An Intranet Based Multi-Agent System for Corporate Memory Management

Agostino Poggi, Giovanni Rimassa and Paola Turci
AOT Lab - Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma
Parco Area delle Scienze 181A, 43100 Parma, Italy
{Poggi,Rimassa,Turci}@CE.UniPR.IT

Abstract

In this paper, we present the multi-agent system (MAS) architecture of CoMMA (Corporate Memory Management through Agents), a system for the management of a corporate memory. In particular we describe the agents roles and interactions, illustrate the design choices for societies and discuss the configuration and implementation issues. CoMMA was realized integrating several emerging technologies: agent technology, knowledge modeling, XML technology, information retrieval and machine learning techniques. The system aims at helping users in the management of an organization corporate memory and facilitating the creation, dissemination, transmission and reuse of knowledge in an organization. CoMMA system is a FIPA compliant agent system and has been implemented by using the JADE agent development software framework. CoMMA is the result of an international project funded by European Commission.

1 Introduction

Companies in the ICT sector 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. These companies "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 important for the insertion of new employees in the company and to diffuse innovative ideas among employees particularly when dealing with technology monitoring activities.

Corporate knowledge management aims at facilitating creation, dissemination, transmission and reuse of knowledge in an organization [Rabarijaona *et al.*, 1999]. It can rely on an organizational memory (or corporate memory), i.e., an explicit, persistent representation of crucial knowledge and information of the organization. This corporate memory can be materialized by a knowledge base, a case base or by documents, made accessible through the company's intranet.

In order to realise such kind of system, some recent works shown that organisations can take advantages of internet and intranet technologies and of agents. Internet and intranet technologies simplify the diffusion of knowledge. In fact, the Web can serve as a basis to distribute information in a uniform way independently of the way the information is stored [Corby and Dieng, 1997]. Agents can help the user supporting retrieval of the relevant information from the corporate memory and adapting the interaction with the system to the user's preferences [Berney and Ferneley, 1999].

In this paper, we describe the multi-agent system architecture of CoMMA, a system for the management of a corporate memory. The next section shows a brief introduction on related work on multi-agent systems for information management. Section three describes the general architecture of CoMMA MAS and introduces the different kinds of agent that populate the system. Section four describes how the system carries out push diffusion of information functionality, that is how the system informs users of the presence of documents likely to interest them and which have recently entered the system. Section five introduces some implementation notes. Section six describes how the system is configured and shows a possible deployment. Finally, section six concludes with a discussion about CoMMA MAS features.

2 Related Work

The research on multi-agent systems for information retrieval and management is very active and some systems have already been realized.

Most of such systems are specialized for information retrieval from heterogeneous information repositories. Some of these systems as, for example, SIMS [Arens *et*

al., 1996] and InfoMaster [Genessereth *et al*., 1997], are composed of agents that wrap these information repositories, and combine and translate information through mediation techniques based on a local schema within the mediator agents themselves.

Other systems as, for example, RETSINA [Decker and Sycara, 1997] and InfoSleuth [Nodine *et al*.] are based on resource agents having in charge the translation of information. Another important set of systems, see, for example, CASMIR [Berney and Ferneley, 1999], is specialized to make easy the gathering of information. In these systems, agents help the user supporting retrieval of the relevant information from one or more information repositories and adapting the interaction with the system to the user's preferences.

Finally, a third interesting set of systems is represented by agent based digital libraries, see, for example, SAIRE [Odubiyi *et al*., 1997] and UMDL [Weinstein *et al*., 1999]. These systems use agents to realize decentralized digital libraries and to enhance scaling, flexibility and extensibility using market based or other distributed artificial intelligence techniques for agents coordination.

3 Multi-Agent System Architecture

The main objective of the CoMMA 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 process 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.

Agents are used for wrapping information repositories (i.e., the corporate memory), for enhancing scaling, flexibility and extensibility of the corporate memory and to adapt the system interface to the users. 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.

In particular, CoMMA system helps the users in performing three main tasks: insertion of XML annotations of new documents, search of existing documents, and autonomous document delivery in a push fashion to provide her/him with information about new interesting documents. These tasks 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.

3.1 Dedicated Sub-Societies

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 must play two roles defined in a FIPA compliant platform [FIPA, 2000]:

- The Agent Management System (AMS) has the duty to maintain system white pages where agents may register themselves and ask for addresses of other agents.
- The Directory Facilitator (DF) 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 organized into graphs, with the possibility of searches of arbitrary depth.

The agents from the ontology dedicated sub-society are concerned with the management of the ontological aspects of the information retrieval activity, especially the queries about the hierarchy of concepts and the different views.

The ontology repository maintains a set of concepts and their relationships that *"attempt to constrain the intended meaning of a vocabulary by eliminated unintended models in the representation…"* [Guarino, 1998]. Documents of the community are annotated using these ontologies and ontologies are used to search documents into the corporate memory and to navigate into it. In particular, the CoMMA ontology describes the documents maintained in the organization corporate memory and the enterprise model describes the structure of the organization ruling, for example, the access to the different type of documents of the corporate memory. A replicated organization for the ontology sub-society is conceivable because ontologies shared by users should be stable and most of the queries will need the whole ontology to apply inference algorithms. The agents of the ontology sub-society must play two roles:

- The Ontology Archivist (OA) has the duty to store and retrieve information to/from the CoMMA ontology repository.
- The Enterprise Model Archivist (EMA) has the duty to store and retrieve information to/from the CoMMA enterprise model repository.

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 organization for the document sub-society is appropriate because separates 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 must 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 query 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 document from different repositories.

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 dedicated sub-society must play four roles:

- User Profile Manager (UPM) is in charge of updating and exploiting the user's profile when the user is logged on to the system. This agent uses machine learning techniques to adapt the system to the user; for example, it is able to know if a document is of interest for a user and to order a list of documents on the basis of user preferences.
- 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. During system usage, the IC interacts directly with the agents belonging to the ontology sub-society, whereas it relies on a responsible UPM agent to deal with the document sub-society. A distinguished feature of IC agent is its GUI, through which users can exploit the CoMMA MAS. The GUI makes the user look as an agent to the rest of the system, because as soon as the GUI layer is crossed, all the information passing happens by means of FIPA ACL messages.
- 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. Only this agent has the access to the user profile repository of a specific sub-net.
- The User Profile Processor (UPP) has the duty to perform proactive queries on the corporate memory on the basis of the user profiles and information about new documents.

4 Push Diffusion of Information

CoMMA behavior may be shown through the scenarios describing how agents interact to perform the different tasks. These scenarios are: u*ser login*, a*dd a new annotation*, e*dit an annotation*, *information retrieval*, *update user profile*, *push diffusion of information*. In the following, we introduce the description of one of the most relevant scenarios: push diffusion of information.

The agent roles taking part to this scenario are: Interface Controller, User Profile Manager, User Profile Processor, User Profile Archivist and Annotation Mediator. This scenario is played whenever a new annotation appears within the corporate memory or a pre-existing annotation is modified. The purpose of this scenario is to retrieve relevant information contained within the corporate memory, and present it to the user in the most comfortable way.

Even if CoMMA users connect to the system intermittently, in order to carry out the push diffusion of information functionality, there is an obvious need to persistently store information about them and remembering it across all the sessions. Therefore, user dedicated agents are continuously running, gathering information from the user to improve system performance. Moreover, CoMMA system supports the user capability of logging into the system from different parts of the corporate intranet, while preserving his/her *organization group* (the group of people he/she belongs to) and his/her *home location* (the network host he/she usually logs on).

At any point in time, a UPM agent handles all the users currently connected to the system from within its *jurisdiction domain*. This means that the whole corporate network is partitioned in subsets and there is one and only one UPM for each subset. The UPP has the duty to perform proactive queries on the corporate memory on the basis of user profiles and information about new documents. Concerning push mode, the UPP has 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 pushing of information involves two phases. In the first phase user agents compute the list of users that might be interested in the information. In the second phase, user agents inform the possible interested users about the information.

Figure 1 shows the *protocol diagram* (a restriction of the UML collaboration diagram, proposed within the AUML initiative [Bergenti and Poggi, 2000]) of that scenario when the user is connected; the name *Mgr* denotes the concrete agent that plays UPM and UPP roles.

The task is performed in three steps:

1. The AM informs all UPPs a new document is annotated or a pre-existing annotation is modified.
2. Each UPP generates a request to its local UPA, using the annotation and an inference mechanism

in order to receive the list of users (users profiles) that can be interested.

3. Each UPP sends information to each user interested. There are two possible cases:
   a. The user is connected. The UPP informs the IC of the pending notification; the effective information is sent on explicit user demand (*fipa-request* interaction protocol).
   b. The user is not connected. The UPP requests the UPA to store the pending information in the user profile. When the user logs in, the user will be notified of the stored pending information.

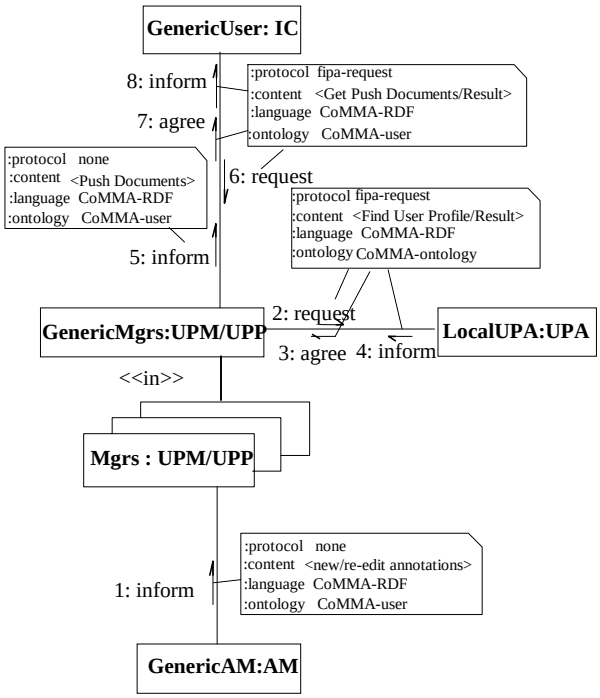

Figure 1: Push Diffusion of Information (user connected)

5 System Implementation

The realization of the system was carried out by using a pre-existing software framework for the development of agent applications called JADE [Bellifemine *et al.*, 2001] and by using a semantic search engine prototype, called CORESE [Corby *et al.*, 2000].

CORESE is a prototype of a search engine enabling inferences on Resource Description Framework (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.

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 [JADE, 1999]. 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.

6 Configuration and Deployment

The CoMMA system must adapt its configuration to a specific corporate environment at deployment time. CoMMA system explicitly targets corporate networks, and the network topology of a corporate intranet is deeply influenced by the company organizational structure, so that agent acquaintance and reachability will vary, depending upon specific company policies. In fact, the deployment of the CoMMA MAS should 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 topologies characteristics, the network maps and data rates, data servers location, gateways, firewalls, etc) and by interests area, 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.

Even if the static model (agent roles) and the dynamic model (agent interactions) describe quite well the CoMMA MAS, it is useful to provide a third view that is the *Deployment Model*. This model can help a lot in focusing on the actual system, because the CoMMA KM solution is made by a highly distributed system deployed over a structured, managed corporate network.

6.1 Three-Tier Deployment

The Figure 2 shows a possible deployment, called *Three-Tier Deployment* because the physical computers hosting the agents are divided into clients (*client tier*), application servers (*middle tier*) and database servers (*DB tier*). Such a deployment strategy has the main advantage of matching one of the most popular intranet structures.

In this diagram, boxes represent physical hosts and generic components are replaced by agents (with bufferfly shape) and knowledge bases (represented using the familiar mass storage icon). An important observation to make is that this diagram contains physical agent instances and not agent roles. This means that a single butterfly icon can represent many agent roles that will be played together by a single physical component. A directed graph shows the agent acquaintance graph. Moreover, agents belonging to the same JADE platform are grouped together (notice that belonging to a platform implies the a priori knowledge of the Agent IDs for the AMS and the default DF).

According to the three-tier architecture, only agents that play the role of Interface Controller live on the client

machines (performing pure presentation) and only agents that play the role of Annotation Archivist live on the database servers (performing pure data management). The middle tier, as in classical three-tier architectures, connects the client tier and the data base tier (through DF and AM agents) and is where the most interesting stuff happens (ontology and enterprise/user model management). In this deployment, the ontology, the enterprise model and the user model belong to the middle tier and not to the data base tier because they are supposed to be much smaller in size than the annotation base and they are not distributed.

In this deployment example, there is a JADE platform spanning the whole *Lab A*, another the *Lab B* and a third that contains all the agents that play the role of Annotation Archivist, managing the distributed annotation base.

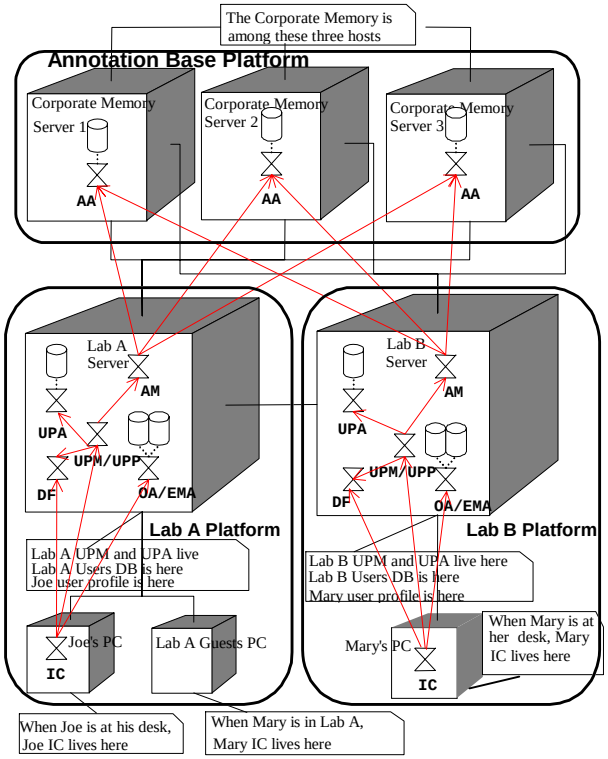

Figure 2: CoMMA Three-Tier Deployment

The user dedicated sub-society is deployed according to the following rules: the IC is active for a session and runs on the host where the user logs for the current session; the UPM is active continuously and runs on a server that manages the *current location* of the user (i.e. the machine he or she is logged on); the UPA is active continuously and runs on a server that manages the *home location* of the user (i.e. the machine where he or she belongs and where he or she usually logs on).

6.2 Configuration and System Flexibility

The CoMMA system was targeted right from the start at corporate intranets, so it relies on assumptions such having a common, agreed upon ontology, or having a network whose topology can be derived from the enterprise model. However, even when requiring a global ontology or an administered network, the fact that those entities change with time needs to be taken into account. The CoMMA system needs to be adaptable to organizational changes of the corporation using it.

Several aspects of the CoMMA system work together to achieve this goal: firstly, the O'CoMMA ontology is a separate asset, materialized into RDF Schema files and completely decoupled from the other software artifacts. The ontology can be evolved after the initial deployment, with no need for system-wide operations: agents in the ontology subsociety will be able to reload the new version of the ontology. Of course, some rigidity remains due to the need for an ontology to capture semantic information, that cannot be produced automatically but needs to be extracted from humans. The CoMMA solution tries to ease this demanding task by defining a methodology to build and refine ontologies.

Even if the previous subsection presented a classical three-tier deployment for the CoMMA system, it should be emphasized that this was only a sample deployment, chosen to set up a simple example in a familiar environment. Moreover, the enterprise model is just another set of RDF annotations, so it can be augmented and modified during the system usage. 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. So, the network structural changes due to corporate reorganizations can be dealt with rather seamlessly.

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 user feedback through questionnaires and walkthroughs. A first version of the system was released to end users after one year of development, and a first system trial was carried out. The feedback gathered was used in building the second version of the system, featuring such non trivial changes as a complete 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 to a complete RDF-based content language, relying on a more advanced metamodel introduced by the latest JADE versions. All this happened while some major new features were introduced in the system (information push, user-defined inference rules in CORESE); nevertheless, the loosely coupled component model promoted by multi agent systems proved itself extremely important in minimizing interference between different change efforts.

Moreover, relying on a framework as JADE helped a lot in managing the content language replacement, an infrastructural change that could have affected every soft-

ware component in the system, but which remained well hidden behind JADE abstract APIs.

8 Conclusions

In this paper, we presented a multi-agent system for the management of a corporate memory. The innovative aspect of the system is the integration of several emerging technologies that were generally used separately until now: agent technology, knowledge modeling, XML and RDF technologies, information retrieval techniques and machine learning techniques.

CoMMA MAS architecture showed good modularity and flexibility. In fact, during its implementation we needed to modify some parts of the system to cope to new user requirements and it does not generate side effects on the other parts of the CoMMA system. The use of JADE increases system modularity and flexibility. The separation between the software platform infrastructure managing agent life-cycle, distribution and communication and the software implementing agent tasks decouples modifications in these two parts.

The system was developed within an international research project, funded by the European Commission IST program [CoMMA, 2000]. The project lasted two years, ending in January 2002, and it spanned two iterations of a complete development life cycle, including two field trials, carried out by end user project partners in order to validate the overall system and to provide feedback to the project partners more involved in the development. As it often happens in this kind of applied research projects, the software developed within the CoMMA effort is not a mere proof of concept, but rather a prototype that is expected to be further evolved into a product or a solution after the project conclusion. An exploitation plan was part of the CoMMA project right from the start, and several avenues were envisaged, involving the whole CoMMA KM solution or separate components (e.g. the ontology, the MAS architecture). The project final report is a public document containing explicit information about project result exploitation strategies.

Acknowledgments

Thanks to all CoMMA partners who contributed to the development of the project. This work is partially supported by the European Commission through the contract IST-1999-12217, *CoMMA – Corporate Memory Management through Agents*.

References

[Arens *et al.*, 1996] Y. Arens, C.A. Knoblock, W. Shen. Query reformulation for dynamic information integration. Journal of Intelligent Information Systems, 6(2):99-130, 1996.

[Bellifemine *et al.*, 2001] F. Bellifemine, A. Poggi, G. Rimassa. Developing multi agent systems with a FIPA-compliant agent framework. Software Practice & Experience, 31:103-128, 2001.

[Bergenti and Poggi, 2000] F. Bergenti, A. Poggi. Using UML to Model Agent Interfaces. In Proc. 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2000), pp. 645-648, Orlando, FL, 2000

[Berney and Ferneley, 1999] B. Berney and E. Ferneley. CASMIR: Information retrieval Based on Collaborative User Profiling. 1999.

[CoMMA, 2000] CoMMA Home Page. Available at http://www.ii.atosgroup.com/sophia/comma/HomePage.htm.

[Corby and Dieng, 1997] O. Corby and R. Dieng. A CommonKADS Expertise Model Web Server. In Proc. Fifth Int. Symp. on the Management of Industrial and Corporate Knowledge, 1997.

[Corby *et al.*, 2000] 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.

[Decker and Sycara, 1997] K. Decker, K.P. Sycara. Intelligent adaptive information agents. Journal of Intelligent Information Systems, 9(3):239-260, 1997.

[FIPA, 2000] FIPA Home Page. Available at http://www.fipa.org.

[Genessereth *et al.*, 1997] M.R. Genessereth, A. Keller, O.M. Duschka. Infomaster: An Information Integration System. In Proc. Of the ACM SIGMOD Int. Conf. On Management of Data, pp. 539-542, Tucson, AR, 1997.

[Guarino, 1998] N. Guarino. Some Ontological Principles for Designing Upper Level Lexical Resources, Proc. First International Conf. Language Resources Assoc., Granada, Spain, 1998, pp.527-534.

[JADE, 1999] JADE Home Page, 1999. Available at http://jade.cselt.it.

[Nodine *et al.*] M. Nodine, J. Fowler, T. Ksiezyk, B. Perry, M. Taylor, A. Unruh. Active Information Gathering in InfoSleuth. International Journal of Cooperative Information Systems.

[Odubiyi *et al.*, 1997] J.B. Odubiyi, D.J. Kocur, S.M. Weinstein, N. Wakim, S. Srivastava, C. Gokey, J. Graham. SAIRE – A Scalable Agent-Based Information Retrieval Engine, Proc. Autonomous Agents 97, Marina Del Rey, CA.

[Rabarijaona *et al.*, 1999] Rabarijaona, R. Dieng and O. Corby. Building a XML-based Corporate Memory. In Proc. European Knowledge Acquisition Workshop (EKAW'99). LNAI 1621, Springer-Verlag, 1999.

[Weinstein *et al.*, 1999] P.C. Weinstein, W.P. Birmingham, E.H. Durfee. Agent-Based Digital Libraries: Decentralization and Coordination. IEEE Communication Magazine, pp. 110-115, 1999