Corporate Memory Management through Agents

Philippe Perez, Hervé Karp, Rose Dieng, Olivier Corby, Alain Giboin, Fabien
Gandon, Joël Quinqueton, Agostino Poggi, Giovanni Rimassa, Claudio Fietta,
et al.

Corporate Memory Management through Agents

CoMMA consortium *comma-it@atos-group.com*
Philippe PEREZ, Hervé KARP *Atos Integration*; Rose DIENG, Olivier CORBY, Alain GIBOIN, Fabien GANDON *INRIA*; Joel QUINQUETON *LIRMM*; Agostino POGGI, Giovanni RIMASSA *University of Parma*; Claudio FIETTA *CSELT*; Juergen MUELLER, Joachim HACKSTEIN *T-Nova*

Abstract. The CoMMA project (Corporate Memory Management through Agents) aims at developing an open, agent-based platform for the management of a corporate memory by using the most advanced results on the technical, the content, and the user interaction level. We focus here on methodologies for the set-up of multi-agent systems, requirement engineering and knowledge acquisition approaches.

1. Introduction

How to improve access, share and reuse of both internal and external knowledge in a company? How to improve newcomers' learning and integration in a company? How to enhance technology monitoring in a company? Knowledge Management (KM) aims at solving such problems. Different research communities offer - partial - solutions for supporting KM. The integration of results from these different research fields seems to be a promising approach. This is the motivation of the CoMMA IST project -funded by the European Commission- which started February 2000. The main objective is to implement and test a Corporate Memory (CM) management framework integrating several emerging technologies: agent technology, knowledge modeling, XML technology, information retrieval and machine learning techniques (MLT). Integration of these technologies in one system is already a challenge yet another is the definition of the methodology supporting the whole design process. The project intends to implement the system in the context of two scenarios (1) the insertion of *new employees* (NE) in the company and (2) the support of *technology monitoring* (TM). This paper will present the technical choices and their motivations, the methodological approach and finally the first achievements of CoMMA.

2. Technical Approach

The two implementation scenarios address the following common issue: How to retrieve relevant knowledge within an important mass of information? The solution proposed by CoMMA is based on a *multi-agents architecture* of cooperating agents, being able to adapt to the user, to the context, and supporting retrieval of relevant information in the CM. These agents will be able to (a) communicate with the others to delegate tasks, and to (b) make elementary reasoning and decisions, supporting the choice between several documents. They will have *inference mechanisms* exploiting *ontologies*. They may help authors of documents to annotate the documents, to perform technological monitoring on the Internet and to diffuse the acquired innovative ideas in the interest of employees in the company. The project will focus on the case where the CM is materialized by *XML documents*, annotated by *meta-information in RDF* in order to offer intelligent search functionalities and improve document retrieval. Finally, the project will exploit *machine*

learning techniques (MLT) in order to make agents adaptive to their users and the context. Figure 1 shows a complete schema of the system and the following parts describe more precisely each technical component of the proposed solution.
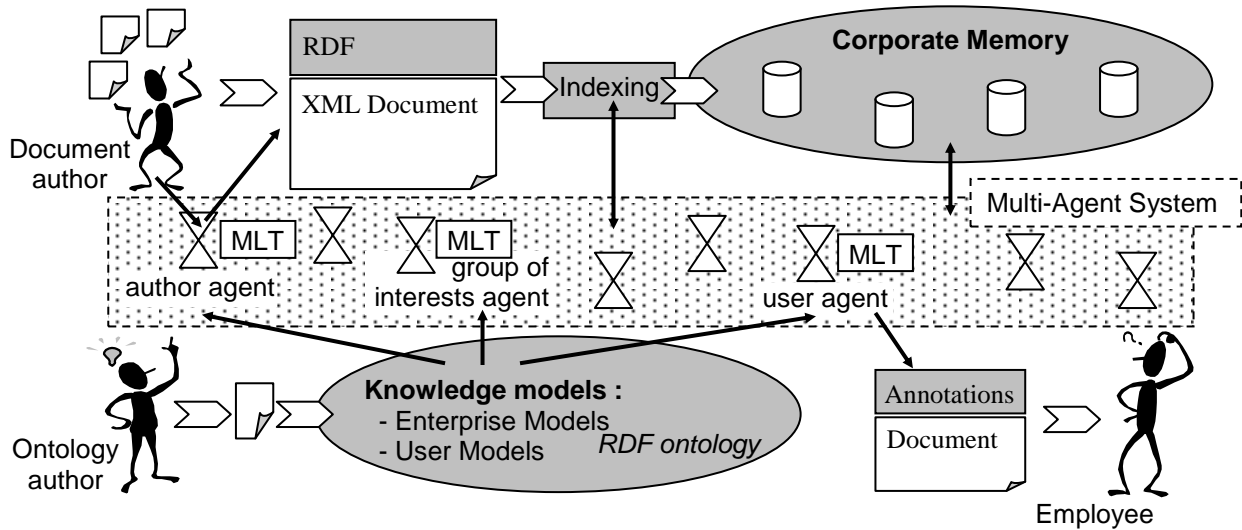

Figure 1. Schematic view of the CoMMA solution for Knowledge Management

2.1 Agent Technology

The realization of the *multi-agent system* (MAS) will be simplified by using a pre-existing *software framework for the development* of agent applications called JADE [2]. JADE makes the development of agent applications easy and compliant with the FIPA specifications [7]. JADE offers an extensible agent model and an agent platform trying to keep high the performance of a distributed agent system implemented with the Java language. In particular, the communication architecture of the agent platform tries to offer flexible and efficient messaging, transparently choosing the best transport available and leveraging state-of-the-art distributed object technology embedded within the Java runtime environment. JADE is an Open Source project and it can be downloaded from [8].

2.2 XML, RDF, RDFS and Ontologies

The *Extensible Markup* Language (XML) is a description language recommended by the World Wide Web Consortium for creating and accessing structured data and structured documents in text format over the Internet. The CoMMA project especially exploits the *Resource Description Framework* (RDF) that uses a simple data model expressed in XML syntax for representing properties of Web resources (documents, images,…) and their relationships. RDF enables us to describe the contents of documents through *semantic annotations* and use them to search for information. RDF makes no assumptions about a particular application domain, nor does it define its semantics a priori. The annotations are based upon an *ontology* that can be described and shared thanks to RDF Schema (RDFS). The idea is that (a) a community specifies concepts and their relationships in ontologies (b) documents of the community are annotated using these ontologies (c) annotations are used to search the memory and navigate into it. See [20] for details.

2.3 Machine learning techniques

Machine Learning methods aim to generalize sets of examples to produce general knowledge to be applied to new situations. The examples on which these learning capabilities will rely are chronicles made of events representing the successive actions of

the user, sampled at a level depending on the goal of learning (actions on page, change of page, change of site, change of user's goal). Within the CoMMA project, three complementary way of using Learning Techniques will be considered [13]: (a*) Learning on the fly*: for a quick adaptation to the user during a session (b) *Remote learning*: to enhance the behavior of the system between sessions (c) *Lazy learning* i.e. case-based reasoning on a base of indexed cases: to take into account very specific cases or aspects.

3. Methodological Approach

3.1 Starting from the user end

The need for a tool which supports KM within a company is becoming more and more inevitable. To structure the knowledge which will be maintained by such a tool different kinds of models are useful. On one hand the system users will be modeled to support specific needs for specific "types" of users. On the other hand the enterprise will be modeled to handle the differences existing between the various activities of the enterprise. Information for user models is acquired through interviews with the knowledge keepers (i.e.: system users). Information for enterprise models is acquired from interviews with knowledgeable people and organizational documents. Many trials for an overall solution of a KM system have failed due to the complexity of the problem. To reduce this complexity we limited the scope of the project to the two scenarios NE and TM.

3.2 Starting from the technical end

The system, as it is envisaged, comprises three main components: user interfaces, CM and MAS. The MAS has the main responsibility of the system: it manages the CM, it adapts the user interfaces and it connects the CM with user interfaces. One of the methods employed to design the MAS is AWIC (for Agents, World, Interoperability, and Coordination) used in domains such as transport management, traffic control, distributed databases, and cooperating expert systems [15]. AWIC is a hybrid design method based on various software engineering ideas including strategies of simulation systems building [14].

A model: Identification of the agents in the problem domain. Specification of their tasks, perception, activities, world knowledge, and basic planning abilities.

W model: Representation of the world (environment) the agents will act in. The activities defined in the A-model have to be offered as services and general laws have to be provided which prevent the agents from acting harmfully.

I model: Definition of the interoperability between the world and the agents. i.e. the action requests (from agent to world) and perceptual information (from world to agent).

C model: Specification of the agent coordination through inter-agent communication, extension of the planning abilities towards joint planning and joint activities.

The general idea is to fill these models in a sequence of iterations. Each iteration is followed by a cross-checking phase where the models are analyzed concerning consistency and over-specification.

A task oriented design method is used to define the A-model and to check the consistency of the interaction (coordination) activities in the C-model. This method comprises four steps: (1) The *identification* of the tasks that agents must perform (this step copes with the problem of defining the interface between the agents and the external world, i.e., the CM and the user interfaces); (2) The *decomposition* of complex and distributed tasks in tasks that can be managed by single agents; (3) The *assignment* of the tasks to the appropriate agents; and (4) The *definition of the interactions* among the agents to perform complex and distributed tasks and to avoid conflicts between different tasks.

UML-Use Case diagrams may be used in the I-model and the C-model in order to describe the interactions between the respective entities. UML is suitable to define traditional systems but because of the proactive behavior of some agents, the behavior of the system itself is not completely driven by user actions. For this reason we will use dynamic diagrams (Use Case, Sequence, Collaboration and State Diagrams) with agents as external actors when they show a pro-active behavior.

4. Know-how and first achievements

4.1 Scenarios and Interviews

These interviews mainly involved employees or groups of employees from T-Nova and CSELT, and were semi-structured with a free opening discussion and a second part more structured for clarification and synthesis. The results from these interviews are used to build scenarios and models, and for requirement analysis.

Concerning the NE scenario, one of the problems encountered was the fact that KM was something new in these companies and a lot of people we viewed were uninterested in spending time on such an idea. In this scenario, we distinguish two general types of interviewees: the NEs themselves and the personnel department peoples responsible for NEs. First results underlined the need for functionalities such as: interactive organizational chart and phonebook, up to date and just in time information, enhancement through personalized agents, indication about where to find what.

Complementary interviews and observations performed at INRIA, revealed other points to be taken into consideration. Concerning the functionalities, it seems that users would appreciate a system that support capitalization of retrieved information for further use, sharing the retrieved information and collective search of information. Another interesting ability of the system would be to provide an indication of its capability to reply to a query in a timely fashion. Concerning the information manipulated by the system different types and characteristics emerged: information helping to clarify the problem itself, organizational information necessary to be learned by an NE, task-oriented or task-related information, context-sensitive information to automatically contextualize and refine queries from the user, information rich in relevant pointers to organizational information, information provided electronically (e-information) and information provided by members of the organization (human sources and expert matching), and finally, relevant information answering narrowly to a specific query.

To complement the interviews, we started to perform a tour of some newcomers' Web sites and some Web site handling annotations. This tour allowed us to illustrate, identify, and imagine interesting features for the CoMMA scenarios, ontologies, and interfaces. For instance: (a) Present information in terms of task: provide a topic list elaborated in terms of tasks, and task steps, relating the document to the current task of the user. One can find pointers such as "writing a monthly report" (b) Customize information: presenting information in terms of [day, month]-long presence of NEs in the organization (e.g., providing information about "the most important tasks" NEs have to do "within the first 3 months after their arrival" such as "get your access card") (c) Allow NEs to provide feedback to information providers (e.g. they may suggest updates or corrections).

Concerning the TM scenario, almost-free interviews have been held, mainly to identify roles, tasks and goals of the people involved and the information they manage. In this case we isolated three kinds of interviewees involved in the TM process at different operational levels: technical experts (researchers involved in technical projects), project managers (researchers in charge of collecting information in specific technical areas) and process owners (TM managers and coordinators). So far, the following requirements were collected

to be taken into consideration for system development: the necessity to be able to monitor several sources of information, tools and taxonomy for document classification (technicality and relevance), repository of up to date information and analysis reports, knowledge of who does what, user profiling through personalized agents, support to communities of interest.

4.2 Agent typology & System Architecture for information retrieval

The outputs from the requirements analysis were exploited to identify the major tasks users require the system to accomplish. These tasks were mapped to suitable members of the agent society of the CoMMA MAS.

An important issue was the need for configuration adaptability of the system: since CoMMA is supposed to be installed on a corporate intranet, a complex, structured and carefully administered network environment is to be expected. Moreover, different corporations will have their own structure and policy for their network. Therefore, the CoMMA system will look different when installed on different environments; the agent acquaintance relation graph will be particularly sensitive to corporate administrative policies. So, the first description of CoMMA multi-agent architecture does not focus on actual agents, but rather on agent categories and roles. Obtaining a system that keeps a defined architecture while adapting to different heterogeneous configurations is one of the major benefits that CoMMA project expects from agent technology.

As is the case with many information retrieval systems [16], CoMMA agent society can be divided in three main areas or sub-societies: (1) *User Agents*: these agents deal directly with human users, often they can adapt to individual users, maybe learning from their past actions and acting on their behalf. (2) *Resource Agents*: these agents manage some kind of resource or service, making it available to the agent society under a suitably thin agent disguise. (3) *Mediator Agents*: these agents act somehow to get user and resource agents together. Most of the responsibility for yielding a configurable and adaptable system falls on mediator agents.

Dealing with user agents, CoMMA keeps a separation of concerns between the user interface aspects (information presentation, deployment technology to access the system from many different client machines, etc.) and user support aspects (user modeling, proactive action on user's behalf). User Interface Agents will take care of dialogues with human users, whereas User Profile Agents will perform dynamic user modeling and proactive actions. The User Interface Agent for a given user will be active only while the user is logged on to the CoMMA system; on the other hand, User Profile Agents will persist across logon sessions.

So far the CoMMA MAS features two kinds of resource agents: Ontology Server Agents and Annotation Server Agents. These agents do not have a particularly high autonomy level, but simply act in response to other agents. However, since one of CoMMA field trials involves pushing selected information to interested users, Ontology Server and Annotation Server agents are likely to have some sort of proactive behavior, too.

Some other agent categories are foreseen such as agents to help users in annotating new documents with RDF, agents that try to discover users with common interests, and various flavors of mediator agents (brokers, facilitators, matchmakers, etc.).

4.3 Merging RDF and Conceptual Graphs

In the final MAS system, the central elementary sequence of interaction between agents when searching in the base of RDF annotations will be : (a) the sending of a query from a requesting agent to an archiving agent, followed by (b) the reply of the latter.

We will reuse CORESE [5], a prototype of a search engine enabling inferences on RDF annotations by translating the RDF triples to Conceptual Graphs (CGs) and vice versa. CORESE combines the advantages of using the standard language RDF for expressing and exchanging metadata, and the querying and inferencing mechanisms available in CG formalism. Among Artificial Intelligence knowledge representation formalisms, CGs are widely appreciated for being based on a strong formal model and for providing a powerful means of expression and very good readability. Moreover, inference and query mechanisms have been developed and tested, and are available to manipulate CGs. For all these reasons CGs appear to be well suited for implementing KM systems. On the other hand, RDF was introduced and recommended by the World Wide Web Consortium and is likely to become a widely used standard enabling the description of a resource by attaching semantic annotations to it [9] [10]. Another reason for using a mapping between RDF and CG is the adequacy between the two models: RDFS classes and properties [4] smoothly map onto CG concept types and relation types. More precisely, RDF statements [10] are mapped to a base of CG facts, the class hierarchy defined in an RDF schema is mapped to a concept type hierarchy in the CG formalism and the hierarchy of properties described in the RDF schema is mapped to a relation type hierarchy in CG. The concept type hierarchy and the relation type hierarchy constitute what is called a support in CG formalism: they define the conceptual vocabulary to be used in the CGs for the considered application.

The CORESE prototype was built using Notio CG platform [18], and the VRP RDF parser [19]. Queries are RDF statements with wildcard characters to describe the pattern to be found and the values to be returned. The RDF query is translated into a CG which is projected onto the CG base to isolate the matching graphs and extract the requested values that are then translated back into RDF. The projection mechanism takes into account the hierarchies and specialization relations described by the CG support obtained from the RDF schemas. It also allows for tuning the matching processes, enabling approximate matching or generalization. We are currently investigating the development of a complete query language based on RDF and its mapping to CG projection. Another on going work is the extension of the functionalities previously developed for the engine in order to implement the agent behaviors related to archiving and searching the documents in the CM.

5. Conclusions

Even though the CoMMA project has only started recently, we can try to summarize on the fly the lessons learnt so far: (1) Interest of relying on XML and RDF for materializing the CM as suggested in [17] (2) Flexibility and adaptability enabled by the MAS approach and MLT techniques (3) Complexity of building ontologies that need to be both human and agent understandable (4) Interest and difficulty of building explicit models of the future uses of the system (5) Organizational problems to solve in order to specify a realistic system and to integrate it later on with a good adequacy with the corporate culture.

The CoMMA project is innovative by its approach to corporate KM (XML + MAS + MLT). There exists very recent studies on the use of ontologies in order to guide information retrieval from the Web: Ontobroker [6], Shoe [11], WebKB [12]. It is also one basis of the OntoKnowledge IST project. But CoMMA will rely on the enterprise model and on the user models to retrieve information from the CM. It is close to the process-

oriented approach proposed in [1] where the KM system exploits a model of the corporate work processes and a model of users' tasks.

Even if agents should be one of the most interesting key items for the software market of the future, the only available examples are systems where agents perform very simple tasks with limited interactions with users and other agents. The realised system could be a first example of a system where the use of agents allows the execution of the main part of the work and might be a guide from which commercial systems should be realised by using agents for helping users in their computer-based tasks.

Moreover, the CoMMA project will be able to support and promote standards and give feedback to some standardisation organizations. The project will use FIPA agent technology specifications, and therefore it may give some useful feedback to FIPA. The project will use XML documents annotated with RDF meta-information and it should give some useful feedback for the next version of these languages.

As a conclusion, the CoMMA project fits very well to the current evolution of the World Wide Web towards the Semantic Web [3].

References

[1] A. Abecker, A. Bernardi, K. Hinkelmann, O. Kühn, M. Sintek. Towards a Technology for Organizational Memories. IEEE Intelligent Systems, 13(3):40-48, May-June 1998.

[2] F. Bellifemine, A. Poggi and G. Rimassa. JADE - A FIPA-compliant Agent Framework. In Proc. Fourth International Conference on the Practical Application of Intelligent Agent and Multi Agent Technology (PAAM99), pp. 97-108, London, UK, 1999.

[3] T. Berners-Lee. Semantic Web Road map. http://www.w3.org/DesignIssues/Semantic.html

[4] D. Brickley and R.V. Guha eds. Resource Description Framework; (RDF) Schema Specification, W3C Proposed; Recommendation 03 March 1999, http://www.w3.org/TR/PR-rdf-schema

[5] O. Corby, R. Dieng, C. Hébert. A Conceptual Graph Model for W3C Resource Description Framework. To appear in Proc. of the 8th International Conference on Conceptual Structures (ICCS'2000), August 2000; Darmstadt, Germany.

[6] D. Fensel, S. Decker, M. Erdmann, and R. Studer. Ontobroker: Or How to Enable Intelligent Access to the WWW. In B. Gaines, M. Musen eds, Proc of KAW'98, Banff, Canada, April 1998,http://ksi.cpsc.ucalgary.ca/KAW/KAW98/KAW98Proc.html.

[7] FIPA Specifications 1997. 1997. Available at http://www.fipa.org.

[8] The JADE Project Home Page. 2000. Available at http://sharon.cselt.it/projects/jade.

[9] O. Lassila 1998. Web Metadata: A Matter of Semantics. IEEE; Internet Computing, July-August 1998, Vol. 2, No. 4.

[10] O. Lassila and R. R. Swick eds. Resource Description Framework; (RDF) Model and Syntax Specification, W3C Recommendation 22 February 1999; http://www.w3.org/TR/REC-rdf-syntax.

[11] S. Luke, L. Spector, D. Rager, J. Hendler. Ontology-based Web Agents. In Proc. of the First Int. Conference on Autonomous Agents, 1997.

[12] P. Martin, P. Eklund. Embedding Knowledge in Web Documents. The Eighth International World Wide Web Conference, pp. 324--341, Elsevier, 1999.

[13] T.M. Mitchell. Machine Learning, (1997), 414 p., McGraw-Hill.

[14] H.J. Mueller. Towards agent systems engineering, International Journal on Data and Knowledge Engineering, Special Issue on Distributed Expertise, 23 (1997), pp. 217 - 245, Elsevier

[15] H.J. Mueller. Towards a Formal Multi-Agent System Specification for Flexible Manufacturing Systems, 9th Symposium of the International Federation of Automatic Control on Information Control in Manufacturing, 1998.

[16] J. Odubiyi, D. Kocur, S. Weinstein, N. Wakim, S. Srivastava, C. Gokey, J. Graham. Scalable Agent-based Information Retrieval Engine. In Proceedings of the First Annual Conference on Autonomous Agents. Marina del Rey, California USA February 5-8, 1997 ACM Press / ACM SIGART, p. 292-299.

[17] A. Rabarijaona, R. Dieng, O. Corby. Exploitation of XML for Corporate Knowledge Management. Proc. of the 11th European Workshop on Knowledge Acquisition, Modeling, and Management (EKAW'99), Dagstuhl, Germany, Springer Verlag, LNAI 1621, May 1999, p. 373-378.

[18] F. Southey. Univ. of Guelf, Canada, http://dorian.cis.uoguelph.ca/CG/

[19] K. Tolle. The VRP RDF Parser, ICS Forth, Greece, http://www.ics.forth.gr/proj/isst/RDF

[20] References about XML: http://www.w3.org/; http://www.oasis-open.org/cover/