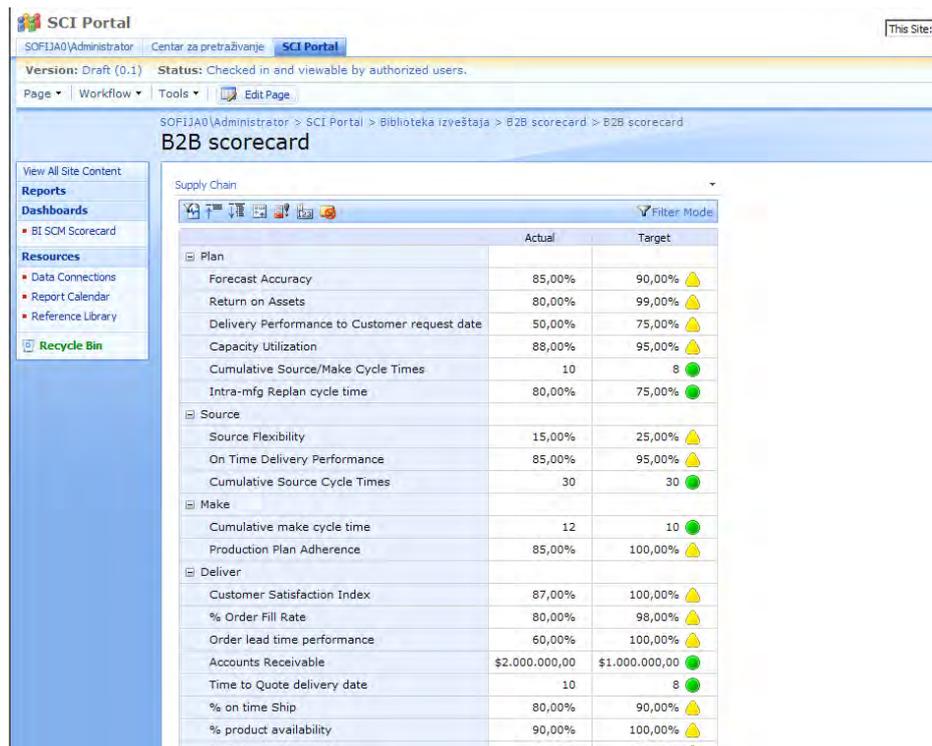


Fig. 5. Supply Chain Scorecard

It is constructed based on top of the OLAP KPIs, which are again based on the SCM process model and metrics. KPI are created by SCM segments (plan, source, make, deliver and return) as hierarchies, so it is possible to perform drill-down analysis and track performance against defined goals.

The presented solution is very flexible in terms of presenting the KPIs. Owing to several specific BI web parts, the portal can display KPIs from the OLAP server, spreadsheets, and other sources (portals, report servers, etc.).

Besides, the portal also support KPI management directly from the portal, meaning that the user can use the KPI web part template to create the new module on the certain web page and manually define KPIs, target, goals, graphic indicators, data sources, etc. This way, the end-users can define, update and track the KPIs without the help of the IT specialist.

The portal also supports the concept of strategy maps provide a hierarchical view of the KPI measures across levels of the organization by presenting relationships, priorities and perspectives. Figure 6 illustrates a PM web page with two web parts.

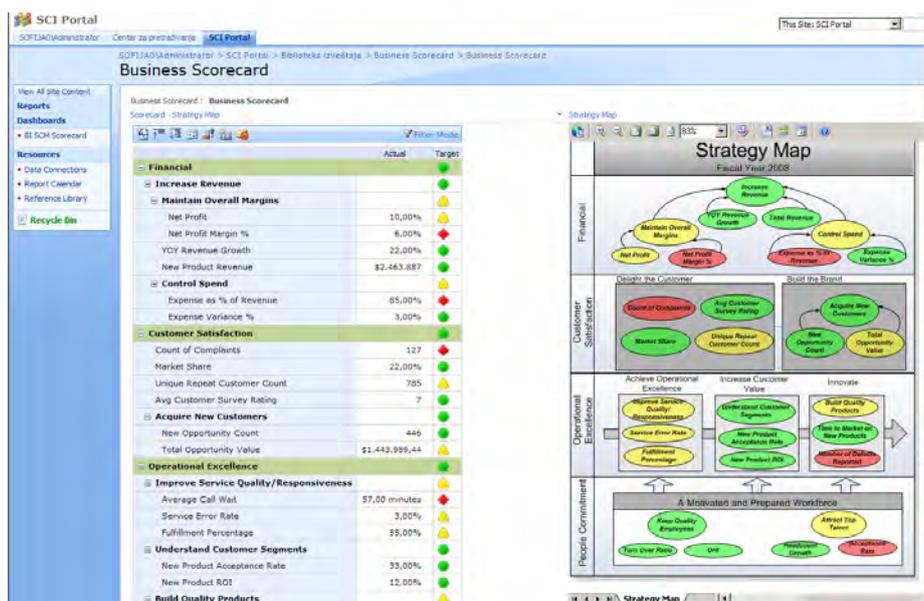


Fig. 6. PM Scorecard and Strategy Map

The first web part renders balanced scorecard with relevant KPIs, and the other web part shows an automatically generated strategy map related to the balanced scorecard. Each element on the map is highlighted with appropriate color. This enables visual performance tracking in relation to predefined strategy. Modular architecture of the portal enables creation of various PM mashups tailored to specific business needs. Web parts, pages and also the complete portals can be saved as templates and reused many times.

These electronic and dynamic strategy maps help explain high-level causes and effects that facilitate making choices. With strategy maps and their resultant choices of strategic objectives and the action items to attain them, managers and employees easily see the priorities and adjust their plans accordingly.

The dashboard page can display numerous metrics and views business on a single screen. Portal supports quick deployment of dashboards assembled from web parts. Each web part can contain a particular view or metric, and users can customize their individual dashboards to display the views that are most meaningful to them, such as those with the metrics they need to monitor on a daily basis.

Additionally, portal supports events and automatic alerting. Users can subscribe to specific documents or keywords and categories, to be notified when metrics are updated or new intelligence becomes available. They can also use other features, such as planning, enterprise search, subscription, and routing functions, to work with team members on a single item (i.e. scorecard, KPI, etc.), and to automate collaborative performance management processes. Portal also provides fine-grained authentication and authorization.

PM portal enables business users to define and use scorecards and key performance indicators to drive accountability and alignment across the supply chain. Scorecards and KPIs reflect planning, budgeting, and forecasting changes in real time to help users understand the business drivers, challenges, and opportunities they face. Monitoring becomes a part of the regular, day-to-day management process.

PM portal enables business users to define and use scorecards and key performance indicators to drive accountability and alignment across the supply chain. Scorecards and KPIs reflect planning, budgeting, and forecasting changes in real time to help users understand the business drivers, challenges, and opportunities they face. Monitoring becomes a part of the regular, day-to-day management process.

Visualization elements and up-to-date performance dashboards and scorecards help business users align their actions with strategic goals. All users can gain insight into business drivers and collaborate more efficiently using detailed, contextual supply chain analysis gained from structured information as well as unstructured information. Rich collaboration capabilities, strategy maps, accountability mapping, and spreadsheet integration, make business information more actionable because all users better understand the context of performance. These PM and analytical functionalities help users rapidly identify trends, opportunities, and even threats hidden within large quantities of data.

Because PM portal gives users easy-to-use functionality for setting up their own performance dashboards and scorecards, IT department is freed for more critical functions such as meeting security and compliance requirements.

4.2. Business Activity Monitoring Web Portal

Most businesses are probably not using BI to continually and automatically monitor events in their operational business processes as their businesses operate to rapidly respond to detected problems or to predict if problems lie ahead. In general, therefore, companies have no active real-time element to their BI systems. The consequences are that nothing is helping the business to automatically respond immediately when problems occur or opportunities arise. Also, there is no automatic notification or flagging of alerts to take action that may avoid unnecessary costs, business disruption, operational mistakes and unhappy customers in the future.

Business Activity Monitoring (BAM) is a collection of tools that allow you to manage aggregations, alerts, and profiles to monitor relevant business metrics (Key Performance Indicators - KPIs). It gives users end-to-end visibility into business processes, providing accurate information about the status and results of various operations, processes, and transactions so they can address problem areas and resolve issues within your business. BAM software products incorporate concepts from — and sometimes are built on — ERP, business intelligence, business process management and enterprise application integration (EAI) software.

We propose the specialized BAM model and architecture based on the data warehouse technology, messaging, business process orchestration and web portals. The Figure 7 shows BAM architecture and how data and messages flow within the BAM system [17].

BAM gives a different perspective on a business process. For example, a BAM system might provide graphical depictions of per-product sales trends or current inventory levels or other key performance indicators. The information might be updated every day, every hour, or more frequently.

There are two important BAM components that we should examine first: activities and views.

BAM relies on one or more BAM activities. A BAM activity refers to the encapsulation and representation of an end-to-end process. An activity could be a single process or a collection of many orchestrations or applications.

Within the activity definition, you generally find two entities:

- Milestones represent how long it takes to process something.
- Data of interest is made up of individual data points that we want to monitor (price, stock level, and so on).

By assembling the milestones and/or the data of interest, we create a BAM activity.

A BAM view is a tailored representation of the data involved similar to the database views. By implementing BAM views, we can disseminate our processes in a manner that is appropriate for the end user. For instance, we may have payroll logic for the accounting department that shelters the actual individual pay rates but presents the final cost of labor for a particular department.

Supply Chain Performance Measurement System Based on Scorecards and Web Portals

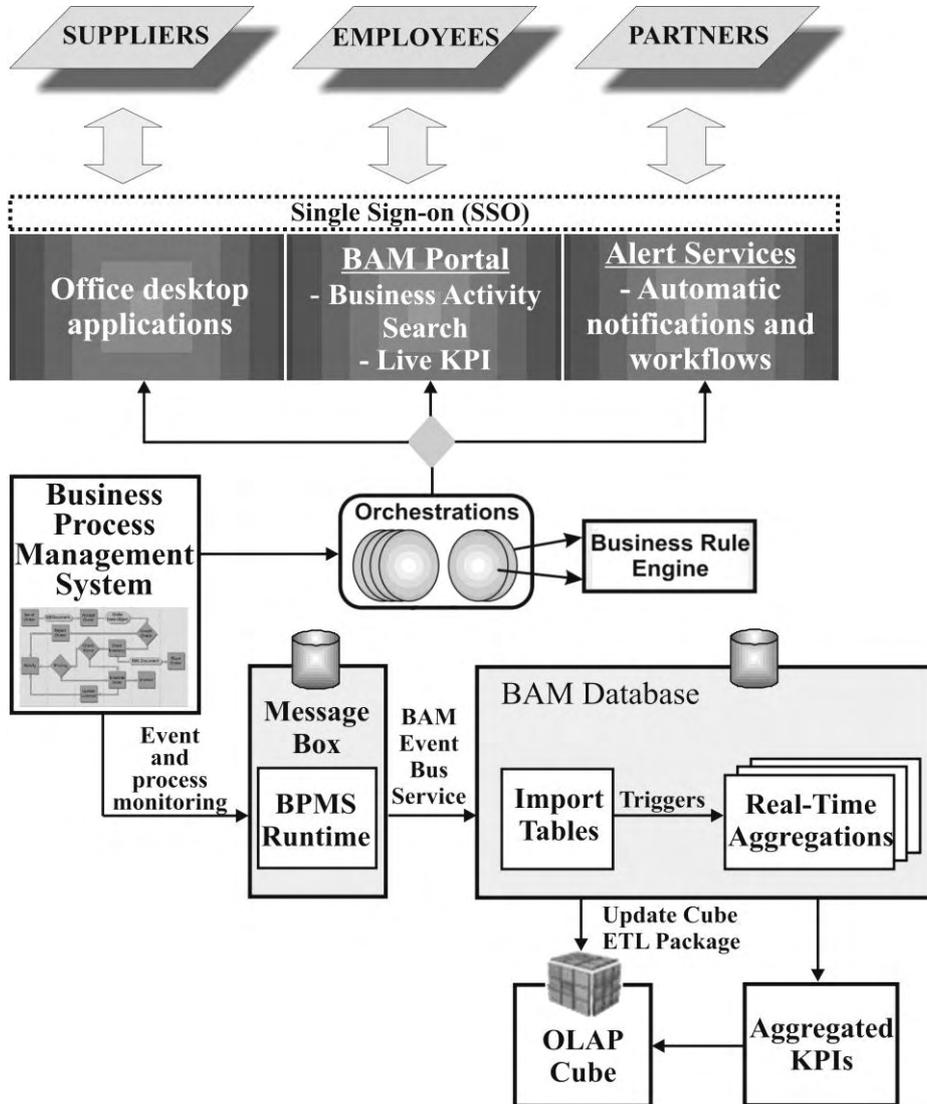


Fig. 7. BAM Architecture

The need for automation and interaction of business processes necessitate the use of modern technologies for managing business process, trading partner relationships and monitoring and analyzing in real-time. For these purposes we have designed two specialized web portals - Business Activity Services (BAS) and Business Activities Monitoring (BAM) portal.

Business Activity Services

Business Activity Services (BAS) provides an interaction and collaboration self-service portal Web site for information workers. BAS enables you to easily configure and interact with business processes and collaborate with trading partners.

BAS provides the infrastructure to capture business user input into a business process easily. Then based on the human input, the business process (defined and automated as BPMS-Business Process Management System orchestration) can continue with the subsequent steps in the pre-defined workflow.

BAS web portal architecture consists of the following modules:

- **Business User Portal** - The self-service Web site that enables business users to interact with partners and business processes through familiar metaphors such as Mailboxes. The business process delivers message documents into the mailboxes where the business user can manage them. A business user can respond to a business process, by saving the document in the Outbox. Then the message is picked by the business process as a response to the previous Inbox message.
- **Trading Partner Management (TPM)** - A set of interactive tools and forms that enable the business user to manage online interactions with trading partners. These include the following: creating and maintaining the partner profiles and the agreements, tracking the daily interactions with them such as the exchange of purchase orders, invoices, payments, and others, and monitoring the status and the statistics regarding the interactions and the documents.
- **Business Process Configuration** - This primarily includes the design and programming of the Orchestrations and TPM elements on the BAS site in such a way that the business users can interact with them. Orchestrations are complex technical workflow schedules. Developers typically create orchestrations. The developer first defines parameter constructs in the orchestration and in TPM. This configuration enables BAS to combine the collaboration portal with the friendly InfoPath forms, with the power of the BPMS Workflow and Messaging engines, creating an effective self-service solution for the business users. Business users can then set the values for these parameters as needed, for different business partners or for different parts of the organization.
- **Business User Interaction and Collaboration** - This is the ultimate business goal of BAS. As soon as the orchestrations and TPM elements go through the configuration process by using the key parameters, the business users can use the end-to-end infrastructure to perform the daily interactions with the trading partners. Business users set the values of the pre-defined parameters that enable them to control and adjust the behavior of the business process orchestrations.

The following are the brief descriptions of the key concepts in BAS:

- **Trading Partner** - An external or internal organization with which your organization exchanges electronic documents and data to perform

Supply Chain Performance Measurement System Based on Scorecards and Web Portals

business operations and transactions. For example, a trading partner could be a supplier, a customer, or an internal department.

- Partner Profile - The basic business information about the partner such as the company name and related contact details. A profile is conceptually similar to a business card. BAS takes advantage of InfoPath forms and templates to manage partner profiles effectively. Figure 8 shows web page for Partner Profiles.
- Self Profile - The self profile contains information about your own company in a way such that when a document sends to a partner you can use the information to identify your company's details to the partner.

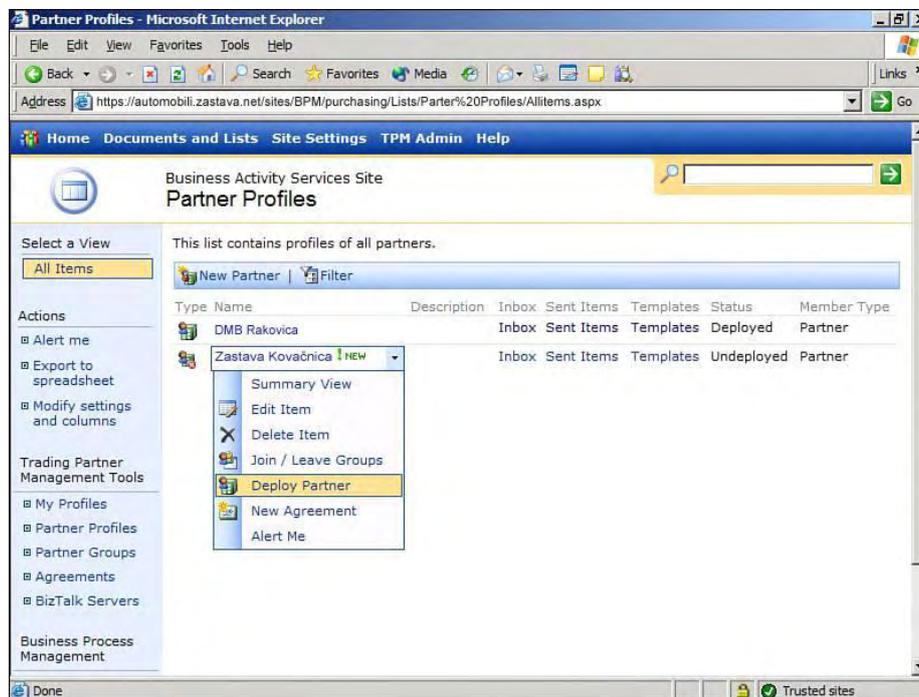


Fig. 8. BAS Portal

- Agreement - The key concept in BAS for defining the relationship with a trading partner. Agreement organizes the definition of business terms in an intuitive manner so that business users can change it with ease. Then the business parameters defined in an agreement translate into the technical elements that drive the business process orchestrations in BPMS.
- Addendum - Addendum is a segment in an agreement that defines the orchestration, policies and parameters involved in the trading relationship.
- Inbox - A document library that BAS uses for receiving messages and documents from a running business process (BPMS Orchestration). This is conceptually analogous to the Inbox used in a typical e-mail system.

- Outbox - A document library that BAS uses for sending messages and documents to a running business process (BPMS Orchestration). This is conceptually analogous to the Outbox used in a typical e-mail system.
- Sent Items - A document library that BAS uses for archiving messages and sent documents. The Sent Items folder serves as the storing mechanism for tracking the sent business documents for future reference and auditing needs.

Business Activity Monitoring

One of the newest technology trends is toward the integration of BAM systems with web portals. Business end users can use the BAM portals to monitor KPIs, which measure progress toward a business goal, as well as other information about their business process. Information workers use BAM portal to gain a real-time holistic view of business processes that span heterogeneous applications, regardless of the infrastructure implementation.

There are two ways information workers can use BAM to view business processes: using the spreadsheet application and through BAM web portal.

Each view gives a different perspective on a business process. For example, a BAM view might provide graphical depictions of per-product sales trends or current inventory levels or other key performance indicators. The information in these views might be updated every day, every hour, or more frequently.

Each BAM view relies on one or more BAM activities. A BAM activity represents a specific business process, such as handling purchase orders or shipping a product, and each one has a defined set of milestones and business data. For example, a purchase order activity might have milestones such as Approved, Denied, and Delivered along with business data like Customer Name and Product.

The following list describes how other ways information workers can use BAM features [18]:

- View a single activity instance such as a purchase order or loan (process) in real-time or as historical data.
- This view shows only the data relevant to the business process the knowledge worker is concerned with and hides complexity of the heterogeneous implementation.
- Search for activity instances based on their progress or business data (Figure 9). For example, we can search for loans that are waiting for customer signature and the dollar amount is greater than a given value.
- Browse aggregations (which are key performance indicators) around all the business activities that are currently being processed or have already happened. The aggregations can be done in real-time or can be based on a snapshot of the activities taken at specific time.
- Navigate to the related activity instances such as shipments associated with given purchase order, or the Invoice in which it is included.

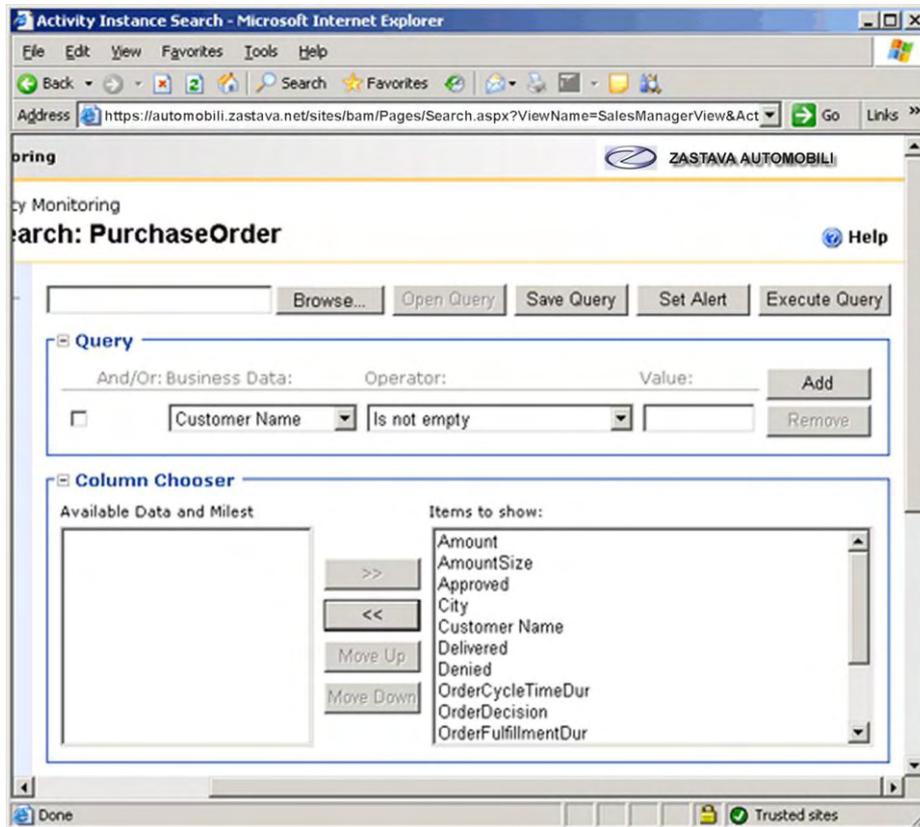


Fig. 9. BAM search interface

Additionally, it is possible to create different activity-related alerts. Alerts allow us to define important events about business processes, such as Key Performance Indicators, that can be delivered to users on a real-time basis. Users subscribe to alerts to receive notification of the business event that the alert monitors.

There are two types of alerts, aggregate and instance. An aggregate alert allows users to specify threshold data across a time frame whereas an instance alert is based on specific qualifying data points.

Alerts are delivered in one of two methods, through an e-mail message to the subscribers or as a file written to a system-specified location (e.g. network file share or web folder).

The key benefit of the presented BAM environment is that operational processes can be monitored and exceptions acted on in close to real time. Other benefits include better supply chain visibility, operational productivity, risk avoidance and business continuity.

Some of the supply chain processes that are the best candidates for the BAM are: order management, product recalls, quality control and monitoring,

sales forecasting and pipeline monitoring, just-in-time inventory monitoring and parts receivable, real-time inventory analysis, real-time marketing and promotions, transportation tracking, and product recalls.

For example, in an order management process, we can define KPIs for how many orders are received, their money amounts, and whether they are processed successfully. It is also possible to define KPIs that compare order volumes by channel, customer type, or other categorizations.

Another good example of the BAM use case is one related to the supplier delivery. It is possible to specify that if the activity confirming the receipt of the order is not finished by five days after the order has been initiated, then the order will most likely be late. In that case, the user should be notified via some graphical indicators (typically a red bar to indicate that a threshold has been exceeded). Additionally, the BAM solution can be configured to notify the user via email, desktop message or even SMS, so that he can then notify the affected parties or consider an alternative solution. Afterward, he can drill down to the individual activity to get additional details on why the order was late (lack of resources, for example).

5. Conclusion

Supply chain performance measurement is vital for a company in order to survive in today's competitive business environment. Supply chain performance measurement should be a business-critical process, driven by metrics and supported by business intelligence. With increasing competition and changing market forces, tapping into this critical asset is essential in sustaining competitive advantage in the global space.

PM model presented in this paper fuses all relevant elements such as people, business processes, metrics, KPIs and tools into a single comprehensive system. The underlying architecture supports the complete cycle of BI/PM processes like ETL, DW, OLAP and reporting.

The unified data model helps to establish a single version of the truth on supply chain performance, which is quantifiable and understood by all entities in the supply chain.

Specialized PM web portal offers the following benefits:

- Real-time supply chain monitoring and alerting.
- Flexibility, personalization and customization.
- Integration with existing transactional, BI and collaboration systems
- Built-in knowledge and best practices through predefined web parts and templates.

The presented BAM portal provides a rich view into data collected via BAS system and enables activity data to be searched and viewed in a variety of ways. Because the BAM data is held in database tables and views, it's easy to access the information from a variety of tools, including different reporting tools, which can produce a highly detailed tracking portal providing very rich business intelligence.

With PM and BAM systems in place, all parties in a supply chain network can track the real-time flow of goods, money, and information across the network. They can also answer customer questions on the state of the process, no matter who is handling the process at a given point in time.

References

1. Parker, B.: EPM: From Concept to Implementation, AMR Research's Spring Executive Conference: Creating the Real-Time Enterprise, May 29-31. (2002)
2. Mittlender, D.: Pervasive Business Intelligence: Enhancing Key Performance Indicators, DM Review. (August 2005)
3. Vriens, D.: Information and Communications Technology for Competitive Intelligence, Idea Group Publishing. (2004)
4. Bello, D. C., Gilliland, D. I.: The Effects of Output Controls, Process Controls, and Flexibility on Export Channel Performance, *Journal of Marketing* 61 p. 22 (Winter 1997); and Stank, T. P., Lackey Jr., C. W.: Enhancing Performance Through Logistical Capabilities in Mexican Maquiladora Firms, *Journal of Business Logistics* Vol. 18, No. 1, 91–123. (1997)
5. Foster, T. A.: It Pays to Measure Performance: Logistics Performance Compensation Programs, *Chilton's Distribution* 90. (September 1991)
6. Lee, H. L., Billington, C.: Managing Supply chain Inventory: Pitfalls and Opportunities, *Sloan Management Review*, Vol. 33, No. 3, 65–73. (1992)
7. Lappide, L.: What About Measuring Supply Chain Performance, *ASCET*, Vol. 3, Montgomery Research. (2002)
8. Gintic.: Measuring supply chain performance using a SCOR-based approach, March. (2002)
9. Kaplan, R. and Norton, D.: *The Balanced Scorecard*, Harvard Business School Press, Boston. (1996)
10. Hammer, M.: *The Agenda*, Chapter 6: Measure Like You Mean It, Crown Business, New York. (2001)
11. Kaplan, R., Norton, D.: *Strategy-Focused Organization How Balanced Scorecard Companies Thrive in the New Business Environment*, Boston, MA : Harvard Business School Press. (2006)
12. Leahy, T.: The Balanced Scorecard Meets BPM. *Business Finance*, Vol. 33. (June 2003)
13. Stefanovic, D., Majstorovic, V., Stefanovic, N.: Methodology for Process Integration in Supply Networks, 38th CIRP Manufacturing Systems Seminar, Brazil, May 16-18. (2005)
14. Stefanovic, D. and Stefanovic, N.: Methodology for Modeling and Analysis of Supply Networks, *Journal of Intelligent Manufacturing*, Springer, Vol. 19, No. 4, 485-503. (2008)
15. Stefanovic, N., Radenkovic, B., Stefanovic, D.: Designing OLAP Multidimensional Systems For Supply Chain Management, *International Journal of Pure and Applied Mathematics*, IJPAM, ISSN 1311-8080. (2007).
16. Stefanovic, N., Radenkovic, B., Stefanovic, D.: Supply chain intelligence, In: D.T. Pham, D.T., Eldukhri, E.E., Soroka, A.J. (eds.): *Intelligent Production Machines and Systems*, Whittles Publishing, Whittles Publishing, 292-297. (2008)
17. Stefanovic, N., Stefanovic, D., Mistic, M.: Application of Business Intelligence for Business Process Management. In: Bramer, M. (ed.): *Artificial Intelligence in*

Nenad Stefanović and Dušan Stefanović

Theory and Practice II, IFIP State of the Art Series, Vol. 276. Springer Boston, 445–449. (2008)

18. Stefanovic, N. and Stefanovic, D.: Methodology for BPM in Supply Networks. 5th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering, Ischia, Italy, July 25-28. (2006)

Nenad Stefanović completed his BSc in Mechanical Engineering with the University of Kragujevac, where he also got his MSc in Information Systems. He received his PhD in Electronic Business from the Faculty of science, University of Belgrade. Nenad works as an ICT Operation & Architectures Manager at the Fiat Automobiles Serbia. He is also an assistant professor at the Faculty of Science, Department of Mathematics and Informatics, University of Kragujevac. His research interests include business intelligence, supply chain management, management information systems, and web portals. Nenad is a member of several national and international ICT and business related societies and research projects. He has published more than 60 research papers, book chapters and invited lectures, and published in national and international journals, conference proceedings and books.

Dušan Stefanović is a professor of computer science at the Faculty of Science, University of Kragujevac, Serbia. His main research interests include supply chain management, business process simulation, databases and information systems development. Dusan has published more than 80 research papers and authored/coauthored several books. He designed and developed a number of enterprise information systems and participated in many national and international research projects.

Received: June 08, 2009; Accepted: February 19, 2010.

Usage of Agents in Document Management

Dragoslav Pešović¹, Milan Vidaković², Mirjana Ivanović¹, Zoran Budimac¹,
and Jovana Vidaković¹

¹Department of Mathematics and Informatics, Faculty of Science,
University of Novi Sad, Trg D. Obradovića 4,
21000 Novi Sad, Serbia
{dragoslav,mira,zjb,jovana}@dmi.uns.ac.rs

²Computing and Control Department, Faculty of Technical Sciences,
University of Novi Sad, Trg D. Obradovića 6,
21000 Novi Sad, Serbia
minja@uns.ac.rs

Abstract. EXtensible Java-based Agent Framework (XJAF) is a pluggable architecture of the hierarchical intelligent agents system with communication based on KQML. Workers, Inc. is a workflow management system implemented using mobile agents. It is especially suited for highly distributed and heterogeneous environments. The application of the above-mentioned systems will be considered in the area of Document Management Systems.

Keywords: Mobile Agents, Workflow Management Systems, Document Management.

1. Introduction

According to the most general definition, a mobile agent is a program that is able to stop its execution at one node in a computer network, and to transfer itself to another node where its execution continues. An important feature of mobile agents is their autonomous behavior: a mobile agent autonomously decides when and where it will be transferred.

EXtensible Java-based Agent Framework (XJAF) [32] is a pluggable architecture of the hierarchical intelligent agents system with communication based on KQML. This framework supports pluggable software managers that are dealing with a particular job. The system is designed so that it is possible to choose an arbitrary manager when configuring provided that it implements the given interface. This enables the use of arbitrary managers whose existence is not necessary at compile-time. The system is compliant to the FIPA specification and has been implemented using Java Enterprise Edition (JEE) technology.

Workflow [13, 37, 38] can be defined as the automated part of a business process, organized as a collection of activities, where documents, information or tasks are passed between participants according to a set of procedural

Dragoslav Pešović at all.

rules. A workflow management system (WFMS) provides for defining, creating, and managing of workflow instances.

The usage of mobile agents [14] in modeling and implementation of a workflow [37] simplifies the workflow management. Workers, Inc. [25] consists of individual agents with autonomous behavior. Mobile agents carrying out workflow instances (the so-called workers) have the ability to move to different users, where they can interact with them locally, autonomously taking care of their current position, state, and further itinerary. In order to achieve the flow of work, workers split the work in logical parts, cooperate together, and synchronize themselves.

To allow the exchange of process definitions with various other workflow products (ranging from other workflow management systems to modeling and simulation tools), the system had to be made compliant with XML Process Definition Language (XPDL) [39], the proposed standard in the area of workflow definition languages. In order to comply with XPDL, the system first had to be modified to conform to the basic constructs of XPDL and the underlying meta-model. Moreover, a system-specific import layer had to be provided to allow the translation of XPDL process definitions, generated using a visual modeling tool, into worker execution contexts, their internal system representations.

The application of the above-mentioned systems will be considered in the area of Document Management Systems.

The rest of the paper is organized as follows. In the next section, the related work is presented. Main concepts of agent frameworks are described in the third section, introducing basic technologies that can be used for their implementation. The section 4 presents the architecture of XJAF, while sections 5 and 6 describe the architecture of Workers, Inc. The seventh section provides a brief introduction into the area of document management systems, discussing possibilities of an agent-oriented approach to the design and implementation of such systems. Finally, the eighth section concludes the paper.

2. Related Work

Agent frameworks can be analyzed from several points of view. From the problem domain point of view, frameworks can be general-purpose [1, 2, 8, 9], or specialized ones, which solve particular problems [18, 36]. Also, from the technology point of view, agent frameworks are based on either proprietary solutions or on solutions based on the distributed components technology. Agent frameworks like JAF (Java Agent Framework) [8] and JAT (Java Agent Template) [9] are based on proprietary solutions, while Aglets [1] and JADE (Java Agent DEvelopment framework) [2] are based on the RMI, CORBA and Java EE technology.

The large number of papers is related to the security issues in agent frameworks [3, 12, 30, 35]. Security issues regarding agent frameworks

include: providing message integrity, code protection during agent migration and protecting agent frameworks from malicious agents.

This paper presents an implementation of an agent framework which is based on the Java EE technology. All important elements of this framework are implemented as plug-ins, which provides for flexibility in both design and implementation.

Several authors have recently suggested a usage of agents in workflow and document management.

Rather than going top-down in describing possible use of mobile agents in workflow management, we take bottom-up approach. Workers, Inc. is highly decentralized and consists solely of individual agents with autonomous behavior, which differentiates it from approaches in [5, 6, 15, 20]. The only centralized control in our system is the control of user rights to create, access, and change agents and templates, while all mentioned papers describe systems that had some forms of centralized control or services.

With respect to decentralization, our system resembles [28] that is based on static CORBA objects. While decentralization in [28] was one of explicit design goals and had to be explicitly implemented, decentralization in our system comes for free as a natural consequence of agent mobility and autonomous behavior. Moreover, our system is uniform – it is designed to use only one mechanism (mobile agents), without the need for additional mechanisms (transactions, HTTP protocol, HTML documents, Web browsers, CORBA, etc.)

While in [16] full decentralization of WfMS using mobile agents is shortly mentioned, the paper in fact describes the usage of mobile agents in centralized WfMS and only for external parties. Moreover, agents in [16] are specialized and created to one task only. Itinerary is saved on hosts instead in agents themselves which reduces the agent autonomy. Our system completely relies on agents and is fully distributed with autonomous agents.

Stromer in [29] describes similar goals and advantages of using mobile agents in WfMS as we are, but his implementation is different.

Among document management systems proposed over the years, there are some that are agent-based [7, 27] or only agent-enhanced (where an agent layer is added on top of existing infrastructure) [21]. However, none of the mentioned systems emphasizes the benefits of using agent mobility.

Proposed workflow and document management systems bring some fresh views not only in particular fields of workflow and document management, but in mobile computing as well. In both fields, the advantages of highly decentralized and distributed approach in designing a system have not been often recognized. The most cited advantages of mobile agents with respect to classical techniques of distributed programming are:

1. Potentially better efficiency of the whole system. A client program migrates to a server node, locally communicates with a server program, and returns to the original node with a result. In that way, the overall network traffic, including the number of remote interactions and the amount of data communicated over the network, is therefore potentially reduced.

Dragoslav Pešović at all.

2. Greater reliability, because the connection between nodes must not be established all the time.

Our workflow and document management systems emphasize the fact that mobile agent has organizational advantages as well. Mobile agent systems can be regarded as a separate programming paradigm and not only as an improvement of distributed programming style. Solutions to some problems are easier to program, understand, and maintain, if implemented using mobile agents.

3. Agent Frameworks

Agent technology [14] represents one of the most consistent approaches in distributed systems implementation. Software agents realize distributed component concept entirely. This means that besides solving the problem, agents utilize a certain degree of intelligence and autonomy that are needed to solve the problem. Therefore, agents represent software entities capable of searching and processing the large quantity of information, utilizing a certain degree of intelligence, autonomy and communication.

Agents need a programming environment which will create and enable agents to execute tasks. Agent framework [14] represents programming environment that controls agent life cycle and provides all necessary mechanisms for task execution (communication, agent mobility, services and security). Besides controlling life cycle of an agent, an agent framework also provides messaging and service subsystems to effectively support agents. Messaging allows agents to communicate to each other, and service subsystem gives them the possibility of accessing various resources or executing complex algorithms that are not needed to be implemented in the agent itself. An agent framework also provides agent mobility and security. Agent mobility allows agents to migrate from one agent framework to another. The security subsystem provides security mechanisms which protect both agents and frameworks. It is also necessary to provide mechanism of searching agents and services present in agent framework. This mechanism is called a directory and it represents searchable repositories of agents and services which can be used by both agents and their clients.

Agent framework implementations rely on Object Oriented (OO) techniques. It is possible to implement agents easily and effectively taking advantage of its features (encapsulation, inheritance, polymorphism, dynamic binding and persistence). Most of the existing agent frameworks are implemented using Java programming language [11]. Java-based agent frameworks usually use RMI (Remote Method Invocation) [11], CORBA (Common Object Request Broker Architecture) [19] and Java EE (Java Enterprise Edition) [10] technologies for distribute code execution.

CORBA is a standard created by the Object Management Group (OMG) consortium. This standard defines the framework for creating objects being executed on the server side, and it also defines the servers themselves. It is

based on the Internet Inter-ORB Protocol (IIOP). This standard anticipates the execution of components written in all supporting programming languages. Besides above mentioned concepts, CORBA supports transactions and its own components naming and search system (Object Naming Service - COS Naming).

The Java EE technology is particularly useful for agent framework implementation because it comprises a large set of technologies and provides for scalability, reliability and has a large number of implementations. One element of the Java EE technology is particularly useful – the EJB (Enterprise JavaBeans) technology. This is a technology of distributed software components which are created, executed and destroyed in the application servers. All performance-related issues like load-balancing, distribution-per-server, etc. are left to the implementation of the application server. Besides supporting distributed components, Java EE also has all other technologies for the agent framework implementation: JMS (Java Message Service) for message exchange, JNDI (Java Naming and Directory Interface) for directory implementation, Java Security, etc.

4. XJAF

The EXtensible Java-based Agent Framework (XJAF) [32] is based on the J2EE technology. The system consists of clients and facilitators. The clients refer to the facilitators for task execution. The task is being executed by the agents engaged by the facilitator. The Figure 1 shows the link between a client and an agent framework.



Figure 1. Client and Agent framework link

The client assigns the task to the facilitator; the facilitator engages an agent to execute the task and returns the result to the client. The FacilitatorProxy class ensures that the client application can access the facilitator. It also hides all techniques necessary for work with agents from the client. The client only needs to create an object of the FacilitatorProxy class and to pass it the class representing the task or the KQML message, as well as the corresponding listener, which would notify it of the result. All other details are managed by the FacilitatorProxy class.

The client is any Java application. The facilitator is an instance of the Facilitator class. The Facilitator class realizes the facilitator functionality. The agent is an instance of a class implementing the Agent interface and realizing the functionality of an individual agent.

Dragoslav Pešović at all.

Extensibility of this framework is based on the plug-in concept. Plug-ins are realized as pluggable managers. The facilitator forwards the parts of its job to the corresponding pluggable managers. The managers are instances of classes implementing the corresponding managerial interfaces. The AgentManager interface is responsible for allocating and releasing agents. The TaskManager interface manages the tasks. The MessageManager interface is responsible for interagent communication. The ConnectionManager interface manages facilitator connection and relations. The SecurityManager handles security of inter-agent communication.

The classes that implement the mentioned interfaces implement the corresponding algorithms for individual functions. The system is designed so that it is possible to choose an arbitrary manager when configuring provided that it implements the given interface. This enables the use of arbitrary managers whose existence is not necessary at compile-time, but not until initialization (plug-in concept). The Figure 2 lists all the managers in the framework.

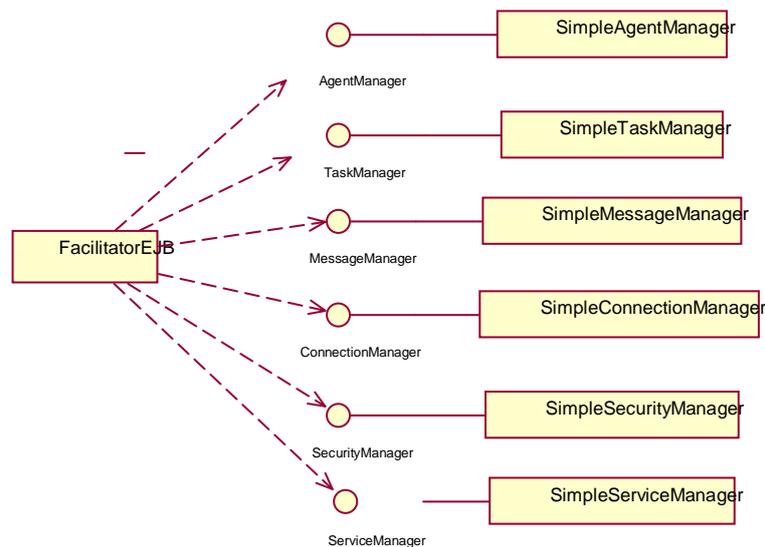


Figure 2. Functionality of individual parts is assigned to managers

4.1. Agent Manager

Agent management is done using the AgentManager component. Controlling the agent life cycle means creating and destroying an agent.

This component is also used as an agent directory. All relevant data is kept in a repository. The repository can be a database, or an LDAP server, or any other data storage. This manager also keeps track of all local agents required by external facilitators, and of all agents that have been moved to another facilitator.

4.2. Task Manager

The TaskManager component manages tasks to be performed by the agent framework. It is realized through the class which implements the TaskManager interface. It also provides a way of notifying the client about the task execution progress.

Each task is stored in this component. When completed, it is removed from it. Tasks are instances of classes which implement the AgentTask interface. There are two types of task execution: programmatically or by sending a KQML message to the agent.

The method execute() is run asynchronously, i.e. it does not block the execution of the client's code for the time of task execution, but it calls the instance of the class inheriting the interface AgentListener created by the client application upon and during the task execution. In this way the client application can proceed with the code execution and it will be notified of the agent's results by the listener reference.

When executing a task by sending a KQML message to the agent, the client application sends the KQML message to the Facilitator component. This component looks for the appropriate agent and sends the message to it. When the task is completed, the agent replies to the original message and the message is forwarded to the client using the FacilitatorProxy component.

4.3. Message Manager

The exchange of messages between the agents is actually KQML messages exchange. These messages are encapsulated in the base class - the class KQMLMessage. The messages are exchanged by passing the KQMLMessage class objects between the dialog participants. All the communication is done by the MessageManager component.

When an agent sends a KQML message to another agent, it is embedded into a JMS message. The JMS message is sent to all agent frameworks subscribed to this service, but only the agent framework having the destination agent will receive the message and extract the KQML message from it. This KQML message is then sent to the agent.

4.4. Connection Manager

The ConnectionManager component defines an inter-facilitator connectivity mechanism. This mechanism defines how separate facilitators form a network. Each facilitator is a node in this network and is automatically registered on the network at the initialization time. This means that the programmer does not have to know the exact address of an arbitrary facilitator and does not have to maintain the list of all available facilitators.

Dragoslav Pešović at all.

Instead, the nodes are registered automatically and the list of all available facilitators is maintained automatically.

The facilitators form a certain hierarchy structure. One approach is to form a tree structure with the primary facilitator in the root. The Figure 3 shows this organization.

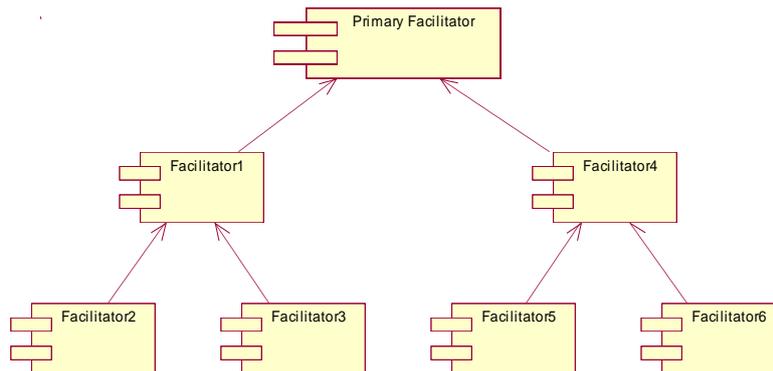


Figure 3. Component diagram of facilitator hierarchy

4.5. Security Manager

SecurityManager component [33] handles security issues. It provides encryption, decryption, signature generation and verification for all messages passing through the framework. Also, this manager handles access to local resources. Access control segment of security subsystem insures integrity of data and code. It provides for integrity of data exchanged between agents and also protects agent framework from malicious agents. Any cryptographic system can be used since this manager is proposed by the SecurityManager interface, and implementation is left to the developer.

4.6. Service Manager

The ServiceManager component implements service directory subsystem. This component manages the set of services available to agents.

The ServiceManager component includes the service repository which holds all available services. Services can be added, removed, searched and used. When the service is not needed anymore, it must be returned to the repository. Services are implemented as Java classes which implement the Service interface.

5. Workflow Management System Using Mobile Agents

Workers, Inc. [4, 22, 23, 24, 25, 26] is a workflow management system under development at the University of Novi Sad. The system is implemented using the technology of mobile agents and is therefore especially suited for highly distributed and heterogeneous environments.

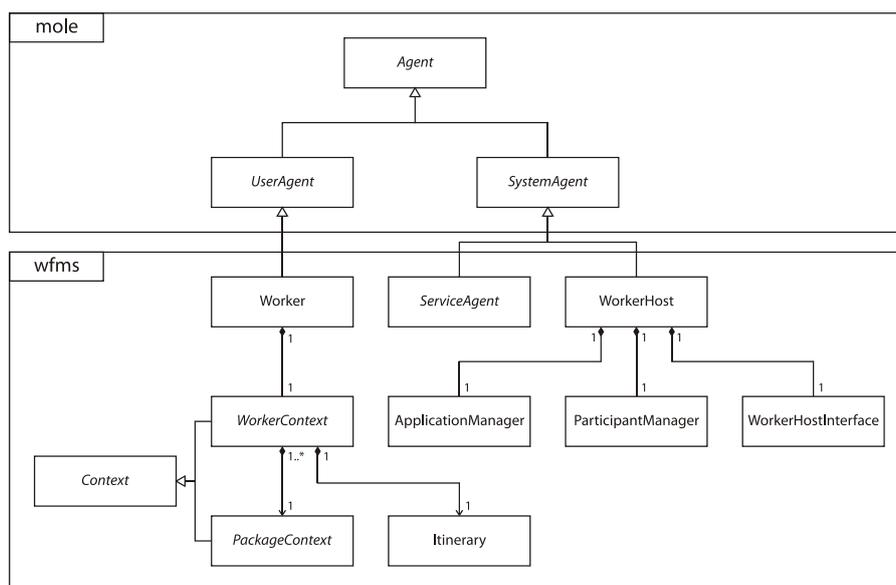


Figure 4. The architecture of Workers, Inc.

Workers, Inc. is envisioned as a community of cooperative agents, its main characteristics being full decentralization and distribution of workflow functions. The current architecture is essentially two-part, consisting of work-agents (workers) and host-agents (worker hosts). Workers, Inc. is built on top of a Java-based mobile agent system, and uses Java as the language of implementation as well as of the agent development. Agent migration and inter-agent communication benefit from Java RMI and class serialization, and Java sandbox security model is the basis for providing secure agent execution environment. Java API for XML Processing (JAXP) is used for XPDL document parsing.

Process definitions are being completely handled by workers, while the enactment is achieved through the cooperation of a worker carrying a process definition (or a set of workers when concurrency or subprocesses are involved) and worker hosts residing at every node of the network. Worker hosts represent central components of the system mediating between the underlying system, workers, and human users.

Dragoslav Pešović at all.

5.1. Workers

A worker is the key system component encapsulating both the process definition and the execution state of a workflow. While performing a workflow, a worker itinerates among distributed resources carrying process-specific information and autonomously taking care of its execution state. In that way, workers manage not only to perform workflow activities locally with respect to assigned resources, but to avoid the need to consult a central server or the originating machine at every step.

A worker's behavior is entirely defined by its execution context. A worker context is an executable process definition, a worker being just a medium through which its context is transmitted and accomplished. When a worker migrates, its entire execution context as an object net is being encompassed by object serialization, and then transported and reconstructed at the target location.

The most important part of a context is the worker itinerary, which represents a flow of a worker through a network. By representing itineraries with directed graphs we are able to represent complex flow patterns that could be needed by workflow applications.

To allow concurrent activity execution, agent social abilities are employed. When a single thread of control needs to split into two or more threads, which can be executed in parallel, the worker context is cloned and multiple worker instances are allowed to be executed simultaneously. On the other hand, when multiple parallel threads of execution need to converge into a single thread, agent coordination mechanisms and synchronization techniques are employed.

To strengthen security of the system, mobile agents and thus workers are forbidden to access any system resources directly. Critical resources can be accessed only by communicating with system agents, i.e. worker-hosts.

5.2. Worker Hosts

Every node in the network contains a worker host, which is implemented as a stationary system agent, having special privileges for the access to host system resources. A worker host is a passive entity, which spends most of its lifetime receiving requests from workers or users and coordinating their actions. There are three main subcomponents of a worker host: an application manager, a participant manager, and a user interface.

5.3. Other Specialized Agents

Although workers are almost fully autonomous, they may need additional services to finish their work. Those services cannot be embedded directly into the workers as this would prevent keeping workers as small as possible.

Services are therefore implemented separately, as specialized stationary agents.

Workers, Inc. is a fully distributed system, without central administration, control, and maintenance. All reports, control, and management can be achieved by creating and sending specialized agents that will communicate with other agents in the system and achieve the intended results. Those agents may be mobile or stationary, depending on the nature of the task they are intended to accomplish.

6. Worker Execution Contexts

The design of an execution context is done so as to comply with the workflow meta-model specification. From the control-flow perspective, the itinerary is the most important part of a context.

6.1. Itinerary

The itinerary has the structure of an arbitrary complex directed graph, where vertices of the graph represent process activities, and edges of the graph correspond to process transitions.

6.2. Activities

An activity is the smallest, atomic unit of work in a business process. The three main properties of an activity specification, which can be seen as answers to the accompanied questions, are:

- Performer assignment (Where?) – It specifies the performer of the activity. In the process of workflow participant resolution, the actual location of a participant is determined. By evaluating a performer expression, a worker knows where its activity needs to be carried out, and will transfer itself over the network accordingly.
- Implementation specification (What?) – It specifies what the concrete realization of the activity is. It can be a call to a declared application, another workflow process, or an embedded activity set. Also, the activity may have no implementation at all, in which case it supports complex flow transitions or manually performed activities.
- Automation modes (How?) – Information on whether the activity is to be started / finished manually by the user or automatically by the worker itself.

published, and consumed, and how they are ultimately disposed of or retained.

A well-designed document management system promotes finding and sharing information easily. It organizes content in a logical way, and makes it easy to standardize content creation and presentation across an enterprise. It promotes knowledge management and information mining. It provides features at each stage of a document's life cycle, from template creation to document authoring, reviewing, publishing, auditing, and ultimately destroying or archiving.

7.1. Main Features of Document Management Systems

There are several common issues that are involved in the document management. Document management systems commonly address the following issues:

- Location. Where will documents be stored? Where will people need to go to access documents? How content moves between locations? It may be necessary to move or copy a document from one site or library to another at different stages of its life cycle. For example, the publishing process may include moving a document from a staging site to a public Internet site. If content needs to be converted from one format to another as it moves from site to site, content conversions must be specified.
- Filing. How will documents be filed? What methods will be used to organize or index the documents to assist in later retrieval? Documents can be organized in free-form document libraries for ad-hoc document creation and collaboration, or specialized sites such as team sites and portal sites. Databases can be used to store filing information.
- Retrieval. How will documents be found? Typically, retrieval encompasses both browsing through documents and searching for specific information.
- Security. How will documents be kept secure? How will unauthorized personnel be prevented from reading, modifying or destroying documents?
- Retention period. How long should documents be kept, i.e. retained?
- Archiving. How can documents be preserved for future readability?
- Distribution. How can documents be available to the people that need them?
- Workflow. By planning workflows, one can control and track how documents move from one team member to another as each participant collaborates in a document's life cycle. A system may include workflows for common team tasks such as reviewing and approving documents. It may also support creating and installing custom workflows.
- Creation. How are documents created? This question becomes important when multiple people need to collaborate, and the logistics of version control and authoring arise.
- Authentication. Is there a way to vouch for the authenticity of a document?

Dragoslav Pešović at all.

- Content types. Content types can be used to organize information about types of documents, such as metadata, document templates, policies, and workflow processes.

7.2. Agent-Oriented Approach

The usage of software agents in modeling and implementation of a document management system simplifies the document management. The organization and implementation of the system are easy to understand and follow, because most of its parts are uniformly implemented as (mobile) agents:

- User agents to assist individual users (with incorporated access rights). Every user of the document management system would have a devoted user agent to assist him/her in the authoring and access processes. Those user agents would communicate with other agents in the system directly, or create and send specialized mobile agents in order to achieve the intended results. Specific user's rights will be incorporated in his/her agent, controlling access to other agents or registered system services. Every user agent will need a user interface for communication with the user.
- Specialized agents for document retrieval, indexing, archiving, etc. Those agents may be mobile or stationary, depending on the nature of the task they are intended to accomplish. For example, a retrieval agent would be mobile, searching for documents on every site or library in the system, and gathering a report. Once back, it will present the report of the status and location of all found documents.
- Workflow agents to support all kinds of workflows within the system, including collaboration, versioning, and publishing. For example, publishing a document may involve the procedures of proofreading, peer or public reviewing, authorizing, printing and approving etc. Collaboration procedures, on the other hand, define how a group of users can work on the same document(s). Workgroups can benefit from agents to coordinate their access efforts. The Joint Paper Worker, presented in [24, 25], is an example of a collaboration workflow. A number of common workflows can be provided in advance, while keeping the possibility to create custom workflows at any time.

Since the system consists of many autonomous agents, the system is easily changed, extended, and improved. It is often needed just to introduce new agents, without the need to change and even to understand the rest of the system.

8. Concluding Remarks

The idea of implementing a document management system involved two modern, attractive and promising fields in computer science:

1. Software agents with the attributes they possess: autonomy, social ability, responsiveness, proactiveness. Also, two distinguishing characteristics are very important and make agents more promising for application in different areas:
 - high-level tasks can be delegated to agents who will autonomously carry them out,
 - agents are situated in an environment which can dynamically affect their problem solving behavior and strategy.
2. Workflow is concerned with automation of procedures where documents, information or tasks are passed between participants according to a defined set of rules to achieve or contribute to an overall business goal.

The approach to an agent framework implementation using the Java EE technology provides for scalability and reliability. This approach offers agent and service directory services, security, message exchange and agent mobility. The EJB technology, as a part of the J2EE concept, offers the simple use of all technologies necessary to implement this approach. The future work will include defining and implementing the specialized agent language, based on KQML. Also, all concepts handled by managers will be developed further. The research in the field of interoperability among different platforms (including web services) and improvement of security will be taken into consideration for further real implementation.

The main characteristics of a workflow system suggested in this paper are almost full decentralization and distribution of workflow functions. The proposed organization mimics usual user activities in a real flow of work. Moreover, it relieves them (or any centralized control) from the need to know what to do next with the work-agent. Every user takes care only of work-agents that are currently on its node. Where they came from, why they are here, and where they will go later, is not concern of the user.

Finally, the paper presented a possibility of using the two very attractive fields (agent technology and workflow) for the document management implementation, emphasizing the basic features of the proposed system. The work in these directions has already been started and more significant results are expected in the following period.

References

1. Aglets Home Page, <http://www.trl.ibm.com/aglets/>
2. Bellifemine, F., Poggi, A., Rimassa, G.: "JADE – A FIPA-compliant agent framework", Proceedings of Practical Applications of Intelligent Agents (PAAM'99), London, April 1999, pp. 97-108.
3. Binder, W., Roth, V.: "Secure mobile agent systems using Java: where are we heading?", Proceedings of the 2002 ACM symposium on Applied computing, 2002, Madrid, Spain, ISBN:1-58113-445-2, pp. 115-119.
4. Budimac, Z., Ivanović, M., Popović, A.: "Workflow Management System Using Mobile Agents", Proc. of ADBIS '99, Lecture Notes in Computer Science 1691, Springer Verlag, Berlin, (Maribor, Slovenia), pp. 169 - 178, 1999.

Dragoslav Pešović at all.

5. Cai, T., Gloor, P.A., Nog S.: "Dartflow: a workflow management system on the web using transportable agents", Technical report PCS-TR96-186, 1996.
6. Chang, W., Scott, C.: "Agent-based workflow: TRP support environment", Computer networks and ISDN systems vol. 28, issues 7-11, 1997.
7. Ginsburg, M.: "An Agent Framework for Intranet Document Management", Journal of Autonomous Agents and Multi-Agent Systems, volume 2, issue 3, pp. 271-286, 1999.
8. Java Agent Framework Home Page, <http://mas.cs.umass.edu>
9. Java Agent Template Home Page, <http://java.stanford.edu>
10. Java Enterprise Edition Homepage, <http://java.sun.com/javae>
11. Java homepage, <http://java.sun.com>
12. Kim Tan, H., Moreau, L.: "Certificates for mobile code security", Proceedings of the 17th symposium on Proceedings of the 2002 ACM symposium on applied computing, 2002, Madrid, Spain, ISBN:1-58113-445-2, pp. 76-81.
13. Leymann, F., Roller, D.: "Production Workflow – Concepts and Techniques", Prentice Hall PTR, New Jersey, 2000.
14. M. Knapik, J. Johnson. Developing Intelligent Agents for Distributed Systems. McGraw-Hill, 1998, pp. 3, 37-39.
15. Meng, J., Helal, S., Su, S.: "An ad-hoc workflow system architecture based on mobile agents and rule-based reasoning", Proc. of Int. Conf. on parallel, and distributed computing techniques and applications, Las Vegas, 2000.
16. Merz, M., Liberman, B., Lamersdorf, W.: "Using mobile agents to support inter-organizational workflow management", Int. J. on applied artificial intelligence 11(6), pp. 551-572, 1997.
17. Microsoft TechNet Library, <http://technet.microsoft.com/en-us/library/cc261933.aspx>[5]
18. Nardi, B., Miller, J., Wright, D.: "Collaborative, programmable intelligent agents", Communications of the ACM, Volume 41 , Issue 3 (March 1998), pp. 96-104.
19. OMG. Common Object Request Broker: Architecture and Specification. OMG Specification Revision 2.0, July 1995.
20. Padalkra, A., Nabar, P., Arora, S., Naik, P.: "SWIFT: Scalable workflow management system using mobile Agents", <http://www.iitb.ac.in/~pranav/php/paper.pdf>, 2000.
21. Papaspyrou, N., Sgouropoulou, C., Skordalakis, E.: "A Model of Collaborating Agents for Content-Based Electronic Document Filtering", Journal of Intelligent and Robotic Systems, vol. 26, no. 2, pp. 199–213, October 1999.
22. Pešović, D., Budimac, Z., Ivanović, M.: "Distributed Mobile Agent Workflow - Activity Coordination Constructs in Workflow Process Graphs", Proc. of 11th International Multiconference Information Society 2008, Ljubljana, Slovenia, Oct 13-17, 2008.
23. Pešović, D., Budimac, Z., Ivanović, M.: "Towards a Visual Definition of a Process in a Distributed Environment", 2nd International Symposium on Intelligent Distributed Computing - IDC'2008, Catania, Italy, 2008.
24. Pešović, D., Budimac, Z.: "Advanced Joint-Paper Worker", CCS Journal, 4th issue, pp. 44-46, March 2003.
25. Pešović, D.: "A High-Level Language for Defining Business Processes", PhD Thesis, University of Novi Sad, Novi Sad, 2007.
26. Pešović, D.: "The Implementation of a Workflow Management System Using Mobile Agents", Master Thesis, University of Novi Sad, Novi Sad, 2002.
27. Roberto, V., Della Mea, V., Di Gaspero, L., Conti, A.: "MANTHA: Agent-based Management of Hypermedia Documents", Proceedings of 6th IEEE Int. Conf. on

- Multimedia Computing and Systems (IEEE ICMCS '99), Firenze, IEEE Computer Society, vol II, pp. 814-818, 1999.
28. Sheth, A., Kochut, K., Miller, J., Worah, D., Das, S., Lin, C., Palaniswami, D., Lynch, J., Shevchenko, I.: "Supporting State-wide Immunization Tracking using Multi-Paradigm Workflow Technology", Proc. of 22nd VLDB Conference (Bombay, India), 1996.
 29. Stormer, H.: "A flexible agent-based workflow system", Workshop on Agent-based approaches to B2B, 2001.
 30. Varadharajan, V.: "Security enhanced mobile agents", Proceedings of the 7th ACM conference on Computer and communications security, Greece, Athens, November 2000, pp. 200-209.
 31. Vidaković M., Konjović Z.: "One Implementation of Virtual Library Based on Agent Technology", Proceedings of the 4th Conference on Informatics and Information Technology, Bitola, 11-14 of December, 2003
 32. Vidaković, M., Konjović, Z., "EJB Based Intelligent Agents Framework", Proceedings of the 6th IASTED International Conference on Software Engineering and Applications (SEA 2002), Cambridge, USA, November 4-6, 2002, pp. 343-348
 33. Vidaković, M., Sladić, G., Konjović, Z., "Security Management In J2EE Based Intelligent Agent Framework", Proceedings of the 7th IASTED International Conference on Software Engineering and Applications (SEA 2003), Marina Del Rey, USA, November 3-5, 2003, pp. 128-133
 34. Vidaković, M., Sladić, G., Zarić, M.: "Metadata Harvesting Using Agent Technology", Proceedings of the 8th IASTED International Conference on Software Engineering and Applications (SEA 2004), Cambridge, USA, November 9-11, 2004., pp. 489-493
 35. Vuong, S., Fu, P.: "A security architecture and design for mobile intelligent agent systems", ACM SIGAPP Applied Computing Review archive, Volume 9 , Issue 3 (Fall 2001), pp. 21-30.
 36. Wilson, L., Burroughs, D., Sucharitaves, J., Kumar, A.: "An agent-based framework for linking distributed simulations", Proceedings of the 32nd conference on Winter simulation, 2000 , Orlando, Florida, ISBN:1-23456-789-0, pp. 1713 – 1721.
 37. Workflow Management Coalition, Hollingsworth, D.: "The Workflow Reference Model", Homepage of Workflow Management Coalition, 1995.
 38. Workflow Management Coalition: "Terminology and Glossary", Homepage of Workflow Management Coalition, 1999.
 39. Workflow Management Coalition: "Workflow Process Definition Interface – XML Process Definition Language, Version 1.0", Homepage of Workflow Management Coalition, 2002.

Dragoslav Pešović is an assistant professor at Faculty of Science, Department of Mathematics and Informatics, University of Novi Sad. His research interests include: mobile agents and distributed systems. He has published 10 scientific papers in proceedings of international conferences, has written 2 university textbooks.

Dragoslav Pešović at all.

Milan Vidaković is holding the associate professor position at the Faculty of Technical Sciences, Novi Sad, Serbia. He received his PhD degree (2003) in Computer Science from the University of Novi Sad, Faculty of Technical Sciences. Since 1998 he has been with the Faculty of Technical Sciences in Novi Sad. Mr. Vidaković participated in several science projects. He published more than 60 scientific and professional papers. His main research interests include web and internet programming, distributed computing, software agents, and language internationalization and localization.

Zoran Budimac is a professor at Faculty of Science, Department of Mathematics and Informatics, University of Novi Sad. He graduated in 1983 (informatics), received master's degree (computer science) in 1991 and doctor's degree (computer science) in 1994. His research interests include: mobile agents, e-learning, software engineering, case-based reasoning, implementation of programming languages. He has been project leader for several international and several national projects. He has published over 180 scientific papers in proceedings of international conferences and journals, has written more than 12 university textbooks in different fields of informatics.

Mirjana Ivanović is a professor at Faculty of Sciences, Department of Mathematics and Informatics, University of Novi Sad. She graduated in 1983 (informatics), received master's degree (discrete mathematics and programming) in 1988 and doctor's degree (computer science) in 1992. Her research interests include: multi-agent systems, e-learning and web-based learning, data mining, case-based reasoning, programming languages and tools. She actively participates in more than 10 international and several national projects. She has published over 190 scientific papers in proceedings of international conferences and journals, has written more than 10 university textbooks in the field of informatics and ICT. She is the Head of Computer Science Chair.

Jovana Vidaković is a teaching assistant at Faculty of Sciences, Department of Mathematics and Informatics, University of Novi Sad. She graduated in 1999 (informatics), received master's degree in 2003. Her main research interests include databases, XML technologies, web and internet programming. She has published 7 scientific papers and has been coauthor of 1 university textbook.

Received: June 08, 2010; Accepted: February 04, 2010.

Integration of Recommendations and Adaptive Hypermedia into Java Tutoring System

Aleksandra Klašnja-Milićević¹, Boban Vesin¹, Mirjana Ivanović²,
and Zoran Budimac²

¹Higher School of Professional Business Studies
Vladimira Perića Valtera 4, 21000 Novi Sad, Serbia
{aklasnja, bobanvesin}@yahoo.com

²Faculty of Science, Department of Mathematics and Informatics
Trg Dositeja Obradovica 3, 21000 Novi Sad, Serbia
{mira, zjb}@dmi.uns.ac.rs

Abstract. A way to improve the effectiveness in e-learning is to offer the personalized approach to the learner. Adaptive e-learning system needs to use different strategies and technologies to predict and recommend the most likely preferred options for further learning material. This can be achieved by recommending and adapting the appearance of hyperlinks or simply by recommending actions and resources. This paper presents an idea for integration of such recommender system into existing web-based Java tutoring system in order to provide various adaptive programming courses.

Key words: e-learning, recommender systems, adaptive hypermedia, Java tutoring system

1. Introduction

Contemporary e-Learning systems provide useful tools for computer-supported learning, such as forums, chat rooms, discussion groups and e-mail. However, most of them display content and educational material in the same way to all learners, allowing them to choose their own learning pathway through the course, which is not necessarily the most effective one in terms of their previous knowledge or needs. One possible solution to this problem is to use adaptive and intelligent web-based educational systems. These systems can use different recommendation techniques in order to suggest online learning activities or optimal browsing pathways to learners, based on their preferences, knowledge and the browsing history of other learners with similar characteristics. Their main objective is to adapt and personalize learning to the needs of each learner [29].

The task of delivering personalized content is often framed in terms of a recommendation task in which the system recommends items to an active user [15]. Recommender systems help users find and evaluate items of

interest. Such systems have become powerful tools in many domains from electronic commerce to digital libraries and knowledge management [18]. Some recommender systems have also been applied to e-Learning systems for recommending lessons (learning objects or concepts) that learners should study next [11] or for providing course recommendation about courses offered that contribute to the learner's progress towards particular goals [7].

Recommender systems can use data mining techniques for making recommendations using knowledge learnt from the action and attributes of users [18]. The objective of data mining is to discover new, interesting and useful knowledge using a variety of techniques such as prediction, classification, clustering, association rule mining and sequential pattern discovery. Currently, there is an increasing interest in data mining and educational systems, making educational data mining a new and growing research community [17]. The data mining approach to personalization uses all the available information about learners on the web site (in the web course) in order to create user/learner models and to use these models for adaptation of content.

Another approach for course personalization is the use of *adaptive hypermedia* methods and techniques that are used in *Adaptive educational hypermedia systems* [5]. Adaptive educational hypermedia systems can adaptively sort, annotate, or partly hide the links on web pages to make it easier to choose or to recommend to the learners where they should go from a certain point based on his/her goals, preferences and level of knowledge. Learners can be informed about importance and relevance of certain links.

We propose an architecture framework for new web-based learning system that will provide adaptive courses using hypermedia and recommender systems, based on existing web-based Java tutoring system called Mag [26]. To personalize the learning process for each individual learner, our system needs to use strategies of both recommender systems and adaptive hypermedia.

The previous version of Mag [26] has been used by our students during one semester as a learning tool. Although they used system in informal manner, we gained positive response and comments. It gave us opportunity to evaluate the main features of the system, the results obtained by students and the educational objectives in order to improve its functionalities and characteristics. The new architecture and supplements to Mag system will be proposed based on the opinion of students in order to increase efficiency of this kind of adaptive Java programming course and benefits that it could offer to a learner.

The rest of this paper is organized as follows. Section 2 presents related work from the area of research. Section 3 explains the general proposed architecture and system design for combining recommender systems and adaptive hypermedia. Finally, in section 4, conclusions are drawn and future work is considered.

2. Related Work

Two important ways of increasing the quality of service in e-Learning systems are to make them intelligent and adaptive. These goals are likely to be achieved by different approaches. In some systems, developers apply different forms of learner models to recommend the content and the links of hypermedia course pages to the learner. These systems are named *Adaptive educational hypermedia systems* [5]. Different approach uses aggregation of the recommended materials from other e-Learning web sites and prediction of more suitable material to learners.

Most of the tutoring systems for learning programming languages found on the Web are more or less only well formatted versions of lecture notes or textbooks. Consequently, these systems do not have implemented interactivity and adaptivity. The functions that such systems can perform vary. Some of them are used for learner assessment like JavaBugs [22] and JITS [24], [23], or basic tutoring like Jeliot 3 and Logic-ITA [1], while some of them are adaptive web-based tutorials [16], [21]. One step further in implementation of adaptation was made by systems like JOSH-online [2], iWeaver [28] and CIMEL ITS [27], [8]. Systems are presented in order from those that do not implement any adaptation, towards systems with considerable amount of intelligence implemented.

JavaBugs examines a complete Java program and identifies the most similar correct program to the learner's solution among a collection of correct solutions. After that, it builds trees of misconceptions using similarity measures and background knowledge [22]. They focused on the construction of a bug library for novice Java programmer errors, which is a collection of commonly occurring errors and misconceptions.

Java Intelligent Tutoring System - JITS is a tutoring system designed for learning Java programming [24]. JITS implements JECA (*Java Error Correction Algorithm*), an algorithm for a compiler that enables error correction intelligently changing code, and identifies errors more clearly than other compilers. This practical compiler intelligently learns and corrects errors in learners' program.

Jeliot 3 is a program animation tool aimed to support novices in their learning to program [1]. It graphically displays the execution of object-oriented programs written in Java. The Logic-ITA is a web-based intelligent teaching assistant system that allows students to practice formal proofs in propositional logic whilst receiving feedback [13].

Romero et al. developed a personalized recommender system that uses web mining techniques for recommending a learner which (next) links to visit within an adaptable educational hypermedia system [16]. They presented a specific mining tool and a recommender engine that they have integrated in the AHA! system, in order to help the teacher to carry out the whole web mining process. They made several experiments with real data in order to show the suitability of using both clustering and sequential pattern mining algorithms together for discovering personalized recommendation links.

Another system implemented by Soonthornphisaj et al. [21] allows all learners to collaborate their expertise in order to predict the most suitable learning materials to each learner. This smart e-Learning system applies the collaborative filtering [15] approach that has an ability to predict the most suitable documents to the learner. All learners have the chance to introduce new material by uploading the documents to the server or pointing out the web link from the Internet and rate the currently available materials.

JOSH is an interpreter for the Java programming language [27] originally designed to make easier teaching Java to beginners. Recently the interpreter was restructured into a server based interpreter applet and integrated into an online tutorial on Java programming called *JOSH-online*.

iWeaver is an interactive web-based adaptive learning environment, developed as a multidisciplinary research project at RMIT University Melbourne, Australia [28]. *iWeaver* was designed to provide an environment for the learner by implementing adaptive hypermedia techniques to teach the Java programming language. It implements several established adaptation techniques, including link sorting, link hiding and conditional page content.

CIMEL ITS is an intelligent tutoring system that provides one-on-one tutoring to help beginners in learning object-oriented analysis and design. It uses elements of UML before implementing any code [27]. A three-layered *Learner Model* is included which supports adaptive tutoring by deducing the problem-specific knowledge state from learner solutions, the historical knowledge state of the learner and cognitive reasons about why the learner makes an error [8]. This *Learner Model* provides an accurate profile of a learner so that the ITS can support adaptive tutoring.

None of the above stated systems is complete web-based tutoring system with personalisation options. Some of them are not web-based and are only executed on stand alone machine (JavaBugs, JITS, CIMEL ITS, Jeliot 3) and some of them had just basic interactivity and adaptivity implemented (*JOSH-online*, JavaBugs, Logic-ITA). Systems like Logic-ITA [13] and Jeliot 3 [1] offered us good ideas and perspective which functionalities could be included in new web-based tutoring system.

This new version of Mag system will integrate content and link adaptation in order to accomplish completely functional web-based tutoring system with personalization possibilities.

3. Proposed Architecture and System Design

In this section, architecture for new web-based learning system that will provide adaptive courses using hypermedia and recommender systems will be presented. The architecture is an extension of existing web-based Java tutoring system called Mag [9] that is developed at Department of Mathematics and Informatics, Faculty of Science, Novi Sad.

3.1. Mag System

Mag is a tutoring system designed to help learners in learning programming languages in different courses [25]. It is an interactive system that allows learners to use teaching material prepared for programming languages and to test acquired knowledge within appropriate courses.

Mag is multifunctional educational system that fulfills three primary goals, identified by earlier exploration in this field [10]. The first goal is to provide intelligent tutoring system for learners in a platform independent manner. The second goal is to provide the teachers with useful reports identifying the strengths and weaknesses of learner's learning process. Finally, the third goal is to provide a rapid development tool for creating basic elements of tutoring system: new learning objects, units, tutorials and tests.

In spite of fact that this system is designed and implemented as a general tutoring system for different programming languages, the first completely proposed and tested version was used for introductory Java programming course.

Preliminary design of the *Mag* system was influenced by several basic system requirements that every on-line learning system for a programming language should have [8]:

- separated user interfaces for learners and their mentors
- easy-to-access tutorials for learners
- various examples for every particular lesson (learning module)
- different tests for every particular lesson that can be adjusted to particular learner
- online programming, compiling and running of programs
- summaries and reports about learner's work
- functionalities for easy monitoring of learner's work
- functionalities for adding new lessons, examples, and tests
- possibilities for communication between learners and mentors.

Two main roles exist in the system, intended for two types of users [26]:

- *learners* - they are taking the Java programming course and will be using the system in order to gain certain knowledge and
- *instructors* (learners' mentors) - their role is to be the lesson and learner database administrator, to track progress of learners learning and to help them with their assignments.

Therefore, two separated user interfaces are provided for both roles: learner (student) and instructor (learner's mentor). Instructor's interface helps in process of managing data about a learner and course material. Learner's interface is a series of web pages that provide two options: taking lessons and testing learner's knowledge. All data about learner and his progress in the course, as well as data about tutorials, tests and examples are stored in the system's server.

3.2. Adaptive Learning and Personalization of a Content Delivery

The ultimate goal for developing Mag system is increasing the learning opportunities, challenges and efficiency. Two important ways of increasing the quality of service of Mag are to make it intelligent and adaptive. Different techniques must be implemented to adapt content delivery to individual learners according to their learning characteristics, preferences, styles, and goals. In order to support adaptive learning and personalization of a content delivery, we must constantly measure the learner's knowledge and progress, build learner model and possibly redirect the course accordingly.

Our goal is to provide two general categories of personalization in the system, based on previously mentioned levels of personalization:

- *Content adaptation* - presenting the content in different ways, according to the domain model and information from the learner model. All learners and contents will be grouped into classes of similar objects in order to recommend optimum resources and pathways. The principle of clustering is maximizing the similarity inside an object group and minimizing the similarity between the different object groups. Such clusters need to be defined in Mag system in order to provide learner with the most suitable learning material and to form the most suitable pathway.
- *Link adaptation* - the system modifies the appearance and/or availability of every link that appears in a course web page, in order to show the learner whether the link leads to interesting new information, to new information the learner is not ready for, or to a page that provides no new knowledge.

Three different levels of personalization (based on the levels of increasing abstraction and sophistication) must be included in the Mag system, which are suggested in [5] and [21]: self-described personalization, segmented personalization and cognitive-based personalization.

Self-described Personalization

The learners will describe their preferences and common attributes with use of surveys or questionnaire, as well as identify their backgrounds and previous experiences. These create the initial learner model to start with in the instruction to follow. Preferred style for every learner is determined with questionnaire filled at the beginning of the course and with optional questionnaire filled at the end of every completed lesson. We consider the fact that surveys and questionnaires are intrusive and distracting in a learning environment [29]. Therefore, Mag also needs to track learner's achievement and update learner model accordingly.

Segmented Personalization

Learners will be grouped into smaller, identifiable and manageable clusters, based on their common attributes (e.g., class, age), preferences and results

of surveys. Parts of the instruction are then tailored to the groups, and are applied in the same or similar way to all members of a segmented group. Learning material must also be clustered by its purpose, based on benefits it delivers to learners. Therefore, tests in our system will contain three types of questions [9]:

- *Multiple – choice of syntax check.* This type of test is used to ask the learner to trace the correct sample code.
- *Multiple – choice of execution results.* This type of test is used to ask the learner to choose correct result after execution of offered piece of code.
- *Code completion.* Problem is presented in form of skeleton program with the specific missed parts of code. The learner is expected to enter appropriate code snippet according to program specification.

Multiple-choice of syntax check questions are used to grade learner's understanding of syntax rules, while multiple-choice of execution results questions shows how learner understands existing code. Code completions tasks are used to check application skills of a learner. If learner has difficulties with a particular kind of questions or tasks, the system will increase their number in next session in order to provide learner with additional opportunities to improve particular skills.

Cognitive-based Personalization

Cognitive-based personalization represents process of adapting and delivering content and instruction to specific types of learners, defined according to information about their capabilities, and preferences. These may include, for example: a learner's preference for specific type of tests or tasks, or linear sequencing over grouping of hyperlinks, as well as recognition of the learner's reasoning capacity and capability for inductive reasoning. This type of personalization is more complex for implementation than the previous types, as it requires collecting data, monitoring the learner's activities, comparing it to other learners' behavior, building a learner model and predicting and recommending what the learner would like to do or see next.

System will consider preferred learning styles for every particular learner before presenting him/her appropriate material. Every learner has his/her own learning style that indicates a preference for some media type(s) over others. Mag will distinguish three learner's learning styles (among others identified by Dunn and Dunn [6]) and therefore use three different presentation methods:

- *Textual.* Learners that prefer to perceive materials as text are provided with lessons, which are in form of text pages with rich formatting and highlighted source code.
- *Visual.* Learners that prefer to perceive materials in form of pictures are provided with illustrations, figures, diagrams, flowcharts, etc.
- *Interactive.* Learners that prefer to interact physically with learning material are provided with interactive flesh animation.

Mag will track improvement that specific learners make while using specific presentation method and update learner model accordingly. In every moment learner can manually switch between different presentation methods.

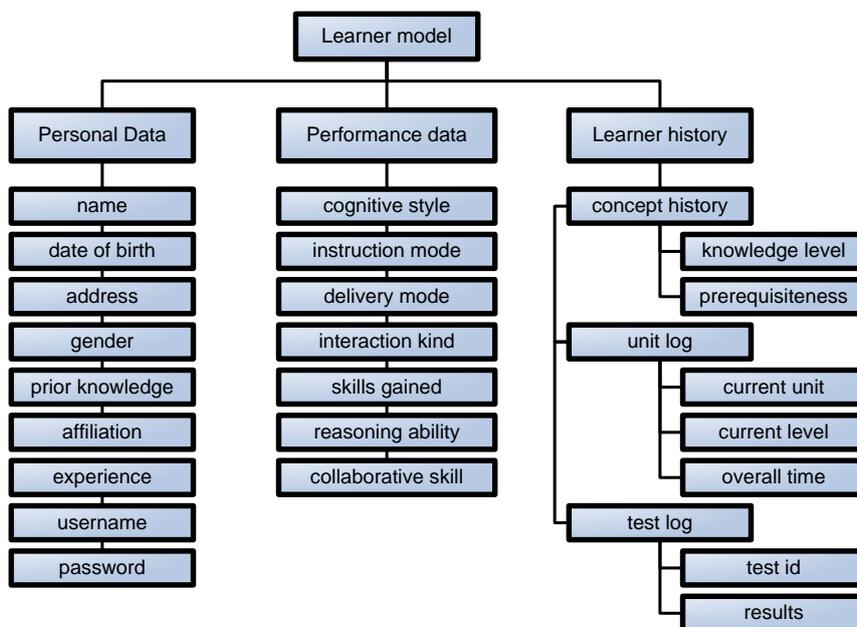


Fig 1. Learner model

We need to add new categories of questions, tests, tasks and tutorials that will provide variety of testing and presenting possibilities during usage of the system. In addition, learners must be categorized into clusters based on the preferable categories of content delivery. That categorization will be accomplished by different surveys (that the learner will be prompt to fill during the registration with the system and optionally after every lesson) and by monitoring the learner’s actions, progress and overall performance. In this vein, we plan to track characteristics of the learner and collect a variety of useful information:

- information about the learner, including cognitive, affective and social characteristics,
- information about the learner’s perspectives on the content itself, including the learner’s feedback on the content, the learner’s knowledge of the content (as determined, for example, by a test administered during the learner’s interactions with the system),
- information about the technical context of use, including characteristics of the learner’s software and hardware environment,

- information about that how the learner interacts with content, including observed metrics such as dwell time, number of learner keystrokes, patterns of access.

Gathered information will be classified along three layers in Learner model that is presented in figure 1.

3.3. Recommendation possibilities

Recommender systems are becoming very popular in e-commerce applications to recommend the online purchase of some products [29]. These systems can be very useful in an e-learning environment to recommend actions, resources or simply links to be followed.

Recommendation system integrated in a learning environment recommend next task to a learner based on the tasks already done by the learner and his/her success, and based on tasks made by other "similar" learners. The similarity between learners is established using learner profiles, or is based on common previous access patterns [16]. There are several ways to implement recommendation in e-learning systems [29]. We plan to implement two the most common recommender techniques in the new version of Mag: collaborative filtering and association rule mining [17].

Collaborative filtering. When learner needs suggestion about which location to visit or which test or example will provide the most benefits, the learner profile is compared to the other profiles to find similar [29]. A selection from these similar profiles will be used to produce recommendation. For this mater, we will need good and trusted ratings entered by the learners. The learners will be prompt to fill short questionnaire after every lesson. This questionnaire will be optional because entering ratings could be considered intrusive. If learner refuses to fill the questionnaire, system will use history logs by other learners as input for his/her profile.

Association rule mining. Authors of courses, when setting up the structure of the course, have a certain navigation pattern in mind and assume that the most of the learners would follow a consistent path, materialized by some hyperlinks. Learners, on the other hand, could follow different paths based on their preferences and generate a variety of series of learning activities. Often this series are not the optimum series, and probably they differ from series intended by the designer. All those variety of series of learning activities are noted down, ordered by success of learners that performed them and recommended to the future learners. Therefore, the automatic recommendation in Mag will be based on the author's intended sequence of navigation in the course material, or based on navigation patterns of other successful learners. This technique is used for recommending shortcuts or jumps to some resources to help learners better navigate the course materials [29].

Those two recommender techniques present a means for the personalization implemented by the Mag adaptation model. This model is part of general architecture of Mag and will be addressed in next section.

3.4. Proposed Architecture of the Hybrid System

The proposed architecture is concerned with producing previously mentioned types of personalization that will be implemented in the already existing learning environment and used by a large number of learners. Open standards, like XML, RDF and OWL [14], [19] needed to be used in order to allow the specification of ontologies to standardize and formalize meaning and to enable the reuse and interoperability. Figure 2 shows a graphical representation of the proposed architecture. This architecture presents adapted architecture of first version of Mag [25] based on experiences of similar web-based learning systems [3], [20], [14] and architecture for ontology-supported adaptive web-based education systems [5] and [4].

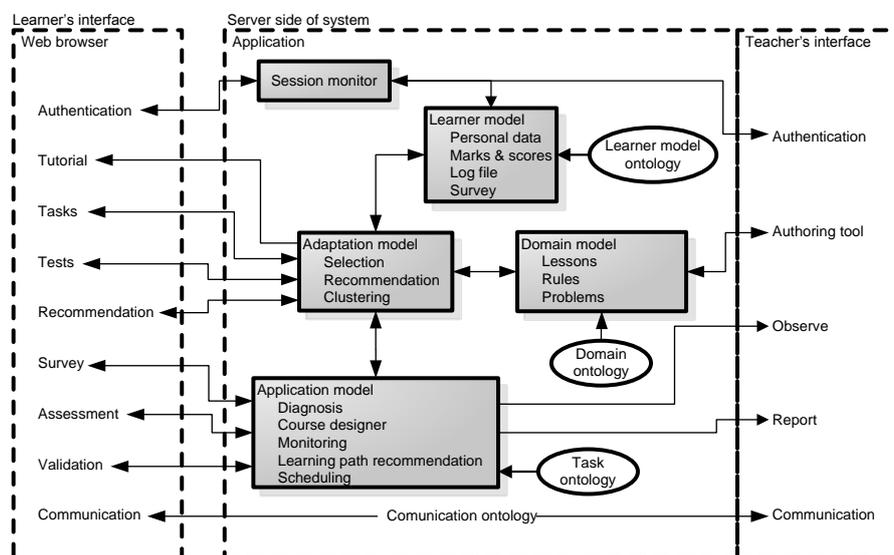


Fig. 2. Proposed system architecture

This is essentially a centralized architecture. The core of the system includes the *adaptation model*, *learner model*, *application model* and *domain model*, all of them stored on a central server. *Domain model* presents storage for all essential concepts in the domain, tutorials and tests. It describes how the information content is structured. Instructors (learner's mentors) can create *domain model* using appropriate authoring tool.

The *application model* applies different strategies and techniques to ensure efficient tailoring of the learning content to the individual learners and personalized task and navigation sequencing. It supports a given pedagogical strategy. For example, that strategy often consists of selecting or computing a specific navigation sequences among the resources based on the information contained in the learner model. The *adaptation model* follows the instructional

directions specified by the *application model* and creates navigation sequence of resources recommended for the particular learner. These two components are separated in order to make easier adding new content clusters and adaptation functionalities.

The adaptation model is also responsible for building and updating learner model characteristics and for personalization of the application to the learner. It processes changing of learner's characteristics based on learner's activities and it provides an adaptation of visible aspects of the system for specific learner. Its main tasks also include storage and management of course data, ways of presenting courses to learners, provision of reports and test results etc.

Each *learner model* is a collection of both static (personal data, specific course objectives, etc.) and dynamic (marks, scores, time spent on specific lesson, etc.) data about the learner, as well as a representation of the learner's performances and learning history. The system uses that information in order to predict the learner's behavior, and thereby adapt course to his/her individual needs.

Within *session monitor* component, the system gradually builds the *learner model* during each session, in order to keep track of the learner's actions and his/her progress, detect and correct his/her errors and possibly redirect the session accordingly. In the end of the session, the *learner model* is updated. It is then used along with other information and knowledge to initialize the next session with the same learner.

Ontology engineering is a key aspect for the success of proposed web-based educational system. Educational ontologies for different purposes must be included, such as for presenting a domain (*domain ontologies*), for building learner model (*learner model ontologies*) or for presenting activities in the system (*task ontologies*) [20]. A repository of ontologies must be built to achieve easier knowledge sharing and reuse, more effective learner modeling and easier extension of a system. These ontologies will be built based on the e-Learning standards defined in SCORM (*Sharable Content Object Reference Model*) [19]. This ontological representation (OWL/RDF) will enable not only to represent meta-data but also reasoning in order to provide the best solution for each individual learner.

4. Conclusion and Future Work

E-learning environments can use different recommendation techniques in order to suggest the most appropriate online learning activities to learners, based on their preferences, knowledge and the browsing history of other learners with similar characteristics. The ultimate goal is improving the teaching and learning process by providing the learners with personalized courses.

In this paper, we presented the idea for integration of recommender system into an existing web-based Java tutoring system [25] in order to introduce

intelligence in the system and to make it adaptive to individual learner's needs and interactions. The proposed architecture contains elements of two different categories of e-Learning systems. First category is one that applies different forms of learner models to adapt the content and the links of hypermedia course pages to the learner, called *Adaptive educational hypermedia systems (AEHSs)*. Recommender systems for classifying learners and contents in order to recommend optimum resources and pathways are the second category of e-learning systems.

The existing Mag system will be extended according to proposed architecture. New learning objects will be added to system in form of the educational ontologies. Appropriate clustering of both learning material and learners will be implemented in order to introduce recommendation of appropriate material to specific types of learners. We plan to track characteristics of the learner and collect a variety of useful information about the learner's perspectives on the content itself.

References

1. Bednarik R., Moreno A., Myller N.: Program Visualization for Programming Education – Case of Jeliot 3. Association for Computing Machinery New Zealand Bulletin, p.8 (2006)
2. Bieg C., Diehl S.: Education and technical design of a Web-based interactive tutorial on programming in Java. Science of Computer Programming, pp. 25-36 (2004)
3. Chen C.M.: Ontology-based concept map for planning a personalized learning path. British Journal of Educational Technology, Blackwell publishing, pp. 1-31 (2008)
4. De Bra P., Aroyo L., Chepegin V.: The next big thing: adaptive Web-based systems. Journal of Digital Information 5, Article No. 247, p.12 (2004)
5. Devedžić V.: Education and the Semantic Web. International Journal of Artificial Intelligence in Education 14, IOS Press, pp 39-65 (2004)
6. Dunn R., Dunn K.: Teaching Students Through Their Individual Learning Styles: A Practical Approach. Reston Publishing: Reston, VA, (1978)
7. Farzan, R., Brusilovsky, P.: Social Navigation Support in a Course Recommendation System. In proceedings of 4th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Dublin, pp. 91-100 (2006)
8. Glenn B., Shahida P., Wei F. Moritz S.: A Web-based ITS for OO Design. 12th International Conference on Artificial Intelligence in Education, Amsterdam, pp. 59-64 (2004)
9. Ivanović M., Pribela I., Vesin B., Budimac Z.: Multifunctional Environment For E-Learning Purposes. Novi Sad Journal of Mathematics, NSJOM Vol. 38, No. 2, pp. 153-170 (2008)
10. Jones N., Macasek M., Walonoski J., Rasmussen K., Heffernan N.: Common Tutor Object Platform – an e-Learning Software Development Strategy. Proceedings of the 15th international conference on World Wide Web, Edinburgh, Scotland, pp. 307-316. (2006)

11. Ksristofic, A.: Recommender System for Adaptive Hypermedia Applications. In Proceeding of Informatics and Information Technology Student Research Conference, Bratislava, pp. 229-234 (2005)
12. Martinez, M.: Designing learning objects to mass customize and personalize learning. in: The Instructional Use of Learning Objects, D. A. Wiley, ed. Article 3.1, p. 29 (2000)
13. Merceron A., Yacef K.: A Web-Based Tutoring Tool with Mining Facilities to Improve Learning and Teaching. Artificial Intelligence in Education, IOS Press, pp. 201-208 (2003)
14. Merino P.J.M., Kloos C.D.: An Architecture for Combining Semantic Web Tehniques with Intelligent Tutoring Systems. ITS 2008, Springer-Verlag Berlin Heideberg, pp. 540-550 (2008)
15. Mobasher, B.: Data Mining for Personalization. In The Adaptive Web: Methods and Strategies of Web Personalization. Springer-Verlag, Berlin Heidelberg, pp. 1-46 (2007)
16. Romero C., Ventura S., Delgado J.A., De Bra P.: Personalized Links Recommendation Based on Data Mining in Adaptive Educational Hypermedia Systems. Creating New Learning Experiences on a Global Scale pp. 292-306 (2007)
17. Romero, C., Ventura, S.: Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with Applications. Elsevier 1:33, pp. 135-146 (2007)
18. Schafer, J.B.: The application of data-mining to recommender systems, Encyclopedia of data warehousing and mining. Hershey, PA. Idea Group, pp. 44-48 (2005)
19. SCORM - <http://www.adlnet.gov/scorm/>
20. Šimić G.: The Multi-courses Tutoring System Design. ComSIS Vol. 1, No. 1, pp. 141-155 (2004)
21. Soonthornphisaj N., Rojsattarat E., Yim-ngam S.: Smart E-Learning Using Recommender System. ICIC Springer-Verlag Berlin Heidelberg, pp 518-523 (2006)
22. Suarez M., Sison R.: Automatic Construction of a Bug Library for Objected-Oriented Novice Java Programmer Errors. ITS 2008, Springer-Verlag Berlin Heideberg, pp. 184-193 (2008)
23. Sykes E.R., Franek F.: An Intelligent Tutoring System Prototype for Learning to Program Java. The third IEEE International Conference on Advanced Learning Technologies (ICALT'03) Athens, Greece, pp 485-492 (2003)
24. Sykes E.R., Franek F.: Presenting JECA: A Java Error Correcting Algorithm for the Java Intelligent Tutoring System. Advances in Computer Science and Technology - 2004, St. Thomas, US Virgin Islands (2004)
25. Vesin B. Ivanović M., Budimac Z., Pribela I.: Tutoring System for Distance Learning of Java Programming Language. 10th Symposium on Programming Languages and Software Tools SPLST, Dobogókő, Hungary, pp. 310-320. (2007)
26. Vesin B., Ivanović M., Budimac Z.: Learning Management System for Programming in Java. Annales Universitatis Scientiarum De Rolando Eötvös Nominatae, Sectio - Computatorica, vol. 31., pp. 75-92 (2009)
27. Wei F., Moritz S.H., Parvez S.M., Blank G.D.: A Student Model for Object-Oriented Design and Programming. □ Journal of Computing Sciences in Colleges, pp. 260 - 273 (2005)
28. Wolf C.: iWeaver: Towards 'Learning Style' - based e-Learning in Computer Science Education. Australasian Computing Education Conference, Adelaide, Australia, vol. 20 (2003)
29. Zaiane O.R.: Recommender systems for e-learning: toward non-intrusive web mining. Data mining in e-learning, WIT press, pp. 79-96 (2006)

Aleksandra Klašnja-Milićević, Boban Vesin, Mirjana Ivanović, and Zoran Budimac

Aleksandra Klašnja-Milićević is a Lecturer of Computer Science in Higher School of Professional Business Studies at University of Novi Sad, Serbia. She received her Diploma in Electrical and Computer Engineering from Faculty of Technical Sciences at the University of Novi Sad in 2002. She joined the graduate program in Computer Sciences at Faculty of Science, Department of Mathematics and Informatics, University of Novi Sad in 2003, where she received her M.Sc. degree (2007). Her research interests include recommender systems, information retrieval, user modeling and personalization, and electronic commerce. She has published more than 20 referred papers on her work in national and international conferences and journals.

Boban Vesin has graduated at the Faculty of Science, University of Novi Sad, in 2002. He got his master degree at the same Faculty in 2007 and is working on his PhD thesis. Currently he is a lecturer at Higher School of Professional Business Studies, University of Novi Sad. His major research interests are e-Learning and personalization in intelligent tutoring systems. He has published a number of scientific papers in the area.

Mirjana Ivanović is a Professor at Faculty of Science, Department of Mathematics and Informatics, University of Novi Sad, Serbia. Professor Ivanović's research addresses the issues of multi-agent systems, e-learning and web-based learning, data mining, case-based reasoning, programming languages and tools. She actively participates in more than 10 international and several national projects. She has been a program committee member of more the 50 international conferences. Prof. Ivanović is Editor-in-Chief of ComSIS (Computer Science and Information Systems) Journal. She has published over 200 scientific papers in proceedings of international conferences and journals, has written more than 10 university textbooks in the field of informatics and ICT.

Zoran Budimac is a professor at Faculty of Science, Department of Mathematics and Informatics, University of Novi Sad. He graduated in 1983 (informatics), received master's degree (computer science) in 1991 and doctor's degree (computer science) in 1994. His research interests include: mobile agents, e-learning, software engineering, case-based reasoning, implementation of programming languages. He has been project leader for several international and several national projects. He has published over 180 scientific papers in proceedings of international conferences and journals, has written more than 12 university textbooks in different fields of informatics.

Received: June 08, 2009; Accepted: February 17, 2010.

CIP – Каталогизacija y publikaciji
Народна библиотека Србије, Београд

004

COMPUTER Science and Information
Systems : the International journal /
Editor-in-Chief **Mirjana Ivanović**. – Vol. 8,
No 1 (2011) - . – Novi Sad (**Trg D. Obradovića**
3): ComSIS Consortium, 2011 - (Belgrade
: Sgra star). –30 cm

Polugodišnje. – Tekst na engleskom jeziku

ISSN 1820-0214 = Computer Science and
Information Systems
COBISS.SR-ID 112261644

Cover design: V. Štavljanin
Library administration: M. Petković
Printed by: Sgra star, Beograd



Contents

Editorial

Papers

- 1 VPN Solutions and Network Monitoring to Support Virtual Teams Work in Virtual Enterprises
Sebastian Marius Roşu, George Drăgoi
- 27 Towards the Methodology for Development of Fuzzy Relational Database Applications
Srđan Skrbic, Miloš Racković, Aleksandar Takači
- 41 Design Pattern Instantiation Directed by Concretization and Specialization
Peter Kajsá, Lubomir Majtas, Pavol Navrat
- 73 A Grammar-based Model for the Semantic Web
Hyosook Jung, Seongbin Park
- 101 Ontology-based Multi-label Classification of Economic Articles
Sergeja Vogrinčić, Zoran Bosnić
- 121 An Analysis of an Assessment Model for Participation in Online Forums
Mee-Chin Wee, A. Abrizah
- 141 A New Method for Constructing Kernel Vectors in Morphological Associative Memories of Binary Patterns
Yiannis.S. Boutalis

Papers selected from "The 4th International Conference on Information Technology", June 3 - 5, Amman, Jordan 2009

- 167 Supply Chain Performance Measurement System Based on Scorecards and Web Portals
Nenad Stefanovic, Dusan Stefanovic
- 193 Usage of Agents in Document Management
Dragoslav Pešović, Milan Vidaković, Mirjana Ivanović, Zoran Budimac, Jovana Vidaković
- 211 Integration of Recommendations and Adaptive Hypermedia into Java Tutoring System
Aleksandra Klačnja-Milićević, Boban Vesin, Mirjana Ivanović, Zoran Budimac