

SYSTEME MULTI-AGENTS POUR UN WEB SEMANTIQUE D'ENTREPRISE

Fabien Gandon, Olivier Corby
Projet ACACIA – INRIA Sophia Antipolis
Fabien.Gandon@sophia.inria.fr

1 Introduction

CoMMA (Corporate Memory Management through Agents) est un projet IST (1999-12217), subventionné par la Commission Européenne, rassemblant des industriels et des organismes de recherche pour développer et tester un environnement de gestion de la mémoire d'entreprise en s'intéressant en particulier à deux scénarios : l'aide à l'insertion d'un nouvel employé et le support à la veille technologique. Pour atteindre ses objectifs, CoMMA a entrepris de conjuguer au sein d'un système multi-agents (SMA) des techniques de l'ingénierie des connaissances, de la galaxie XML, de la recherche d'information et de l'apprentissage symbolique. Nous présentons ici trois prototypes issus respectivement de nos travaux sur le Web Sémantique et les ontologies, l'utilisation des graphes conceptuels pour manipuler RDF et RDFS et l'intégration de ces technologies aux systèmes multi-agents.

2 Ontologies en RDFS et utilisation de feuilles de style

Extensible Markup Language (XML) est en passe de devenir un standard pour l'échange et le stockage d'informations dans l'entreprise. A la différence de l'HyperText Markup Language (HTML), les balises de XML visent à permettre de décrire la structure des documents, plutôt que leur présentation. La présentation et la structuration du contenu sont complètement indépendantes. Extensible Stylesheet Language Transformations (XSLT) peut être utilisé pour décrire des feuilles de style. Il fournit des outils de manipulation de la structure des documents et des données qui vont au-delà du simple formatage. Nous verrons que ces feuilles de style permettent notamment de naviguer dans l'ontologie en manipulant le RDFS.

RDF (Resource Description Framework) fait partie de la galaxie XML: ce langage utilise une syntaxe XML pour décrire des triplets représentant les propriétés de ressources telles que des images, des pages Web, etc. disponibles sur le réseau et les relations qui existent entre elles. On peut décrire les documents par des annotations sémantiques puis utiliser ces annotations pour fouiller plus efficacement la masse d'information de la mémoire. Les annotations sont basées sur une ontologie décrite et partagée dans le langage RDFS (RDF Schema) proche des langages de représentation par objets. Nous utilisons RDFS pour formaliser le niveau terminologique de l'ontologie (où s'organisent les termes recueillis) et le niveau intensionnel où s'inscrit la structure logique de l'ontologie. Le niveau extensionnel se compose des annotations RDF qui structurent la mémoire, formalisent le modèle d'entreprise et les profils utilisateurs.

Grâce aux feuilles de style XLST nous pouvons générer en permanence des vues de l'ontologie et naviguer à l'intérieur : le lexique de la mémoire, les tables des concepts, des relations et des attributs, un arbre indenté des intensions etc. Un jeu de feuilles permet notamment de rechercher des termes et de naviguer au niveau terminologique ou dans la taxonomie des intensions des concepts ou des relations, représentées à l'écran par les ensembles de termes synonymes qui leurs sont associés. On peut aussi consulter les relations dont la signature porte sur un concept. L'aspect terminologique permet une meilleure navigation ainsi qu'un accès multilingue à l'ontologie et même à la mémoire.

3 Prototype de moteur de recherche sémantique

Afin de manipuler l'ontologie, les annotations, et d'effectuer des inférences, le projet ACACIA a développé CORESE, un prototype de moteur de recherche permettant des inférences sur des annotations RDF en traduisant les triplets RDF en graphes conceptuels

(GC) et vice versa. CORESE combine les avantages d'utiliser d'une part le langage standard RDF pour exprimer et échanger les annotations, et de l'autre les mécanismes de requête et d'inférence disponibles dans le formalisme des GC. Il y a une bonne adéquation entre les deux modèles: aux classes et propriétés de RDFS correspondent les types de concept et les types de relation des GC. Les annotations sont traduites en graphes faits et le schéma en un support dans le formalisme des GC. Les requêtes sont des énoncés RDF avec des caractères joker décrivant l'information cherchée et les valeurs à retourner. La requête de RDF est traduite en GC et projetée sur la base de graphes. Des graphes isolés par projection, sont extraites les valeurs demandées qui sont ensuite retraduites en RDF. Le mécanisme de projection tient compte des hiérarchies et des relations de spécialisation décrites par le support obtenu à partir du schéma en RDFS. Les feuilles de styles sont utilisées pour présenter et documenter le résultat des requêtes en utilisant les définitions et les termes associés aux concepts dans l'ontologie. Dans CoMMA, des modules de CORESE sont intégrés aux agents dédiés à l'ontologie et aux documents, pour leur fournir les capacités techniques nécessaires à leur rôle.

4 Le paradigme multi-agents pour la mémoire d'entreprise

Dans l'optique de CoMMA, la mémoire organisationnelle est distribuée et hétérogène, et 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 d'une part se calquer sur le paysage d'information, et d'autre part s'adapter aux utilisateurs. Le paradigme des SMA représente une nouvelle étape dans l'abstraction de la conception de logiciels, qui peut être employé pour comprendre, modéliser, et développer des systèmes distribués complexes. Le paradigme des SMA apparaît particulièrement adapté au déploiement d'une architecture logicielle au-dessus de ce paysage d'information distribué.

D'un côté, les agents sont les acteurs d'une société qui par leur collaboration et leur coopération basées sur l'échange de messages au niveau sémantique, résolvent des problèmes complexes nécessitant une vue globale et intégrée de la mémoire d'entreprise. *La collaboration intelligente des agents du SMA permet de réaliser une capitalisation globale de la connaissance de l'entreprise.* D'un autre côté, les agents sont des entités localisées 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 possèdent des capacités d'inférences sur les annotations, les modèles, l'ontologie ou les profils utilisateurs et des interfaces graphiques et ils intègrent ses capacités à leur comportement social. *L'autonomie et l'individualité de chaque agent permet de s'adapter localement à la spécificité des ressources et des utilisateurs individuels, tout en en faisant bénéficier l'ensemble de la société.*

Les SMA offrent la perspective d'un réseau de composants logiciels faiblement couplés, les agents, qui fonctionnent ensemble comme une société visant à résoudre des problèmes qui seraient hors d'atteinte pour n'importe quel agent pris individuellement. Pour ce faire, les agents logiciels doivent avoir la capacité d'acquérir des informations sémantiques utiles à partir du contexte du monde dans lequel ils évoluent pour devenir rapidement des acteurs intelligents : les paysages d'information annotés sont, en l'état de l'art actuel, une façon rapide de rendre les agents d'information plus "intelligents". L'idée derrière CoMMA est que si la mémoire d'entreprise devient un monde annoté (en RDF), les agents peuvent utiliser la sémantique des annotations et à l'aide d'inférences assister les utilisateurs dans leurs activités.

Pointeurs

- CoMMA Corporate Memory Management through Agents, *Actes de la Conférence E-Work & E-Business*, Madrid, 2000
- Corby O., Dieng R., Hébert C. A Conceptual Graph Model for W3C Resource Description Framework. *Actes ICCS'2000*, Springer LNAI 1927, p.468-482
- Gandon F., Dieng R., Corby O., Giboin A., A Multi-Agents System to Support Exploiting an XML-based Corporate Memory, *Actes 3rd International Conference on Practical Aspects of Knowledge Management 2000*, p.10-1 to 10-12