

Ernestina Menasalvas
Javier Segovia
Piotr S. Szczepaniak (Eds.)

LNAI 2663

Advances in Web Intelligence

First International Atlantic Web Intelligence Conference
AWIC 2003, Madrid, Spain, May 2003
Proceedings



Springer

Ernestina Menasalvas Javier Segovia
Piotr S. Szczepaniak (Eds.)

Advances in Web Intelligence

First International Atlantic Web Intelligence Conference
AWIC 2003
Madrid, Spain, May 5-6, 2003
Proceedings

Springer

Berlin
Heidelberg
New York
Hong Kong
London
Mannheim
Paris
Tokyo



Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Ernestina Menasalvas
Javier Segovia
Universidad Politécnica de Madrid
Facultad de Informática
Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain
E-mail: {emenasalvas/fsegovia}@fi.upm.es

Piotr S. Szczepaniak
Technical University of Lodz
Institute of Computer Science
ul. Sterlinga 16/18, 90-217 Lodz, Poland
E-mail: office@ics.p.lodz.pl

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

CR Subject Classification (1998): I.2, H.3, H.4, C.2, H.5, J.1, J.2

ISSN 0302-9743

ISBN 3-540-40124-5 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2003
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin GmbH

Preface

We are pleased to present the proceedings of the 2003 Atlantic Web Intelligence Conference, AWIC 2003. The conference was located in Madrid, Spain during May 5–6, 2003, organized locally by the Technical University of Madrid.

AWIC 2003 aimed to be the first of a series of conferences on Web Intelligence, to be celebrated annually, alternatively in Europe and America, starting in Madrid. It was born as an activity of the recently created WIC-Poland Research Centre and the WIC-Spain Research Centre, both belonging to the Web Intelligence Consortium (WIC) (<http://wi-consortium.org>). AWIC 2003 was supported with grants from the Spanish Ministry for Science and Technology and the European Network of Excellence in Knowledge Discovery, KDNes.

AWIC 2003 brought together scientists, engineers, computer users, and students to exchange and share their experiences, new ideas, and research results about all aspects (theory, applications, and tools) of artificial intelligence techniques applied to Web-based systems, and to discuss the practical challenges encountered and the solutions adopted. Almost 70 contributions were submitted. After a preliminary evaluation, 60 of these papers were accepted to the conference and were assigned at least two reviewers from the international program committee. Out of this 60, 33 were conditionally accepted, and 32 of them were finally accepted after the conditions set by the reviewers had been met, which resulted in an acceptance ratio of 45%.

AWIC 2003 had a general session with 9 papers, and the others were organized into special sessions on areas such as:

- Web security,
- Semantic Web,
- Web authoring and design,
- Web information retrieval,
- Agents for the web, and
- User behavior

During AWIC 2003 we were pleased to have some important invited speakers. Prof. Lotfi Zadeh, who does not need any introduction, was the honorary chair and gave a talk about the transition of search engines to question-answering systems. This book contains what could be considered an extended abstract of his actual lecture. Prof. Ning Zhong, President of WIC, who can be considered as one of the fathers of the term “Web Intelligence,” gave a talk about the past, present, and future of the area. Finally, Prof. Ryszard Tadeusiewicz spoke about the application of artificial intelligence techniques for information retrieval.

We thank all the authors for contributing to what in our minds was a more than interesting program in Web Intelligence. We also want to thank the remarkable effort made by Coro Perez, secretary of the conference, and the rest of the local committee. We hope to see all of them, and many others at AWIC 2004.

May 2003

E. Menasalvas
J. Segovia

AWIC 2003 Organization

Honorary Chair

Lotfi A. Zadeh (University of California, USA)

Conference Co-chairs

Javier Segovia (Universidad Politécnica de Madrid, Spain)

Piotr Szczepaniak (Technical University of Lodz, Poland)

Steering Committee

Ning Zhong (Maebashi Institute of Technology, Japan)

Enric Trillas (Universidad Politécnica de Madrid, Spain)

Jiming Liu (Hong Kong Baptist University)

Janusz Kacprzyk (Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland)

Organizing Committee

Ernestina Menasalvas (Universidad Politécnica de Madrid, Spain)

Jesús Favela (Centro de Investigación Científica y de Educación Superior de Ensenada, México)

Local Committee

José M. Peña (Universidad Politécnica de Madrid, Spain)

Oscar Marbán (Universidad Carlos III de Madrid, Spain)

Ángel Lucas (Universidad Politécnica de Madrid, Spain)

Sonia Frutos (Universidad Politécnica de Madrid, Spain)

Javier Crespo (Universidad Carlos III de Madrid, Spain)

Eugenio Santos (Universidad Politécnica de Madrid, Spain)

Organization of the Special Track on Uncertainty, Hypermedia and the Web: Search, Personalization and Usability

José Angel Olivas (Universidad de Castilla-La Mancha, Spain)

Miguel-Ángel Sicilia (Universidad Carlos III, Spain)

AWIC 2003 Program Committee

S.S. Anand, UK	B. Mobasher, USA
R. Baeza-Yates, Northern Ireland, UK	C. Montes, Spain
T. Bao Ho, Japan	O. Nasraoui, USA
B. Berendt, Germany	M. Piattini, Spain
A. Büchner, Northern Ireland, UK	P. Quaresma, Portugal
J. Favela, México	R. Rastogi, USA
U. Fayyad, USA	V. Robles, Spain
W. Gaul, Germany	M. Rudnicki, Poland
J. Ghosh, USA	J.W. Shavlik, USA
O. Guenther, Germany	A. Skowron, Poland
M. Hadjimichael, USA	M. Spiliopoulou, Germany
E. Hochsztain, Uruguay	R. Tadeusiewicz, Poland
P. Isasi, Spain	M.A. Vila, Spain
L. Joyanes Aguilar, Spain	A. Wasilewska, USA
J. Kacprzyk, Poland	X. Wu, USA
S. Kaski, Finland	R. Yager, USA
C. A. Knoblock, USA	Y. Yao, Canada
D. H. Kraft, USA	P.S. Yu, USA
R. Kruse, Germany	S. Zadrozny, Poland
C. Y. Lin, USA	M. Zaki, USA
H. Liu, Hong Kong	W. Ziarko, Canada
L. Mason, USA	Z. Zheng, USA
K. Manolopoulos, Greece	N. Zhong, Japan
S. Millán, Colombia	

Table of Contents

Invited Speakers

Toward Web Intelligence	1
<i>Ning Zhong</i>	
From Search Engines to Question-Answering Systems – The Need for New Tools	15
<i>Lotfi A. Zadeh</i>	
Artificial Intelligence Techniques in Retrieval of Visual Data Semantic Information	18
<i>Ryszard Tadeusiewicz, Marek R. Ogiela</i>	

General Session

A Framework to Integrate Business Goals in Web Usage Mining	28
<i>E. Hochsztain, S. Millán, B. Pardo, J.M. Peña, E. Menasalvas</i>	
Clustered Organized Conceptual Queries in the Internet Using Fuzzy Interrelations	37
<i>Pablo J. Garcés, José A. Olivas, Francisco P. Romero</i>	
Collaborative Filtering Using Interval Estimation Naïve Bayes	46
<i>V. Robles, P. Larrañaga, J.M. Peña, O. Marbán, J. Crespo, M.S. Pérez</i>	
DJ-Boids: Flocks of DJ's to Program Internet Multichannel Radio Stations	54
<i>Jesus Ibanez, Antonio Gomez-Skarmeta, Josep Blat</i>	
Evaluating the Informative Quality of Web Sites by Fuzzy Computing with Words	62
<i>Enrique Herrera-Viedma, Eduardo Peis, María Dolores Olvera, Juan Carlos Herrera, Yusef Hassan Montero</i>	
Mining Association Rules Using Fuzzy Inference on Web Data	73
<i>Mariluz Martínez, Gélver Vargas, Andrés Dorado, Marta Millán</i>	
The "Who Is It?" Problem Application for Customizable Web Sites	83
<i>Pierre-François Lanier, Jean-Michel Lacroix</i>	

Using Case-Based Reasoning to Improve Information Retrieval in Knowledge Management Systems	94
<i>Norbert Gronau, Frank Laskowski</i>	

Web Page Classification: A Soft Computing Approach	103
<i>Angela Ribeiro, Víctor Fresno, María C. Garcia-Alegre, Domingo Guinea</i>	

Special Session: Web Security

Content-Based Methodology for Anomaly Detection on the Web	113
<i>Mark Last, Bracha Shapira, Yuval Elovici, Omer Zaafrany, Abraham Kandel</i>	

Designing Reliable Web Security Systems Using Rule-Based Systems Approach	124
<i>Grzegorz J. Nalepa, Antoni Ligeza</i>	

Secure Intelligent Agents Based on Formal Description Techniques	134
<i>L. Mengual and C. de la Puente</i>	

Special Session: Semantic Web

A Natural Language Interface for Information Retrieval on Semantic Web Documents	142
<i>Paulo Quaresma, Irene Pimenta Rodrigues</i>	

Conceptual User Tracking	155
<i>Daniel Oberle, Bettina Berendt, Andreas Hotho, Jorge Gonzalez</i>	

Mapping with Web Knowledge	165
<i>J.L. Arjona, R. Corchuelo, J. Peña, D. Ruiz</i>	

Formalization of Web Design Patterns Using Ontologies	179
<i>Susana Montero, Paloma Díaz, Ignacio Aedo</i>	

Special Session: Web Authoring and Design

A Multi-agent System for Web Document Authoring	189
<i>M. Pérez-Coutiño, A. López-López, M. Montes-y-Gómez, C. Villaseñor-Pineda</i>	

On the Vague Modelling of Web Page Characteristics Regarding Usability	199
<i>Elena García Barriocanal, Miguel A. Sicilia Urbán, J. Antonio Gutiérrez</i>	

Semantic Interoperability Based on Ontology Mapping in Distributed Collaborative Design Environment	208
<i>Jingyi Zhang, Shensheng Zhang, Chun Chen, Bo Wang</i>	

The Use of Data Mining Techniques to Classify Styles and Types of Web-Based Information: Generation of Style Sheets Using Data Mining	218
<i>Kathryn E.B. Thornton, Mark Carrington</i>	

Special Session: Web Information Retrieval

Building Topic Profiles Based on Expert Profile Aggregation	230
<i>Miguel Delgado, María J. Martín-Bautista, Daniel Sánchez, José M. Serrano, María-Amparo Vila</i>	

Carrot ² and Language Properties in Web Search Results Clustering	240
<i>Jerzy Stefanowski, Dawid Weiss</i>	

Practical Evaluation of Textual Fuzzy Similarity as a Tool for Information Retrieval	250
<i>Piotr S. Szczepaniak, Marcin Gil</i>	

Rough – Fuzzy Reasoning for Customized Text Information Retrieval	258
<i>Shailendra Singh, P. Dhanalakshmi, Lipika Dey</i>	

Special Session: Agents for the Web

A Machine Learning Based Evaluation of a Negotiation between Agents Involving Fuzzy Counter-Offers	268
<i>Javier Carbo, Agapito Ledezma</i>	

Autonomous Agents to Support Interoperability and Physical Integration in Pervasive Environments	278
<i>Marcela Rodríguez, Jesus Favela</i>	

Feature Selection Algorithms to Improve Documents' Classification Performance	288
<i>Pedro A.C. Sousa, João Paulo Pimentão, Bruno René D. Santos, Fernando Moura-Pires</i>	

ROSA – Multi-agent System for Web Services Personalization	297
<i>Przemysław Kazienko, Maciej Kisielecki</i>	

Supporting Software Maintenance in Web Repositories through a Multi-agent System	307
<i>Aurora Vizcaino, Jesús Favela, Mario Piattini, Felix García</i>	
<hr/>	
Special Session: User Behavior	
<hr/>	
A Hybrid Method for Patterns Mining and Outliers Detection in the Web Usage Log	318
<i>Mikhail Petrovskiy</i>	
Analysis of User's Behaviour in Very Large Business Application Systems with Methods of the Web Usage Mining – A Case Study on SAP® R/3®	329
<i>Jorge Marx Gómez, Gamal Kassem, Claus Rautenstrauch, Markus Melato</i>	
Automation of the Deep Web with User Defined Behaviours	339
<i>Vicente Luque Centeno, Carlos Delgado Kloos, Peter T. Breuer, Luis Sánchez Fernández, Ma. Eugenia Gonzalo Cabellos, Juan Antonio Herráiz Pérez</i>	
Author Index	349

Toward Web Intelligence

Ning Zhong

Department of Information Engineering
 Maebashi Institute of Technology
 460-1 Kamisadori-Cho, Maebashi-City 371-0816, Japan
 zhong@maebashi-it.ac.jp

Abstract. Web Intelligence (WI) presents excellent opportunities and challenges for the research and development of new generation of Web-based information processing technology, as well as for exploiting Web-based advanced applications. Based on two perspectives of WI research: an intelligent Web-based business-centric schematic diagram and the conceptual levels of WI, we investigate various ways to study WI and potential applications.

1 Introduction

The concept of Web Intelligence (WI for short) was first introduced in our papers and books [19,35,38,41,45,46]. Broadly speaking, Web Intelligence is a new direction for scientific research and development that explores the fundamental roles as well as practical impacts of Artificial Intelligence (AI)¹ and advanced Information Technology (IT) on the next generation of Web-empowered systems, services, and environments. The WI technology revolutionizes the way in which information is gathered, stored, processed, presented, shared, and used by e-commerce, virtualization, globalization, standardization, personalization, and portals.

With the rapid growth of the Web, research and development on WI have received much attention. There is great potential for WI to make useful contributions to e-business (include e-commerce), e-science, e-learning, e-government, and so on. Many specific applications and systems have been proposed and studied. In particular, the e-business activity that involves the end user is undergoing a significant revolution [29]. The ability to track users' browsing behavior down to individual mouse clicks has brought the vendor and end customer closer than ever before. It is now possible for a vendor to personalize his product message for individual customers at a massive scale. This is called *direct marketing* (or *targeted marketing*) [32,47]. Web mining and Web usage analysis play an important role in e-business for customer relationship management (CRM) and direct marketing. Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services [15,29,36,

¹ Here the term of AI includes classical AI, computational intelligence, and soft computing.

Supporting Software Maintenance in Web Repositories through a Multi-agent System

Aurora Vizcaino¹, Jesús Favela², Mario Piattini¹, and Felix García¹

¹Escuela Superior de Informática. Universidad de Castilla-La Mancha
13071, Paseo de la Universidad, 4, Ciudad Real, Spain
{avizcain, mpiattini, fgarcia}@inf-cr.uclm.es

²CICESE, Ensenada, México
favela@cicese.mx

Abstract. Software Maintenance (SM) is a knowledge-intensive activity. A suitable management of this knowledge could decrease the high costs (economic and in effort) of software maintenance tasks.

The challenge of managing this knowledge increases as the distributed development of software becomes more popular, and developers as well as knowledge are distributed worldwide. Increasingly web repositories are being used for the coordination of development tasks among software and maintenance engineers. Thus, an appropriate technical solution to this problem should be based on a web architecture and associated protocols.

On the other hand, in order to work with all the concepts related to SM is advisable to establish different levels of abstraction, thus the complexity of the concepts, and their management, are simplified. This work presents a system that, by storing information in XMI documents, manages the data and metadata generated during SM, facilitating the work of SM engineers.

1 Introduction

Many studies [7,15,18] provide evidence that the majority of the overall expenses incurred during the life-cycle of a software product occur during the maintenance stages. Thus, in recent years researchers have focused their attention on proposing methods and techniques which help increase the efficiency of the Software Maintenance Process (SMP).

One way to improve maintenance quality and decrease software maintenance costs is to reuse previous information and knowledge [10]. However, for information to be usable it needs to be modeled, structured, generalised and stored in a reusable form, to allow for effective retrieval [1].

In order to decrease the efforts and cost in the SMP we developed a knowledge management system called KM-MANTIS. The system is in charge of storing and reusing information, knowledge and expertise generated during the SMP. KM-MANTIS is based on the experience factory concept [4,5] since it is known that an organizational memory must be maintained by an organizational unit separate from the project organizations because it is mainly concerned to keeping schedules and cost constraints, and proving knowledge would imply an extra effort for the project organizations employees.

KM-MANTIS is a multi-agent system where different types of agent manage the diverse types of information generated during SMP. Agents interchange data and take advantage of the information and experience acquired by other agents.

The main feature of KM-MANTIS is that it stores data and metadata in XMI (XML metadata interchange) [13] documents. It is critical to manage different levels of abstraction to carry out an integral management of a software process. Moreover, being XMI an open format fosters the interchange of data between distributed and heterogeneous repositories which also use XMI.

The contents of this paper are divided as follows. Section 2 discusses the importance of managing the knowledge generated during the software maintenance process. Section 3 describes the features of KM-MANTIS. Section 4 explains how and why information is stored in a XMI repository, and shows, through an example, the use of the system. Finally, conclusions are presented in Sect. 5.

2 The Need for Knowledge Management in Software Maintenance

Software maintenance is an activity where different kinds of knowledge are generated from different sources. This knowledge comes not only from the expertise of the professionals involved in the process, but it is also intrinsic to the product being maintained, and to the reasons that motivate the maintenance (new requirements, user complaints, etc). Moreover, the diverse types of knowledge are produced at different stages of the maintenance process. For instance, [6] claims that during Initial Development the maintenance staff acquires knowledge about the application domain, user requirements, roles of the application, solutions and algorithms, data formats, strengths and weaknesses of the program architecture, operating environment, etc. In the software Evolution stage new users and environmental requirements arise and changes are produced which generate new knowledge, and this process continues in the rest of the stages.

Furthermore, this knowledge is used by different persons at different stages. Each person has partial information that is required by other members of the group. If the knowledge only exists in the software engineers and there is no system in charge of transferring the tacit knowledge (contained in the employees) to explicit knowledge (stored on paper, in files, etc) when an employee abandons the organization a significant part of the intellectual capital goes with him/her.

One of the five factors that have been identified as having a major impact on the productivity of software maintenance is the expertise of the staff members [2]. It has been found that systems maintained by relatively inexperienced programmers average significantly higher error rates [3].

Another well-known issue that complicates the SMP is the scarce documentation that exists related to a specific software system. Even if detailed documentation was produced when the original system was developed, it is seldom updated as the system evolves. For example, legacy software written by other units often has little or no documentation describing the features of the software.

In addition, organizations still lack a culture of reuse and information sharing. As [12] claims, companies only use a quarter part of its intellectual capital.

On the other hand, thanks to technology advances it is frequent for maintenance

- Coordination of distributed work.
- Providing access to knowledge to the different sub-units.
- Managing experience compiled from previous projects and making it available and reusable for new projects.

A knowledge management system could satisfy these needs so as to avoid some of the issues commented previously. For instance, if organizations store their information and knowledge in a KM system, they would own this intellectual capital. Therefore, even if experts left the organization their expertise would remain within the companies.

In addition, a suitable storage of information facilitates its reuse. For these reasons we decided to develop a KM system where the diverse kinds of knowledge generated from different stages of SMP were stored, analysed, shared and reused.

Another advantage of a KM system is that staff may also be informed about the location of information. A critical factor for maintenance engineers is to have access to the knowledge the organizations have. In [22] a study was conducted, which found that the number one barrier to knowledge sharing was "ignorance": the sub-units are ignorant of the knowledge that exists in the organizations, or the sub-units possessing the knowledge are ignorant of the fact that another sub-unit needs such knowledge. Sometimes the organization itself is not aware of the location of the pockets of knowledge or expertise [11].

3 The KM-MANTIS Multi-agent System

This section describes the features of KM-MANTIS and its architecture. However, before presenting the system we explain why a multi-agent architecture for the management of information was chosen.

3.1 Why Agents?

Previous works such as [21, 19] have used agents for knowledge management. There are several reasons why agents are recommendable for this task. First of all, agents can manage both distributed and local information. This is an important feature since, as it was explained in Sect. 2, the software maintenance information is generated by different sources and often from different places.

In a distributed architecture each agent knows about its local environment, and it is close to the place where the information is generated. When global knowledge is required the agents can act together and share knowledge. The alternative will be to have a centralized system that knows "everything". However, this is more difficult to implement and also less robust.

An aspect related to the previous one is that agents may cooperate and interchange information. Thus, each agent can share its knowledge with others or ask them advice, benefiting from the other agents' experience. Therefore, there is reuse and knowledge management in the architecture of the system itself.

Another important issue is that agents can learn from their own experience. Consequently, the system is expected to become more efficient with time since the agents have learnt from their previous mistakes and successes.

On the other hand, agents may utilize different reasoning techniques depending on the situation. For instance, they can use ID3 algorithms to learn from previous experiences and use case-based reasoning to advise a client how to solve a problem.

3.2 The KM-MANTIS Architecture

The system is formed of a set of agent communities which manage different types of knowledge. When the ontology to represent the maintenance domain was developed, different types of information were detected, thus the system has a community per type of information detected since each one has its specific features. The three more relevant types of information identified were: information related to the products to be maintained; information referent to the activities to perform in order to maintain the products; and, peopleware involved during software maintenance process [23].

Therefore, KM-MANTIS has three communities: a community termed "products community", another called "activities community", and the last community denoted "peopleware community". In what follows, we describe each community in more detail.

Products Community: This community manages the information related to the products to be maintained. Since each product has its own features and follows a specific evolution this community has one agent per product. The agents have information about the initial requirements, changes made to the product, and about metrics that evaluate features related to the maintainability of the product (this information is obtained from different documents such as modification requests, preventive, corrective or preventive actions performed or product measurements). Therefore, the agents monitor the product's evolution in order to have up to date information about it at each moment.

Each time an agent detects that information about its product is being introduced or modified in KM-MANTIS (the agent detects it when the product identification number that it represents is introduced or displayed in the interface of KM-MANTIS) the agent starts to work analysing the new information (in the case of there being any), or comparing it to that previously held in order to detect inconsistencies, or checking the differences and storing the relevant information in order to have up-to-date information.

Besides information relevant to each product, which is stored in a database that each agent has, general information about products such as type of products that the system manages should be considered. This information which can be considered as metadata is stored in a XMI repository common for the community, called "community repository" (see Fig. 1). The decision to use XMI documents based on the MOF (Meta-Object Facility) standard (they are described in more detail in Sect. 4) makes it possible for agents of this community to have access to the different levels of information that they need to process and classify their information.

The products and activities communities have a common repository where general information and metadata are stored. For instance, an agent can detect that whenever a change labelled "A" is demanded another change called "B" is required. This

will be shared with the rest of the product agents. In a previous design of this system a single agent was in charge of this repository but this fact increased considerably the traffic of messages between agents, and so it was eliminated, producing thereby an improvement in the speed of the system.

Besides the repository that each community has, agents can also consult a general repository where the documents generated during the SMP are stored in format XML. This repository is usually used when agents need additional information for answering queries or performing their tasks. Later, in Sect. 4, we provide examples of this.

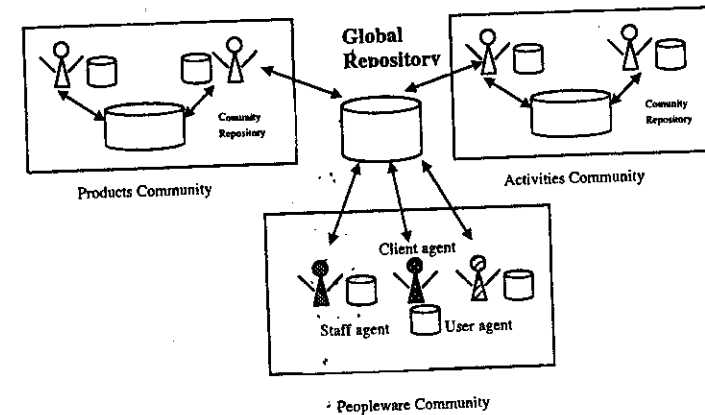


Fig. 1. KM-MANTIS architecture

Activities Community: Each new change demanded implies performing one or more activities. This community, which has one agent per activity, is in charge of managing the knowledge related to the different activities such as methods, techniques and resources used to perform an activity.

Activities agents can also obtain new knowledge from their experience and learning. For instance, an activity agent can learn what person usually carries out a specific activity.

As in the previous case each agent has a database to store its information, and a community repository to access it. This repository contains metadata and data important for all the community. For instance, a taxonomy of activities.

Peopleware Community: Three profiles of people can be clearly differentiated in MP [16]: the maintainer, the customer and the user. The Peopleware community has three types of agent, one per profile detected. One agent is in charge of the information related to staff (maintainers). This is the staff agent. Another manages information related to the clients (customers) and is called the client agent. The last one is in charge of the users and is termed the user agent.

The staff agent knows the personal data of the employees, in what activities they have worked, and what product they have maintained. Of course, the agent also has current information about each member of the staff. Therefore, it knows where each person is working at each moment.

The agent can utilise the information that it has to generate new knowledge. For instance the staff agent may calculate statistics that indicate the time that an employee took to perform a task or calculate the performance graph of each member.

The client agent stores the information of each client, their requirements (even the initial requirements if they are available) and requests. The client agent also tries to obtain new knowledge. For instance, it tries to guess future requirements depending on previous requirements or it estimates the costs of changes that the client wants to make, warning him for instance of the high costs associated with a specific change request.

The server agent is in charge of knowing the requirements of the users of each product, their background and also their complaints and comments about the products. New knowledge could be generated from this information, for example by testing to what degree the users' characteristics influence the maintenance of the product.

Each type of agent has a database but in this case there is not a community repository because there is no data common to the three types of agents. However, as arrows indicate in Fig. 1 agents can access the global repository.

Agents exchange information by using ACL (Agent Communication Language). This language provides agents with the means to share information and knowledge through a set of communicative acts.

The platform used is JADE (Java Agent Development Framework) since it simplifies the implementation of multi-agent systems through a set of tools that supports the debugging and deployment phase.

4 Web Repositories of Information

The World Wide Web has moved from its origins as a medium for the dissemination of information to a middleware platform for the development of sophisticated distributed applications. This trend has given rise to a new branch of Software Engineering, named Web Engineering, concerned with establishing sound principles, techniques, and tools for the development and maintenance of systems, services and applications over the Web.

As the web becomes the preferred medium for the deployment of software applications several CASE tools have migrated to the web and several others have been developed. The use of the web as a platform for the support of software development offers several advantages:

- Ubiquitous access. Web repositories can be accessed from any computer connected to the Internet, including portable devices.
- Simple integration of tools. The web is based on simple, open protocols that can be easily integrated in CASE tools to allow their interoperability. At the lowest level of integration a tool can export content to HTML or semantically richer XML documents.
- Support for distributed software development. Large-scale software systems are increasingly being developed by distributed teams of specialists that need to communicate and coordinate their activities. The web offers a simple middleware for the deployment of workflow and groupware tools that support collaboration.

Of particular interest to us is the use of web servers as repositories for product and process information. This information is not only essential to the production of software, but to its maintenance as well. The centralized storage of this information facilitates configuration management, quality control, project tracking, and the management of requirements.

An important effort in this area is the IETF's DeltaV project [24], an effort to extend the WebDAV and HTTP protocols to support document versioning and configuration control for shared web documents and resources.

On the other hand, applications manipulating Web data require both, documents or information retrieved from the Web, and metadata about this information [8]. In order to specify and manage meta-data we used the standard MOF proposed by the OMG [13].

This standard defines four conceptual levels (see Table 1): the base level, called M0, is where the instances or real data are. For example in the KM-MANTIS repository at this level we find the concrete changes that have been performed. Data managed in this level are instances of the concepts defined in the superior level called M1. The specific model that we used in level M1 is based on the MANTEMA methodology [17] and on a set of techniques adapted to maintenance, such as effort or risk estimations [20]. An example of information at this level are the different types of *support activities* that exist according to MANTEMA. The next level is the M2, which corresponds to the meta-models, in our case to the software maintenance process meta-model. And the last level, M3, is where the meta-meta model is represented, this is the MOF model itself.

Table 1. Conceptual levels in MOF and KM-MANTIS

Level	MOF	KM-MANTIS
M3	MOF model (Meta-metamodel)	MOF Model
M2	Meta-model	SMP metamodel
M1	Model	MANTEMA
M0	Data	SMP instances (products, activities)

An example will help illustrate this classification. The generic concept "Maintenance Activity" represents the correspondence between M3 and M2 (see Fig. 2). This concept is instantiated in a specific type of activity in the correspondence M2 and M1. For instance, in "Urgent Corrective Intervention" activity. When we have real instances of this type of activity it will be stored in the level M0, for example "Intervention number 3 performed in the project Counts.

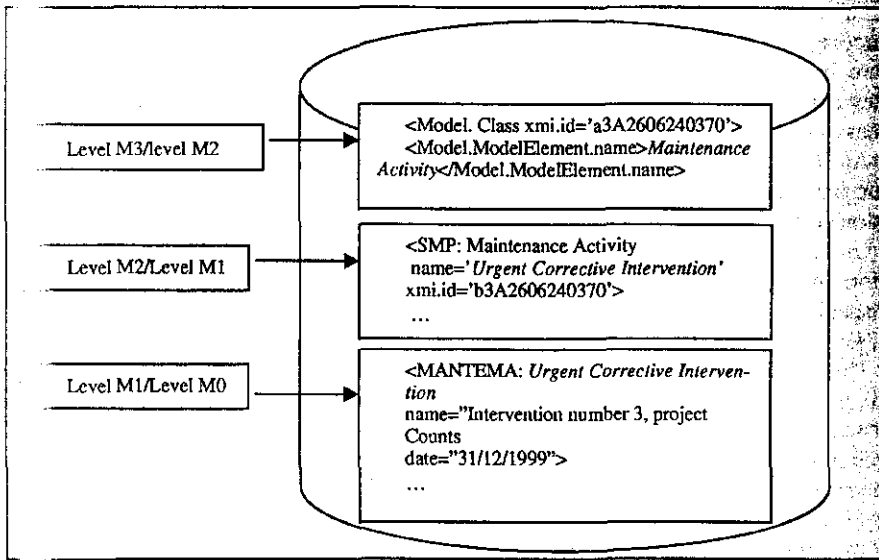


Fig. 2. Example of the XMI repository

As Fig. 2 shows, data from an Mi level are converted into metadata in the inferior level Mi-1. Thus, "Maintenance Activity", that is a piece of data in the XMI document which represents the correspondence between M3-M2, is converted into a label (metadata) in the XMI document which represents the correspondence between the levels M2 and M1.

Using XMI enables us to represent different levels of abstraction and their relationship. Knowing the relationship between these levels helps to deduce new knowledge from the previous one. Moreover, XMI permits to obtain information from other repositories, systems or information sources which support XMI [14]. And it also facilitates the interchange of information between agents since all use the same information representation.

In order to demonstrate how agents use the repositories we describe a scenario. Suppose a user consults in KM-MANTIS, when and by whom the intervention number 3, which was an Urgent Corrective Intervention, was performed. In this case the agent in charge of this type of activity should act. As Fig. 3 shows, the urgent corrective intervention agent will consult in its database the date and name of the person who carried out the intervention number 3. If a second query was performed asking the name of the maintenance activities the agent would have to consult the community repository where the information requested is. This type of information is in a higher level in the MOF standard, which corresponds to M2-M1, therefore XMI documents of this representation would be consulted in order to find all the names of maintenance activities.

To show an access to the global repository, imagine that a user needs to see the intervention number 3 request document. In this case the agent will access the global repository in order to get the XMI document that represents that request document.

A last example showing collaboration between agents is exposed. Once it is known who performed the intervention number 3 the user asks for more information about this person. The agent would contact the staff agent which will answer this request providing all information related to John Cruise such as telephone number, e-mail, his office location, etc.

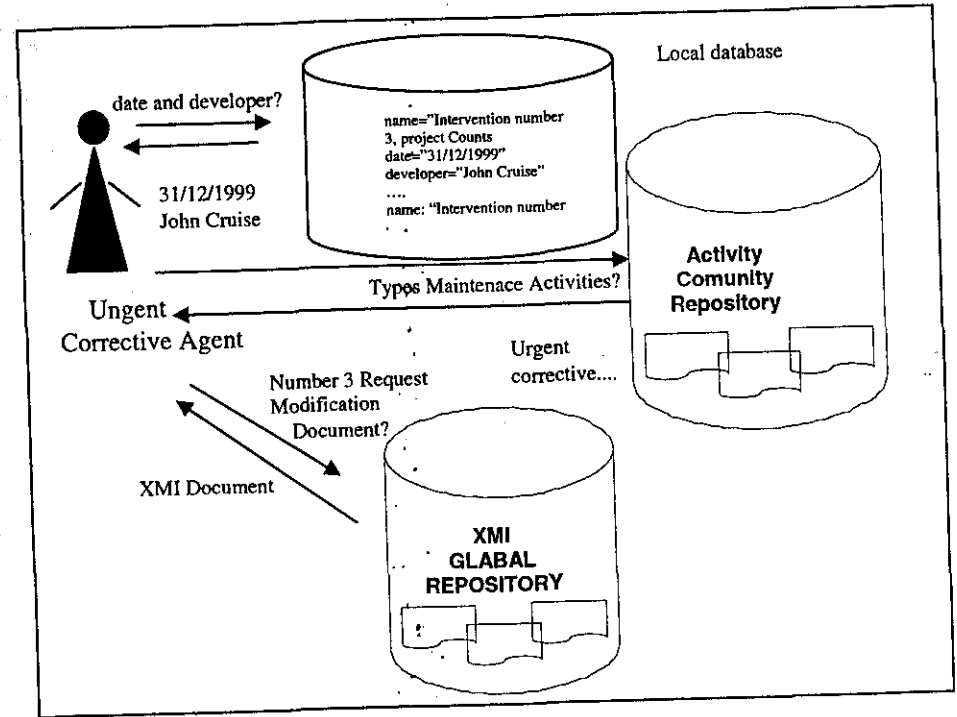


Fig. 3. Example of consults in KM-MANTIS

5 Conclusions

The complexity of the SM process demands a model of the information with a suitable level of abstraction for dealing with the different concepts that the process involves. XMI and MOF enable us to define and store a model with different levels of metadata and data. Several examples have clarified how agents utilise the levels and how diverse abstraction queries can be solved.

A future work will be the evaluation of KM-MANTIS following the DESMET [9] method. DESMET proposes a technique of evaluation divided into two types: Evaluation focused on the measurable effects of using a method or tool and the second evaluation centred on establishing criteria to measure in what degree a method or tool is suitable for an organisation taking into account its necessities and its culture.

References

1. Althoff, K-D, Birk, A., and Tautz, C. : The Experience Factory Approach: Realizing Learning from Experience in Software Development Organizations. In proceedings of the 9th German Workshop on Machine Learning (FGML 1997), University of Karlsruhe (1997), 6-8
2. Banker, R.D., Datar, S.M. and Kemerer, C.F. A model to evaluate variables impacting the productivity of software Maintenance Project. *Management Science* 37 (1), (1991), 1-18
3. Banker, R.D., Datar, S.M. and Kemerer, C.F., and Zweig, D. Software Errors and Software Maintenance Management. *Journal of Information Technology & Management*. Pirkul, N., and Varghese S.J (Eds.). Vol. 03. ISSN 1385-951X. Kluwer Academic Publishers, (2002) 35-41
4. Basili, V.R., and Rombach, H.D. The TAME project: Towards improvement-oriented software environments. *IEEE Transactions on Software Engineering*, SE-14(6):(1988): 688-773
5. Basili, V.R., Caldiera, G., and Rombach, H.D.: Experience Factory. In John J. Marciniak (Ed.) *Encyclopedia of Software Engineering*, v. 1, pp. 469-476. John Wiley & Sons.(1994)
6. Bennett K.H., and Rajlich V.T. Software Maintenance and Evolution: a Roadmap, ininkelstein, A. (Ed.), *The Future of Software Engineering*, ICSE 2000, June 4-11, Limerick, Ireland, (2000) 75-87
7. Bird, D.N., and Glass, R.L. *Measuring Software Design Quality*. Prentice Hall, Englewood Cliffs, NJ (1990)
8. Gomes, D., Campos, J.P., Silva, M.J: Versus: A Web Repository. *Workshop on Distributed Data and Structures (WDAS)* (2002)
9. Hitchenham, B. Evaluating Software Methods and Tools Using the DESMET Methodology. In *Proc. Of Empirical Assesment in Software Engineering (EASE)*, University of Keele, U.K. (1997).
10. Loeff, L.A., *Information Systems Outsourcing Decision Making: a Managerial Approach*. Hershey, PA: Idea Group Publishing, (1990)
11. Rebus, J. Framing the Knowledge Search Problem: Whom Do We Contact, and Why Do We Contact Them? *Academy of Management Best Papers Proceedings*, (2001) h1-h7
12. Nonaka, Ikujiro and Takeuchi, Hirotaka.; *The knowledge creating company*. Oxford University Press. ISBN 0-19-509269-4. (1995)
13. OMG Meta Object Facility (MOF) Specification, v. 1.3 RTF, Sep. 1999. In <http://www.omg.org>.
14. OMG XML Metadata Interchange (XMI), v. 1.1, Nov-2000
15. Goski, T.M.: *Practical Software Maintenance. Best Practices for Managing Your Investment*. Ed. John Wiley & Sons, USA, (1997)
16. Polo, M., Piattini, M., Ruiz, F. and Calero, C.: Roles in the Maintenance Process. *Software Engineering Notes*; vol 24, no. 4, 84-86. ACM., (1999)
17. Polo, M., Piattini, M., Ruiz, F. and Calero, C.: MANTEMA: A Complete Rigorous methodology for Supporting Maintenance based on the ISO/IEC 12207 Standard. *Third Euro-micro Conference on Software Maintenance and Reengineering (CSMR'99)*. Amsterdam (The Netherlands). IEEE Computer Society, (1999) 178-181
18. Polo, M., Piattini, M., and Ruiz, F.: Using a Qualitative Research Method for Building A Software Maintenance Methodology. In *Software Practice & Experience*. John Wiley and Sons. Vol. 32, No. 13, (2002): 1239-1260
19. Rumus, D.W. Mind Tools: Connecting to Groupware. *PC AI*, September/October, (1996) 2-36
20. Ruiz, F.; Piattini, M.; Polo, M.; Calero, C.; *Audit of Software Maintenance*. In "Auditing Information Systems". Idea Group Publishing, USA (2000)
21. Overman, B.G., Bedewi, N. and Morales, A. *Intelligent Agents in Software Reuse Repositories*. CIKM Workshop on Intelligent Information Agents. Baltimore, (1995)
22. Szulanski, G.: *Intra-Firm Transfer of Best Practices Project*. American Productivity and Quality Centre, Houston, Texas, (1994) 2-19
23. Vizcaino, A., Ruiz, F., Favela, J., and Piattini, M. A Multi-Agent Architecture for Knowledge Management in Software Maintenance. In *Proceedings of International Workshop on Practical Applications of Agents and Multiagent Systems (IWPAAMS'02)*, Salamanca, Spain 23-25 October, (2002) 39-52
24. Whitehead, J. "The Future of Distributed Software Development on the Internet." *Web Techniques*, Vol. 4, No. 10, October, (1999), 57-63