
Le projet GDST

Bernard Lefebvre*, Serge Tadié, Omar Cherkaoui, Gilles Gauthier, Olivier Gerbé, Jean-Guy Meunier

**Laboratoire GDAC
Département d'informatique, Université du Québec à Montréal
Case Postale 8888, succursale Centre-Ville
Montréal (Québec) H3C 3P8
lefebvre.bernard@uqam.ca*

RÉSUMÉ: Le projet GDST (Gestion et Diffusion du Savoir en Télécommunication) a pour principal objectif de mettre en place un prototype de gestion informatisée de documents relatifs à un domaine qui puisse permettre le développement des compétences des ressources humaines pour les besoins d'une entreprise. Le domaine considéré dans le cadre de cette expérimentation est celui des télécommunications sans fil. La gestion repose sur un environnement informatique qui intègre des outils d'édition, de maintenance, d'interrogation et de diffusion documentaire. Les connaissances nécessaires sont organisées en quatre niveaux (meta structure, structure, données et privé) et elles référencent une base de ressources documentaires (HTML ou PDF). Les deux premiers niveaux sont constitués d'ontologies qui décrivent les concepts relatifs au domaine de la connaissance, les rôles et processus d'affaire de l'entreprise, les compétences et caractéristiques des employés. Les niveaux inférieurs sont formés d'instanciations de concepts décrits dans les niveaux supérieurs et relatifs aux employés et à l'entreprise. Le niveau privé est relatif à des ressources qui sont personnelles à un usager et non accessibles aux autres.

MOTS CLÉS : Web Sémantique, Ontologie, DAML-OIL, gestion des connaissances.

1. Introduction

De récentes études de cas [COURT 98] ont montré l'importance de la connaissance située localement dans l'activité de conception ou de réalisation des ingénieurs. Par connaissance locale on entend celle qui est détenue dans la mémoire de la personne, celle dont on peut se remémorer la source, celle qui peut être localisée grossièrement à l'aide d'une connaissance vague mémorisée et celle que l'on peut obtenir à partir d'un processus d'acquisition dont on se souvient.

La connaissance locale peut aussi être considérée comme tacite [NONAKA 91], si elle est imbriquée dans des expériences individuelles, ou explicites, si elle est transmissible en langage formel. Dans ce cas, elle peut facilement se partager et s'articuler parce qu'elle est indépendante de l'esprit individuel. L'entreposage de connaissances explicites qui sont souvent dispersées et disparates au sein d'une entreprise est appelé mémoire corporative ou mémoire organisationnelle (MO). Cette mémoire est la source d'information qu'un système de gestion de la connaissance doit structurer.

Caractériser la mémoire organisationnelle c'est aussi définir le lien qui existe entre connaissance, information et données [TREMBLAY 2002]. Les données sont de simples éléments qui deviennent information lorsqu'ils sont combinés en structures significatives. Cette information devient ultérieurement de la connaissance lorsqu'elle est mise dans un contexte où elle peut être employée à des fins particulières. Selon cette perspective, les données sont donc

un préalable à l'information, et l'information est un préalable à la connaissance. Pour Peter Drucker [DRUCKER 94; DRUCKER 99] l'information serait ainsi des données dotées d'utilité et ayant une raison d'être donc des données mises ensemble pour un destinataire et qui auront une utilité pour celui-ci. Pour devenir connaissance il est ainsi nécessaire que l'information soit cataloguée et structurée, disponible d'accès par les bonnes personnes, et ce, au bon moment.

L'implantation de solutions informatiques pour la gestion de la connaissance qui répondent à ces objectifs est un phénomène plutôt récent [RUBENSTEIN-MONTANO et al. 2001]. Ceci implique en effet l'intégration plutôt difficile de concepts et techniques issus de différents domaines comme l'intelligence artificielle, l'ingénierie des systèmes d'informations, la réingénierie des processus ou le comportement des organisations et de leurs ressources humaines [LIEBOWITZ 99].

Le projet GDST (Gestion et Diffusion de Savoir en Télécommunication) entre dans ce cadre. Il a pour principal objectif de faciliter le développement des compétences des ressources humaines pour les besoins d'une entreprise. Ceci nécessite la création d'un environnement informatique permettant d'intégrer les outils d'édition, de représentation des connaissances et de diffusion documentaire nécessaires à cette gestion.

2. Aspects Méthodologiques

La réalisation du système GDST repose sur des travaux relatifs à l'utilisation d'ontologies pour les entreprises [BARRY 98; USCHOLD M. et al. 98] et les domaines de connaissances. Beaucoup de ceux-ci [ABECKER et al. 2000; FOX et al. 96] se réfèrent au travail d'Abecker, Bernardi, Hinkelmann, Kühn et Sintek [ABECKER et al. 98] qui dans le cadre du projet « KnowMore » ont développé une architecture de mémoire organisationnelle répondant aux besoins d'interrogation contextuelle à une tâche.

Ces dernières années, la nécessité d'augmenter l'exactitude des moteurs de recherche sur le Web a motivé une quantité énorme de recherche autour de ce qui s'est appelé le « Web sémantique ». Ces recherches ont plusieurs objectifs :

- Définir des langages basés sur XML afin de représenter une connaissance qui peut être traitée par un ordinateur.
- Construire des ontologies relativement à des domaines de connaissance.
- Utiliser ces ontologies pour produire une annotation sémantique des documents sur le Web.
- Définir des explorateurs et moteurs de recherche qui exploitent le potentiel de ces annotations sémantiques.

Le système GDST se distingue des réalisations existantes comme [FOX et al. 96; GRUNINGER et al. 2000] par l'usage de ces développements récents et de ces langages en plus d'offrir une architecture originale qui répond aux besoins spécifiques d'une entreprise. Il offre notamment à ses usagers des moyens comme :

- un système de diffusion qui fournit une nouvelle information documentaire en fonction de leurs besoins et du niveau d'intérêt,
- des systèmes d'exploration et d'interrogation qui peuvent les aider dans le contexte de la réalisation d'une tâche.

3. L'architecture du système

Le système de GDST réalise et intègre les aspects mentionnés dans le contexte et les langages du Web sémantique. Par rapport aux desseins du Web Sémantique, la portée est cependant considérablement réduite. Une autre différence essentielle est que le contexte d'utilisation du système peut être défini avec précision ainsi que les caractéristiques des utilisateurs. Celles-ci seront mises à contribution afin de focaliser les outils du système.

L'architecture est organisée en trois blocs comme le représente la figure 1. Les outils d'édition et de maintenance qui opèrent sur les connaissances qui sont elles mêmes utilisées par les outils d'exploitation.

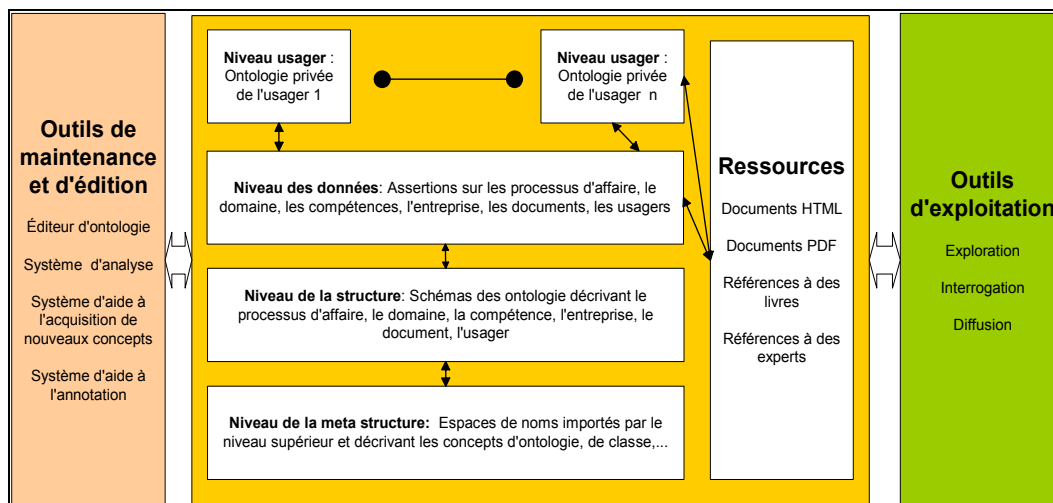


Figure 1. L'architecture du système GDST

3.1. La modélisation des connaissances

L'objet de cette modélisation est de caractériser le contenu des documents ainsi que le contexte dans lequel ils sont utilisés. Cette modélisation contextuelle est réalisée au moyen de langages développés pour le Web Sémantique selon les recommandations du consortium W3. Étant donnée la panoplie d'outils existants le choix s'est porté sur le langage DAML+OIL puisque son successeur OWL n'a pas encore atteint un degré de maturité suffisant. La modélisation repose sur une hiérarchie à 4 niveaux de bases de connaissances qui sont principalement des ontologies et des assertions sur ces ontologies.

3.1.1. Le niveau de la meta structure

Ce niveau correspond au meta modèle relatif aux ontologies. Il est défini au moyen de documents standard fournis par le consortium W3. Il décrit les notions permettant de définir une ontologie, ce que sont notamment une classe ou une propriété. Il exprime également la nature générale d'un domaine de connaissance par le biais de la notion générale de concept.

Les espaces de noms correspondants sont utilisés par le niveau suivant.

3.1.2. Le niveau de la structure

Ce niveau importe le précédent et il l'instancie. Il est formé de 6 classes principales (« PAT », « Concept », « DocumentResource », « ExpertiseLevel », « Employee » and « Company ») qui sont associés comme le montre la figure 2. Chacune de ces classes est décrite et détaillée dans le cadre d'une ontologie spécifique.

- « PAT »: Cette entité est de nature abstraite. Elle définit les propriétés communes aux processus, activités, tâches et sous tâches qui caractérisent l'activité professionnelle des catégories de personnels d'une entreprise pour lesquelles le système est conçu.

- « Concept » : Cette entité du niveau supérieur généralise ici l'ensemble des concepts d'un domaine de connaissances qui sont également ses instances. C'est en tant qu'instance de cette

classe que les concepts du domaine peuvent être cités dans une ressource documentaire. Cette dualité, sous classe, instance, permet également, par héritage, aux instances d'un concept du domaine de connaissance de pouvoir aussi être citées.

- « DocumentResource » : Cette classe abstraite introduit les propriétés relatives à toutes les ressources documentaires. Elle se spécialise en « Document » et « DocumentPart ».

- « ExpertiseLevel »: Cette classe est au centre du modèle. Elle permet d'associer des documents et des processus aux utilisateurs. Ces niveaux d'expertise sont associés à des compétences qui elles mêmes sont caractérisées par des qualifications. La caractérisation de la notion de compétence est assez complète et repose sur des propositions recensées dans l'ouvrage de Gilbert Paquette [PAQUETTE 2002].

- « Employee »: Cette classe caractérise l'entité employé d'une entreprise. Un employé d'une entreprise y exerce un rôle et possède des niveaux d'expertise par rapport à certaines compétences.

- « Company »: Cette classe permet de décrire une entreprise en termes, par exemple, de rôles que peuvent exercer des employés. L'exercice de ces rôles requiert la possession de compétences particulières à certains niveaux.

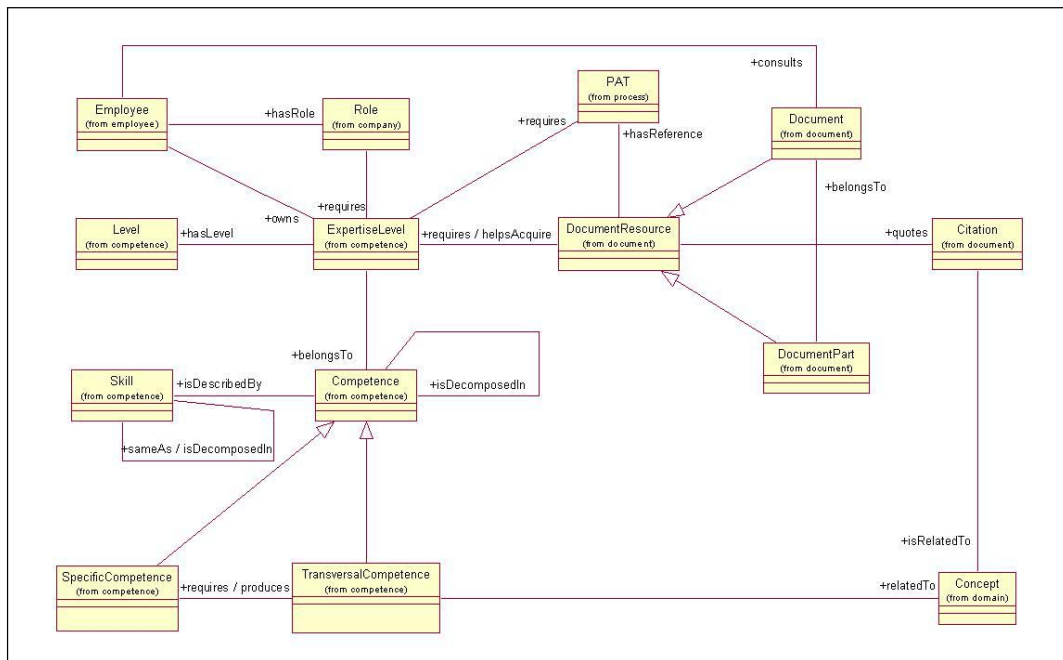


Figure 2. Le modèle simplifié de la connaissance

3.1.3. Le niveau des données

Il est le troisième niveau de l'architecture. Il contient des affirmations au sujet des ontologies du niveau deux. À l'aide de ces ontologies il décrit une entreprise en particulier, ses employés et leurs compétences, son processus d'affaire, ses documents utiles à la réalisation de tâches.

Dans le cas du projet GDST, la compagnie œuvre dans le domaine de la télécommunication, et l'objectif du développement est d'aider ses ingénieurs en systèmes de télécommunication à exploiter les ressources documentaires nécessaires à la réalisation du processus de mise en place d'une solution réseau sans fil pour un client.

Le domaine des connaissances est donc celui des réseaux sans fil. Il décrit, par exemple, le concept de point d'accès ou d'antenne. Ces entités du niveau deux sont des spécialisations de la classe « Concept » et aussi des instances de cette classe de niveau un. Comme toute instance de

concept, elles peuvent être citées dans des ressources documentaires ce qui se matérialise par l'existence d'éléments citations qui indiquent l'importance du traitement de la notion dans le document. L'instance d'une antenne de marque et type particulier peut, elle aussi, en tant qu'instance de concept être citée avec un certain degré d'importance dans un document. Ces éléments constituent les annotations documentaires relatives au domaine. Une ressource documentaire peut aussi être nécessaire à la réalisation d'une tâche ou permettre l'acquisition d'un niveau d'expertise. Ce niveau contient donc aussi les instanciations des propriétés correspondantes.

3.1.4. Les ressources documentaires

Certaines ressources documentaires peuvent être stockées localement dans une base sous une forme électronique, HTML ou PDF. Ces documents ou leurs parties, comme d'autres documents laissés sur des sites distants, sont repérés par un URI et ont été annotés pour pouvoir être exploités par les différents outils du système.

3.1.5. Le niveau privé de l'utilisateur

Le niveau quatre est le niveau privé de l'utilisateur. Il contient des affirmations spécifiques à un utilisateur. Il est actuellement exploité par certains outils du système. Il permet de savoir, par exemple si un employé a déjà pris connaissance de tel document. Ultérieurement il pourra contenir également les documents produits par l'utilisateur et qui sont destinés à un usage exclusif de même que les ressources qu'il aura documentées et annotées pour ses propres besoins.

3.2. Outils d'édition et de maintenance

Créer, contrôler et maintenir la connaissance dans cette architecture est une entreprise difficile qui peut être facilitée par l'utilisation d'outils.

3.2.1. Éditeur d'ontologies

Le langage utilisé pour décrire les ontologies est DAM+OIL. De nombreux éditeurs existent pour ce langage dont Protege-2000 [NOY et al. 2001] ou Oiled [FENSEL et al. 2001]. Ces éditeurs sont très versatiles par vocation ce qui rend leur usage délicat et réservé à des spécialistes du génie de la connaissance. Des éditeurs plus conviviaux et plus adaptés devront être développés pour certaines ontologies comme celles du niveau privé.

3.2.2. Système d'analyse de documents et d'aide à la découverte de concepts

Il se compose d'un ensemble d'outils intégrés dans une plateforme nommée SATIM [BISKRI & MEUNIER 2002] qui a été développé au laboratoire de LANCI de l'UQAM. Le rôle de ces outils et principalement du module GRAMEXCO est d'assister la découverte de notions dans des documents ceci afin de mettre à jour l'ontologie du domaine en tenant compte de l'évolution des connaissances. Le processus [GARGOURI et al.] consiste d'abord à extraire le vocabulaire approprié à partir des documents HTML ou PDF qui ont été divisés en segments ayant une longueur approximative d'une page. Ces segments élémentaires sont ensuite classifiés selon le vocabulaire pertinent qu'ils contiennent à l'aide d'un réseau de neurones de type non supervisé basé sur ART (Adaptive Resonance Theory) [GROSSBERG 88]. Les regroupements donnés par le classificateur peuvent être liés aux concepts qu'ils représentent. Ceci est fait sur la base des termes pertinents communs à tous les documents d'un groupe. Se baser sur un groupe plutôt que sur un segment unique permet de limiter le nombre de termes et d'en augmenter la pertinence tout en simplifiant le traitement d'un ensemble de segments. Par exemple, si le vocabulaire commun d'un groupe contient les mots : « access », « ansiiiee », « control », « edition »,

« ieee », « isoiec » et « specifications » il est possible pour un spécialiste d'inférer que tous les segments de ce groupe sont liés au thème « Access Control Specifications ». Dans une deuxième étape, une méthode d'indexation de type « analyse de la sémantique latente » permet d'extraire au sein d'un groupe quels sont les termes les plus souvent associés et qui sont potentiellement révélateurs de propriétés relative à un concept. Ceux-ci peuvent être utilisés par un expert pour valider l'association et mettre à jour l'ontologie.

3.2.3. Système d'aide à l'annotation de documents

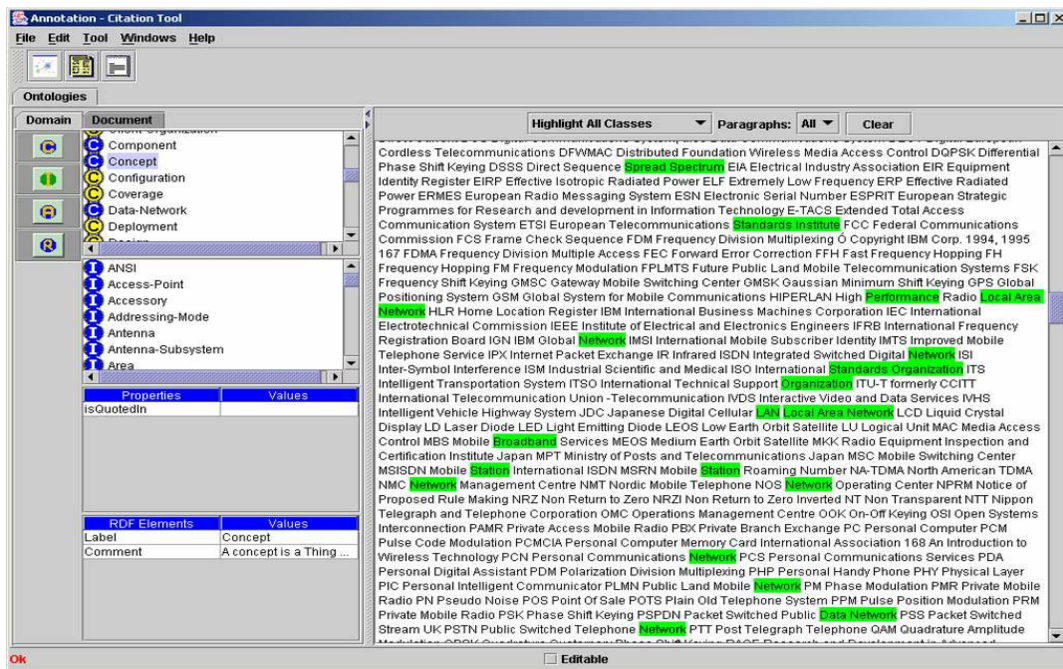


Figure 3. L'outil d'annotation

Cet outil semi-automatique nommé « AnnoCitaTool » permet de faciliter les processus de l'annotation de documents par rapport à l'ontologie du domaine. Les annotations sont constituées d'instances de la classe « Citation » qui mettent en relation une ressource documentaire et un concept avec un certain degré d'importance. L'outil assiste la génération de ces liens. Il affiche la ressource documentaire (segment d'environ une page en général) et met en relief les termes de ce document qui apparaissent dans l'ontologie du domaine et en particulier dans les descriptions de concepts sélectionnés qui, par l'intermédiaire des commentaires associés ou des propriétés, fournissent un vocabulaire pertinent.

3.3. Les outils d'exploitation

Ces outils sont disponibles à l'utilisateur. Ils lui fournissent certains moyens d'accès personnalisés à l'information contenue dans les ontologies. Ces outils ont été développés en Java. Ils utilisent la bibliothèque de classes Jena [MCBRIDE et al. 2002]. Ce sont des servlets qui sont donc accessibles par le Web.

3.3.1. L'outil d'exploration

Cet outil permet d'explorer les ontologies sous une forme simplifiée et adaptée aux exigences et particularités du projet. Un utilisateur peut ainsi naviguer d'une ontologie à une autre en fonction des liens qui les unissent. Il peut, par exemple, partant de l'ontologie du processus d'affaire et d'une instance de processus accéder à un document pertinent et le visualiser.

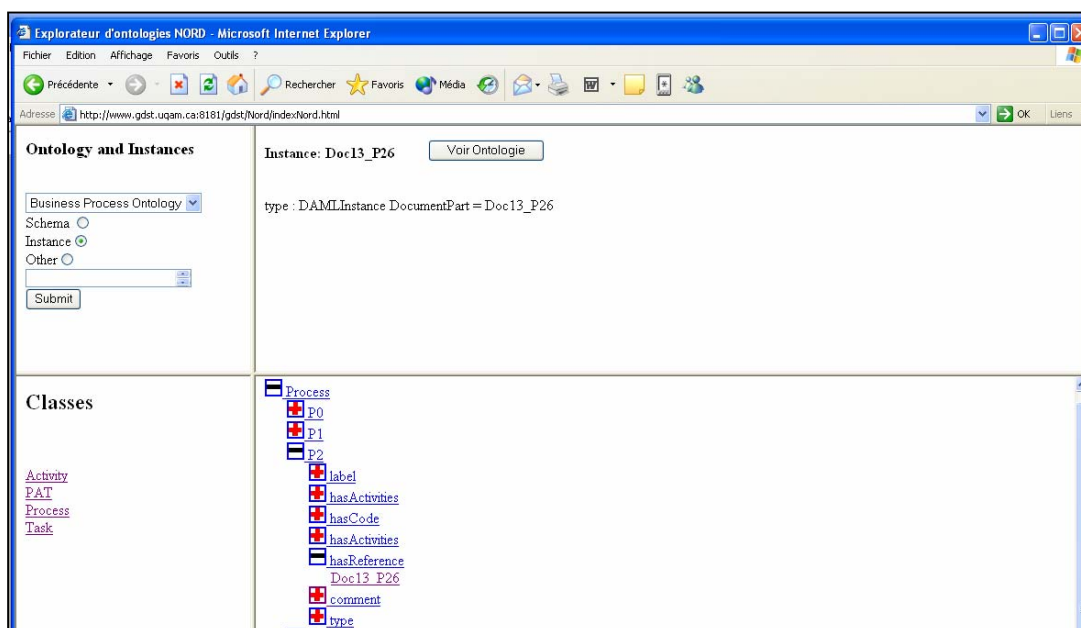


Figure 4. Exemple d'exploration avec NORD

3.3.2. L'outil de diffusion

Cet outil prend en charge la diffusion auprès d'utilisateurs de nouveaux documents. Ceux-ci doivent au préalable avoir été annotés et donc caractérisés non seulement en fonction des concepts que ces documents décrivent mais aussi en fonction des compétences requises pour les comprendre ainsi que des compétences qui pourraient être acquises à un certain niveau d'expertise par l'assimilation de leur contenu. En fonction de ces informations, l'outil de diffusion sélectionne les usagers potentiellement concernés par ces documents en considérant leur « zone proximale de développement » [VYGOTSKII & PIAGET 85]. Un usager est retenu par ce processus de sélection s'il a les compétences requises au bon niveau d'expertise pour assimiler le contenu sans avoir toutes celles qui peuvent être obtenues après assimilation.

3.3.3. Outil d'interrogation

Le système intelligent d'interrogation est un système de recherche documentaire basé sur des ontologies. Il recherche des documents non seulement sur la base de critères qu'un utilisateur propose, mais il prend également en considération le contexte du document et le modèle de compétence de cet utilisateur de la même manière que l'outil de diffusion. Un utilisateur peut ainsi demander les documents qui peuvent l'aider à réaliser un processus de travail. Partant de cette demande, le système peut obtenir la hiérarchie entière du processus et toutes les compétences requises pour maîtriser l'ensemble de ce processus. En comparant cette liste de compétences nécessaires à celles déjà acquises par l'utilisateur, le système peut déterminer celles que l'utilisateur devrait acquérir. Il peut alors proposer les documents que l'utilisateur devrait consulter pour maîtriser ces dernières.

4. Conclusion

Le projet de GDST a été conçu pour fournir aux entreprises les moyens leur permettant de résoudre ce problème d'acquisition et de transmission de connaissance lié à l'innovation technologique.

Les défis que ce projet adresse sont nombreux et ils ont été partiellement surmontés. Les ontologies ont été implantées sur le plan de leur structure. Elles ont été instanciées de sorte que les fonctionnalités du système ont pu être testées. Les premières versions des outils d'exploitation ont été développées et sont fonctionnelles. Cette instanciation partielle est cependant insuffisante pour que le système puisse être utilisé et validé dans le cadre réel d'une entreprise en prenant en compte l'ensemble du processus d'affaire. Cet important travail reste à entreprendre.

Les outils de maintenance du système sont indispensables à son maintien et à son évolution future. Les difficultés méthodologiques relatives à leur mise en œuvre ont été en bonne partie résolues. Des propositions de solutions existent qu'il faut encore achever et tester dans un contexte d'application réel.

Remerciements

Ces travaux sont réalisés dans le cadre du projet de GDST (Gestion et Diffusion du Savoir en Télécommunication), par des chercheurs de l'UQAM (Université du Québec à Montréal) et de l'UdeM (Université de Montréal) : Bernard Lefebvre, Jean-Guy Meunier, Gilles Gauthier, Omar Cherkaoui, Olivier Gerbé, Serge Tadié et d'autres étudiants en maîtrise et au doctorat.

Nous voudrions remercier le LUB (Laboratoire Universitaire Bell) et le CRSNG (Conseil de Recherche en Sciences Naturelles et en Génie du Canada) pour leur aide financière.

Bibliographie

- [ABECKER et al. 98] ABECKER A., BERNARDI A., HINKELMANN K., KUHN O., Sintek M., "*Toward a Technology for Organizational Memories*", *IEEE Intelligent Systems*, vol. 13, 1998, p. 40-48.
- [ABECKER et al. 00] ABECKER A., BERNARDI A., HINKELMANN K., KÜHN O., Sintek M., "*Context-Aware, Proactive Delivery of Task-Specific Knowledge: The KnowMore Project*", *Int. Journal on Information Systems Frontiers (ISF)*, vol. 2, n° (3/4), 2000, p. 139-162.
- [BARRY 98] BARRY R. E., "*Document Management for the Enterprise: Principles, Techniques and Applications*", *Journal of the American Society for Information Science*, vol. 49, n° 1, January 1998, p. 94.
- [BISKRI & MEUNIER 02] BISKRI I., MEUNIER J. G., "*SATIM : Système d'Analyse et de Traitement de l'Information Multidimensionnelle*", *JADT 2002*, St-Malo, France, 13 2002.
- [COURT 98] COURT A. W., "*Issues for integrating knowledge in new product development: Reflections from an empirical study*", *Knowledge Based Systems*, vol. 11, n° 7, 1998, p. 391-398.
- [DRUCKER 94] DRUCKER P. F., "*The age of social transformation*", *The Atlantic Monthly*, vol. 274, n° 5, November 1994, p. 53.
- [DRUCKER 99] DRUCKER P. F., "*Knowledge-worker productivity: The biggest challenge*", *California Management Review*, vol. 41, n° 2, Winter 1999, p. 79.

- [FENSEL et al. 01] FENSEL D., VAN HARMELEN F., HORROCKS I., MCGUINNESS D. L., Patel-Schneider P. F., "OIL: An ontology infrastructure for the semantic Web ", *IEEE Intelligent Systems*, vol. 16, n° 2, March 2001-April 2001, p. 38.
- [FOX et al. 96] FOX M. S., BARBUCEANU M., GRUNINGER M., "An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour", *Computers in Industry*, vol. 29, n° 1,2, July 1996, p. 123.
- [GARGOURI et al. 03] GARGOURI Y., LEFEBVRE B., MEUNIER J.-G., "Ontology Maintenance using Textual Analysis", *7th World Multiconference on Systemics, Cybernetics and Informatics, SCI 2003*, Orlando, Florida, July 2003.
- [GROSSBERG 88] GROSSBERG S., *Neural networks and natural intelligence*, Cambridge, Mass., MIT Press, 88, xi, 637 p. p.
- [GRUNINGER et al. 00] GRUNINGER M., ATEFI K., FOX M. S., "Ontologies to Support Process Integration in Enterprise Engineering", *Computational and Mathematical Organization Theory*, vol. 6, n° 4, December 2000, p. 381.
- [LIEBOWITZ 99] LIEBOWITZ J., *Knowledge management handbook*, Boca Raton, Fla, CRC Press, 99.
- [MCBRIDE et al. 02] MCBRIDE B., SEABORNE A., CARROLL J., "Jena Tutorial for Release 1.4.0", <http://www.hpl.hp.com/semweb/doc/tutorial/index.html>, 2003.
- [NONAKA 91] NONAKA I., "The Knowledge-Creating Company", *Harvard Business Review*, vol. 69, n° 6, November 1991-December 1991, p. 96.
- [NOY et al. 01] NOY N. F., SINTEK M., DECKER S., CRUBEZY M. and others, "Creating semantic Web contents with protege-2000", *IEEE Intelligent Systems*, vol. 16, n° 2, March 2001-April 2001, p. 60.
- [PAQUETTE 02] PAQUETTE G., *Modélisation des connaissances et des compétences : un langage graphique pour concevoir et apprendre*, Sainte-Foy, Presses de l'Université du Québec, 2002, xxx, 357 p. ill. p.
- [RUBENSTEIN-MONTANO et al. 01] RUBENSTEIN-MONTANO B., LIEBOWITZ J., BUCHWALTER J., MCCAW D., NEWMAN B. and others, "A systems thinking framework for knowledge management", *Decision Support Systems*, vol. 31, n° 1, May 2001, p. 5.
- [TREMBLAY 02] TREMBLAY J.-F., *Gestion et diffusion automatisées de l'Information, Mémoire de maîtrise*, École des Hautes Études Commerciales, Université de Montréal.
- [USCHOLD et al. 98] USCHOLD M., KING M., MORALEE S., ZORGIOS Y., "The Enterprise Ontology " , *The Knowledge Engineering Review*, vol. 13, n° Special Issue on Putting Ontologies to Use (eds. Mike Uschold and Austin Tate), 1998.
- [VYGOTSKII & PIAGET 85] VYGOTSKII L. S., PIAGET J., *Pensée et langage*, Paris, Éditions sociales, 85.