



Web Sémantique et Approche Multi-Agents pour la Gestion d'une Mémoire Organisationnelle Distribuée

Fabien Gandon, Rose Dieng-Kuntz, Olivier Corby, Alain Giboin

► To cite this version:

Fabien Gandon, Rose Dieng-Kuntz, Olivier Corby, Alain Giboin. Web Sémantique et Approche Multi-Agents pour la Gestion d'une Mémoire Organisationnelle Distribuée. 13eme IC, Conférence Ingénierie des Connaissances, May 2002, Rouen, France. hal-01146235

HAL Id: hal-01146235

<https://hal.inria.fr/hal-01146235>

Submitted on 28 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Web Sémantique et Approche Multi-Agents pour la Gestion d'une Mémoire Organisationnelle Distribuée

Fabien Gandon, Rose Dieng-Kuntz, Olivier Corby, Alain Giboin

INRIA - Projet ACACIA, 2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis, FRANCE

Résumé

Les organisations ont à gérer une quantité de documents hétérogènes et dispersés de plus en plus importante, et essaient de la transformer en une mémoire organisationnelle active et utile. Nous présentons une approche pour implanter une plate-forme de gestion des ressources d'un intranet, basée sur les technologies issues de travaux sur le Web sémantique et les systèmes multi-agents. La clef de voûte de cette intégration est une ontologie commune et partagée assurant la cohérence de la mémoire et des communications entre agents.

Mots clés : ontologie, Web sémantique, mémoire organisationnelle, système multi-agents.

1 Introduction

En raison d'un renouvellement du personnel de plus en plus rapide, d'un environnement globalisé et changeant qui nécessite une restructuration perpétuelle et une gestion orientée projet, et en raison aussi d'infrastructures grandissantes, les organisations ont besoin d'outils et de méthodologies pour contrôler une mémoire active et persistante de leurs expériences antérieures. Cette mémoire repose de plus en plus souvent sur un usage interne à l'organisation des technologies de l'Internet (les intranets) et du Web (les Webs internes ou intrawebs). Cela mène à des quantités de données et d'informations semi-structurées disponibles en ligne, mais enterrées et dormantes dans leur masse. Dans le projet européen CoMMA, nous évaluons les contributions de plusieurs technologies émergentes au problème d'un intraweb distribué et hétérogène. Des méthodologies et un prototype ont été développés et évalués pour deux scénarios d'application: (1) l'aide à l'insertion de nouveaux employés dans une organisation et (2) l'assistance à la veille technologique. Une ontologie a été conçue et utilisée pour annoter sémantiquement les ressources des intrawebs d'une entreprise dans le domaine des Télécommunications (Deutsch Telekom) et d'un organisme de recherche sur le Bâtiment (CSTB). Nous avons étudié des fonctionnalités de recherche et de diffusion proactive d'informations pour assister l'exploitation de la mémoire. Les différentes tâches d'exploitation ont été assignées aux différents types d'agent développés, dont les instances sont réparties sur l'intranet.

Cet article décrit notre approche en deux sections: la première section présente la notion d'un Web sémantique d'entreprise et la deuxième décrit l'architecture multi-agents dédiée à la gestion de la mémoire. Tout le long de l'article, nous nous attacherons à souligner le rôle crucial joué par l'ontologie constituant une base commune transversale au système CoMMA, tant lors de sa conception que lors de son utilisation.

2 Vers un intraweb sémantique

Une mémoire organisationnelle est, par nature, un paysage d'informations hétérogènes et distribuées. Les mémoires organisationnelles ont maintenant à faire face aux mêmes problèmes que le Web en matière d'exhaustivité et de précision lors de recherche d'information. Cette section montre comment les techniques du Web sémantique peuvent aider à résoudre ces problèmes.

2.1 Le concept d'un Web sémantique

Le Web sémantique [3] est une approche prometteuse où la sémantique du contenu des documents est rendue explicite par des annotations utilisées pour guider une exploitation ultérieure, par exemple pour la recherche d'informations sur le Web. RDF (Resource Description Framework [22]) est à la base du Web sémantique et est doté d'une syntaxe XML.

RDF nous permet d'annoter sémantiquement les ressources de la mémoire: il propose un modèle de données simple fournissant un langage de représentation des propriétés et des relations des ressources du Web. Nous étudions alors la mémoire d'une entreprise comme un Web sémantique d'entreprise ou intraweb sémantique: nous décrivons le contenu sémantique des documents de l'entreprise par des annotations sémantiques utilisées ensuite avec des inférences pour fouiller plus intelligemment la masse d'informations de la mémoire organisationnelle. De même qu'une caractéristique importante d'un nouveau logiciel est sa capacité à intégrer ou s'interfacer avec des systèmes existants, une caractéristique importante d'un système de gestion d'une mémoire organisationnelle est sa capacité à intégrer des archives existantes. Puisque les annotations de RDF peuvent être internes ou externes aux documents, des documents existants de la mémoire peuvent être gardés intacts et dotés d'annotations externes.

RDF ne fait aucune hypothèse sur le domaine d'application ; les annotations sont basées sur une ontologie formalisée en RDF Schema [6] (RDFS). L'approche consiste à (1) indiquer les concepts de la mémoire envisagée et leurs relations dans une ontologie formalisée dans un schéma RDFS, (2) annoter les documents de la mémoire en RDF en utilisant le vocabulaire conceptuel fourni par le schéma, (3) exploiter les annotations pour fouiller la mémoire organisationnelle. Nous allons détailler chacune de ces trois étapes.

2.2 L'ontologie O'CoMMA

Nous avons proposé une méthode pour construire des ontologies qui a été appliquée pour obtenir l'ontologie O'CoMMA [17][18]. Le résultat est une ontologie contenant 470 types de concepts organisés en une hiérarchie d'une profondeur maximale de 13 liens de subsomption, 79 types de relations formant une hiérarchie d'une profondeur maximale de 2 liens de subsomption. O'CoMMA utilise 715 termes anglais et 699 termes français pour étiqueter ces primitives ainsi que 547 définitions en Français et 550 définitions en Anglais.

La figure 1 montre les trois couches structurant l'ontologie:

- une couche supérieure très générale qui ressemble aux autres ontologies de haut niveau ;
- une couche médiane assez importante, divisée en une partie générique concernant le domaine des mémoires organisationnelles (documents, organisations, personnes...) et une autre consacrée au domaine d'application (ex. pour les télécommunications: WAP, mobile, etc.) ;
- une couche d'extensions spécifiques au scénario et à l'entreprise, avec des concepts complexes (ex. rapport d'analyse de tendance, carte d'itinéraire du nouvel employé).

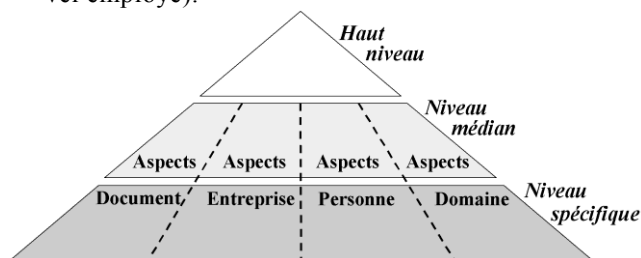


FIG. 1 – Structure de O'CoMMA

La partie supérieure, qui est très abstraite, et la première partie de la couche médiane, qui décrit des concepts communs aux applications de mémoire d'entreprise, sont réutilisables dans d'autres scénarios d'applications concernant la mémoire d'entreprise. La deuxième partie de la couche médiane, qui traite du domaine d'application, n'est réutilisable que pour des scénarios dans le même domaine d'application. La dernière couche contenant des concepts spécifiques n'est plus réutilisable dès que l'organisation, le scénario ou le domaine d'application change. Cependant, cette dernière couche est de loin la plus intéressante pour l'utilisateur

car elle recèle les concepts qu'il manipule dans son travail quotidien.

En RDFS, les concepts sont formalisés comme des classes et les relations ou attributs sont formalisés comme des propriétés. Des instances de ces classes et propriétés sont créées pour formuler des annotations. Etant donné que la multi-instanciation est autorisée, une ressource peut être une instance de plusieurs classes. Un objet global tel que l'activité "la recherche" est formalisé comme une ressource dont l'identificateur (URI) est celui de sa classe. Cette convention assure une unicité universelle. Les termes sont formalisés en RDFS comme les labels des classes et des propriétés et sont indépendants de l'identificateur système unique de ces notions. De même, les définitions en langue naturelle sont capturées en RDFS comme des commentaires sur les classes ou propriétés. La possibilité d'indiquer la langue naturelle utilisée nous permet d'avoir une ontologie multilingue. Une notion liée à plusieurs termes est caractéristique de la synonymie de ces termes. Un terme lié à plusieurs notions est ambigu.

En utilisant des feuilles de style XSLT, nous reproduisons des vues de documents intermédiaires créés lors de la construction de l'ontologie et nous offrons à l'utilisateur de nouvelles vues permettant la navigation dans l'ontologie:

- le tableau terminologique initial représentant un lexique de la mémoire ;
- les tableaux de concepts et de propriétés ;
- des pages de navigation et de recherche dans les notions aux niveaux conceptuel et terminologique: elles permettent de chercher des notions dénotées par un terme, de naviguer dans leurs descriptions (figure 3) en suivant les liens de subsomption, de chercher des relations ayant une signature compatible avec un concept donné ;
- une vue de l'ontologie filtrée par un profil utilisateur pour proposer des points d'entrée préférés dans l'ontologie afin d'amorcer une navigation sur le modèle de MyYahoo ;
- l'arbre indenté des notions (figure 2).

L'utilisateur peut également demander la liste des instances d'une notion, un échantillon d'instances jouant le rôle d'exemples pour améliorer la compréhension d'une définition.

Le choix de XML nous permet de baser notre système sur une norme qui bénéficie de toutes les technologies du Web pour la gestion des aspects réseau, affichage et navigation, et ceci est un atout majeur pour l'intégration à l'intraweb d'une entreprise.

2.3 Intra-Web basé sur un modèle

Les utilisateurs de la mémoire sont, par nature, hétérogènes dans leur profil et répartis dans l'organisation. Afin de donner au système un aperçu de son environnement et des utilisateurs avec lesquels il interagit, la mémoire est basée sur des modèles de la structure organisationnelle et sur des profils utilisateur afin. Ces modèles permettent la personnalisation et l'apprentissage des

préférences des utilisateurs, et une diffusion intelligente et proactive de l'information, basée sur les centres d'intérêt des utilisateurs. Nous verrons que l'ontologie fournit les primitives pour la description des utilisateurs et de l'entreprise.

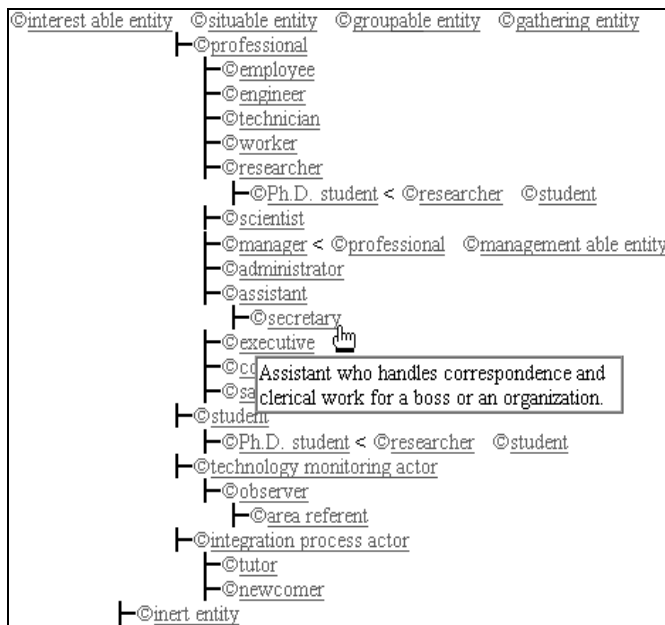


FIG. 2 – Navigation dans les hiérarchies de O'CoMMA

personne : personne, humain, etre humain,

Inherits from : @chose @entite @entite de role @entite dirigeable

Class ID : Person - See Instances - See available relations on that concept

Switch to Francais - -

Natural Language definition :
Entite vivante appartenant a l'humanite, un etre humain individuel.

More general notion : (8)

- © entite dirigeable : entite dirigeable,
- © entite capable d administrer : entite capable d administrer,
- © entite capable d activite : entite capable d activite,
- © etre vivant : etre vivant, entite vivante,
- © entite capable d etre interessee : entite capable d etre interessee,
- © entite localisable : entite localisable,
- © entite groupable : entite groupable,
- © entite de rassemblement : entite de rassemblement,

More precise notions : (4)

- © acteur de la veille technologique : acteur de la veille technologique,
- © acteur du processus d integration : acteur du processus d integration,
- © etudiant : etudiant,
- © professionnel : professionnel,

FIG. 3 – Affichage de la classe 'Personne' de O'CoMMA

2.3.1 Description des profils utilisateur : "annoter les personnes"

Un profil d'utilisateur capture les aspects de l'utilisateur qui ont été identifiés comme pertinents (et formalisables) pour le comportement du système. Il contient des informations administratives et les préférences explicites qui vont de la personnalisation d'interface aux centres d'intérêt. Il positionne également l'utilisateur dans l'organisation (rôle, position, réseau de connaissance, etc.) permettant au système de cibler ses actions de diffusion. En outre, le système déduit des informations à partir d'une session d'utilisation: pour ce faire il mémorise l'historique des documents visités et, le cas échéant, l'évaluation effectuée par l'utilisateur sur la pertinence et la qualité des documents proposés. Grâce à l'exploitation de ces informations, le système apprend certaines préférences de l'utilisateur. Il les utilise ensuite pour la présentation de résultats ou pour la diffusion proactive d'informations. L'adaptation à l'utilisateur se fonde sur des techniques d'apprentissage symbolique [21] et le profil est matérialisé et échangé sous la forme d'une annotation RDF sur l'utilisateur concerné. Le vocabulaire conceptuel utilisé par cette annotation est donné par l'ontologie O'CoMMA.

2.3.2 Description de l'entreprise : "annoter l'organisation"

Un modèle d'entreprise est une représentation explicite et focalisée de l'organisation. Jusque très récemment, la modélisation d'entreprises était principalement destinée à la simulation et l'optimisation des systèmes de production. Elle fournissait des bancs d'essai pour les processus d'affaires et était utilisée pour leur rétro-ingénierie. Mais le changement des règles du jeu des marchés a amené les organisations à se rendre compte de la valeur de leur mémoire et du fait que les modèles d'organisation ont aussi un rôle à jouer dans cette application [24].

Dans CoMMA, le modèle vise à supporter les activités de la mémoire d'entreprise survenant dans nos scénarios d'application. Ce modèle donne au système une vue de son contexte de fonctionnement et de l'environnement organisationnel. Ainsi, il peut exploiter les aspects décrits dans ce modèle pour l'interaction entre les agents et surtout entre les agents et les utilisateurs. Il est nécessaire, pour le système, d'avoir une compréhension de l'environnement qu'est l'organisation, de ses buts et de sa politique, de sorte que le système résultant fonctionne effectivement en collaboration avec les agents humains afin d'atteindre un objectif commun [23].

Nous utilisons RDF pour implanter notre description de l'organisation, en annotant les entités organisationnelles (services, laboratoires, activités...) avec leurs relations (contrôle, utilise, est inclus, s'intéresse à, etc.). Là encore, les concepts et les relations utilisés sont définis par l'ontologie.

2.3.3 Architecture de la mémoire

Dans leur article à propos des agents dans des mondes virtuels annotés, Doyle et Hayes-Roth [12] ont expliqué

que les environnements annotés contenant des explications à propos des buts, des utilisations et des activités possibles dans les mondes où ils évoluent, permettent à des agents de devenir très rapidement des acteurs intelligents dans ces espaces. Les paysages d'information annotés permettent ainsi de rendre des agents d'information plus intelligents. Si la mémoire organisationnelle devient un monde annoté, les agents peuvent utiliser la sémantique des annotations et par des inférences, assister les utilisateurs dans leur exploitation de la mémoire d'entreprise.

Avec RDF(S), nous formalisons une ontologie, nous décrivons la teneur des documents et des descriptions de l'environnement organisationnel à travers des annotations sémantiques en RDF (cf. figure 4) puis nous utilisons et nous inférons à partir de ces annotations pour intelligemment explorer la masse d'informations de la mémoire.

Cette approche amène la mémoire d'entreprise à être matérialisée comme un Web sémantique d'entreprise ou intraweb sémantique. Les vignettes de la figure 5 décrivent et commentent une telle structure.

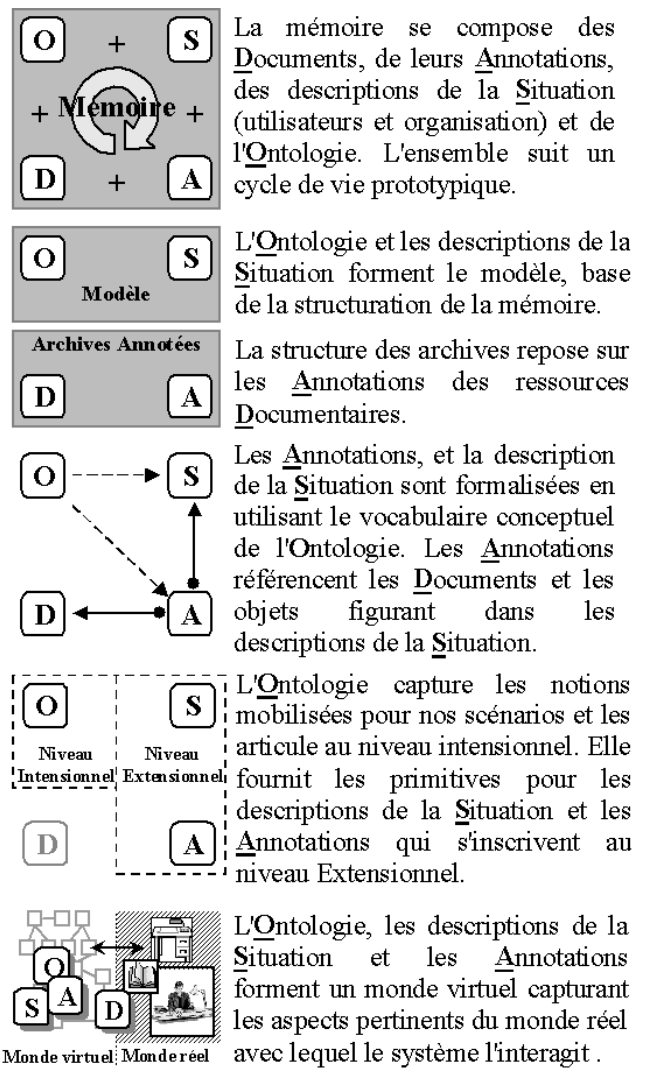


FIG. 5 – La structure de la mémoire

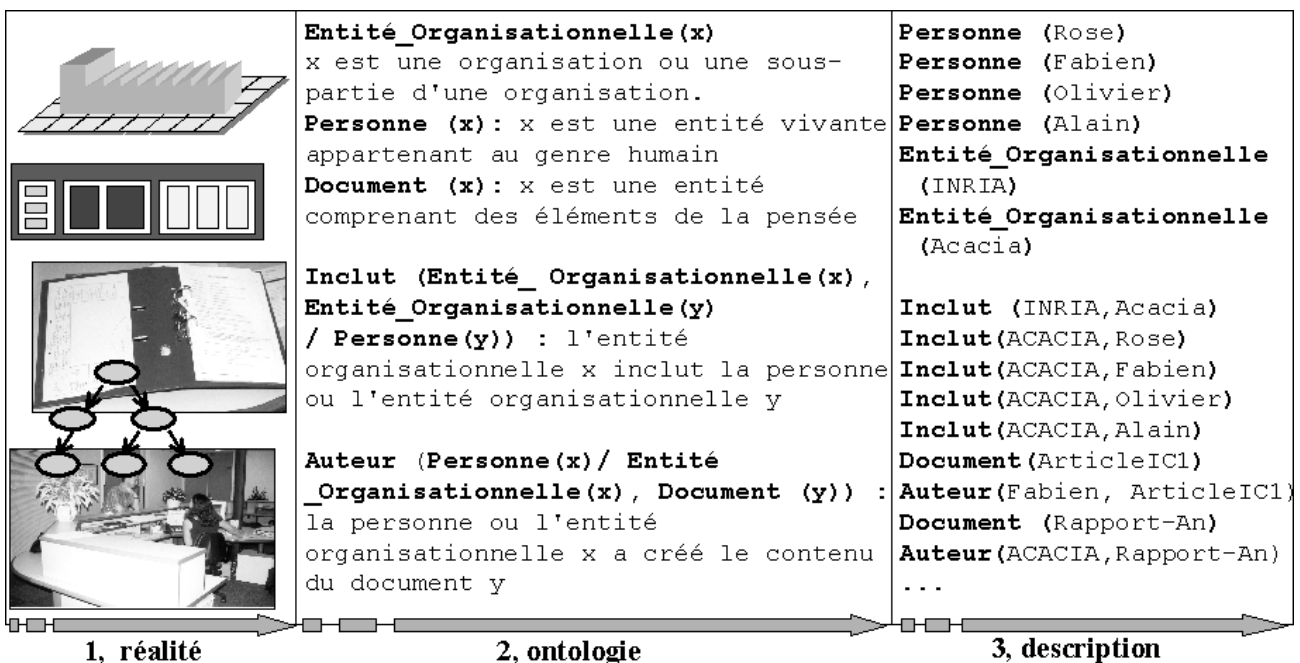


FIG. 4 – Etapes de modélisation et structuration

2.3.4 CORESE: moteur de recherche sémantique

Les moteurs de recherche par mots-clés sont limités aux termes, partie visible mais ambiguë, dénotant l'utilisation des notions. L'introduction des ontologies permet aux logiciels de raisonner au niveau sémantique. Afin de manipuler l'ontologie et les annotations, nous avons développé CORESE [7] un prototype de moteur de recherche permettant de faire des inférences sur des annotations RDF. Contrairement à SiLRi [9] qui repose sur une traduction de RDF(S) en Frame Logic, CORESE combine les avantages d'utiliser le langage RDF(S)/XML pour exprimer et échanger des métadonnées, et les mécanismes de requête et d'inférence disponibles pour le formalisme des Graphes Conceptuels (GCs) [25]. Il existe une véritable adéquation entre les deux modèles: les annotations RDF sont traduites en graphes-faits dans les GCs en n'utilisant que des relations binaires ; la hiérarchie des classes et celle des propriétés décrites dans un schéma RDFS sont traduites en une hiérarchie de types de concepts et une hiérarchie de types de relation dans le formalisme des GCs.

Une requête CORESE est un énoncé RDF utilisant des variables pour décrire les énoncés recherchés, les valeurs à retourner et les liens de co-références contraignant ces valeurs ; les expressions régulières sont utilisées pour contraindre les valeurs littérales et des opérateurs supplémentaires sont utilisés pour exprimer la disjonction et la négation. La requête RDF est traduite en un graphe requête que l'on projette sur la base des graphes-faits. On isole des graphes correspondants pour en extraire les valeurs demandées; ces graphes réponse sont alors traduits de nouveau en RDF (Figure 6).

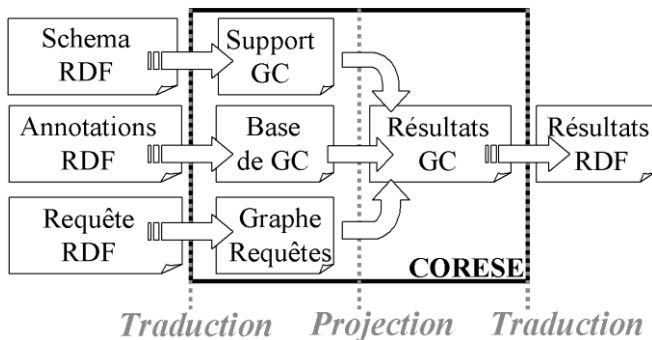


FIG. 6 – Principe de CORESE

Le mécanisme de projection tient compte des liens de spécialisation décrits dans les hiérarchies traduites à partir du schéma. Ainsi le moteur de recherche exploite les connaissances ontologiques pour améliorer le taux rappel de sa recherche, par exemple: si un utilisateur demande les documents concernant l'intelligence artificielle et que le système connaît un document concernant les systèmes multi-agents, il sera capable d'utiliser l'ontologie pour inférer que les systèmes multi-agents sont une branche de l'intelligence artificielle et donc que ce document est une réponse pertinente.

CORESE apporte aussi une solution aux limitations d'expressivité qui ont été rencontrées lors de la formalisation de l'ontologie en RDFS, en particulier pour formaliser des connaissances de base ou connaissances implicites.

Par exemple, quand nous déclarons que quelqu'un dirige un groupe il est implicite que cette personne est un directeur. Ainsi le concept de 'directeur' devrait être un concept défini, c'est-à-dire un concept ayant une définition formelle explicite lui permettant d'être dérivé d'autres concepts existants. Cependant la notion de concept défini n'existe pas en RDFS.

Un deuxième exemple est la notion de 'collègue', collectée pour le scénario de veille technologique. Cette notion a d'abord été intégrée dans l'ontologie en tant que concept, mais personne n'est un 'collègue' en soi ! Nous avons donc introduit une relation 'x est un collègue de y'. Cette relation nécessitait d'être typée 'relation symétrique', mais RDFS ne fournit pas cette primitive. Enfin, en considérant les scénarios d'application, il s'est avéré que cette relation ne serait pas utilisée pour l'annotation mais qu'elle serait inférée de ce qui est décrit de la structure organisationnelle (ex: M. Corby et Mme Dieng-Kuntz appartiennent au projet ACACIA donc il existe une relation 'collègue' entre M. Corby et Mme Dieng-Kuntz). Nous avons donc besoin de pouvoir définir formellement les conditions nécessaires et suffisantes pour que deux personnes soient collègues et de pouvoir changer cette définition si nécessaire (par exemple pour une autre organisation). Ces limitations de RDFS deviennent rapidement un problème puisque la capacité à factoriser la connaissance dans une ontologie exige la possibilité de décrire des définitions formelles.

Dans la version actuelle, les définitions formelles sont représentées dans des règles écrites dans un langage de règles RDF/XML spécialement créé pour RDF(S) et CORESE. Un moteur d'inférence exploite ces règles pour compléter la base d'annotations avec des faits implicites déductibles. Les caractéristiques de symétrie, de transitivité et de réflexivité des propriétés ont exigé des extensions de RDFS spécifiques à CORESE. Au lieu d'utiliser des règles nous pourrions également étendre le modèle de RDFS pour ajouter l'expressivité manquante comme le propose DRDF(S) [10], OIL [15], ou DAML+OIL [28]. La figure 7 montre la règle utilisée pour définir une condition suffisante de la propriété 'Collègue'.

L'ontologie est un puissant objet conceptuel pour la modélisation. Sa complexité nous a amenés à développer des méthodologies et des outils d'aide à sa conception qui doivent tendre vers des plates-formes intégrées supportant le cycle de vie complet de ces objets vivants. Cependant cette complexité se manifeste aussi par une augmentation de la complexité des solutions dans lesquelles les ontologies sont introduites. En particulier le fossé entre les modèles conceptuels sous-jacents à un système et les préoccupations quotidiennes d'un utilisateur lambda de cette solution se creuse dangereusement.

```

<cos:rule>
  <cos:if>
    <rdf:RDF>
      <CoMMA:OrganizationalEntity>
        <CoMMA:Include> <CoMMA:Person rdf:about="?x"/> </CoMMA:Include>
        <CoMMA:Include> <CoMMA:Person rdf:about="?y"/> </CoMMA:Include>
      </CoMMA:OrganizationalEntity>
    </rdf:RDF>
  </cos:if>

  <cos:then>
    <rdf:RDF>
      <CoMMA:Person rdf:about="?x">
        <CoMMA:Colleague> <CoMMA:Person rdf:about="?y"/></CoMMA:Colleague>
      </CoMMA:Person>
    </rdf:RDF>
  </cos:then>
</cos:rule>

```

FIG. 7 – Règle définissant la propriété 'Collègue'

En réponse, plusieurs interfaces ont été étudiées pour CORESE afin de permettre à un utilisateur de manipuler l'ontologie pour différentes tâches des scénarios d'application. Deux d'entre elles ont inspiré l'interface de formulation d'une annotation ou d'une requête de CoMMA.

La figure 8 montre la première interface exploitant la métaphore du shopping en ligne et la technique du "drag & drop": l'utilisateur remplit son panier avec des notions qu'il sélectionne par une recherche par mot-clefs dans l'ontologie. Le panier donne ensuite une requête par conjonction des notions sélectionnées.

La figure 9 montre la deuxième interface, dite experte, permettant de construire une requête quelconque. Elle est intéressante pour des utilisateurs habitués à la recherche de documents, par exemple dans le cadre du scénario de veille technologique, les documentalistes ont demandé une interface comme celle-ci pour construire leurs requêtes.

L'interface finale de CoMMA (figure 12) est un croisement entre ces deux interfaces. L'utilisateur construit sa requête ou son annotation par parcours en Z de l'interface au cours duquel: (1) il donne un mot-clef (2) il choisit une notion parmi celles qui peuvent être dénotées par ce terme (3) il positionne cette notion dans la structure de son annotation/ sa requête (4) choisit des attributs littéraux et leur valeurs (ex. nom, titre, etc.)

Bien que CORESE puisse être utilisé en mode client-serveur, il offre également une API ; ainsi, dans CoMMA, des modules de CORESE sont intégrés dans les agents dédiés à l'ontologie, aux modèles et aux documents, afin de leur procurer les compétences nécessaires pour leurs rôles. La section suivante va détailler l'architecture multi-agents choisie pour CoMMA et le rôle de l'ontologie dans ce cadre.

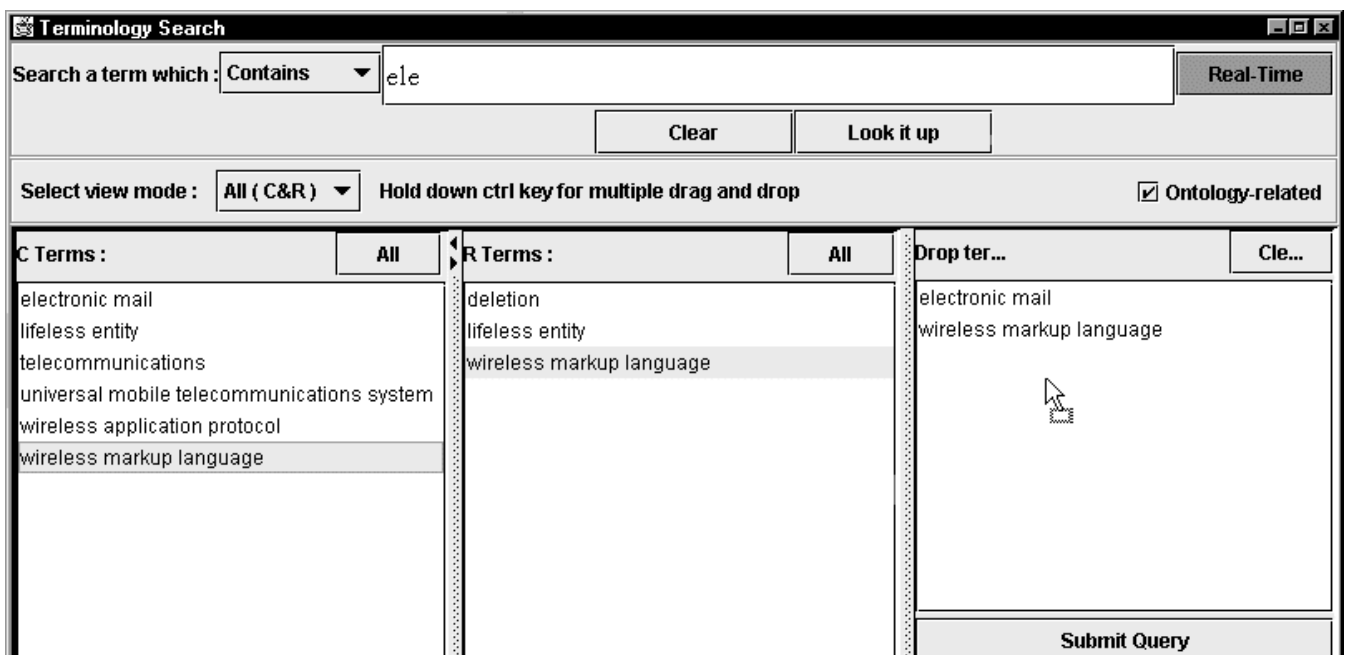


FIG. 8 – Interface 'Shopping dans l'ontologie'

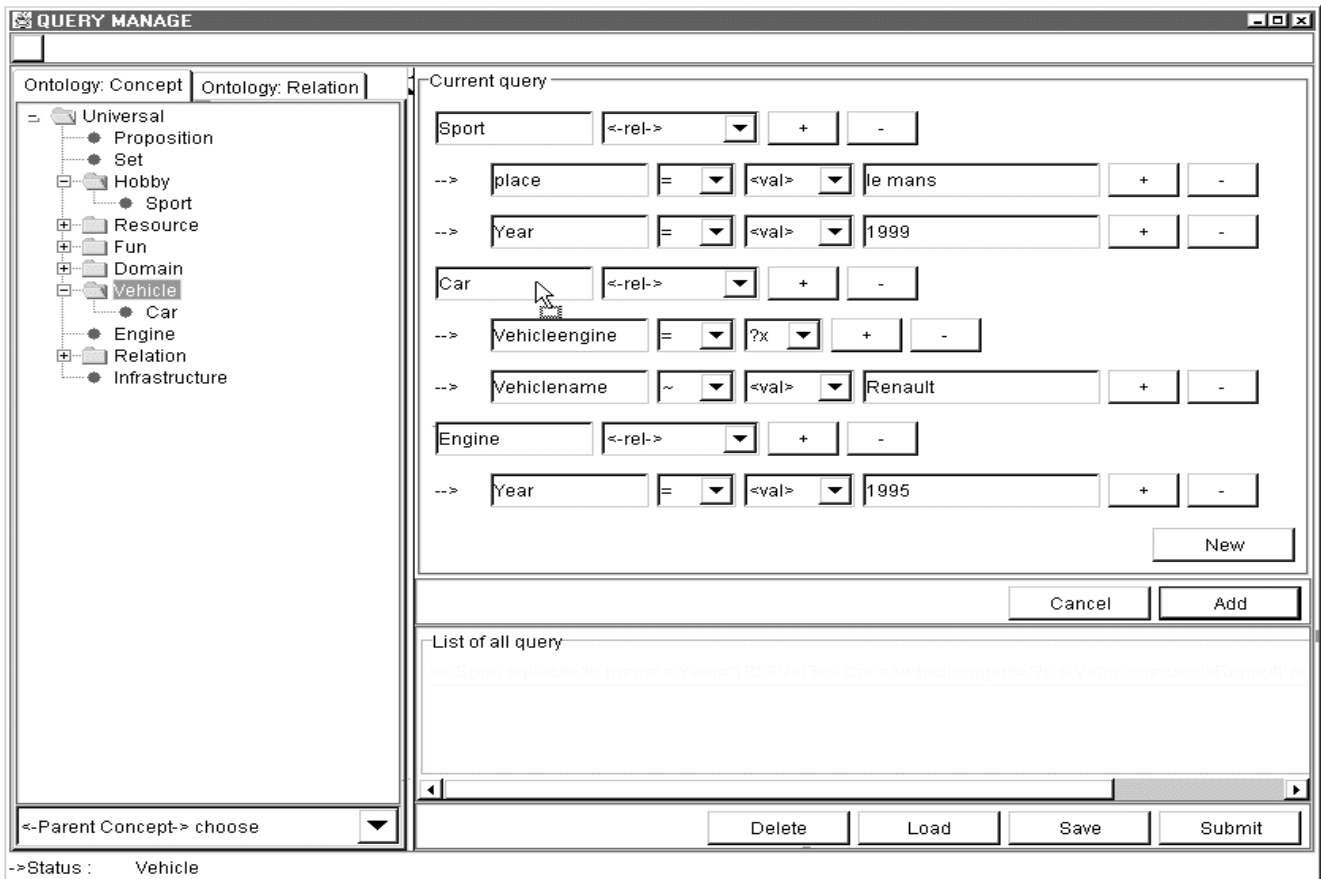


FIG. 9 – Interface complexe mais complète

3 Une Architecture Logicielle Multi-Agents

Les tâches à exécuter sur la mémoire organisationnelle, la mémoire elle-même et la population des utilisateurs sont distribuées et hétérogènes. Par conséquent, il est intéressant que l'architecture logicielle du système soit elle-même hétérogène et distribuée.

Les progrès de la programmation ont été réalisés à travers des abstractions de plus en plus élevées nous permettant de modéliser des systèmes de plus en plus complexes. Les Systèmes Multi-Agents (SMA) sont candidats à être une nouvelle étape dans les niveaux d'abstraction, pour comprendre, modéliser et développer des systèmes répartis [29]. De plus, les SMA sont reconnus comme une architecture logicielle possible pour supporter le déploiement du Web sémantique [3].

Dans CoMMA, le paradigme SMA apparaît très adapté au déploiement d'une architecture logicielle au-dessus du paysage d'informations distribuées qu'est la mémoire d'entreprise: d'une part, les différents agents s'adaptent localement aux utilisateurs et aux ressources auxquels ils sont dédiés ; d'autre part, grâce à leur coopération, les agents logiciels répartis sur l'intranet nous proposent une vue intégrée et globale de la mémoire organisationnelle matérialisée comme un intraweb sémantique.

3.1 Système d'information multi-agents

Les agents d'information sont une sous-catégorie des agents intelligents. Un SMA est un réseau d'agents faiblement couplés, qui fonctionnent ensemble comme une société visant à résoudre des problèmes qui seraient généralement au-delà des capacités de tout agent pris individuellement. Un SMA est hétérogène quand il inclut des agents d'au moins deux types. Un Système d'Information Multi-Agents (SIMA) est un SMA visant à fournir une partie ou une gamme complète de fonctionnalités pour gérer et exploiter des ressources d'information. Lors d l'application des SIMA aux mémoires organisationnelles, la coopération des agents vise à améliorer la capitalisation de la connaissance dans l'organisation. *L'architecture logicielle de CoMMA est ainsi celle d'un SIMA hétérogène dédié à la gestion d'un intraweb sémantique.*

Une architecture SMA est une structure qui dépeint les différentes familles d'agents et leurs rapports. Une configuration est l'instanciation d'une architecture avec un agencement choisi et un nombre approprié d'agents de chaque type. Pour une architecture donnée, on peut générer plusieurs configurations, et une configuration donnée est étroitement liée à la topographie et au contexte de l'endroit où elle est déployée (structure de l'or-

ganisation, caractéristiques de l'Intranet, localisation des intéressés, etc.). Ainsi, l'architecture doit être conçue de sorte que l'ensemble des configurations possibles couvre les différents contextes organisationnels envisageables. La configuration est étudiée et documentée lors du déploiement alors que la description architecturale est étudiée et fixée lors de la conception.

3.2 L'architecture de CoMMA

L'analyse architecturale part du niveau d'abstraction le plus élevé (*i.e.* la société) et par raffinements successifs (*i.e.* sociétés imbriquées), elle descend jusqu'au point où les rôles et les interactions entre agents peuvent être identifiés.

3.2.1 Du niveau macroscopique du SMA au niveau microscopique des agents

Nous avons adopté une approche organisationnelle : l'architecture SMA est abordée, comme une société humaine, en termes de rôles et de rapports. Dans cette société artificielle, l'objectif commun est la gestion et la circulation des connaissances distribuées, rendues possibles par une ontologie commune et partagée.

Les spécifications fonctionnelles du système ne se calquent pas simplement sur des fonctionnalités d'agents mais ont une influence sur les interactions sociales des différents agents et sur l'ensemble des capacités, des rôles et des comportements qui leur sont attachés. En considérant les fonctionnalités du système, nous avons identifié trois sociétés d'agents dédiées aux ressources (l'ontologie et les modèles ; les annotations ; les pages jaunes pour gérer les interconnexions entre agents) et une dédiée aux utilisateurs (figure 10).

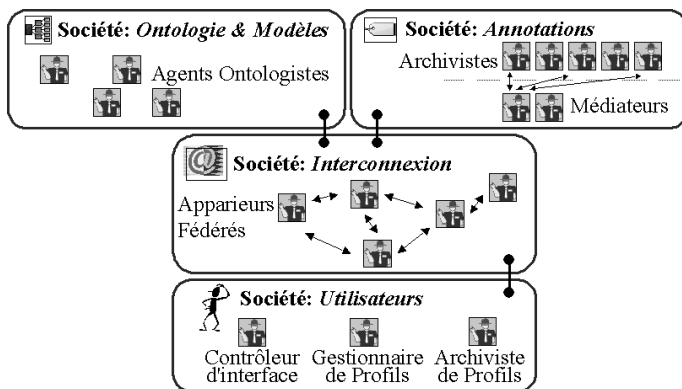


FIG. 10 – Sociétés pour la gestion de la mémoire

En analysant les sociétés dédiées aux ressources, nous avons identifié un ensemble récurrent d'organisations possibles: hiérarchique, égalitaire ou réplication. Selon le type de tâches devant être exécutées, et selon la taille et la complexité des ressources manipulées, nous avons préféré telle ou telle organisation de la société d'agents.

La société dédiée à l'ontologie et aux modèles est actuellement organisée comme une société de réplication : elle rassemble des agents ontologistes ayant chacun une copie complète de l'ontologie.

La société dédiée aux annotations est une organisation hiérarchique : des agents médiateurs d'annotations sont responsables d'allouer les nouvelles annotations et les tâches de résolution distribuée des requêtes aux agents archivistes d'annotations.

Les agents dédiés aux pages jaunes sont dans une organisation égalitaire : ils sont fournis par la plate-forme JADE [2] utilisée pour CoMMA. Les agents de cette société sont appelés 'apparieurs fédérables' (c.f. figure 14) : ils coopèrent pour apparier les agents demandeurs d'un service de la mémoire et les agents susceptibles de fournir un tel service. La description de ces services repose elle aussi sur une ontologie consensuelle aux développeurs.

Les agents de la société dédiée à l'utilisateur sont concernés par l'interface, la surveillance, l'aide et l'adaptation à l'utilisateur. Ils sont typiquement demandeurs de services et de ressources. Puisqu'ils ne sont pas liés à un type de ressource comme dans les sociétés précédentes, ils ne peuvent pas être étudiés en utilisant notre typologie. Nous pouvons cependant distinguer au moins deux rôles récurrents dans ce type de société:

- la gestion d'interface utilisateur: pour dialoguer avec les utilisateurs, leur permettre d'exprimer leurs requêtes puis de les raffiner, et présenter les résultats dans un format adéquat ;
- la gestion du profil utilisateur: pour archiver et rendre les profils disponibles afin qu'ils soient utilisables pour l'interface, les techniques d'apprentissage symbolique et les processus de diffusion proactive de l'information.

3.2.2 Rôles, interactions et protocoles

L'analyse de l'architecture nous a permis d'identifier des rôles d'agents et nous pouvons alors étudier leurs caractéristiques et leurs interactions pour pouvoir implanter les comportements correspondants dans un ensemble de types d'agent.

Les rôles représentent la position d'un agent dans une société, les responsabilités et les activités assignées à cette position et considérées par les autres agents comme étant remplies. Ces activités, couplées aux interactions, participent à la gestion de la mémoire distribuée. Dans [19], nous donnons une description des rôles identifiés et leurs caractéristiques selon des critères encore largement discutés dans la communauté des systèmes multi-agents. La notion d'ontologie peut aider à définir les caractéristiques des rôles. Ceci est particulièrement intéressant lorsque le système envisagé est un système ouvert, c'est-à-dire que des agents de type inconnu peuvent se joindre au système. Il est alors nécessaire de disposer de primitives de description des services offerts et des besoins émis par les agents afin de leur permettre de s'identifier et de se contacter. L'ontologie est typiquement un objet capturant ce type de primitives.

Après l'identification des rôles vient la spécification des interactions entre rôles. Le modèle de chaque conversation doit être indiqué avec son protocole. Les agents doivent se conformer aux protocoles afin que le

SMA fonctionne correctement. La définition d'un protocole commence par un graphe d'accointances au niveau des rôles, celui-ci est un graphe orienté identifiant des canaux de communication entre les agents jouant les rôles considérés. Nous indiquons alors l'ordre possible des messages échangés. Le réseau d'accointances et les protocoles dérivent de l'analyse architecturale et des cas d'utilisation isolés à partir des scénarios d'application.

Le protocoles utilisent des primitives (ex: inform, cfp, accept, reject...) qui dénotent des actes du langage spécifiés et normalisés dans une ontologie recommandée par le consortium FIPA [16]. Tout agent utilisant et suivant ces protocoles est en droit d'attendre un comportement 'compréhensible' de la part de ses interlocuteurs.

Plusieurs ontologies sont utilisées pour la rédaction d'un message. Un message envoyé par un agent est décomposable en trois niveaux:

- le premier niveau utilise des primitives standardisées par l'ontologie de FIPA pour décrire la nature d'un message entre deux agents.
- le deuxième niveau utilise des primitives consensuelles entre les programmeurs du prototype de CoMMA ; elles sont capturées par une ontologie gérée par la plate-forme JADE [2] et fournissent un vocabulaire spécialisé pour les actes du langage utilisés dans la gestion d'un intraweb sémantique.
- le troisième niveau utilise des primitives de O'CoMMA pour décrire les annotations recherchées.

En partant des descriptions des rôles et des interactions, les différents membres du consortium CoMMA ont proposé et implanté des types d'agent qui remplissent un ou plusieurs rôles impliqués dans la gestion et l'exploitation de la mémoire.

L'aspect faiblement couplé des agents et le consensus ontologique supportant un couplage au niveau sémantique ont permis de répartir le développement tout en assurant une bonne facilité d'intégration pour le prototype final. Les comportements des agents proviennent des choix d'implantation déterminant les réponses, les actions et les réactions de l'agent. L'implantation d'un comportement d'agent est contrainte par le rôle qu'il joue et est sujette à la boîte à outils des techniques disponibles aux concepteurs.

Le faible couplage a permis à chacun des partenaires d'intégrer dans ses agents des techniques de son domaine d'expertise (ex: apprentissage symbolique, etc.) tout en permettant au système intégré de bénéficier dans son ensemble de l'association de ces techniques pour un objectif commun: gérer et diffuser la connaissance manipulée. Chaque partenaire est tenu de programmer le comportement de son agent afin qu'il réponde à tous les messages qui peuvent lui être envoyés, sachant que ceux-ci utiliseront toujours les primitives de l'ontologie qui est disponible pour tout agent par simple demande auprès de l'agent ontologiste.

3.2.3 Exemple: la société dédiée aux annotations

L'enjeu de cette société est de trouver des mécanismes pour décider où stocker les annotations nouvellement soumises et comment distribuer une requête pour ne pas laisser échapper des réponses simplement parce que l'information nécessaire est dispersée entre plusieurs bases d'annotation. Ces deux facettes de la distribution sont liées puisque le processus de résolution distribuée d'une requête dépend étroitement du choix fait pour la distribution des annotations. Dans cette société le médiateur d'annotation (AM) est responsable de la gestion de la dispersion des annotations ; ces annotations étant réparties entre des archivistes d'annotation (AAs). Les mécanismes décrits ici sont détaillés dans [20].

Pour allouer une annotation nouvellement soumise, l'AM émet un appel à proposition à l'attention des AAs. Un protocole d'interaction correspondant à un 'contract-net' [16] imbriqué est utilisé pour déterminer lequel des AAs gagnera l'annotation nouvellement soumise. En réponse, chaque AA mesure la proximité sémantique entre l'annotation et les types des concepts et de relations actuellement dans ses archives. L'AA le plus proche gagne le contrat. Ce protocole permet de spécialiser les bases d'annotations sémantiques de la mémoire et d'entretenir cette spécialisation. Pour ce protocole, nous avons donc défini une pseudo-distance en utilisant la hiérarchie de l'ontologie et une distance lexicographique pour les valeurs littérales. Nous l'employons pour comparer les offres des différents AAs. *L'ontologie est ainsi utilisée comme un espace commun permettant de définir des (pseudo-)distances communes dont les résultats sont comparables.* Le consensus ontologique fournit donc une base pour d'autres consensus (ex: un consensus sur le calcul d'une métrique partagée). *L'objet 'ontologie' est donc la pierre de touche de tous les mécanismes distribués d'une gestion intelligente des connaissances dispersées.*

La résolution d'une requête peut impliquer plusieurs bases d'annotations réparties entre plusieurs AAs ; le résultat est une fusion de résultats partiels. Pour déterminer si et quand un AA doit participer à la résolution d'une requête, les AAs calculent le recouvrement entre la liste des notions actuellement utilisées dans leur base et la liste de celles utilisées dans la requête (en prenant en compte les liens de subsomption dans l'ontologie). En utilisant ces descriptions de recouvrement, l'AM peut identifier à chaque étape de son algorithme de décomposition et pour chaque requête intermédiaire qu'il produit, les AAs à consulter. *L'ontologie partagée fournit ici les primitives permettant de décrire des connaissances allouées à chaque agent et permet ainsi de statuer sur la pertinence d'une participation d'un agent à une tâche donnée.*

Une fois que les rôles d'AA et d'AM ont été spécifiés avec leurs interactions, des modules de CORESE [7] ont été intégrés dans les types d'agent implantant ces rôles afin de leur fournir les compétences nécessaires.

Nous avons détaillé l'exemple de cette société pour montrer comment les techniques de l'ingénierie des connaissances pouvaient fournir les outils nécessaires

pour que d'autres domaines puissent apporter leur contribution à la résolution de problèmes complexes. Ainsi, l'approche reposant sur la mise en place d'une ontologie partagée permet à un SMA de participer à la gestion de connaissances distribuées. Inversement, le projet FRODO [14] montre comment un SMA peut assister la mise en place et le maintien d'un consensus ontologique à l'intérieur d'une communauté et entre des communautés.

4 Conclusion et Discussion

Nous avons présenté une approche innovatrice pour la gestion d'une mémoire organisationnelle combinant l'ingénierie d'ontologie, le Web sémantique et les systèmes multi-agents dans une solution intégrée. CoMMA appartient donc à la famille des SIMA qui s'intéressent à la gestion des connaissances dans une entreprise. Les projets CASMIR [4] et Ricochet[5], par exemple, se concentrent sur la collecte d'information et l'adaptation à l'utilisateur, en apprenant les centres des intérêt pour favoriser l'émergence de communauté et le filtrage collaboratif de l'information au sein de l'organisation. KnowWeb [13] exploite des agents mobiles pour palier au problème de connectivité éphémère entre les membres et leur organisation (ex: systèmes portables, disponibilité partielle, etc.). Ce projet repose sur un modèle du domaine pour extraire des concepts associés à un document et les utiliser dans la résolution de requêtes. RICA[1] maintient une taxonomie partagée dont les nœuds sont attachés à des documents et sont utilisés conjointement avec des profils utilisateurs pour faire des suggestions. Enfin FRODO [27] s'intéresse directement à la notion de mémoire d'entreprise en approfondissant le problème posé par la maintenance et la cohabitations de multiples ontologies. CoMMA se concentre sur la gestion d'une mémoire de documents hétérogènes mais avec des annotations homogènes et basées sur une ontologie partagée en reposant sur les technologies du Web sémantique. Les aspects qui y ont été plus particulièrement étudiés sont la recherche, la suggestion et l'archivage des annotations dans un environnement distribué.

Le prototype implanté en JAVA a été évalué par des utilisateurs d'une compagnie de télécommunication (T-System Nova Deutsch Telekom) et d'un organisme de recherche sur le Bâtiment (CSTB) avec des archives contenant jusqu'à 1000 annotations.

La figure 12 montre l'interface de requête ou d'annotation utilisée par les utilisateurs finaux de CoMMA. La résolution d'une requête ou l'allocation d'une nouvelle annotation reposent sur des communications entre agents qui exploitent les primitives définies par l'ontologie O'CoMMA. *Le consensus ontologique est bien la clef de voûte des mécanismes de gestion des connaissances distribuées.* La figure 13 montre une fenêtre de résultat, obtenue en réponse à une requête.

Des problèmes d'interface et d'ergonomie ont été soulevés par les utilisateurs, mais l'utilité et le potentiel des

fonctionnalités offertes par le système ont été unanimement reconnus. En particulier, les apports d'une ontologie et des techniques basées sur les langages de modélisation des connaissances ont été très appréciés par les utilisateurs finaux. De même, les développeurs du prototype ont vu dans le couple agent-ontologie une nouvelle façon d'aborder la spécification et la répartition d'un travail d'implantation.

Nous avons montré ici que l'intégration d'une approche d'ingénierie des connaissances (formalisation de connaissances sur les ressources dispersées dans un intraweb par des annotations sémantiques basées sur des ontologies) et de l'intelligence artificielle distribuée (systèmes d'information de plusieurs agents couplés par une coopération au niveau sémantique) peut constituer un nouveau paradigme puissant pour la résolution d'une classe de problèmes distribués complexes. Nous avons souligné le rôle crucial de l'ontologie à chaque étape et dans chacun des composants du système de gestion des connaissances distribuées.

Il reste, évidemment, d'autres points à approfondir. En premier lieu, dans le cadre de notre vision du cycle de vie de la mémoire d'entreprise [11] (figure 11), nous allons étudier pour chaque étape l'intérêt d'une telle approche.

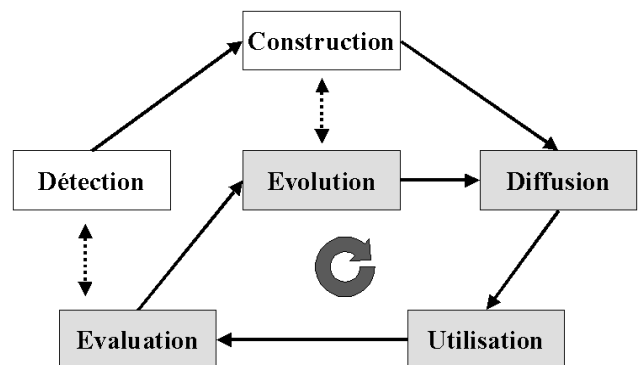


FIG. 11 – Cycle de la mémoire: Les cadres blancs montrent les étapes d'amorçage ; les cadres grisés représentent les étapes qui constituent le cycle de vie à proprement parler.

Enfin, les organisations, nous l'avons dit, ne sont pas des îlots isolés de la société mais participent elles-mêmes à des réseaux d'interactions complexes. Il faudra donc explorer l'extension de solutions similaires aux contextes d'extranet, voire, pour la communauté du Web sémantique, au Web ouvert.

5 Remerciements

Nous remercions chaleureusement nos collègues d'ACACIA et de CoMMA pour les fructueuses discussions que nous avons eues et la Commission Européenne qui a subventionné ce projet (IST-1999-12217).

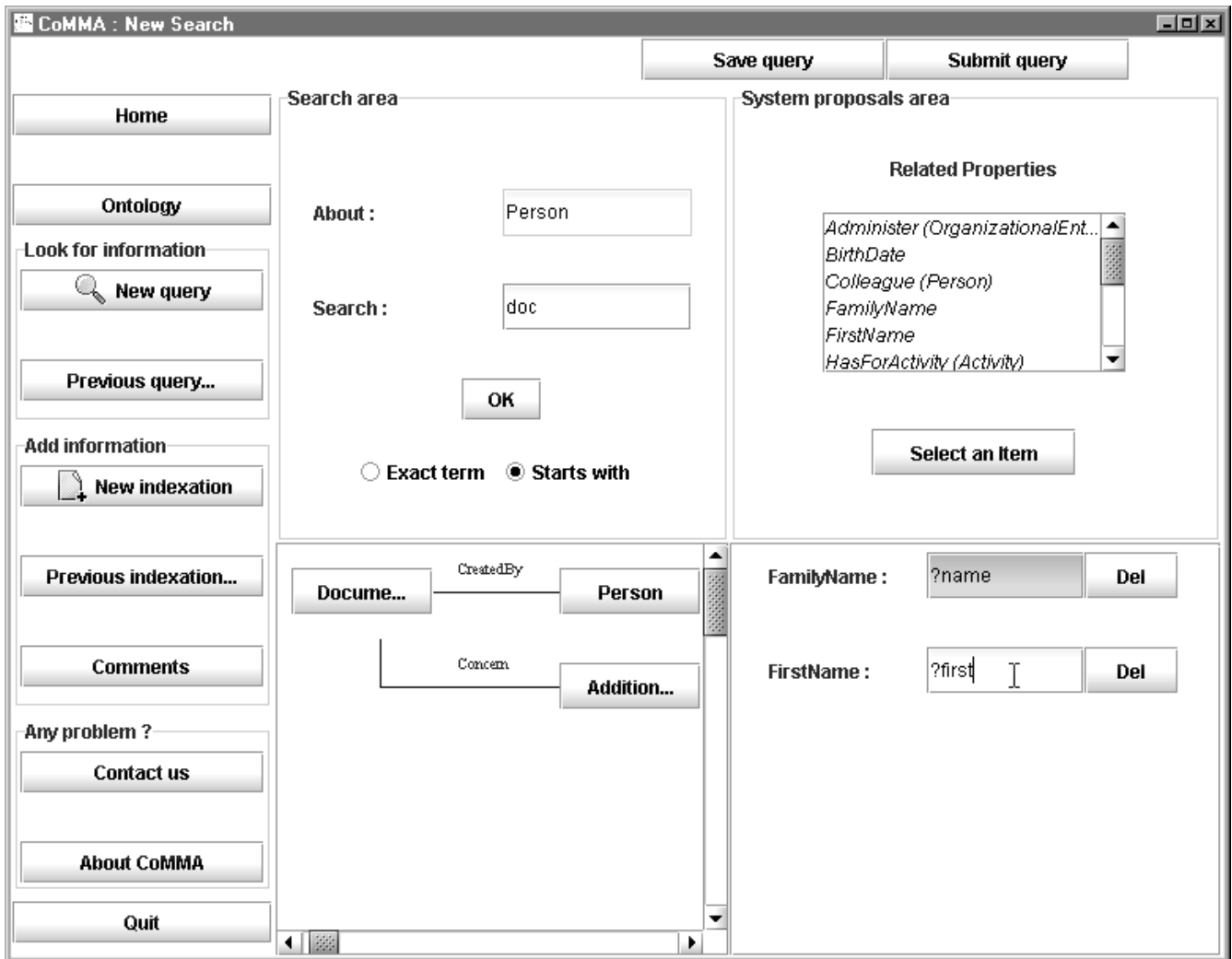


FIG. 12– Interface de requête / annotation de CoMMA

Document url	Type	Required Properties
http://www.fabien.gandon.net/Fabien.Gandon/research/pakm2000/pakm2000.pdf	Article	<p>Concern: <i>KnowledgeManagementTopic</i></p> <p>Concern: <i>MultiAgentSystemTopic</i></p> <p>Concern: <i>XMLTopic</i></p> <p>CreatedBy: <i>Employee</i></p> <p>FamilyNameFirstName: : <i>GANDON</i> : <i>Fabien</i></p>

FIG. 13 – Affichage d'un résultat

Références

- [1] Aguirre, J.L., Brena, R. Cantu-Ortiz, F. (2000). Multiagent-based Knowledge Networks. To appear in the special issue on Knowledge Management of the journal Expert Systems with Applications.
- [2] Bellifemine, Poggi, Rimassa, Developing multi agent systems with a FIPA-compliant agent framework. Software Practice & Experience, (2001) 31:103-128
- [3] Berners-Lee, Hendler, Lassila, The Semantic Web, Scientific American, May 2001:35-43
- [4] Berney, B. and Ferneley, E. (1999), "CASMIIR: Information Retrieval Based on Collaborative User Profiling", In Proceedings of PAAM'99, pp. 41-56. www.casmir.net
- [5] Bothorel, C. and Thomas, H. (1999), "A Distributed Agent Based-Platform for Internet User Communities", In Proceedings of PAAM'99, Lancashire, pp. 23-40.
- [6] D. Brickley, R. Guha, Resource Description Framework Schema Specification 1.0, W3C Candidate Recommendation 27 March 2000
- [7] Corby, Dieng, Hébert, A Conceptual Graph Model for W3C Resource Description Framework. In Proc. ICCS'2000 Darmstadt Germany
- [8] Cyc <http://www.cyc.com/cyc-2-1/cover.html>, Cycorp
- [9] Decker, Brickley, J. Saarela, J. Angele. A Query Service for RDF. Query Languages 98, W3C Workshop.
- [10] Delteil, Faron, Dieng, Extension de RDF(S) pour la représentation de connaissances contextuelles et la construction d'ontologies sur le Web, in J. Charlet coord., Actes de la Conférence Ingénierie des Connaissances (IC'2001), Grenoble, 2001, p. 111-129
- [11] Dieng, Corby, Gandon, Giboin, Golebiowska, Matta, Ribière, Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire pour le "knowledge management", 2ème édition, DUNOD, 2001, INFORMATIQUES, Série Stratégies et systèmes d'information
- [12] Doyle, Hayes-Roth, Agents in Annotated Worlds, In Proc. Autonomous Agents, ACM Press / ACM SIGART, Minneapolis, MN USA (1998) p173-180
- [13] M. Dzbor, J. Paralic and M. Paralic, Knowledge Management in a Distributed Organisation, In Proc. of the BASYS'2000 - 4th IEEE/IFIP International Conference on Information Technology for Balanced Automation Systems in Manufacturing, Kluwer Academic Publishers, London, September 2000, ISBN 0-7923-7958-6, pp. 339-348
- [14] Elst, Abecker, Domain Ontology Agents in Distributed Organizational Memories Workshop on Knowledge Management and Organizational Memories, IJCAI, 2001.
- [15] Fensel, Van Harmelen, Horrocks, McGuinness, Patel-Schneider. OIL: An ontology infrastructure for the semantic web. IEEE Intelligent Systems, 16(2):38-45, 2001.
- [16] FIPA (Foundation for Intelligent Physical Agents) - www.fipa.org
- [17] Gandon, Dieng-Kuntz, Ontologie pour un système multi-agents dédié à une mémoire d'entreprise, in Proceedings-siC'2001, Grenoble, France, June 25-28, 2001, pp 1-20
- [18] Gandon, Ontology Engineering: a survey and a return on experience., Rapport de Recherche INRIA, RR4396, Mars 2002 <http://www.inria.fr/rrrt/rr-4396.html>
- [19] F. Gandon, A Multi-Agent Architecture for Distributed Corporate Memories, Actes du 16ième European Meeting on Cybernetics and Systems Research (EMCSR 2002) April 3 - 5, 2002, Vienne, Autriche, pp 623-628.
- [20] F. Gandon, A Multi-Agent Platform for a Corporate Semantic Web, à paraître dans les actes de AAMAS 2002.
- [21] Kiss, Quinqueton, Multiagent Cooperative Learning of User Preferences, Proc. of European CMLP & PKDD, 2001.
- [22] O. Lassila, R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation 22 February 1999
- [23] Papazoglou, Heuvel, From Business Process to Cooperative Information Systems: An Information Agents Perspective, In Klush Intelligent Information Agent: Agent-Based Information Discovery and Management on the Internet. pp. 10-36, Springer 1999
- [24] Rolstadås, "Development trends to support Enterprise Modeling" ; Chapter 1 pages 3-16 of "Enterprise Modeling: Improving Global Industrial Competitiveness"; Rolstadås and Andersen; Kluwer Academic Publisher; 2000
- [25] Sowa, Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, 1984.
- [26] TOVE <http://www.eil.utoronto.ca/tove/ontoTOC.html> Enterprise Integration Laboratory (EIL), Department of Industrial Engineering University of Toronto
- [27] Van Elst, Abecker, Domain Ontology Agents in Distributed Organizational Memories In Proc. Workshop on Knowledge Management and Organizational Memories, IJCAI, 2001.
- [28] Van Harmelen, Patel-schneider, Horrocks , Reference description of the DAML+OIL ontology markup language. <http://pride.daml.org/2000/12/reference.html>. March 2001.
- [29] Wooldridge, Jennings, Kinny, A Methodology for Agent-Oriented Analysis and Design. In Proc of Autonomous Agents '99 Seattle, ACM 1-58113-066-x/99/05