CoMMA Corporate Memory Management through Agents Corporate Memory Management through Agents: The CoMMA project final report

Laurent Berthelot, Marc Bourdeau, Olivier Corby, Alexandre Delteil, Rose Dieng-Kuntz, Catherine Faron Zucker, Bruno Fies, Fabien Gandon, Alain Giboin, Joachim Hackstein, et al.

CoMMA
Corporate Memory Management through Agents

IST-1999-12217

Key Action II

Action Line: ***Corporate Knowledge Management***

Corporate Memory Management through Agents: The CoMMA project final report

Document Number:	COMMA/WP1/D19
Contractual Date of Delivery to CEC	*January 31st 2002*
Actual Date of Delivery to CEC	*January 18th 2002*
Contributors (Alphabetical Order) :	BERTHELOT Laurent (INRIA)-BOURDEAU Marc (CSTB)-CORBY Olivier (INRIA)-DELTEIL Alexandre (INRIA)-DIENG-KUNTZ Rose (INRIA)-FARON-ZUCKER Catherine (INRIA)-FIES Bruno (CSTB)-GANDON Fabien (INRIA)-GIBOIN Alain (INRIA) - HACKSTEIN Joachim (T-SYSTEM NOVA)-KARP Hervé (ATOS ORIGIN)-KISS Adorjan (LIRMM) LOYSON Emmanuelle (CSTB)-PEREZ Philippe (ATOS ORIGIN) - POGGI Agostino (UNIV. of PARMA) - QUINQUETON Joel (LIRMM)-RIMASSA Giovanni (UNIV. of PARMA)-TURCI Paola (UNIV. of PARMA).
Deliverable Type	Public
Deliverable Nature	R–Report
Work package Contributing	WP1
Version	Version 1

Abstract: This document is the final report of the CoMMA project. It gives an overview of the different search activities that have been achieved through the project. First, a description of the general requirements is proposed through the definition of two scenarios. Then it shows the different technical aspects of the projects and the solution that has been proposed and implemented.

Keywords: Ontology, Agents, Multi agent system, Machine Learning Techniques, New Employee, Semantic Search Engine, Technology Monitoring, Knowledge Management, Corporate Memory, Graphical User Interface, XML, RDF, RDFS, Java.

AUTHOR COPY
Organisations in the CoMMA consortium are:
ATOS, CSTB, INRIA, LIRMM/CNRS, T-SYSTEMS, UNIVERSITY OF PARMA

Acknowledgments

AINOUCHE Salim (ATOS ORIGIN)
BASTIDE Marie (ATOS ORIGIN)
DE FOMBELLE Karim (ATOS ORIGIN)
DELAITRE Sabine (CSTB)
DE SOUSA José (INRIA)
FELT Elisabeth (T-SYSTEM NOVA)
LEIBER Thomas (T-SYSTEM NOVA)
LEYTIER Jérôme (INRIA),
KÖBKE Susanne (T-SYSTEM NOVA)
MEDINA-RAMIREZ Carolina (INRIA)
MÜLLER Prof. Dr. Jürgen (T-SYSTEM NOVA)
REGNIER Florence (CSTB)
VAEZI-NEJAD Sabnam (CSTB)
WUKASCH Jens (T-SYSTEM NOVA)
ZARLI Alain (CSTB)

Executive Summary

This document is the final report of the CoMMA IST project. Its purpose is to give a general overview of the different research, definition and design activities that were achieved during this two years project. It insists on the innovative aspects of the promoted technical solution, and mentions the different contributions that have been brought to the different research areas.

The document is composed of 4 main parts:

- The first part introduces two scenarios of corporate memory uses. Relying on these descriptions, generic functionalities of a corporate memory are proposed and a global system overview is provided.
- The second part describes the technical solution and the main components that compose this solution: the ontology, the multi-agents system, the semantic search engine, the machine learning algorithms and the user environment.
- The third part presents the major CoMMA contributions to the different technological areas addressed by the project.
- The fourth part provides all the dissemination activities (collaboration work with other projects, presentation of articles, contribution to standardisation organisations…) that have been achieved by all the partners of the project.

A conclusion emphasises the CoMMA solution, presents different open issues related to this solution and provides some general perspectives for the future.

Table Of Content

Table of figures

1 Introduction

The main objective of the project is to ***implement and trial a corporate memory management framework based on agent technology***. It addresses particularly the problem of *information retrieval* faced by employees involved in the following scenarios:

- ✓ Enhancing the insertion of new employees in the company,
- ✓ Performing processes that detect, identify and interpret technology movements and interactions for matching technology evolutions with market opportunities to disseminate among employees innovative ideas related to technology monitoring activities.

The project gives opportunities to merge emergent technologies like Multi-Agent Systems, XML and RDF formats, knowledge models based on ontologies, machine learning techniques and to apply them to fully support corporate memory concept.

The exploitable results of the projects are: *new products and tools* providing innovative knowledge management solutions, *new services for knowledge management, gain of expertise on new technologies* from transfer of technologies between academic and industrial partners.

Corporate Memory is becoming increasingly important for a correct, efficient and effective exploitation of the Know How available in an enterprise. Most companies are faced with a high and increasing complexity given by the number of services, a wide variety of customers dispersed over multiple countries, and a large number of rapidly evolving technologies. High complexity implies that a large number of distributed competencies are required and that they are changing dynamically with market, technologies and offered services.

Companies in the ICT sector ”possess” a knowledge that is really large and complex, therefore it becomes fundamental to exploit it, providing employees with effective and efficient means for communicating their knowledge and ideas for the benefit of the whole organisation. This involves the processes of representing, capturing, accumulating and transferring distributed organisational knowledge in working environment.

The communication of knowledge is particularly important for the insertion of new employees in the company and to diffuse innovative ideas among employees especially when dealing with technology monitoring activities.

In this context the project addresses the following issues:

- *Enhancing the insertion of new employees* in the company whatever the cultural extraction they are from,
- Performing processes that detect, identify and interpret technology movements and interactions for matching technology evolutions with market opportunities *to diffuse among employees' innovative ideas related to technology monitoring activities*.

This document corresponds to the CoMMA final report. The aim of this document is to present overall results of this two years project to a large audience.

The next section aims at presenting the two scenarios used for defining user requirements, then functionalities of the CoMMA system are defined. Section 2 gives therefore the user point of view. Section 3 presents the CoMMA system architecture as well as the technical approach. This section gives to readers the technical perspectives. Section 4 is dedicated to the description of the project results. All along the project life, CoMMA partners have been involved in dissemination activities through conferences, workshops and journal articles. Section 5 summarises results of the dissemination activity. The final section concludes this report by presenting CoMMA results in the context of other works and future trends as well as evaluation results from trials carried out with end users.

2 Overview of the project

2.1 Scenarios

To enable the specification and the evaluation of the CoMMA solution, two scenarios have been defined and analysed at the very beginning of the project:

- The New Employee Insertion scenario,
- The Technology Monitoring scenario.

2.1.1 The New Employee Insertion (NEI) scenario

2.1.1.1 Introduction of the scenario

This scenario aims at analysing the different solutions that can be provided by the new technologies to improve the newcomer insertion into a company. The challenge is important: to make a newcomer rapidly efficient in his new job by providing him/her with an optimal working environment that will enable him /her to start his/her productive work for the company as quick possible. This activity is usually supported by the personnel department or by a more experimented and skilled employee who plays the role of tutor.

Therefore the purpose of the scenario is to describe a general human and technical infrastructure where a newcomer gets the right information when he needs it. In this context, the CoMMA system aims at providing an architecture that will support this function.

2.1.1.2 List of the main actors of the scenario

2.1.1.2.1 Human Actors

New employee (NE): The target people of this scenario are the employees just after recruiting. Several kinds of new employee can be considered:

- The beginner who has just been graduated and starts his career in the company,
- The employee who already had experience of one or more other companies,
- The employee who has already worked at the company for a while, and who moves to a different job (i.e. advancement, functional or geographical mutation).

Tutor (or Mentor): Someone who is in charge of supporting the new employee for the first months of his new job. The tutor should be an experienced employee. The tutor is designated by the personnel department, or by the related department or service manager. The scenario envisages the optional possibility to electronically designate him, to relieve busy people to perform this quite difficult task.

Human resource manager: Any person from the personnel department in charge of the workforce management procedure.

Knowledge Manager: This human actor is in charge of defining the different user profiles templates that the system will be able to refer to. This work will be achieved with the contribution of the human resource department.

The last three human roles can be assigned to less than three different persons. In small departments/companies e.g. all three roles can be fulfilled by one person.

2.1.1.2.2 Material actor

The Corporate memory: structured set of all information related to the company and accessible by any employee through a dedicated network (intranet).

User profile: Electronic statement of the employee's personal information, preferences, topics of interest, logged interactions with the system, etc.

2.1.1.3 A global vision of the scenario

The different actors use different functionalities of the system. The tutor and human resource manager use the document annotation system to put information into the system. This information is structured as they are classified with respect to the ontology. For example an interesting web page is annotated as an instance of the *Webpage* concept from the ontology. The annotator fills all properties which are available for web pages in the ontology as far as he is aware of this information. In addition he might say that this web page is of interest for NEs.

To be able to run the NEI scenario several annotations have to be present in the corporate memory. These annotations have to be produced by the tutor/human resource manager. A short list gives an idea of what has been created. During the trials we performed:

- Persons:
  - the NEs (with their predefined profiles)
  - the personnel staff
  - the tutor
  - the CoMMA administrator
- Organisational entities:
  - an organisational chart with the position of the department
- Documents:
  - CoMMA user manual
  - Software available in the company
  - Journals available for subscription
- Interesting web pages:
  - Presentation of company
  - Link to library
  - Link to cafeteria

etc...

The new employee might take advantage from the CoMMA system at several stages of his insertion process. When he just enters the company and accesses the CoMMA system for the first time, interesting information for his arrival is presented to him automatically via the push mechanism. This is maintained via the NEs profile, where is initially stored the fact that he is interested in special annotations that typically are targeting NEs. After the NE has become more familiar with the system, he will be able to use it as the information portal for his everyday work. He can access all stored information with the query interface. In the user profile different standard queries are already available for him, but he can even start to generate his own queries which will typically become more and more complex as the user gets more familiar with the CoMMA system and with the way queries are formulated with the interface.

Users learn the benefits of semantic search as they can express much more specifically what they are searching for than if they would use standard full-text search-engines like Altavista.

Whenever the system displays some results to the user, he is able to rank the answers. This ranking is stored within the corporate memory and used for sorting the results of later queries.

2.1.2 Technological Monitoring (TM) scenario

2.1.2.1 Introduction of the scenario

The specific issue addressed by COMMA within the scope of the technological monitoring could be summarised through the following objectives:

- to diffuse technological knowledge (internal or external) through the whole company,
- to access the relevant information within the corporate memory.

The purpose of the scenario is to describe a general human and technical infrastructure that enhances the knowledge exchange through the enterprise by providing the right person with the right information at the right time.

2.1.2.2 List of the main actors of the scenario

2.1.2.2.1 Human actor

Generally speaking, any employee from the company could be a **TM user,** i.e**.** a consumer of TM (manager, technician, engineer, researcher…). Any TM user can occasionally supply the TM-dedicated base with information. In that case, the writer or the finder (from the Web, from external sources) of an information is considered as an **Author**. By extension, any TM user who needs to provide initial or further annotations to any information source is an author.

More precisely, there are other actors with specific roles in TM System:

- Thematic Groups including:
  1- A responsible manager: expert in a domain, he is in charge of setting up a network of correspondents in the various departments of the company. One of his tasks also consists of regularly writing a trend synthesis that provides state of the art in the monitored domain.
  2- A correspondent of the Research and Development Managing Board, who supervises the research policies of the company and homologous organisations, in France and abroad.
  3- Correspondents in the different departments concerned with the TM topics. Their role is to increase the TM database with information.
- Documentary Unit: in charge of the management, this unit steers the thematic groups and increases the TM-dedicated base with information.

2.1.2.2.2 Material actor

The TM system allows to feed a corporate memory with formal and informal information to answer the needs of :

- Annotating existing documents,
- Sharing information obtained by oral sources (contacts, colleagues…).

The system makes possible to annotate other entities like persons, organisations… in order to identify skills, main actors on a subject…

2.1.2.3 A global vision of the scenario

Knowledge asset annotation:

The system allows to annotate particularly two generic classes of entities:

1. Documents: for example, an author must be able to diffuse a document he has just written down or discovered on the web to relevant people in the company. These annotations are related to the librarian terminology (author, subject, date of writing…).
2. Persons, Organisations or Organisations parts (department…): these annotations are more connected to the enterprise type, the service that produces the document, the author role within the enterprise, the intended readers for the document…. They are specifically dedicated to the context in which the annotation process is achieved and rely on ontological description of the author's world (i.e. enterprise model).

The sub-system in charge to support the annotation process is called the **D**ocument **A**nnotation **S**upport **S**ystem (DASS).

Knowledge asset access:

Two solutions are developed, relying on two different approaches:

- A push approach, in which the information is directly transmitted to the KM user according to his profile: the user just needs to describe himself (profile definition or update) and then the system uses the profile description to determine what kind of document will be of major interest to him. The sub-system that implements this function is called the **P**ushed **D**iffusion of **I**nformation (PDI).

- A second approach is to consider more classical means to access the information, called “pull approach”. In this context, a KM user needs specific information related to a specific field. The system offers him intelligent mechanism to retrieve the information relying on semantic aspect. The search strategy is based on navigation through a hierarchical organisation (taxonomy), coming from general subjects, going to more specific topics (i.e. graphical representation of the enterprise ontology). The pull approach is carried out through a particular sub-system called **D**ocument **R**etrieval **S**upport (DRS).

2.1.2.4 List of the functionalities fitted to the TM case

The functionalities induced by the TM scenario are presented according to the three sets presented above (DASS, PDI and DRS).

2.1.2.4.1 The Document Annotation Support System

Regarding the TM scenario, the following functions have to be supported by the DASS:

- **Graphical navigation** through the ontology description
- **Ontology selection**
- **Graphical entry of annotation** through a graphical user-friendly format based on templates or preformatted annotations to facilitate the process. The template is based on the ontological statement selected during the navigation (cf. Graphical navigation)
- **RDF sources generator** (from the graphical user-friendly format)
- **Identification and Authentication** of the user

2.1.2.4.2 The Pushed Diffusion of Information

The following functionalities are available through the PDI:

- **Updating the user profile:** the TM user will be able to modify some features from his profile in order to have an effect upon the information he receives from the system (for instance: modify the list of topics the TM user is interested in…).
- **Notification to TM user:** The PDI notifies to the TM user the presence of information that can be of interest to him.
- **Presentation of results**: When a list of documents is available, the TM user can navigate through it and access the most important information.
- **Rating of results:** The PDI gives opportunity to the TM user to provide feedback to the system by ranking the relevancy of the PDI outputs.
- **Matching document annotations to user profile:** The PDI is able to compare the annotations to the user profile in order to anticipate TM user expectation and to supply him with relevant information as expected by him.

2.1.2.4.3 The Document Retrieval System

The following functions are provided through the DRS:

- Ontology Navigation
- Building searching query
- Graphical representation of the query
- Refining the query
- Matching query + user profile to documents annotations: the DRS compares the query contents with the different annotations and estimates the relevancy of documents
- Estimation of the relevancy of a document: (cf. Matching query above)
- Presentation of results
- Rating of results

2.2 System functionalities

2.2.1 Figuring out the system genericity

By analysing the previous two scenarios it appears that both have some common aspects that must be figured out to build a generic system. Both scenarios mainly address two specific issues:

An annotation activity that consists of exploiting a framework (i.e. the ontology) to formally represent the content of a document: In the New Employee scenario a privileged employee who could be either the tutor of newcomers or someone working at the company personnel department, has to achieve this task that will make a document accessible by the newcomers. This can concern documents related to the general enterprise life or to the newcomer specific job. In the technology monitoring scenario, dedicated people (information officer, area referents, etc.) can diffuse very innovative information through annotations, using the same process.

A search activity that consists of requesting information from the system (i.e. the corporate memory): As for annotating the query process has to refer to the ontology. In the new employee scenario, a newcomer can address specific requests to the enterprise corporate memory to get very precise information about many subjects (how to have lunch? How to organise professional travels? Who is who? Etc.). In the technology-monitoring scenario, the importance of this function is obvious.

A push mode of information delivering: This function has been proposed for the technology-monitoring scenario, in order to offer a very efficient proactive mechanism to diffuse the documents at the enterprise level taking into account fields of interest. In the new employee case the push mode can offer a flexible mechanism to provide interesting information to newcomers as soon as he is identified by the system. This mode can be used by the tutor to make his/her newcomer access some important documents related to his/her new job.

User profile management: In both scenarios, the user profile will play an important role to memorise important information about the employee.

2.2.2 The whole system

The following use case presents a global vision of the implemented CoMMA system:

(see next page)

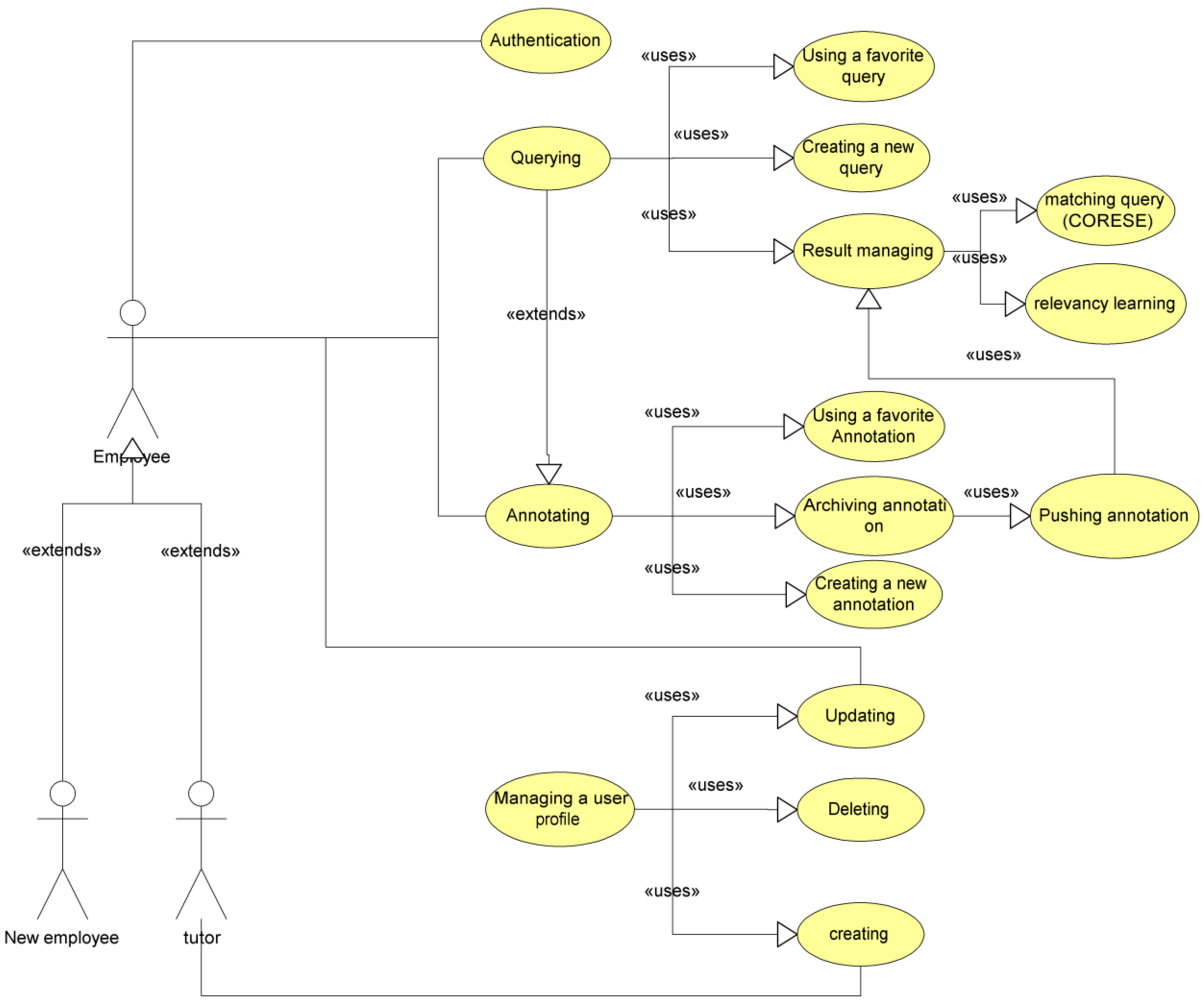

Figure 1 Use case of the existing CoMMA system

The previous use case diagram shows four families of functions:

- Function dedicated to authentication.
- Functions dedicated to annotation activities,
- Functions dedicated to query activities,
- Functions dedicated to user profile management,

Because of the tightness of the project schedule, some functions have not been implemented in the current version. They will be presented in the last section of this chapter.

2.2.3 Authentication

A classical login/password process is used to identify the employee. This process enables to authenticate the logging people as an employee and to load his user profile, particularly the potential push information that can have been loaded into.

2.2.4 Annotation functionalities

The necessity to model document content through formatted annotations requires specific functions that will enable the user:

To create new annotations in a user-friendly way: as the annotation action can be conceptually seen as an instantiation of concepts or properties defined into an ontology, it implies that the user must access this ontology in order to select the required concepts or properties. However the framework provided by the ontology must remain flexible enough:

- Not to oblige the user to know by heart the list of concepts and properties,
- To let the user be free to annotate the document taking into account his/her particular context.

The GUI will play a very important role to provide a flexible process in the compelled ontology framework.

To archive the created annotations in annotation bases: the possibility to have a distributed organisation of these bases imply that a specific algorithm is in charge of selecting the most relevant base where the new annotation will be saved.

To use ready-to-use annotation templates (called favourite annotations): The user has the possibility to access pre-formatted annotations. These templates are saved in the user profile. The user starts with a predefined set of favourite annotations depending on his profile. But he has the possibility to save a new annotation in his user profile as a favourite one.

To trigger the push mode: Each time a new annotation is added to a document, it must be tested if this document can be interesting for other employees of the company. A specific process starts up to compare the new annotation with all the user profiles and to check if some users can be interested in the new document. If a user profile matches, some part of the document annotations can be loaded in. then the pushed information will be sent to him at login time.

The following process illustrates the way the annotation functions are related together:

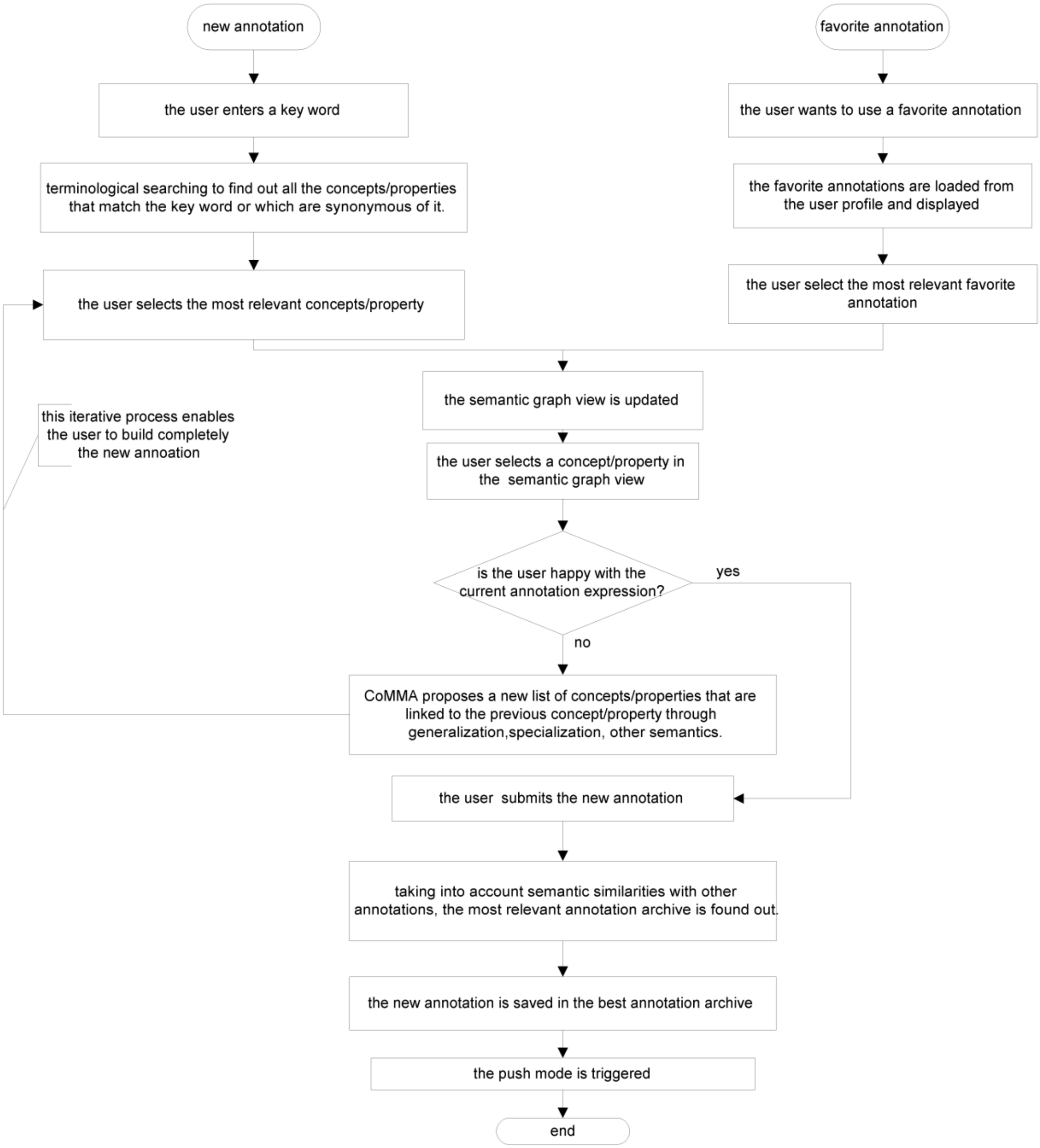

Figure 2 The whole annotation process

2.2.5 Query functionalities

A query can be seen somewhat as an annotation that presents specific degrees of freedom (as particular variables), which the matching process will rely on. That is the reason why this process is defined as an extension of the annotation process in the previous use case diagram (cf. Figure 1 Use case of the existing CoMMA system). Consequently, the GUI for querying is very similar to the GUI for annotating. Particularly, the interactive process to iteratively build a query is exactly the same than the one used to build an annotation (cf. Figure 2 The whole annotation process). Just the interpretation by the system is different.

The query process has some specificities:

- After being submitted, a new query is not archived but it is matched with annotations present in different annotation bases through a semantic search engine. Results of this process are submitted back to the user and a machine learning algorithm is used to improve the relevancy of the results, taking into account the user's areas of interest. Moreover the user can provide feedbacks about the relevancy of the different documents. These feedbacks are the input data of the machine learning algorithm.
- A smart algorithm has been implemented to take into account the fact that an annotation can be split off in different annotation archives. This algorithm relies on a semantic metric.
- No push mode is triggered by the query submission in the current version. It can be envisaged in a next version that a query submission informs some area expert that someone is interested in document related to this area.

2.2.6 User profile management

The current version of the CoMMA system offers particular functions to manage user profiles. A user profile can be envisaged as a specific favourite annotation, which will play a particular role in the CoMMA system. For this reason, the same GUI environment than annotation can be used to visualise, modify, and browse a user profile.

The current version of the CoMMA system offers some usual functions:

- Loading and displaying a user profile,
- Updating a user profile,
- Saving a user profile.

The creation of a new user profile has not been implemented yet in the current version. However, this function could be easily added in a next version for some users who are identified as tutor, or in charge of new employee management activity.

2.2.7 Further functions

The current version of the CoMMA system has been limited to a set of core functionalities that enable to support corporate memory management. However, some new functions or improvement of existing ones can be already envisaged for the next versions of CoMMA.

Creation of new user profiles by privileged users (who can be identified as "tutor" or working at the personnel department): This function has been already mentioned in a previous section.

Adding the push mode for query: Each time some one is looking for some documents, if this person wishes, the query can be pushed to some well identified area experts that can provide to this person more precise information. The set of experts targeted by this process depends on the query content.

Improvement of the push mode: Just one type of implicit query is dealt with by the current version. This query enables to match people interested in some topics, to documents that concern these topics. Other queries of this kind can be proposed (relating document to people's activities,…). Moreover the addition of a new user profile should trigger the push mode as well. This improvement will enable to make a newcomer access all interesting documents already present in the corporate memory. In order to not saturate the newcomer at the first login, a scheduled process can be envisaged.

Management of the ontology: At the very beginning of the project, it has been decided to focus on the corporate memory use and to not address the administration aspect of the corporate memory, particularly the ontology management. However, a new version of the CoMMA system should address this specific issue, particularly:

- Modification of the ontology (addition of new concepts, relations, terms, definitions),
- Management of feedbacks from ontology users,
- Consistency management between past annotations and new ontology version…

2.3 System overview

The CoMMA project aimed at implementing a corporate memory management framework based on several emerging technologies: agents, ontology engineering and knowledge modelling, XML, information retrieval and machine learning techniques. The project intended to implement this system in the context of two scenarios:

- assisting the insertion of new employees in the company
- supporting the technology monitoring process.

The technical choices made in CoMMA were mainly motivated by three observations:

(1) *The corporate memory is, by nature, an heterogeneous and distributed information landscape.* The corporate memories are now facing the same problem of information retrieval and information overload than the Web. To quote John Naisbitt, “we are drowning in information but starved of knowledge”. The initiative of a semantic Web is a promising approach where the semantics of documents is made explicit through metadata and annotations to guide later exploitation, relying on ontologies. XML being likely to become an industry standard for exchanging data, we use it to build and structure the corporate memory. We are especially interested in RDF(S)[1] defined in XML, that allows us to semantically annotate the resources of the memory. The corporate memory is then studied as a "corporate semantic Web".

(2) *The population of the users of the memory is, by nature, heterogeneous and distributed in the corporation*. Agents will be in charge of interfacing users with the system. Adaptation and customisation are a keystone here and CoMMA relies on machine learning techniques in order to make agents adaptive to users and context. This goes from basic customisation to user's habits and learning of preferences, up to push technologies based on interest groups.

(3) *The tasks, as a whole, to be performed on the corporate memory are, by nature, distributed and heterogeneous*. The corporate memory is distributed and heterogeneous. The user population is distributed and heterogeneous. Therefore, it seems interesting that the interface between these two worlds be itself heterogeneous and distributed. Programming progresses were achieved through higher abstraction enabling us to model systems more and more complex; multi-agents systems (MAS) have the potential to become a new stage in abstraction that can be used to understand, to model and to develop a complex distributed systems. The MAS paradigm appears to be suited for the deployment of a software architecture above the distributed information landscape of the corporate memory. On the one hand, individual agents locally adapt to users and resources they are dedicated to. On the other hand, thanks to co-operating software agents distributed over the network we can capitalise an integrated and global view of the corporate memory.

Figure 3 depicts the overall solution adopted in CoMMA to address the application scenarios.

[1] Resource Description Framework (RDF) and RDF Schema

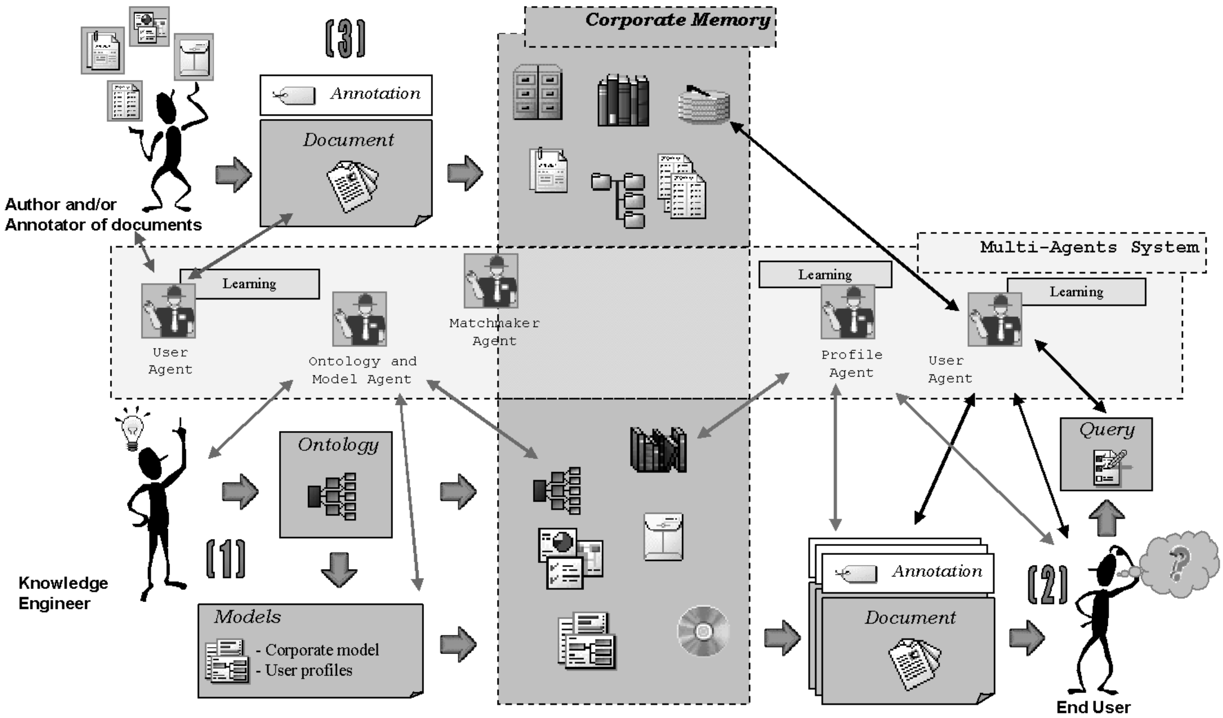

Figure 3 Overview of CoMMA

3 Description of the solution

3.1 Technical Approach

3.1.1 Integration of different technical components

To implement the corporate memory management framework, the project has integrated in a seamless way different technologies:

1) In the framework of knowledge management, retrieval of information can be guided by **ontologies or knowledge models**: corporate document models, domain ontologies, enterprise models, models of the potential users in the company. These knowledge models can index the documents, in order to facilitate their later search by users.

2) The proposed solution is based on a **multi-agents architecture of several co-operating agents**, having adaptive capabilities to the user and to his/her context, and supporting retrieval of the relevant information in the organisational memory. These agents can (a) communicate with the others to delegate tasks, (b) make elementary reasoning and decisions, and help to choose between several documents. They also have inference mechanisms that enable to extend the ontology.

3) The documents are annotated by **meta-information expressed in RDF format**. RDF is a specific language dedicated to metadata representation, in XML galaxy. The project offers reasoning mechanisms on this meta-information in order to improve the document retrieval.

4) The project exploits **machine learning techniques** in order to provide the agents with adaptive capabilities to their users and the context. The project focused particularly on the improvement of search result relevancy taking into account document and user features.

5) The necessity for end-users to refer to an ontology in order to annotate documents or to search them, raises particular tricky **Graphical User Interface** issues. The general usability of the whole system depends on the GUI capacity of tackling this problem. Collaborative work with ergonomists was required to achieve the CoMMA GUI.

These different technologies are integrated through a demonstrator implementing the results of our research.

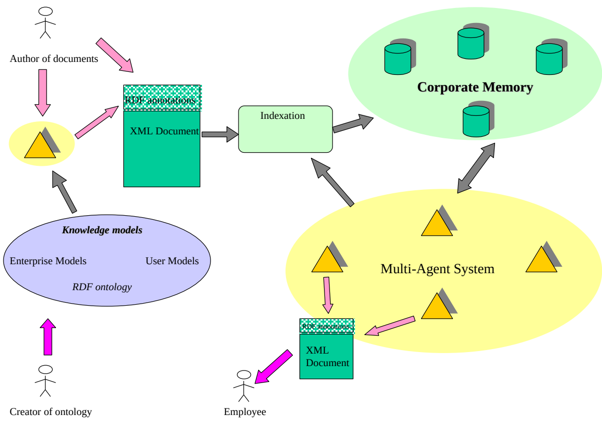

Figure 4 Schematic view of the CoMMA solution for Knowledge Management

This demonstrator has been validated from trials carried out by the industrial partners playing the role of end users. It has also been validated by a group of industrials (Industrial Support Group) from different sectors of activity.

The assessment of the demonstrator was based on two scenarios of particular interest for the industrial partners and allowing to demonstrate the benefits of the CoMMA solution:

The insertion of a new employee scenario,

The technology monitoring scenario.

Both scenarios have been presented at the beginning of this document (c.f. chapter 2).

3.1.2 The project development-cycle

The following figure illustrates the project life-cycle that has been adopted to implement the CoMMA system. Two iterations of this cycle have been achieved during the project, but this cycle could be followed to carry on the development of the next version of the CoMMA system and can be seen as a base of a general corporate memory development methodology.

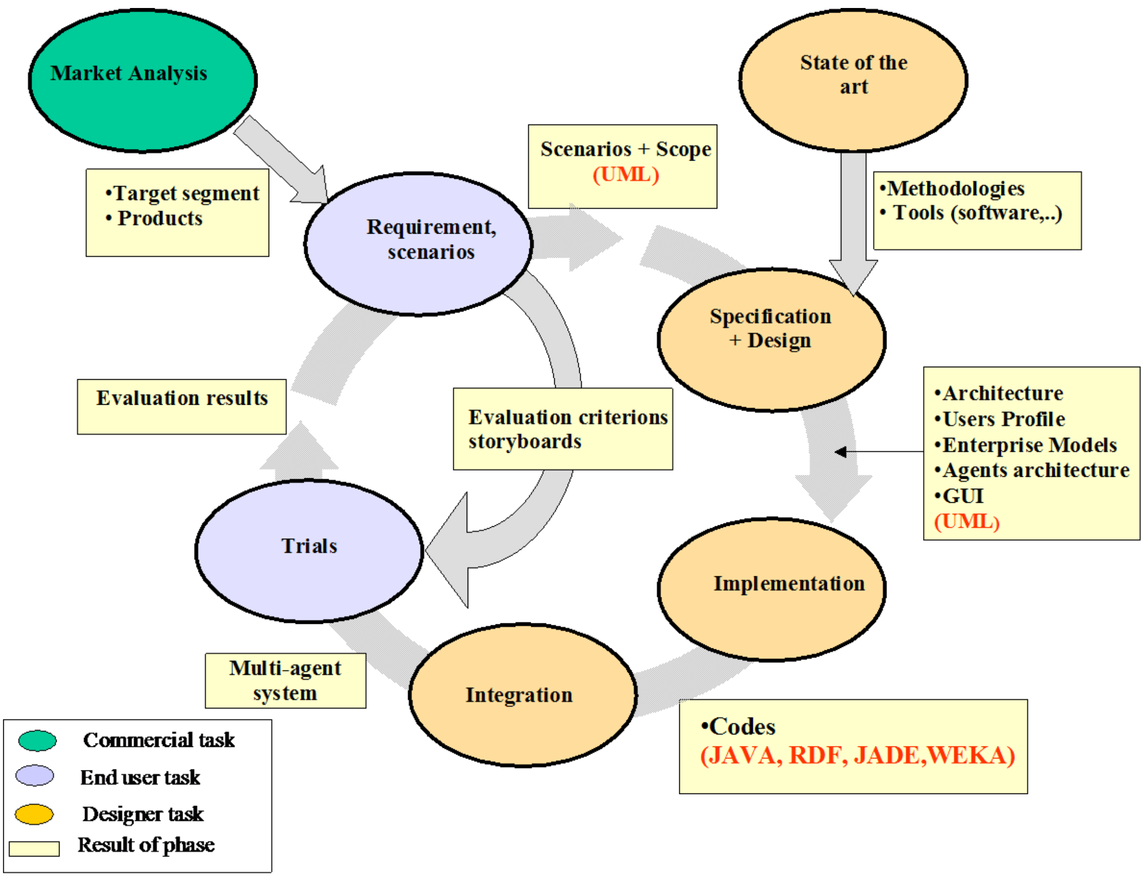

Figure 5 The CoMMA development cycle

3.2 Multi-agent architecture

A multi-agent system (MAS) usually consists of a set of interacting agents in some shared environment, capable of independent operations aimed at meeting their local goals and cooperative actions contributing jointly to the global goal shared across the community.

CoMMA MAS uses agents for wrapping information repositories (i.e., the corporate memory), for information retrieval, for enhancing scaling, flexibility and extensibility of the corporate memory and for adapting the system interface to the users. Several multi-agent systems for information retrieval have been realised in the last years. One of the points that makes CoMMA system different from most multi-agent information systems is that agents are not only used for the retrieval of information, but also for the insertion of new information in the corporate memory.

The tasks carried out by CoMMA system are performed through the cooperation among different kinds of agents that can be divided in four sub-societies: document and annotation management, ontology (Enterprise and User Models) management, user management and agent interconnection and matchmaking.

3.2.1 Dedicated Sub-societies

The agents from the document-dedicated sub-society are concerned with the exploitation of the documents and annotations composing the corporate memory; they will search and retrieve the references matching the query of the user with the help of the ontological agents. A hierarchical organisation for the document sub-society is appropriate because it splits the task of maintaining document repositories from the task of intelligent interface towards the other agents of the system. The agents of the document dedicated sub-society play two roles:

- ✓ The Annotation Archivist (AA): has the duty to store and retrieve information to/from the CoMMA corporate memory.
- ✓ The Annotation Mediator (AM): has the duty to manage queries about the CoMMA corporate memory content. Therefore, it is the interface between the AAs and the other agents of the system making transparent to the other agents the implementation of the corporate memory (i.e., the number of document repositories and of AAs managing them), the policies to store documents in the different repositories, the need of retrieving documents from different repositories.

The agents of the ontology sub-society have to maintain CoMMA ontology repository, composed of RDF schema forms, and CoMMA enterprise model repository, composed of RDF annotations. Moreover, they deliver information about CoMMA ontology and enterprise model to the agents needing it. A replicated organisation for the ontology sub-society is conceivable because the ontology should be rather stable; the ontological commitment is centralised and the image is updated and propagated over the society. The agents of the ontology sub-society play two roles:

- ✓ The Ontology Archivist: has the duty to store and retrieve information to/from the CoMMA ontology repository.
- ✓ The Enterprise Model Archivist: has the duty to store and retrieve information to/from the CoMMA enterprise model repository.

The agents of the user-dedicated sub-society are concerned with the interface, the monitoring, the assistance and the adaptation to the user. Moreover, they have to maintain the user profile repository and distribute information about user profile to the agents needing it. The agents of the user sub-society play four roles:

- ✓ Interface Controller (IC): is in charge of monitoring the user interface and the interaction of the user with the system. The IC agent is the system front end, working in close collaboration with the user. Moreover, it is the only agent in the system that is not running persistently, with its lifetime limited to a single session (IC starts when the user logs into the system and shuts itself down when the user logs out). When IC starts up, it uses the yellow pages services provided by the Directory Facilitator agent to get acquainted with all necessary agents.

- ✓ User Profile Archivist (UPA): has the duty to store and retrieve user profiles to/from CoMMA user profile repository, maintaining the profile of the users belonging to a specific sub-net of the enterprise intranet.
- ✓ User Profile Manager (UPM): is in charge of updating and exploiting the user's profile when the user is logged on to the system.
- ✓ The User Profile Processor: has the duty to perform proactive queries on the corporate memory on the basis of user profiles and information about new documents. The UPP has also the responsibility for each user leaving within its *jurisdiction domain*, to handle Annotation Mediator notifications for new/re-edit annotations and to start the push mode.

The agents of the connection-dedicated sub-society are in charge of the matchmaking of the other agents based upon their respective needs. The agents of the connection sub-society play two roles defined in a FIPA compliant platform:

- ✓ The Agent Management System: has the duty to maintain system white pages where agents may register themselves and ask for addresses of other agents.
- ✓ The Directory Facilitator: has the duty to maintain system yellow pages where agents may register themselves and their capabilities and ask for services of other agents. The agents that play this role can be organised into graphs, with the possibility of searches of arbitrary depth.

3.2.2 Agent Architecture

We decided to develop the CoMMA MAS using JADE. JADE (Java Agent DEvelopment framework) is a software framework to aid the development of agent applications in compliance with the FIPA specifications for interoperable intelligent multi-agent systems. JADE is an Open Source project, and the complete system can be downloaded from JADE Home Page (http://jade.cselt.it).

JADE offers an agent architecture implementing an agent as an *active object*, therefore, this architecture is suitable to realise autonomous agents because active objects are able to start actions without the need of external events triggering the actions, that is, active objects can be proactive. Moreover, JADE agent architecture implements agents using only a Java thread, but able to manage different conversations in parallel. It is possible because JADE agent architecture implements a *'cooperative scheduling on top of the stack*', in which all agent tasks are run from a single stack frame without context saving (*on top of the stack*) and a task continues to run until it returns from its main function and cannot be pre-empted by other tasks (*cooperative schedulin*g). Of course ordinary pre-emption is still active between different agent threads: cooperative scheduling is strictly an intra-agent policy. Agent tasks are encoded into *behaviou*rs: each agent holds a collection of behaviours, which are scheduled and executed to perform agent duties. Behaviours represent the logical threads of a software agent implementation.

3.2.3 Configuration and Deployment

The architecture of a multi-agent system is a structure that portrays the different kinds of agents of its societies and the relationships among them. A configuration is an instantiation of an architecture with a chosen arrangement and an appropriate number of agents. One frozen architecture can lead to several configurations. The configuration is tightly linked to the topology and the context of the place where it is rolled out. The architecture was designed so that the possible configurations cover the different corporate organisational layouts foreseeable in the context of the CoMMA project. We left a manoeuvre margin by having not one configuration per potential corporate instance but a set of possible configurations to be able to adapt to actual corporate policies that will be enforced on-site, while optimising CPU workload, distributed data flow and, in general, system performance. This is an area where the adoption of agent technology can bring significant benefits, because multi-agent systems can be made extremely modular.

CoMMA system explicitly targets corporate networks, and the network topology of a corporate intranet is deeply influenced by the company organisational structure, so that agent acquaintance and “reachability” will vary, depending upon specific company policies. In fact, the deployment of the CoMMA MAS will be driven by an instance of the enterprise model, describing the actual structure of the company where CoMMA is installed. Moreover, the deployment will also be driven by the network topology technical environment and its constraints (such as the topology characteristics, the network maps and data rates, data server location, gateways, firewalls, etc) and by interest areas, where are the stakeholders (users, system managers, content providers…). This means that CoMMA configuration will not just use a single dimension of the enterprise model. Rather, multiple *ontology views* will need to be extracted from the enterprise model instance representing the current company where a CoMMA system is installed.

Due to the highly flexible distributed component infrastructure provided by the JADE environment, the system can be incrementally deployed; new agents or platforms can be added or removed dynamically, without the need to stop normal system operations. The possible network structural changes due to corporate reorganisations can be dealt with rather seamlessly.

3.3 Ontology

We proposed a method to build an ontology and applied it to obtain the Ontology O'CoMMA which one of the results of the project. The methodology is in three stages:

Scenario analysis and Data collection: Scenarios are textual descriptions of the organisational activities and interactions concerning the intended application. They were used for data-collection together with semi-structured interviews, workplace observation and document analysis. This last technique can be coupled with natural language processing tools for scaling-up the approach. We also studied and partially reused existing ontologies whenever possible. Other non company-specific sources or standards helped us structure upper parts of the ontology or list leaves of some precise specialised area.

Terminological stage : The terms denoting notions appearing relevant for the application scenarios are collected, analysed and organised in a set of informal tables forming a lexicon on which the ontology will be built. The synonyms and ambiguous terms are spotted and marked as such. Definitions in natural language are proposed, discussed and refined especially to eliminate fuzziness, circular definitions and incoherence.

Structuring the ontology : Combining bottom-up, top-down and middle-out approaches as three complementary perspectives of a complete methodology, the obtained concepts are iteratively structured in a taxonomy. The initial tables evolve from a semi-informal representation (terminological tables of terms & notions) towards semi-formal representation (taxonomic links, signatures of relations) until each notion has a unique formal identifier and the taxonomic links are sufficiently explicit to be translated in RDFS by scripts. Previous informal views as well as new richer views of the ontology are produced using XSLT style sheets.

As a result, O'CoMMA is an ontology containing 470 concepts organised in a taxonomy with a maximal depth of 13 subsumption hops (using multi-inheritance), 79 relations and 630 terms to label these primitives. In the ontology three layers can be seen: (1) a very general top that roughly looks like other top-ontologies; (2) very large and ever growing middle layer divided in two main branches: one generic to corporate memory domain (documents, organisation, people...) and one dedicated to the topics of the application domain (telecom: wireless technologies, network technologies...); (3) an extension layer, scenario and company specific, with complex concepts (Trend analysis report, New Employee Route Card...).

Concepts are formalised as RDFS classes, relations and attributes are formalised as RDFS properties. Instances of these classes and properties are made to formulate annotations. As multi-instantiation is permitted in RDF, one resource may be an instance of several classes. Terms are formalised as RDFS labels of classes and properties and are independent from the internal unique system label. Likewise the definitions in natural language are captured as RDFS comments. The ability to specify the natural language used, enables us to have multilingual ontologies. A notion with several terms linked to it is characteristic from the synonymy of these terms. A term with several notions linked to it is characteristic of its ambiguity.

Using XSLT style sheets we kept the informal views:

(a) a style sheet recreates the initial terminological table representing a sort of lexicon of the memory;

(b) two style sheets recreate the tables of concepts and properties;

(c) five style sheets propose to navigate and search at the conceptual or terminological levels, search for concepts or relations linked to a term, navigation in the taxonomy, search for relations having a signature compatible with a given concept. The user can also ask for the listing of the instances of a notion ; a sample of these instances can play the role of examples to ease understanding of a notion;

(d) a style sheet filters the ontology with the user profile to propose preferred entrance point in the ontology to start browsing;

(e) two other style sheets propose a view as an indented tree of concepts or relations.

The XML choice enables us to base our system on a standard that benefits from all the web-based technologies for networking, display and navigation, and this is an asset for the integration to a corporate intranet.

3.4 Information retrieval

CoMMA offers several techniques for improving information retrieval from a corporate memory materialised in documents semantically annotated using the concepts of an ontology. In CoMMA, the semantic annotations are expressed in RDF and the ontology in RDF Schema. In order to palliate RDF(S) lack of expressiveness, we propose an extension of RDF(S), called DRDFS, enabling to express class definitions, property definitions, axioms and more generally contextual knowledge.

Improvement of information retrieval can stem from the inferences on the annotation base so as to infer implicit knowledge: these inferences can rely on the reflexivity, symmetry, transitivity of the properties; these inferences can also be performed by an inference engine making deductions from a rule base aimed at completing the annotation base. Therefore, CoMMA relies on extensions of the CORESE engine with new inference capabilities such as exploitation of reflexivity, symmetry, transitivity, inverse properties and exploitation of a rule-based inference engine.

Moreover, an automatic categorisation of the RDF-annotated documents can also help to improve the further information retrieval, by automatic building of a hierarchy of documents.

The CoMMA system also includes techniques for distribution of the annotation base and of the queries: such techniques are also useful for improving the information retrieval from a distributed memory. This distribution of the annotation base among the different Annotation Archivists can help improve further information retrieval: the user's query can then be distributed among different Annotation Archivists according to the base fragmentation and according to the statistics of the different archivists.

Last, in addition to pull mode where information is retrieved after an explicit query of the user, the CoMMA system also enables to push information towards the user according to his/her centers of interest.

3.4.1 Distributed Annotation Management Algorithm

In the Annotation sub-society, the Annotation Mediator (AM) is in charge of handling the annotations distributed over Annotation Archivists (AAs). The stake was to find mechanisms to decide where to store newly submitted annotations and how to distribute a query in order not to miss answers just because the needed information are split over several AAs.

In order to determine which AAs should be involved during the solving of a query or to which one should be given an annotation, we needed a way to compare the specialisation of the archive content of the AAs .

To allocate a newly posted annotation, we designed and implemented a pseudo-distance to measure how close, semantically, an annotation is from the types of concepts and relations present in the archive of a candidate AA. Using this distance, we implemented a contract-net protocol to decide which one of the AAs should win a bid.

When a user expresses a query, its solving may involve several bases distributed over several AAs. The result is a merging of partial results found on the concerned AAs. To determine if and when an AA should participate to the solving of a query, we implemented a calculation of the overlap between the list of types present in the bases of the candidate AAs and the list of types at play in the query. Using the description of the overlaps it received from the AAs, the AM is able to identify at each step of the decomposition algorithm and for each intermediate sub-query it generates, the AAs to be consulted. Using modules of CORESE, we implemented the decomposition of the query in four stages:

1. The query is pre-processed and prepared for distributed solving.
2. The constraints of the query are solved starting from the deepest
ones.
3. The questions of the query are answered.
4. The coreferences are addressed locally in the AM.

3.4.2 CORESE extensions

The RDF semantic search engine of CoMMA is the CORESE engine; it is based on conceptual graphs (CG). More precisely, the CORESE platform implements an RDF/RDFS processor based on CG. It enables the processing of RDF Schemas and RDF statements within the CG formalism. The graph matching algorithm, called projection, enables to retrieve RDF statements according to a query and hence implements a search engine. The projection operation takes advantage of the class and property type hierarchies. When querying on a given resource type (resp. property type), the projection finds instances of the given type as well as instances of the specializations of that type.

In this section, we present the extensions that have been designed for CoMMA.

The principle of information retrieval based on annotations is to describe resources by means of RDF annotations and to search through these annotations. Hence, a query is evaluated against each annotation of the base. In some cases, it is interesting to define knowledge that is true for all annotations and to join this knowledge with each annotation during query processing. CORESE has been augmented with a *global graph* where global knowledge can be defined.

In some cases, it is interesting to declare that a property is transitive and to let the system compute all tuples of the relation. For example, the `relative` property is transitive. If A is relative of B and B is relative of C, then A is relative of C. We have extended the RDF Schema formalism with three meta-properties called *transitive*, *symmetric* and *reflexive* with boolean value `true` or `false` and `rdf:Property` as domain. These properties are defined in the corese namespace. It is also interesting to specify the inverse property of a given property, to infer the range and the domain of the inverse property in the schema, and to generate the inverse property value in annotations. We have defined an extension of the RDF Schema to provide this definition. For example, CORESE can generate the *hasMember* inverse property from the *isMemberOf* property.

The CORESE query language is RDF plus variables and operators for information retrieval. The query processor relies on the conceptual graph projection operation which finds a specialisation (according to the RDF Schema) of a query pattern in a target graph. The query language has been augmented with type operators which enables to specify more precisely the type of the requested resources. For example, one can now look for documents *except* technical reports or for exact instances of *article* and not from subclasses of *article*.

The last extension that has been designed is the inference engine of CORESE. An ontology can contain axioms and rules that enable to deduce new knowledge from existing one. But RDF Schema does not provide such a mechanism. Hence we have proposed an RDF Rule extension to RDF(S).

CORESE now has an inference engine based on forward chaining production rules. The rules apply on Conceptual Graphs and can enrich a graph according to the conclusions. For example, the rule below states that if a person ?p is member of team ?t which is lead by ?m, then ?m manages ?p :

```
IF   [Person:?p]-(memberOf)-[Team:?t]
```

```
AND  [Team:?t]-(leadBy)-[Manager:?m]
THEN [Manager:?m]-(manage)-[Person:?p]
```

The rules are applied once the annotations are loaded and before query processing occurs. Hence, annotations are augmented by rules. An RDF/XML syntax was designed for the rules which are interpreted as conceptual graph rules. The rule above would be written as :

```
<cos:rule>
      <cos:if>
            <c:Person rdf:about='?p'>
                  <c:memberOf>
                        <c:Team rdf:about='?t'/>
                  </c:memberOf>
            </c:Person>
      </cos:if>
      <cos:if>
            <c:Person rdf:about='?m' c:headOf='?t'/>
      </cos:if>
      <cos:then>
            <c:Person rdf:about='?m' c:manage='?p'/>
      </cos:then>
</cos:rule>
```

3.4.3 Extensions of RDFS

When compared to object-oriented knowledge representation languages, description logics, or conceptual graphs, RDF(S) does not enable to define classes or properties nor represent axioms. Now such axiomatic knowledge is the key to discover implicit knowledge in Web page annotations or as corporate memory annotations. As a result, in the framework of the CoMMA project, the RDFS expressiveness appeared too limited to represent the ontological knowledge required for intelligent Information Retrieval in a corporate memory. We thus propose an extension of RDF(S) with class, property and axiom definitions. We call it DRDFS for Defined Resource Description Framework.

DRDFS more generally enables to express contextual knowledge on the Web. The RDF philosophy consists of letting anybody free to declare anything about any resource. Therefore the knowledge of who and in which context an annotation has been stated is crucial. DRDF(S) enables to assign a context to any cluster of annotations. The representation of axioms or class and property definitions is just a particular use of DRDFS contexts. DRDFS is a refinement of the core RDFS which remains totally compliant with the RDF triple model. More precisely, DRDFS is an RDF Schema extending RDFS with new primitives. This extension of RDFS is based on Conceptual Graphs features. DRDFS is built upon the mapping that has been established between RDF(S) and the Simple Conceptual Graph model. A DRDFS context corresponds to a conceptual graph; a DRDFS class (resp. property) definition corresponds to a concept (relation) type definition; a DRDFS axiom correspond to a graph rule.

We hope that DRDF(S) will contribute to the ongoing work of the W3C committee for improving RDF(S) and meet the needs of the Semantic Web community.

3.4.4 Learning concepts from RDF annotations

We offer a method for classifying documents and capturing knowledge by learning concepts from the RDF annotations of the documents. We adopt an approach of concept formation: concept formation or conceptual clustering aims at building hierarchies to cluster similar objects and classify object descriptions. However in these approaches, a single particular hierarchy of classes is built, the best according to a certain criterion. Our approach of concept formation is slightly different since it aims at systematically generating a class for each possible set of objects.

More precisely, it consists of extracting descriptions of the documents from the RDF graph gathering all the document annotations and then forming concepts by generating the most specific generalisation of these descriptions for each possible set of documents. In order to deal with the intrinsic exponential complexity of such a task, the concept hierarchy is incrementally built by increasing at each step the maximum size of the RDF document descriptions that we extract from the whole RDF graph gathering all the annotations.

Moreover, an efficient algorithm enables the building of the concept hierarchy at each step of the process. In particular, the so learned hierarchical structure will be used to organise the documents of the corporate memory, to improve the Information Retrieval process on the corporate memory, and to provide feedback to the ontology designer to refine and enrich the domain ontology with class definitions.

3.5 Machine learning

Machine learning in the CoMMA system is devoted to user adaptation. In the designed scenarios, the system records the opinion of the users on various documents, expressed as a rating or ranking, then it extracts characteristic information of the documents and users who participated in the ranking to create generalisations to be applied in future document searches.

The semantic search engine, that is the core of the CoMMA project, has the ultimate goal to help the user retrieve exactly the information he is looking for, without unexpected answers in the query results. Therefore, at a first guess, there should be no need for a *relevance measure*, in the sense of classical keyword-based search engines, to filter out unwanted results.

Nevertheless, even with semantic search, the user can build generic queries, that return large amounts of information. For example, a query like: "Return all the ***documents*** that were ***created*** by a ***person*** whose ***surname*** is Smith". (Emphasized words in the query are concepts that may be in the ontology which define the semantics) may return a lot of documents, and it is impossible to say that some results satisfy better the search criteria than others, so that we could sort the list of answers.

It is obvious that objective relevance measures (that do not take into account the user's context), like the keyword frequency in keyword-based search engines, will not be of much help in our case. In this situation, it can be imagined that either an interactive negotiation with the user will take place to refine the set of answers before sending them back, or a subjective relevance measure may be used to rank the results.

The use of subjective relevance measures, that consider the user's *context* to rank documents, has been experimented widely also in the domain of classical search-engines. In these cases, the context is characterised usually by one specific attribute or a rather focused view. In some examples of state-of-the-art-systems, the context is defined by other reference documents to which the query results will be compared, collected by the user himself or communities of interest (personal bookmarks, recommending systems, etc.)

In the CoMMA system, the advantage of having semantic annotations and the ontology offering basic concepts for enterprise and user models, became valuable resources for defining a much wider search context. In our approach we have collected a sample set of concepts from the CoMMA ontology, that can describe the search context.

The machine learning component in the CoMMA system was designed to function both in the "pull" and "push" information retrieval scenarios. In the DRS (pull) subsystem, it acts as a mediator in the query processing cycle, using user and context dependent information to filter and augment the query and sort the query results. While in the PDI (push) subsystem, as soon as the confidence on ranking of documents is high enough, the same engine proposes the new documents to the user for which the predicted ranking passes above a certain threshold.

For the implementation of the system, several approaches were experimented. Most of these approaches have been proven to be inadequate due to reasons of flexibility, efficiency, large number of examples needed to function (reinforcement learning,...) or because of the lack of explicit knowledge that was produced (nearest neighbour, etc). Finally the system was based upon a PART method, for which the learning step is quick and produces logical rules, which are readable. This is extremely valuable as we are in a framework concerned of human and machine understandable information: it thus permits a formalisation of the learnt knowledge base in RDF.

The first testing and prototype were implemented using the state-of-the-art WEKA machine-learning package, as it was conform to all the choices done for the implementation of CoMMA. Then finally we decided that the final system will be independent of WEKA, and developed a specific Machine learning module based on the PART method, enhanced to take into account the specific characteristics of the description used in CoMMA.

3.6 User Interfaces

The implementation of a relevant, user-friendly GUI has been one of the critical issues of the CoMMA project. The GUI can be seen as a window through which the user will be able to explore the CoMMA system functionalities. That does mean that the GUI will be a key factor regarding the CoMMA acceptance by the end-users.

3.6.1 Main issues of the GUI implementation

The following issues have been identified during the CoMMA project, particularly after the first uses by end-users.

1) To hide the RDF syntax:

The end-user will not be aware of XML/RDF and will have probably limited skills in computer. Moreover the RDF syntax is quite complicated. So the end-user must be able to handle the CoMMA system without knowing this syntax.

<u>Solution:</u> the RDF will not be visible to the user that will deal with concepts and properties through a graphical representation. The RDF syntax is generated by the GUI when needed.

2) To propose a general and clear view of the query/annotation:

The mechanism used to display the queries or the annotations must offer a clear and immediate understanding of the meaning of them.

<u>Solution:</u> The GUI provides a graph representation of the different query/annotation. Boxes are used to display concepts and graphical links represents the properties. By clicking on a box, the user can display all the information available on the related concept.

3) To facilitate the concept and property selection:

Some guidance must be provided by the system to help the user to build the query or the annotation that could be rather complex. The necessity for the user to refer to dedicated words defined into a huge ontology could be very restrictive. The selection of the relevant conceptual entities must be natural and intuitive.

<u>Solution:</u> The GUI provides a terminological search to enter in the ontology. The user enters a keyword. A comparison is done with all the concepts, the properties and the possible synonyms defined into the ontology. The list of matching terms is proposed to the user who can select the most relevant one. As soon as the user has selected the entrance point within the ontology, the system is able to guide him during the whole process, by providing the list of available concepts and properties that can be selected.

4) To offer an iterative mechanism to build queries/annotation:

To deal with the potential complexity of queries/annotations, the GUI must offer an iterative process that will enable to build the queries/annotations step by step either from scratch or from specific templates.

Solution: the fact that the user has to select the required concept or property among the lists provided by the system (as presented above) offer a natural iterative mechanism to support query or annotation building. The user selects a first concept using a terminological search through the whole ontology, then properties assigned to these concepts in the ontology, then the more suitable concepts pointed out by these properties. Step by step, selecting alternatively either a property or a concept, the user builds his own query.

5) To keep the similarity between the annotation, the query process:

The process to build queries must be the same as the one used to build annotations. Some contextual interpretation must be achieved by the GUI to generate either a query syntax or an annotation syntax. Moreover, as a user profile is just a particular annotation, it can be managed as any other annotation.

Solution: in the query mode, a simple process adds the CORESE variables when required (i.e when the user uses default rdf:ID to point out the resources). No limitation of this process has been found during the trial evaluation. The user profile management environment is exactly the same as the annotation environment except that the user profile is sent to the User Profile Archivist (UPA) and not to the Annotation Archivist (AA) to be saved by the system.

3.6.2 The approach

A task force composed of:

- One ergonomist from INRIA,
- Two Technical end-users from T–Systems and CSTB,
- One Final end-user from CSTB,
- Two technical developers from ATOS.

was in charge of defining, specifying and implementing the GUI.

The milestones of the work were the following:

- A first specification was proposed by INRIA in September using power point annotations to illustrate the different use cases for annotating and querying tasks,
- This specification was discussed and accepted by the task force,
- A first prototype was quickly implemented and presented to end-users at the beginning of November,
- A first version was provided at the end of November before the first trial,
- A second version was supplied at the middle of December, integrating favourite queries/annotation management and user profile management,
- A user manual was attached to each version in order to make end users understand as much as possible how the system works.

3.6.3 Some screen examples

The CoMMA Home page and the push information:

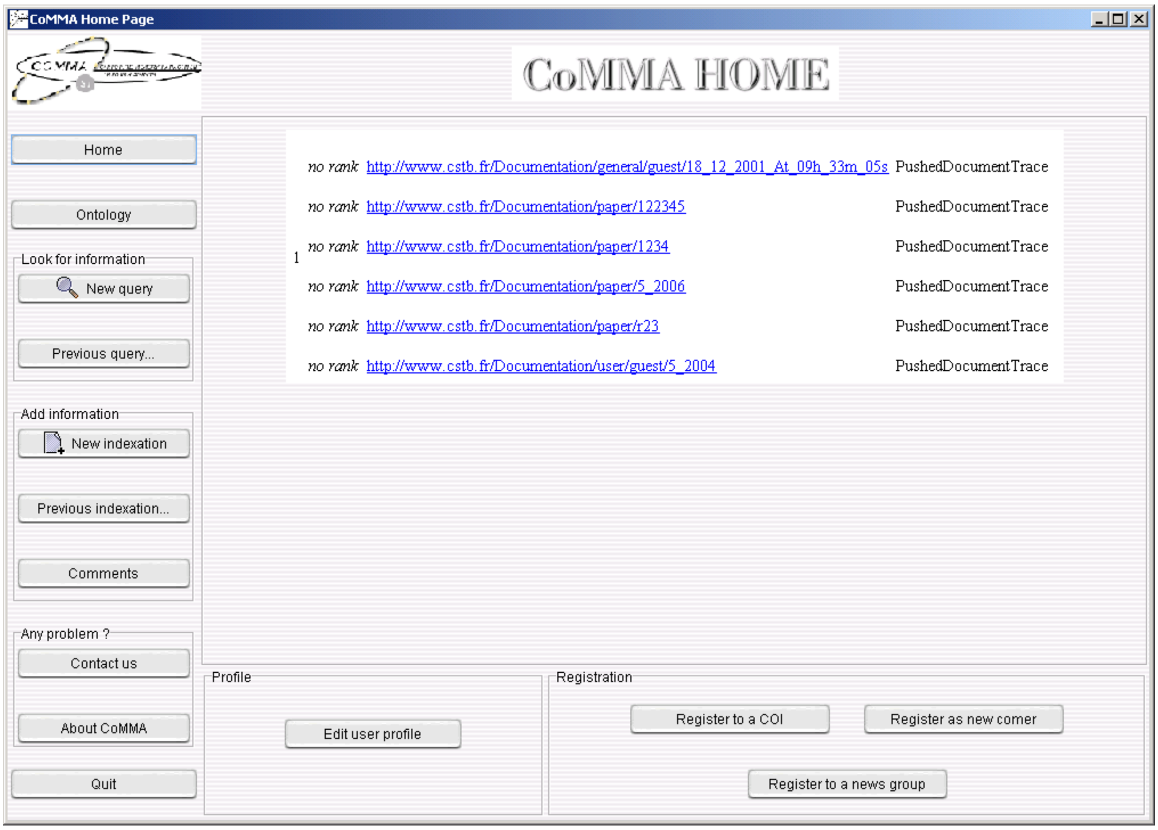

Figure 6 CoMMA home page and push information

The previous figure shows the CoMMA home page, the main functions accessible through the different buttons and the push information: here: 6 document URLs potentially interesting are proposed to the user.

Example of an annotation displayed by CoMMA:

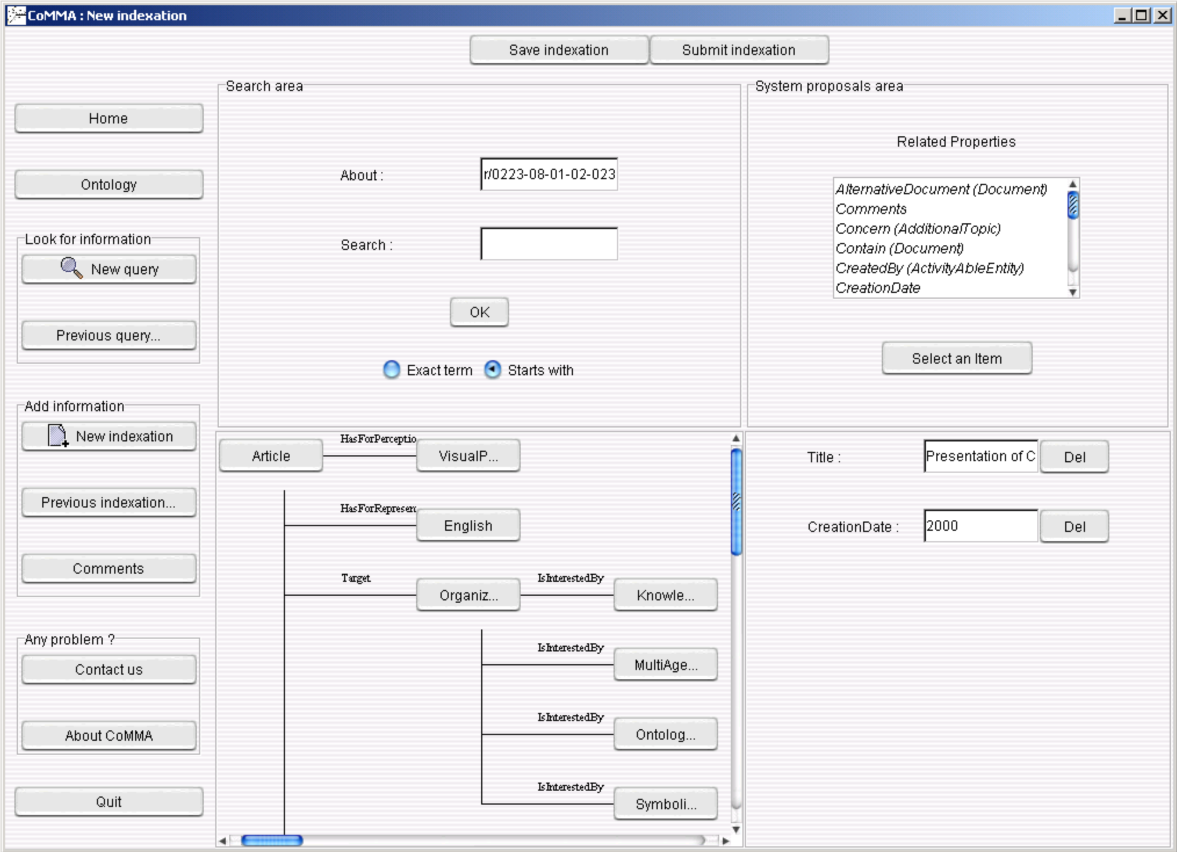

Figure 7 Part of an annotation graph.

The previous figures provides a document annotation view which relates to an article in English which targets an organisation interested in knowledge management, multi-agent system, ontology and symbolic representation.

Example of a query generated by the user:

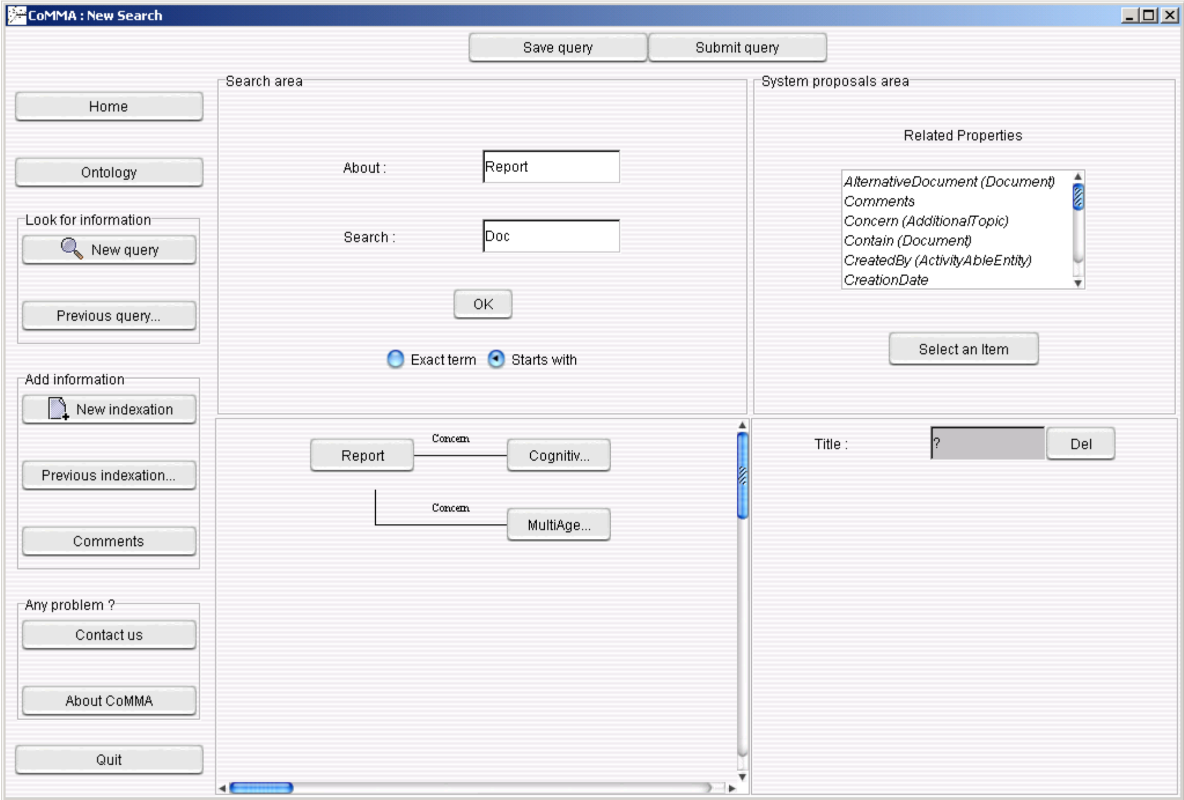

Figure 8 Example of a query

The previous figure shows an example of a graphical query entered by the user. The user wants all the titles of reports that concern Cognitive Science and Multi Agent Technology. All the properties of the "report" concept are available in the "related property" list box. By clicking on the button "submit query", the user will be able to get results from the CoMMA system.

Example of a user profile:

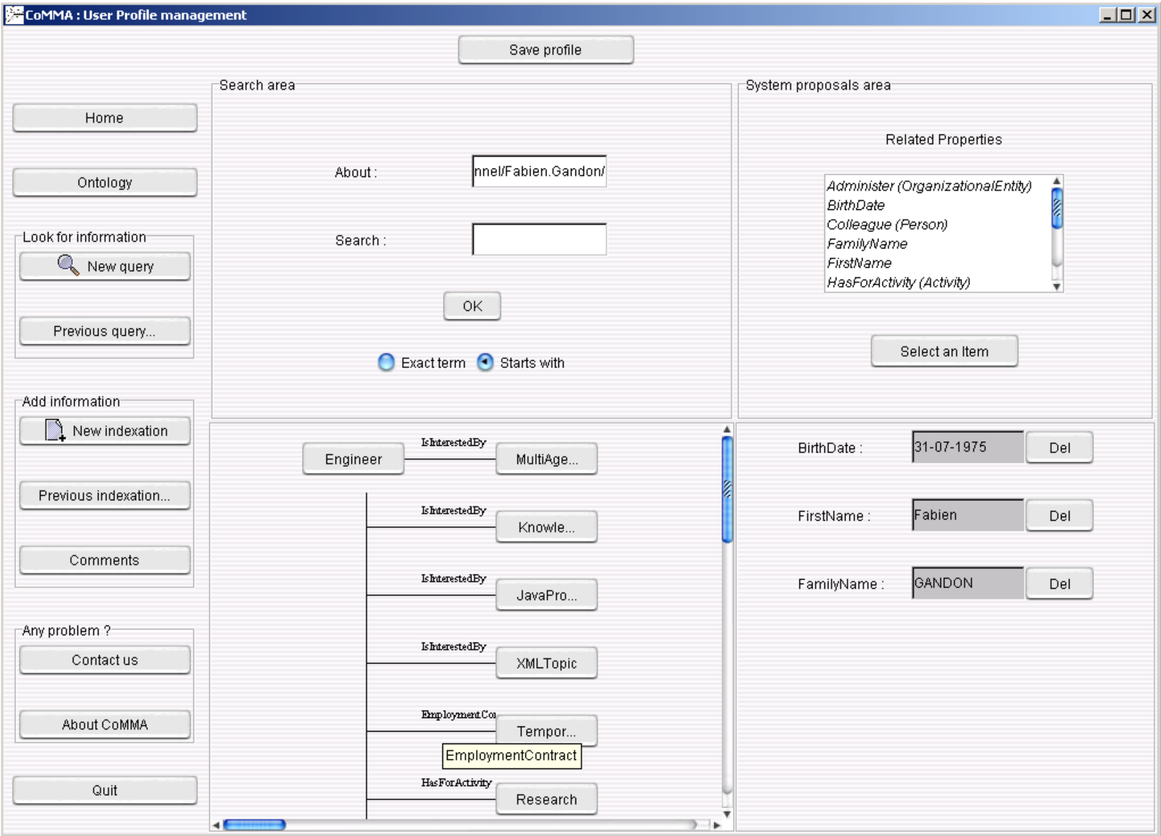

Figure 9 view of a user profile

The previous figure shows a part of a young employee's user profile: This employee is an engineer, interested in Multi-agent system, Knowledge engineering, Java programming, XML. He has a temporary contract and is involved in Research activities.

4 Exploitability of project contributions

4.1 A general solution for corporate memory management

One of the major contributions of the CoMMA project was a general framework to define and implement a distributed corporate memory. The CoMMA solution is based on 5 mainstays:

- The ontology is really the core of the CoMMA solution. The ontology captures the semantic reference of the enterprise and this semantic reference will be used as a reference for the different activities related to the corporate memory. A methodology has been defined to build up the ontology and the result of the work is an ontology for corporate memory (O'CoMMA).
- The multi-agent system that allows the distribution of the corporate memory at the enterprise level. The system is composed of different agents that are specialised in different areas (user management, ontology management, semantic annotation management). These different sub societies defined a general agent architecture for distributed corporate memory. Thanks to the RDF-SL parser extension of JADE, these agents are able to exchange directly RDF data.
- A semantic search engine based on RDF/RDFS and conceptual graph theory, as a very efficient system to improve information retrieval. This semantic search engine brings a lot of innovative aspects, like rules (based on the DRDFS extension) to enrich the ontology, RDF-annotated document classification mechanism to enhance the search activity, distribution algorithm to enable to manage the annotation distribution.
- Machine learning techniques that improve the results of the information retrieval by adapting them to user specificities. An algorithm has been defined to improve the results of information search. It has been applied to pull and push scenarios. This algorithm can be generalised to any kind of data retrieval based on semantics.
- A dedicated graphical user interface to make user able to access very complex notions like ontology and non-readable syntactic objects like RDF annotations, RDF queries, RDF user profiles. The GUI package of the CoMMA solution is composed of an API enabling to manage RDF/RDFS entities from a Java application, an RDF graphical editor enabling to manage a RDF schema and its instantiation, some graphical generic components to build a graph representation of RDF instances.

The following figure illustrates the CoMMA solution and the different technical components that compose this solution. Green boxes point out specific areas where the CoMMA project brings significant exploitable contributions.

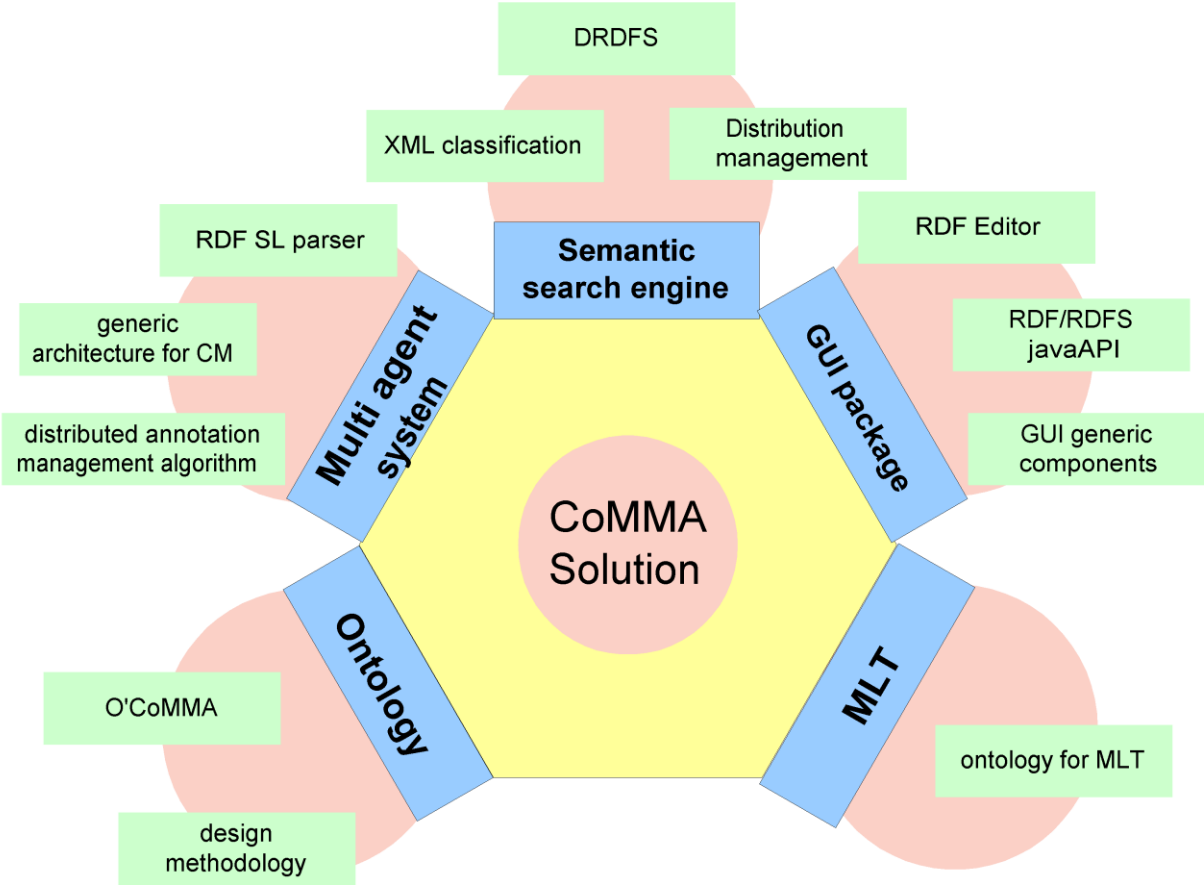

Figure 10 The CoMMA general solution

The following sections will now analyse the exploitability and reusability features of all these major contributions.

4.2 A multi-agent system

Though the Multi Agent System architecture is what keeps the whole CoMMA system together, acting as the backbone of the distributed infrastructure, it can still be used and exploited on its own. In the actual CoMMA KM solution, all the different components such as the GUI, the semantic engine and the user profile processors are plugged into the JADE infrastructure, that is Open Source and available to anyone around the world. Due to the highly flexible software component infrastructure provided by the JADE environment, the system can be incrementally deployed, new agents or platforms can be added or removed dynamically. Above and beyond the qualities of its separate components, the flexibility of the overall CoMMA solution is achieved thanks to the system architecture that holds them together.

Solid software engineering best practices were used, and adapted to multi-agent systems where needed; an iterative development process was used, which performed two complete iterations during the project, gathering users' feedback through questionnaires and walkthroughs. The feedback gathered from the first system trial was used in building the second version of the system, featuring such non trivial changes as a GUI redesign driven by usability results from trial 1 and a change of the content language used by the agents: from FIPA SL with RDF embedded in strings, which was adopted in CoMMAv1, to a complete RDF-based content language, relying on a more advanced metamodel introduced by the latest JADE versions. Moreover, new features were added to the system in the meanwhile, and the loosely coupled component model promoted by multi agent systems and supported by JADE proved itself extremely important in minimizing interference between different change efforts.

Of course, the CoMMA Multi Agent System architecture alone cannot be considered a product. The exploitation path in this case is to use it as a software artefact upon which a complete KM solution, alternative to the actual CoMMA one, can be built. A first option could be to change the ontology or the inference engine: for example it could be possible to adopt OIL as the ontology description language, together with some software component able to reason about OIL statements. In this specific case, since OIL has an RDF Schema representation, the JADE RDF code could also be exploited. Or, new information channels could be added, to deliver knowledge over heterogeneous networks, supporting HTML based interfaces for browser-only clients or audio user interfaces over phones. Again, having CoMMA rely on JADE, which is Open Source and can also run on wireless, resource constrained devices as a result of the LEAP IST project [LEAP], widens potential market for the KM solution (fixed and mobile terminals) and represents a risk reduction strategy (the whole JADE source code is available).

4.3 O'CoMMA

The upper part of O'CoMMA, which is extremely abstract, and the first part of the middle layer, which is describing concepts common to corporate memory applications, are reusable in other corporate memory application scenarios. The second part of the middle layer, which deals with the application domain, is reusable for scenarios only in the same application domain. The last layer containing specific concepts is not reusable as soon as the organisation, the scenario or the application domain changes.

4.4 Methodology for Building Ontologies/Knowledge Models

The methodology we proposed to build O'CoMMA was published in the knowledge engineering and knowledge management communities and is perfectly reusable to build new ontologies for new scenarios. It can be used jointly with natural language processing tools to add the ability to analyse large corpora of documents.

The style sheets produced for the navigation in the ontology are completely generic and reusable for any other ontology formalised in RDFS.

4.5 The generic machine learning ontology

The Machine Learning (ML) engine developed within the CoMMA project exploits the semantic information captured within document annotations and user profiles, in order to increase the efficiency of the knowledge searcher. Thus it uses the O'CoMMA ontology as a basis for the concepts needed to define the machine learning problem.

As the O'CoMMA ontology is structured upon several layers with different reusability levels, so the developed machine learning system should be reusable only in similar conditions. Adapting our first prototype to specific situations (which use different ontological concepts) would require modifying the learning system design (the parts related to the relevant concepts) and rebuilding it.

Therefore, we decided to raise the abstraction level of the learning system, describing it in terms of a more generic "ML ontology". Thus, configuring the system to adapt to different situations (e.g. the specific cases within different enterprises, or major changes inside the same enterprise) will be a question of instantiation of the concepts defining the actual learning problem.

We also decided to use RDF/RDFS to represent the "ML ontology", pointing out some major advantages to have both the problem definition (configuration) and input/output data represented in the same language. Then we defined a first set of abstract concepts starting from the learning attributes used in our first prototype – thus using a "bottom-up" methodology, and succeeded in formalising the prototype problem in terms of these concepts.

Finally we concluded that, even if we are yet far from proposing a completely generic machine learning ontology, we believe that a new methodology based on such a formalisation is the most promising for obtaining well clarified and maintainable systems with the highest degree of reusability. Presently we are committed to consolidate this methodology applying it to other similar problems as well.

4.6 CORESE extensions

In CoMMA, we have explored different approaches for improving information retrieval from the corporate memory. CORESE is thus provided with new inference capabilities enabling the elicitation of implicit knowledge embedded in the semantic annotations of the Web documents of a corporate memory. It thus enables intelligent information retrieval based on both explicit knowledge originally embedded in the semantic annotations of the documents and the implicit knowledge made explicit.

This knowledge elicitation is based on the application of inference rules in forward chaining on the annotations of the document base. These inference rules are formalized in the conceptual graph model upon which the CORESE engine relies. This formalism is provided with a type definition mechanism enabling the representation of class and property definitions. The so-called *type expansion mechanism* enables the elicitation of the definition of a defined class or property in an annotation. The conceptual graph model is also provided with *graph rules* enabling the representation of axioms. A forward chaining mechanism enables the addition of the conclusion of a rule to an annotation when this annotation satisfies the hypothesis of the rule. The graph rules that are considered have no free generic concepts in their conclusion graphs. This ensures the termination of the forward chaining algorithm.

Such extensions of CORESE can be exploited in any application requiring information retrieval from RDF-annotated documents: the CORESE semantic search engine is thus able to perform pre-processing of the annotation base so as to complete this base, thanks to the use of (implicit or explicit) knowledge described in the ontology: such knowledge can consist of the fact that some properties are reflexive, symmetric, transitive or inverse; it can also consist of inferences rules describing some aspects of the enterprise model, useful for improving information retrieval.

4.7 Document classification

Another pre-processing of the RDF annotation base consists of exploiting the RDF annotations so as to learn a document hierarchy: then the information retrieval mechanism can compare the intensions of the nodes of this hierarchy with the user's query instead of working on the whole base of annotations.

These techniques may be reused for applications handling RDF annotations on documents.

4.8 DRDFS

For completing a base of annotations, we can rely on explicit description of contextual knowledge, such as explicit definitions of classes, of properties, of axioms and inference rules. As the present version of RDF(S) does not enable to express such contextual knowledge, we proposed DRDF(S), an extension of RDF(S) inspired of conceptual graph formalism (so as to be able to integrate easily such extensions in our semantic search engine CORESE). We hope that DRDF(S) will contribute to the ongoing work of the W3C committee for improving RDF(S) and meet the needs of the Semantic Web community.

4.9 JADE/RDF Content Language

JADE offers a good support to the usage of content languages and ontologies designed for specific agent applications, in order to manage complex interactions between agents. In particular JADE provides a model of the content language, which is able to describe:

– Object, construct that represents an identifiable entity; this is mainly important to realise a typed knowledge base.
– Proposition, e.g. the content of an "*inform*" communicative-act is a predicate (a subtype of proposition).
– Action, e.g. in the "*request*" communicative-act, tries to express an activity that can be carried out by an object.
– IRE (Identifying Reference Expression) , e.g. in the "*query-ref*" communicative-act.

This model is composed of:

– An abstract content language; ontology independent abstract model, that is a generic model of the concepts that any content language must be able to express (e.g. schemas representation, like predicate, action, etc.)

– A concrete content language (instance of the abstract content language with possible different logic frameworks, like Modal logic, Deontic logic, etc.)

A part of the new JADE APIs, regarding the RDF support, has been developed within the CoMMA project and it is being available as open source, under the *LGPL* license.

The whole framework is characterised by four hot spots, due both to JADE improvements and RDF support. It is possible to have different content language instances (e.g. Deontic logic content language, Modal logic content language…), different content language concrete syntaxes (e.g. RDFS, SL-expression), different ontologies (like CoMMA ontology, Dublin core ontology, …) and different ontology concrete representations (RDFS, SL-expression).

To better clarify the model and the RDF support and to emphasise the available degrees of freedom, in the following figure the message content processing data flow is represented, where CL means Content Language:

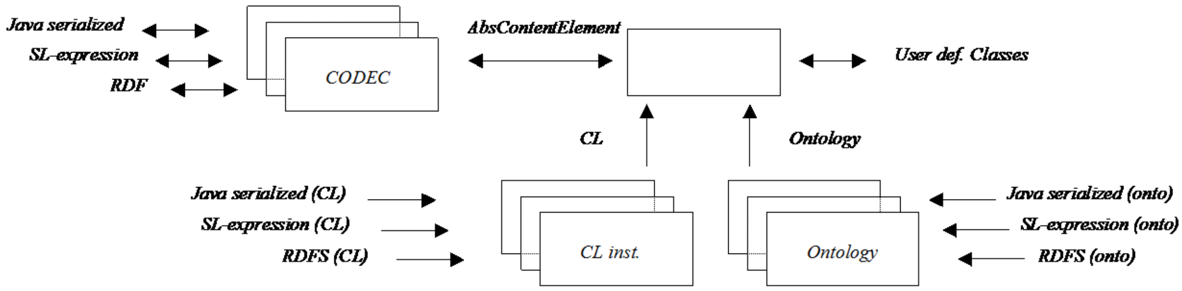

Figure 11 Message content processing data flow

The received messages flow from left to right and they are translated in user defined classes, the sent messages flow from right to left and, in the case of CoMMA project, they are sent in RDF format.

The introduction of an RDF support for the content language and ontology in JADE brings positive benefits to JADE itself and to the FIPA community. JADE project can take advantage of this to improve the level of interoperability between agents developed in different contexts, RDF being a recommendation of W3C; on the other hand, the possibility to use RDF as content language in a FIPA message will contribute to the ongoing work of the FIPA committee of defining a library of FIPA compliant content languages in order to allow more flexibility and efficiency for many applications (e.g., web based).

4.10 The GUI package

The necessity to make the user able to deal with non-tangible concepts like ontology and to handle complex and non-syntactically readable objects like RDFS descriptions, RDF annotations, RDF user profiles, RDF-coded queries, appeared during the CoMMA project like a critical point. The system acceptance by the users would have depended on the capacity to build a user-friendly GUI able to tackle this challenge. The CoMMA consortium developed different graphical components that can be re-used in similar or different situations. These components can be useful as soon as RDF access by human users is required.

4.10.1 The RDF Editor

A tool dedicated to RDF and RDFS file management was implemented through the CoMMA project. This tool offers the following functionalities:

- Loading a given RDF schema: the document annotator is able to load an RDF schema that is relevant to this document, from an RDF file;

- Loading and browsing of an RDF schema; two browsing views are available. The browsing process relies on the "class-property" relationship or the "class-inherited class" relationship, or the "class-parent" class relationship;
- Providing pre formatted windows to help the RDF schema designer to generate valid RDFS syntax:
  - To add/remove/update RDF classes.
  - To add/remove/update RDF properties.
- Providing pre-formatted windows to support the instantiation mechanism;
- Generation of RDF and RDFS syntactic files;
- RDF/RDFS meta-model management.

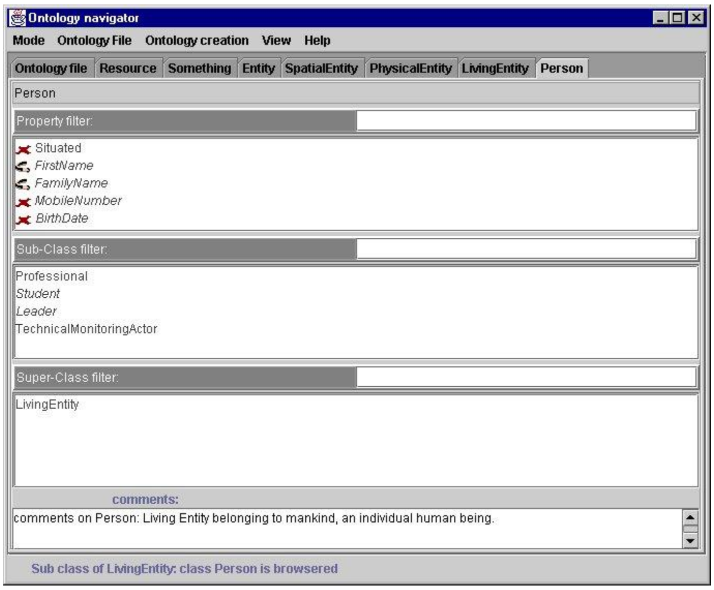

Figure 12 A snapshot of the RDF technical Editor

4.10.2 The RDF/RDFS Java API

A specific java API was developed to deal with RDFS entities like rdfs:class or rdf:property. Rdf instance of rdfs:class objects are managed as well. This API is complementary to and independent from the Sirpac API wich is limited to rdf:resource handling at a low level. This API enables to instantiate JAVA objects that can be a rdfs:class or a rdf:property or a typed resource instance of these entities. This API is composed of three distinct packages:

The package "model.rdfs" that enables to handle rdfs objects in JAVA (class, properties,…)

The package "model.rdf" that concerns RDF instances of rdfs objects.

The package " model.rdfparser" that offers rdf parsing syntax facilities and a default link to the Sirpac parser.

4.10.3 The RDF graph view

The CoMMA projects gave the opportunity to design graphical tools that represent a sexy representation of RDF/RDFS statements. These tools provide particularly a generic graphic view of

RDF syntax that can be used in different contexts than the CoMMA project. This view is active, that is some dedicated events can be dealt with. The figure 9 of the section 3.6 shows an example of such a view of a RDF representation.

5 Dissemination

Dissemination allowed profitable connections and collaboration with other persons or communities (industries or research centres) working in the same field as well as promotion of the results of the project to potential customers. As far as this very innovative project is concerned, where standard and technologies are just emergent, a heavy dissemination has been necessary to bring permanent updating to the whole project achievement.

The following section describes some of the standardisation organisms and industrial forums to which results of the project have been presented in terms of contribution and other dissemination actions.

The last section provides a list of dissemination actions carried out during project life by partners of the CoMMA consortium. Some of the presentations or papers are available from the CoMMA web site.

5.1 Liaisons and dissemination

5.1.1 FIPA

FIPA (Foundation for Intelligent Physical Agents) is an international non-profit association of companies and organisations sharing the effort to produce specifications of generic agent technologies. Generic means here that it is envisaged not just as a technology for one application but as generic technologies that can be integrated by developers to make complex systems with a high degree of interoperability. FIPA was formally established on September 96. It adopts a standardisation process based on work items. Each work item lasts one year and is composed of a Call for Proposals for new Agent Technologies and Applications, a phase of selection and refinement of the received proposals and, finally, the public release of a set of specifications. The subsequent year is instead devoted to the validation of these specifications via field trials. Therefore, FIPA is very interested in projects like CoMMA that could be a real trial of its specifications and give good opportunities to improve them.

The CoMMA project integrates FIPA compliant agent technologies with RDF and RDF Schema, belonging to the XML technology. So, the CoMMA consortium should be able to give valuable feedback to all those FIPA specifications generally dealing with Web technology and with XML in particular.

In its present specifications, FIPA exploits Web technology at various levels. At the Message Transport Service level, an HTTP based MTP is envisaged. This is actually out of CoMMA scope: the project will rely on the underlying JADE platform to exchange ACL messages. At the Message Encoding level, FIPA provides an XML DTD to describe ACL messages. The CoMMA project could contribute to the validation and improvement of this DTD, applying it in a complex application.

The greatest contribution that the CoMMA project is likely to give to FIPA standardisation effort deals with content languages and ontologies. Now, FIPA maintains a Content Language Library which contains, among others, a FIPA-RDF content language using XML Resource Description Format to represent objects, actions and propositions. This matches very well the aims and design approach of the CoMMA project; the CoMMA consortium expects to take advantage of this and to contribute back to FIPA suggestions, improvement and corrections that should validate the standard and further improve its quality.

5.1.2 W3C

The participation of INRIA in W3C enables to give feedback of our work on XML and RDF to W3C.

The information on DRDFS was sent to the RDF special interest group of the W3C. A demonstration of CORESE was performed in presence of Ralph Swick, responsible of W3C group on RDF.

5.1.3 AgentLink Network

AgentLink, European Network of Excellence for agent-based computing, is a network of researchers and developers with a common interest in agent technology. It is funded by the European Commission, and both LIRMM and University of Parma are members of this network. Dissemination of information about the CoMMA project was made within this network (conferences, web site: www.agentlink.org , newsletters...).

5.1.4 OntoWeb

INRIA takes part in the OntoWeb (Ontology-based Information Exchange for Knowledge Management and Electronic Commerce) network of excellence. Dissemination of information about the CoMMA project, the O'CoMMA ontology and the CORESE search engine was performed within OntoWeb.

5.1.5 Liaison with other IST projects

Partners of the CoMMA consortium participated actively to concertation meetings and other conferences organised by Key Action II of the IST programme. This allowed to present CoMMA results to European community and to exchange information with other projects in the same area.

One main result of these concertation actions is the participation of ATOS in the KM Forum project.

5.1.5.1 KM FORUM for the exchange of knowledge management expertise

The Knowledge Management Forum proposes to create a "thematic network" related to KM that will bring together a critical mass of KM experts in Europe. This will enable them to share and exchange the latest developments in the KM domain and to develop visions for the future. KM Forum aims to establish and maintain a well co-ordinated and effective support infrastructure throughout Europe, enabling KM experts to compare their research activities and to network, both on formal and informal level. KM Forum will provide a means for individual organisations to find similarly oriented partners to build special interest groups in order to jointly discuss situations and seek for solutions without losing contact to other broader European view on KM, thus profiting from results achieved and experiences made in other European projects.

Regarding the CoMMA context, KM-FORUM is the perfect place to diffuse the CoMMA results and to make all Europe KM experts be informed about the CoMMA solution. Knowledge board will be used by the CoMMA consortium to communicate about the different components (i.e. the ontology, the role of agents, the machine learning, the GUI,…) and to get fruitful feedbacks from them. Industrial contacts as well as expert recommendations are expected from this dissemination task.

5.1.5.2 e2001 (Venice - 17th-19th of October)

A stand was allocated to the CoMMA project presentation during this event. The CoMMA system was presented and several brochures were given to visitors. A lot of interesting contacts with potential customers have been made.

5.1.5.3 Link with the e-Cognos IST Project (IST-2000-28671)

The e-Cognos project aims at specifying and developing an open model-based infrastructure and a set of tools that promote consistent knowledge management within collaborative construction environments. The work will rely on a deep understanding of knowledge management activities of European construction companies. The analysis of the semantics within and across documents will lead onto the development of ontologies and adaptive mechanisms that can organise documents according to their contents and interdependencies. The web-based infrastructure will include services allowing to create, capture, index, retrieve and disseminate knowledge. It will also favor the integration of proprietary tools. The e-COGNOS approach will be tested and evaluated through a series of field trials. This will be followed by the delivery of business recommendations regarding the deployment of e-COGNOS in the construction sector.

Regarding CoMMA results, the e-Cognos consortium should reuse some parts of O'CoMMA and the methodological aspects, and will investigate the opportunity of a multi-agent system.

5.2 List of actions

Event	Partners	Action	Date	Reference
Knowledge Management symposium in Sophia Antipolis	INRIA, ATOS	Presentation	9/03/00	CoMMA: Corporate Memory through Agents
Knowledge Management Seminar, Dagstuhl (Germany)	INRIA	Presentation	10-15/06/00	
ECAI Workshop on Knowledge Management and Organisational Memories, Berlin	INRIA, T-Systems, U.Parma	Organisation and Presentation	21/08/00	
ICCS 2000 (Darmstadt, Germany)	INRIA	Presentation	Aug. 2000	A Conceptual Graph Model for W3C Resource Description Framework
AICA 2000 Conference, Taormina (Italy)	U.Parma	Presentation	28/09/00	
EKAW'2000 Workshop on Common Approaches of Knowledge Management, Juan Les Pins (France)	INRIA, ATOS	Organisation and Presentations	2/10/00	
European eBusiness eWork 2000, Madrid	INRIA	Presentation	18-20/10/00	Corporate Memory Management through Agents
PAKM2000: Third International Conference on Practical Aspects of Knowledge Management, Basel (Switzerland)	INRIA	Presentation	30-31/10/00	A Multi-Agent System to Support Exploiting and XML-based Corporate Memory
Séminaire sur Systèmes Distribués et Connaissances, Sophia Antipolis (France)	INRIA	Presentation	28-29/11/00	CoMMA : Une approche distribuée de la mémoire organisationnelle
Séminaire sur la Mémoire d'Entreprise, Sophia Antipolis (France)	INRIA	Presentations	30/11/00	
Journée INRIA "Semantic Web", Nancy (France)	INRIA	Presentation & Demonstration	01/01	CORESE, un moteur pour un Web sémantique d'entreprise
Agents 2001, Montreal (Canada)	U.Parma	Presentation and Demonstration	28/05/01	JADE : a FIPA2000 Compliant Agent Development Environment
Conférence sur l'Ingénierie des Connaissances (IC'2001), Grenoble (France)	INRIA	Presentations & Demonstration	25-28/06/01	Ontologie pour un système multi-agents dédié à une mémoire d'entreprise

				Extension de RDF(S) pour la representation de connaissances contextuelles et la construction d'ontologies sur le Web
Conférence Francophone sur l'Apprentissage (CAP'2001), Grenoble (France)	INRIA	Presentation	25-28/06/01	
ICCS'2001 (9th Int. Conf. on Conceptual Structures), Stanford, CA, USA	INRIA	Presentation	30/07 – 3/0801	Extension of RDF(S) based on the CGs Formalisms
IJCAI Workshop on "Ontologies and Information Sharing", Seattle, USA	INRIA	Presentation	4-5/08/01	Experience in Ontology Engineering for a Multi-Agents Corporate Memory System,
IJCAI Workshop on "Ontology Learning ", Seattle, USA	INRIA	Presentation	4/08/01	Learning Ontologies from RDF Annotations
IJCAI Workshop "E-Business and the Intelligent Web", Seattle, USA	INRIA	Presentation	5/08/01	Extending RDF(S) with Contextual and Definitional Knowledge
PKDD/ECML, Freiburg	LIRMM	Presentation	Sept. 2001	Multiagent Cooperative Learning of User Preferences
CEEMAS'2001, Krakow (Poland)	LIRMM	Presentation	Oct. 2001	Learning user preferences in a multiagent system
Synasc Congress, Timisoara (Romania)	LIRMM	Presentation	Oct. 2001	Learning user preferences in Multiagent Knowledge Management
Semantic Web Working Symposium, Stanford, USA	INRIA	Publication	30/07-1/08/01	Extension of RDF(S) with Contextual and Definitional Knowledge
WebNet'01, Orlando, Florida	INRIA	Publication	Oct. 2001	Extension of RDF(S) with Contextual and Definitional Knowledge
ISMICK'2001, Compiègne (France)	LIRMM	Presentation	22-24/08/01	Machine Learning of User Preferences in a Corporate Knowledge Management System
ISMICK'2001, Compiègne (France)	INRIA	Presentation	22-24/08/01	Engineering an Ontology for a Multi-Agents Corporate Memory System
e.2001, Venice (Italy)	ATOS	Demonstration	17-19/10/01	The Project IST CoMMA: Corporate Memory through Agents.

Journée "Ontologies and SemanticWeb", Working group MEMENTO, Sophia-Antipolis (France)	INRIA	Presentations & Demonstration	15/11/01	CORESE, un moteur pour un Web sémantique d'entreprise Ontologie et système multi-agents pour la mémoire d'entreprise distribuée
Séminaire du LIPN "Connaissances, Langue et Informatique", Villetaneuse (France)	INRIA	Presentation	30/11/01	Méthodes et outils pour la gestion des connaissances : l'approche du projet Acacia
Journée INRIA-Industrie "Information retrieval from electronic documents", Rocquencourt (France)	INRIA	Presentation & Demonstration	6/12/01	Capitalisation des connaissances via un Web sémantique d'entreprise
INRIA-INTAP meeting on Semantic Web, Sophia-Antipolis (France)	INRIA	Presentation & Demonstration	15/01/02	Knowledge management & Corporate Semantic Webs

6 Conclusions and perspectives

The efficiency of the CoMMA system can be highlighted through two different points of view:

- The user point of view with a presentation of the CoMMA benefits,
- The technical point of view with a list of the advantages of the CoMMA solution.

6.1 Benefits to user

Regarding the two scenarios envisaged (Insertion of a new employee and Technology monitoring), the CoMMA system leads to some benefits to the user and make him face difficulties that must be considered during the deployment and the learning of the system.

We do not distinguish both scenarios to present the benefits of the CoMMA System. We prefer to focus on the different roles of users, the user as a searcher through the corporate memory, or the user as a contributor to the corporate memory.

6.1.1 The user as a searcher

The user as a searcher can be:

- A new employee,
- More generally, any user requesting some (technical, commercial, organisational, strategic….) information.

Improvement of the document search efficiency:

Compared to classical mechanisms of document search, CoMMA improves the “signal / noise” ratio of a document search:

- Fewer documents will be provided as a result of a search because the CoMMA system offers mechanisms to generate a narrow interpretation of the user’s query,
- The documents proposed through CoMMA will be more relevant with respect to the user’s expectation.

Consideration of the user’s status:

The user’s status is automatically taken into account by the system to improve the relevance of the document delivery; this aspect is more related to the push mode that was implemented in the CoMMA v2 prototype, but it could be also envisaged into the pull mode for the next versions of CoMMA. For instance, when a newcomer queries the corporate memory, or when documents are pushed to him, his status of newcomer will be taken into account. If he is looking for specific documents about Java, maybe some Java tutorials will be provided to him, if he is detected as a Java beginner. When he queries the corporate memory to get information about his new company, documents targeted to newcomer will be sent to him in priority.

Adaptivity to the user and to a community of users:

The system will be able to learn from the user expectations and to adapt further results to the user feedbacks. This way, the relevance of the system answers will be more and more improved for the benefit of the user.

Moreover, CoMMA is able to generalise this learning to a community of users. If a document interests some employee, it can be supposed that it will be interesting for other employees of the same community. Then any employee benefits from the other employee experiences. This is an important point for newcomers to whom documents previously validated by their predecessors will be proposed.

User Guidance and language and concept appropriation by the user: The user will navigate through the ontology to formulate his query. Therefore, he will be somehow guided by this ontology to create the query. If he wants to generalise his query or on the contrary to specialise it, the ontology will propose the most appropriated notions to do that.

One advantage of this aspect is that the user will easily learn the terminology used within the company, and then he will become aware of the company organisation and culture. This aspect is particularly true in the case of a new employee, for whom the navigation through the enterprise could be envisaged as a "company tour". The ontology will lead the user to the conceptual view of the company, and will introduce this conceptual view making him really enter in the world of this company.

Enhancement of the relation between the users and the enterprise:

The ontology definition is a result of a permanent commitment between the employees and their company; this constraint brings some advantages: the relationships between both are reinforced. The mandatory requirement for the ontology appropriation by the user implies a lot of communication, exchanges, and debates between all the different stakeholders of the corporate memory. That should enable the definition of an implicit new contract, based on knowledge sharing.

People, instead of documents:

CoMMA system is flexible enough to enable the implementation of some "query backup modes" that can be used when no document can be provided to classical query (failure of the search). For instance, if no document has been found (i.e. No document annotations match the query), the system can try to find relevant persons (i.e. person whose annotations match the query). This is an interesting aspect of the CoMMA system that could be really valuable, specifically for a newcomer.

6.1.2 The user as a contributor

The user as a contributor to the corporate memory, could be:

- The tutor of a new employee,
- The personal department,
- Any department in charge of communicating documents of interest to the employees,
- The author of a document.

Enhancement of the knowledge diffusion and dissemination:

A document author expects that several interested, motivated and relevant persons will read his documents. The CoMMA system helps to meet perfectly this expectation. By implementing a push mode, by relating it to the user-profiles, the communication of documents to the right employees will be automatically managed by the different sub-communities of agents.

A new way to envisage the diffusion of documents:

Before CoMMA, the document author had to explicitly establish the list of the different persons that could be interested in his document before diffusing it. He had to think in term of "Who will be interested in my documents?" With CoMMA, he will just have to stress the contents of his document through some annotations. Then the document will be automatically pushed to people having declared to be interested in these topics, or having been identified by some CoMMA agent to have a potential interest in the topic.

A knowledge catalyst:

As it improves the way the knowledge circulates through the whole company, providing proactively the right documents to the right persons, the CoMMA system must be seen as a knowledge catalyst. For instance, if a document of interest is provided to concerned experts they will get the opportunity to contribute to the document improvement, so participating to the knowledge refinement process.

6.2 Progress with respect to state of the art

6.2.1 Ontologies

Even if there already exists enterprise ontologies such as AIAI's Enterprise Ontology or TOVE ontology, the O'CoMMA ontology is original by several aspects :

- its structure in three layers (having different reusability/usability features),
- its focus on corporate memory (and in particular, the two scenarios « Integration of new employee » and « Technological watch »),
- its contents : corporate document models, domain ontologies, enterprise models, models of the potential users in the company.,
- its representation in RDF(S),
- the associated tools (in particular the XSL style sheets and the GUI) to browse this ontology.

By the same way, the method proposed for building ontologies is original in comparison to ontology development methods (such as Methontology or Uschold's method) by several aspects:

- it relies on scenarios,
- it includes a terminological stage,
- it combines bottom-up, top-down and middle-out approaches for structuring the ontology.

6.2.2 Multi agent Systems

The CoMMA IST project has designed a specific agent infrastructure to implement corporate memory. This infrastructure is composed of three dedicated sub societies:

- A document (or annotation) sub-society in charge of archiving and managing all the documents,
- An ontology sub-society in charge of managing the ontology and the enterprise model,
- A user ontology sub-society in charge of interacting with the users.

Several generic interaction modes have been provided to support the communication between agents. These interactions use a new Agent Content Language based on RDFS/RDF. Agents are now RDF aware and then able to deal with entities defined into the ontology.

The Multi agent architecture for corporate memory and the new RDF content support API represents the major contribution of the CoMMA project regarding the state of the art.

6.2.3 Semantic search engine

The semantic search engine CORESE is original by:

- Its relying on both RDF(S) and Conceptual Graphs (cf. SirLI relies on RDF(S) and description logics, WebKB relies on conceptual graphs only) ,
- Its inferences on the annotation base so as to infer implicit knowledge: these inferences rely on the reflexivity, symmetry, transitivity of the properties, and CORESE also includes an inference engine making deductions from a rule base aimed at completing the annotation base. Moreover,

class or relation definitions can be expressed through such rules and can be thus processed by the CORESE inference engine.

Several other techniques proposed by CoMMA in order to improve information retrieval are also original:

- The algorithm for an automatic categorisation of the RDF-annotated documents so as to improve the further information retrieval, by automatic building of a hierarchy of documents.
- The algorithms for annotation bases and queries distribution among several agents.

6.2.4 User interface

The CoMMA project brought a new process to manage the ontology through a user-friendly GUI. The ontology is made transparent to the user. This process relies on a terminological matching function, which enables the user to enter any terms he wants to query on. Each potential term relates to the most relevant concepts into the ontology. This process enables the user to enter the ontology without being aware of its existence. Step by step, the system guides the user through the ontology by providing relevant concepts.

6.3 Contribution to CoMMA powerfulness

6.3.1 Ontologies

O'CoMMA plays a central role in CoMMA as all the components exploit the O'CoMMA ontology: the semantic search engine, the agents, the machine learning techniques.

Moreover the ontology is a first citizen component of the memory. So it can be explored by the end-user as the other elements of the corporate memory. It is also useful for improving information retrieval from the corporate memory

Some layers of the O'CoMMA ontology are reusable in different companies interested in building a corporate memory for the two scenarios handled in the CoMMA project but O'CoMMA can also be extended for other scenarios or for other application domains.

6.3.2 The Multi agent system

The CoMMA system as a whole is a Multi Agent System, because it makes available the various system features (ontology-driven inference, learning) as capabilities of a suitable agent or agent group. This of course means that every CoMMA feature is a MAS feature, so that the CoMMA MAS advances the state of the art in MASs by providing an agent society with a unique and novel set of capabilities, well integrated together.

Nevertheless, even the multi agent architecture alone (i.e. with no concern about the internals of the agents, but only about the way they are connected) has a significant impact on CoMMA novelty and effectiveness as a software KM solution. The CoMMA MAS architecture quite straightforward in the Information Agents domain: it can be considered the MAS version of the popular three-tier architecture. The major novelty in CoMMA is that the whole KM process is managed through this architecture, i.e. not only information retrieval, but also annotation creation, information push and users and group management. As far as the effectiveness is concerned, a significant benefit of the MAS architecture lies in the flexible, semantics-driven deployment that exploits the enterprise model to configure the CoMMA system in a concrete corporate environment.

6.3.3 Semantic search engine

The CORESE semantic search engine also plays a central role in CoMMA and contributes to CoMMA powerfulness by improving the precision of information retrieval thanks to:

- its exploitation of the ontology and of the conceptual graph projection enabling to obtain answers more precise than the user's query,
- its inference mechanisms enabling to deduce implicit knowledge and to complete the annotation base (which is crucial in order to ease the work of the people annotating documents in a company and to facilitate CoMMA's approach acceptance).

Thanks to its API, CORESE was easily integrated in the CoMMA agents.

6.3.4 Machine learning algorithms

Ontology-based knowledge management systems, even more than traditional ones, need an especially developed user interfaces for interacting with the knowledge base: feeding with data, maintaining it updated, and efficiently retrieving the information. It is widely accepted that MLT can and will play an important role in accomplishing these objectives of future user interfaces. We have developed the ML solution for CoMMA with this belief in mind, to complement the user interface, using a generic description of the user's context.

Thus, the ML system uses "human and machine understandable" information as input. The natural evolution was that the output of the ML system should also be represented the same way, (same language, ontology) as an augmentation of the information from the knowledge base. Then, the second step was to represent also the ML system description and configuration as well as the internal data used to make decisions, using the same human understandable representation. This resulted in some constraints on the possible choice of classes of ML algorithms, leaving only the ones that produce an interpretable rule base.

We considered in conclusion, that besides the direct benefit of the specific method we implemented (document classification and sorting for the information retrieval), we set the basis of an ontology-based methodology to design configurable machine learning problems applicable to a whole range of situations.

6.3.5 The User interface

The user-friendliness of the proposed GUI contributes to the powerfulness of the CoMMA solution. The possibility to make an end user able to work with a powerful but complex system has been identified as a major challenge after the first trial of the project. The definition of the process that uses the ontology as a support to guide the user through his/her annotating or querying activities was a first key factor. The second trial has shown that end-users were able to understand easily the nature of this process, to annotate documents and to generate queries easily.

6.4 General assessment of CoMMA

A long way has been covered since the beginning of the project. All the different emergent technologies have been successfully integrated, and the CoMMA system has become a reality. Now, *this system can promote a very innovative solution for the Corporate memory* *development, particularly in a distributed environment*. The main advantages of the solution are the following:

- **A unified and committed semantic reference framework that is captured by the ontology**. The ontology takes into account the enterprise specificities, vocabularies and organisation. The CoMMA ontology (called O'CoMMA) is adapted to the corporate memory context and integrates many features like a terminological level, the conceptual ambiguity management, and natural language definitions. This ontology can be partially reused to other CoMMA implementations.
- **A full adaptive and scalable infrastructure** that can be deployed at the enterprise level taking into account local specificities. The CoMMA solution offers a **specific multi-agent architecture adapted to corporate memory implementation**, relying on dedicated agent sub societies. It includes a method to fit the agent deployment to the enterprise organisation.
- **A smart and highly-performant semantic search engine** that offers relevant searching capabilities like generalisation, specialisation, inference rules and logical operators. The engine uses the ontology semantics to provide relevant results.
- **The capability to take into account the employee's specificities**: The multi-agent system provides one agent per-user, which adapts to user's preferences through the user profile content. The user profile can inform the system about the user's background, areas of interests, activities and responsibilities within the enterprise. This information is used by CoMMA to adapt its behaviour to the user's requirements. Moreover, machine-learning algorithms have been implemented to learn from the user's feedbacks. These algorithms are able to generalise user's ratings about document to other employees who have the same field of interests, in order to improve the relevancy of the system results.
- **A user friendly GUI** that hides all the complex aspects of the CoMMA system to employees (notion of ontology, presence of an agent). The GUI provides a natural way to achieve the important cognitive tasks supported by the system like the information retrieval and the resource (document, people…) annotation. The GUI is based on a terminological search layer and uses the ontology to guide the user in his/her different activities.
- ***The CoMMA system provides generic functionalities easily adapted to specific business cases*** such as new employee insertion and technology monitoring. The functionalities provided by the CoMMA system are well adapted to new working methods and provide organisations with a powerful tool enabling all employees to get the right information at the right time.

The CoMMA system is powerful and relevant. In any case, acceptance by end-users of such a system is a major challenge as several new and complex notions are promoted by this solution. Fortunately, the CoMMA project gives the opportunity to assess the CoMMA use at different users' sites and based on different scenarios. Therefore different "non CoMMA-aware" employees from end-user companies have tested the CoMMA system. Their feedbacks were fruitful and will lead to some system improvements. The difficult challenge to put a powerful but complex system in the hand of average employees was successful.

Thanks to the evaluation of the CoMMA system by end users, some possibilities for improvement of the system have been identified:

- The ontology management is not addressed yet by the CoMMA solution. Several maintenance rules must be established in order to facilitate the ontology updating and the synchronisation with existing semantic annotations. It has to be noted that this problem is the general problem for solutions relying on ontologies, some projects are specifically addressing this issue.

- The security aspect is another point that must be tackled. Even if there is no security specificities in a corporate memory implementation, the use of file dedicated to the user description (i.e. the user profile) and on-line access restricted documents raise the well-known issues of computer system like authentication, document confidentiality and user confidentiality.
- The natural language management will bring new facilities to CoMMA end-users. The actual GUI offers a graphical support to handle concepts and properties from the ontologies. The next step should be to provide a natural language processing layer to point out directly the semantic entities and frames after analysing the text entered by the user. The addition of this new process would be a major improvement of the system that will lead to some kind of "universality".

Carrying through these technical challenges will bring major improvements to the current CoMMA solution. However this technical, organisational and human solution is now mature enough to be envisaged in an industrial context and to be applied in different business cases. The areas of application are numerous and will become more and more important with the explosion of Internet and intranet solutions. Now the CoMMA story is about to begin.