

CoMMA : Une approche distribuée de la mémoire organisationnelle

Fabien GANDON

Rose DIENG

Alain GIBOIN

Projet ACACIA – INRIA Sophia Antipolis

{fgandon|dieng|giboin}@sophia.inria.fr

Résumé

La distribution, au sens large, d'une mémoire organisationnelle est une caractéristique fondamentale importante pour le développement et le déploiement de systèmes de gestion des connaissances. Cependant nous montrons dans cet article qu'un certain nombre de technologies émergentes peuvent actuellement faciliter la réalisation d'une solution complète et intégrée à l'organisation.

1. Introduction

La mémoire organisationnelle est un système distribué au sens large du terme. Elle n'est pas seulement distribuée entre des systèmes informatiques, sa distribution inclut d'autres artefacts, ainsi que les membres de l'organisation. Les solutions informatiques pour la gestion de la mémoire, doivent prendre en compte cette "super-distribution" pour s'intégrer aux organisations. Les modèles de la distribution doivent aller au delà de la dimension technique et rendre compte de l'organisation, de son infrastructure, de son environnement et de ses caractéristiques psychosociologiques. Le projet ACACIA fait partie du consortium CoMMA (Corporate Memory Management through Agents cf. [CoM00]). CoMMA est un projet IST subventionné par la commission Européenne visant à développer et tester un environnement de gestion de la mémoire d'entreprise. Le projet s'attache à préserver le contexte de l'existence et de l'utilisation de la mémoire d'entreprise en s'intéressant en particulier à deux scénarios :

- *Aide à l'insertion d'un nouvel employé* : Utiliser la mémoire d'entreprise pour permettre aux nouveaux employés de s'insérer rapidement, de comprendre la politique, le fonctionnement et l'organisation de l'entreprise et les rendre opérationnels le plus rapidement possible en leur permettant de trouver ou en leur suggérant proactivement l'information dont ils ont besoin.
- *Support de la veille technologique* : Utiliser la mémoire d'entreprise pour assister l'identification et l'évaluation de technologies émergentes concernant l'activité de l'entreprise, et diffuser

l'information pertinente aux personnes concernées et compétentes.

Pour atteindre ses objectifs, CoMMA cherche à intégrer un certain nombre de technologies émergentes: l'idée est de conjuguer au sein d'un système multi-agents des techniques et outils issus de l'ingénierie des connaissances, de la galaxie XML, des techniques de recherche d'information et des techniques d'apprentissage machine (Figure 1).

Dans une première partie, nous présenterons les atouts de la galaxie XML pour la matérialisation d'une mémoire organisationnelle. Puis nous aborderons les caractéristiques des systèmes multi-agents qui en font une architecture logicielle et un paradigme séduisants pour la gestion d'une telle mémoire. Enfin nous verrons que l'introduction de techniques d'apprentissage ou de filtrage collectif permet de s'adapter et d'exploiter les caractéristiques des utilisateurs.

2. Une mémoire distribuée et hétérogène

La mémoire organisationnelle a cela en commun avec le Web qu'elle est un paysage d'informations hétérogènes et distribuées. De ce fait, ils partagent aussi le problème de pertinence (bruit et précision) lorsque l'on veut chercher et récupérer leur contenu. Pour reprendre la formule, maintenant consacrée, de John Naisbitt [Naisbitt], les utilisateurs du Web se noient dans l'information alors qu'ils ont soif de connaissance. L'overdose d'information et l'incapacité croissante des moteurs de recherches par mots clefs à couvrir et exploiter le Web est un problème largement reconnu. Parmi les initiatives visant à résoudre ce problème le 'Web sémantique' est une approche prometteuse; l'idée est de rendre explicite la sémantique des documents au travers de meta-données ou d'annotations. Parallèlement, il y a un intérêt croissant des industriels pour capitaliser et faire fructifier les connaissances de l'entreprise, ce qui entraîne le développement et le déploiement de techniques de gestion des connaissances dans de plus en plus de compagnies. Les Intranets reposant sur la technologie internet, peuvent bénéficier des progrès du 'Web sémantique'.

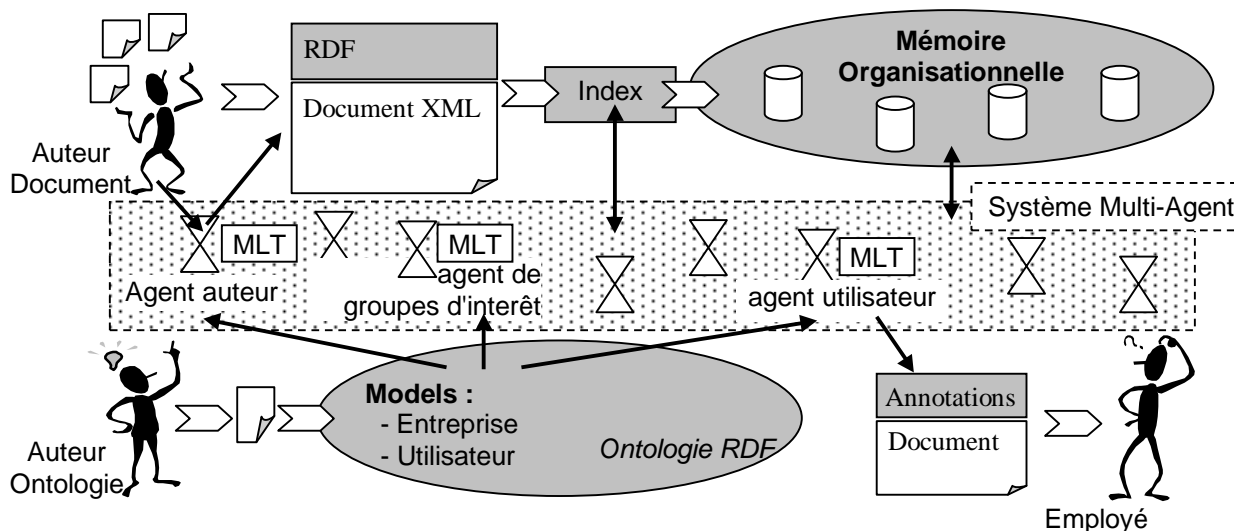


Figure 1. Schéma général de CoMMA

Nous définissons la mémoire organisationnelle comme la représentation persistante, explicite, désincarnée des connaissances et des informations dans une organisation, afin de faciliter leur accès, leur partage et leur réutilisation par les membres adéquats de l'organisation, dans le cadre de leurs tâches [Dieng00]. Dans le projet CoMMA, pour matérialiser la mémoire d'une entreprise nous avons étudié les atouts de la technologie XML (Extensible Markup Language) qui est en passe de devenir un standard accepté et utilisé par l'industrie pour l'échange et le stockage d'informations. C'est aussi dans cette galaxie XML qu'a été proposé le langage RDF (Resource Description Framework) utilisé pour le 'Semantic Web'.

XML est extensible, et permet de structurer des documents et des données dans des fichiers au format texte facilement échangeables et lisibles sur des réseaux basés sur la technologie internet (Internet, Intranet, Extranet...). Le standard rend les données compréhensibles à l'homme et à la machine, même si à terme l'aspect XML sera caché derrière les logiciels applicatifs et les interfaces utilisateur. La structure des documents et des données échangés peut être définie formellement (Document Type Définition - DTD) et permettre ainsi de diffuser un format de document et de vérifier que les instances de documents s'y conforment assurant une certaine cohérence parmi les éléments de la mémoire. Cependant contrairement à SGML les DTD XML ne sont pas obligatoires. Un même document (ou une même source de données) peut ensuite être manipulé, transformé et formaté (XSL Extensible Style Language) en de multiples autres documents pour, par exemple, répondre à différents systèmes logiciels, être présenté à différents utilisateurs, ou s'adapter à différents contextes d'utilisation.

RDF est un langage utilisant la syntaxe XML pour décrire des triplets représentant les propriétés des ressources disponibles sur le réseau et les relations qui

existent entre elles. Ces annotations RDF peuvent être externes aux ressources permettant ainsi d'intégrer dans une nouvelle mémoire des éléments d'archives légataires sans les modifier ou des éléments externes à l'organisation et accessibles en lecture seule. Les annotations reposent sur une ontologie décrite en RDFS (RDF Schema) proche des modèles objet. Dans le cadre de la mémoire organisationnelle, l'ontologie est une spécification explicite mais partielle des concepts, des relations et des règles intervenant dans la structuration et la gestion de la mémoire, et consciencieusement définis avec les termes qui leurs sont associés. En comparaison du Web une mémoire organisationnelle a un contexte délimité et mieux défini : l'organisation. En particulier nous pouvons envisager d'identifier les différents profils d'utilisateurs (intéressés) et surtout, la communauté de l'entreprise partage souvent une certaine vue du monde (politique d'entreprise, pratiques instituées...) qui permet de penser qu'un consensus ontologique est possible dans une certaine mesure. Les particularités de l'organisation et de ses membres peuvent alors être décrits par des modèles utilisant les concepts et les relations décrits dans l'ontologie. De la qualité de l'ontologie dépend par conséquent la qualité et l'exploitabilité des modèles, des meta-données et des annotations. En particulier l'ontologie joue un rôle important dans la cohérence, et la prévention des ambiguïtés et des redondances. RDFS permet de capturer certains aspects sémantiques des concepts et des relations utilisés (héritage, signature) pour structurer la mémoire en l'annotant et, par la suite, inférer sur les annotations pour exploiter la mémoire.

Pour pouvoir exploiter les bases d'annotations en RDF le projet ACACIA a développé CORESE [Corby00] un prototype de moteur de recherche traduisant les annotations RDF en graphes conceptuels et les schémas RDFS en support de graphe conceptuel. CORESE peut ainsi exploiter les mécanismes d'inférence disponibles en graphes conceptuels et réutiliser tout ce qui depuis des années a été fait dans

ce domaine. En particulier l'opérateur de projection permet de fouiller la base des annotations RDF en exploitant les liens de spécialisation définis dans le schéma. Les modules de traduction et d'inférence développés pour CORESE seront intégrés dans des agents pour CoMMA.

3. Une gestion distribuée et hétérogène

La mémoire organisationnelle est distribuée et hétérogène. La population de ses utilisateurs est distribuée et hétérogène. Il semble donc intéressant que l'interface entre ces deux mondes soit elle-même de nature distribuée et hétérogène pour pouvoir se calquer sur le paysage d'information d'un côté, et d'un autre côté s'adapter aux utilisateurs.

L'architecture d'un système multi-agents est une structure qui représente les différentes familles d'agents possibles dans le système et leurs relations. Une configuration est une instance d'une architecture avec un arrangement et un nombre d'agents de chaque type choisis. L'idée est de permettre une intégration en douceur du système de gestion de la mémoire, en optant pour une architecture multi-agents qui permettra plusieurs configurations de déploiement et nous autorisera ainsi à nous adapter à la topographie du paysage d'information, au contexte et à l'environnement de déploiement (agencement de l'organisation, topographie et contraintes réseau, localisation des zones d'intérêt et des membres concernés par le système). Les sources d'informations peuvent rester locales et distribuées d'un point de vue maintenance et stockage, mais les agents en s'organisant en sociétés doivent permettre une vue globale et intégrée de la mémoire.

La modularité et l'hétérogénéité des systèmes multi-agents sont donc des caractéristiques appréciables pour s'adapter aux différents terrains de déploiement. L'hétérogénéité du système multi-agents se traduit par l'existence de plusieurs familles d'agents qui collaborent et coopèrent en se basant sur une connaissance mutuelle des compétences de chacun. La variété des agents permet de placer localement des agents dédiés et adaptés aux ressources et aux utilisateurs et d'en faire bénéficier l'ensemble du système. On distingue actuellement dans CoMMA :

- Des agents dédiés à l'ontologie : ils géreront les aspects ontologiques de la gestion de la mémoire, comme par exemple des requêtes sur la hiérarchie des concepts, les termes et les synonymes associés, les différents points de vue possibles.
- Des agents dédiés aux ressources documentaires ou d'annotation : ils s'occuperont de l'exploitation des bases de documents et d'annotations, cherchant à composer des réponses aux requêtes utilisateur.
- Des agents dédiés aux utilisateurs : ils s'occuperont de l'interface utilisateur, proposeront assistance, adaptation et proactivité.
- Des agents dédiés aux connexions : ces agents aideront les échanges et la communication entre

agents en particulier en maintenant des services de pages jaunes (rechercher un agent pouvant fournir un service donné) et pages blanches (rechercher un agent par son identité).

D'un côté, les agents sont les acteurs d'une société et par leur collaboration et leur coopération basées sur l'échange de messages (en utilisant un langage issu de la théorie des actes du langage), ils peuvent résoudre des problèmes complexes nécessitant une vue globale et intégrée de la mémoire d'entreprise. D'un autre côté, ce sont des entités locales et affectées à un utilisateur pour l'assister dans son interaction avec la mémoire ou dédiées à une ressource pour en permettre l'exploitation au bénéfice de l'ensemble de la société des agents. Pour ce faire, les agents intègrent des capacités d'inférences comme celles sur les annotations décrites dans la partie précédente et d'autres comme celles sur les utilisateurs que nous aborderons dans la partie suivante.

Plus de détails peuvent être trouvés dans [Gandon00] où nous montrons en particulier que les règles et l'organisation des sociétés d'agents (hiérarchie, relations pair-pair, réplication totale) permettent d'adapter le système et de prendre en compte la spécificité de chaque mémoire organisationnelle. Des détails sur la plate-forme multi-agent JADE utilisée peuvent être trouvés sur le site [JADE].

4. Une population utilisateur distribuée et hétérogène

La population des utilisateurs d'une mémoire organisationnelle est :

- Hétérogène : Il existe plusieurs profils de membres concernés par la gestion et l'exploitation de la mémoire. Chacun a ses particularités : ses habitudes, ses aptitudes, ses attentes, sa formation, ses besoins, sa position, sa fonction, son ancienneté, ses interactions avec le système... Différents rôles d'utilisateurs identifiés dans CoMMA sont par exemple : des experts source de connaissances, les auteurs d'ontologies, les auteurs de documents et d'annotations, les veilleurs, les utilisateurs finaux consultant la mémoire,...
- Distribuée : Les utilisateurs sont dispersés dans l'infrastructure de l'organisation. Ils ne connaissent par forcément leur existence, leur fonction ou leurs capacités respectives. Ils ignorent en particulier l'existence, l'emplacement et la disponibilité des artefacts et des gens qui leur sont liés par leurs centres d'intérêts.

Comme nous l'avons vu dans la section 2 nous pouvons envisager d'identifier les différents profils d'utilisateurs (intéressés) de l'organisation et d'intégrer à l'ontologie les concepts et relations nécessaires à la description de modèles utilisateurs. Pour adapter le comportement des agents utilisateurs aux profils de ceux-ci et aux différents contextes d'utilisation de la mémoire, nous nous sommes tournés vers les techniques d'apprentissage. Ces techniques visent à

généraliser des lots d'exemples pour produire une connaissance générale réutilisable lorsque de nouveaux cas se présentent. On peut par exemple étudier les chroniques d'événements représentant les actions d'un utilisateur à des niveaux d'abstraction plus ou moins élevés en fonction de ce que l'on souhaite apprendre. Différents types d'apprentissages existent. Pour CoMMA trois types ont été retenus :

- Un apprentissage au vol pour une adaptation rapide à l'utilisateur lors d'une même session
- Un apprentissage différé pour améliorer le comportement du système entre deux sessions, et par exemple proactivement lancer des recherches potentiellement intéressantes.
- Le raisonnement à partir de cas pour prendre en compte des cas ou des aspects spécifiques.

Pour supporter l'émergence de communautés d'intérêt et la diffusion proactive d'information nous nous sommes aussi intéressés à une exploitation collective des profils et des actions :

- En proposant aux utilisateurs de rejoindre des groupes d'intérêt public, gérés par des agents surveillant la mémoire et en charge de détecter et diffuser tout document correspondant au centre d'intérêt qu'ils représentent.
- En introduisant des agents en charge de comparer les profils existants et détecter les similarités pour conseiller et de manière proactive proposer des documents susceptibles d'intéresser les utilisateurs. Cette technique s'inspire de l'automatisation du 'bouche à oreille' comme on le trouve dans les techniques de filtrage collectif.

Ici encore le comportement local et social des agents peut être enrichi par des techniques empruntées à de multiples domaines et ces améliorations bénéficient immédiatement à l'ensemble du système.

5. Conclusion

Nous avons présenté dans cet article l'approche que nous avons choisie dans le projet CoMMA pour matérialiser, gérer et assister l'exploitation d'une mémoire organisationnelle hétérogène et distribuée. L'approche consistant à utiliser des méta-données reposant sur une ontologie a déjà fait ses preuves dans d'autres projets tels que Ontobroker [Dec99], Shoe [Hef99], WebKB [Mar99] and OSIRIX [Rab00]. D'autres projets IST européens s'intéressent actuellement à ces problèmes de gestion et de diffusion des connaissances (ex: On-To-Knowledge, XML-KM, CORMA, PROMOTE, ASSIST, PICK, DECOR... [Cordis]). CoMMA se distingue par son approche basée sur l'intégration de plusieurs technologies émergentes (XML, Systèmes Multi-Agents, Apprentissage Symbolique, Ingénierie des connaissances). Chacune de ces technologies apporte des éléments de solution pour la réalisation, la gestion et l'exploitation d'une mémoire organisationnelle distribuée et d'hétérogène.

6. Remerciements

Nous remercions nos partenaires du projet CoMMA pour les discussions fructueuses que nous avons eues. Nous remercions la Commission Européenne qui subventionne le projet CoMMA

7. Références

- [CoM00] CoMMA Consortium, *Corporate Memory Management through Agents*, Conférence E-Work & E-Business in Madrid, Octobre 2000
- [Cordis] <http://www.cordis.lu/ist/projects.htm>
- [Corby00] O. Corby, R. Dieng, C. Hébert. *A Conceptual Graph Model for W3C Resource Description Framework*. Presented at the 8th International Conference on Conceptual Structures (ICCS'2000), Août 2000, Springer-Verlag Darmstadt, Allemagne
- [Dec99] Decker S., Erdmann M., Fensel D., Studer R., *Ontobroker: Ontology based Access to Distributed and Semi-Structured Information*. Univ. Karlsruhe, Germany, Kluwer Academic Publishers, Boston, 1999.
- [Dieng00] R. Dieng, O. Corby, A. Giboin, J. Golebiowska, N. Matta, M. Ribière *Méthodes et outils pour la gestion des connaissances*, Dunod 2000
- [Gandon00] F. Gandon, R. Dieng, O. Corby, A. Giboin, *A Multi-Agent System to Support Exploiting an XML-based Corporate Memory* Proceedings of PAKM'2000 (Practical Aspects of Knowledge Management), Basel, Suisse Octobre 2000
- [Hef99] Heflin J., Hendler J., Luke S., *SHOE: A Knowledge Representation Language for Internet Applications*. Institute for Advanced Computer Studies, University of Maryland at College Park. 1999.
- [JADE] The JADE Project Home Page. 2000. Available at <http://sharon.csel.it/projects/jade>.
- [Mar99] Martin P., Eklund P., *Embedding Knowledge in Web Documents*, from Griffith University, Australia, 8th International World Wide Web Conference in Toronto 1999
- [Naisbitt] "We are drowning in information but starved of knowledge" - John Naisbitt of Megatrend
- [Rab00] Rabarijaona A., Dieng R., Corby O., Ouaddari R., *Building a XML-based Corporate Memory*, IEEE Intelligent Systems, Special Issue on Knowledge Management and Internet, May-June 2000, p56-64