

BULLETIN

DE

L'AFIA



**ASSOCIATION FRANCAISE POUR
L'INTELLIGENCE ARTIFICIELLE**

AVRIL 2003

N° 54

Présentation du bulletin

Le **Bulletin** de l'**Association Française pour l'Intelligence Artificielle** vise à fournir un cadre de discussion et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le Bulletin de l'AFIA publie également des dossiers plus substantiels sur différents thèmes liés à l'IA. Le comité de rédaction se réserve le droit de ne pas publier des contributions qu'il jugerait contraire à l'esprit du bulletin ou à sa politique éditoriale. En outre, les articles signés, de même que les contributions aux débats, reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

Pour contacter l'AFIA

Président : Marc Schoenauer
Projet FRACTALES - I.N.R.I.A.
Rocquencourt
B.P. 105 - 78153 LE CHESNAY
Cedex - France
Tel : +33 (0)139 63 50 87
Fax : +33 (0)139 63 59 95
Marc.Schoenauer@inria.fr
<http://www-rocq.inria.fr/fractales/Staff/Schoenauer>

Contributions au bulletin
Jean-Paul SANSONNET
voir ci-contre

Serveur WEB
<http://www.afia.polytechnique.fr/>

Adhésions,
Liens avec les adhérents
Marc AYEL
LIA-ESIGEC- Technolac
Université de Savoie
73376 Le Bourget du Lac CEDEX
Mél. : Marc.Ayel@univ-savoie.fr

Parrainage de manifestations
Sylvie PESTY
Laboratoire Leibniz, Institut IMAG
46, Avenue Félix Viallet
38031 Grenoble CEDEX
Mél. : Sylvie.Pesty@imag.fr

Membres d'honneur

Jaques Pitrat, Jean-Paul Haton, Marie-Odile Cordier

Jean-Marc David, Daniel Kayser, Claude Vogel, Alain Colmerauer

Personnes morales adhérentes à l'AFIA

ADIT, AI*IA, CNET PARIS-A, ENS Mines, ENST-PARIS, ESIEA, IGN,
ILOG, INRETS, INRIA, INSTITUT FRANCAIS DU PÉTROLE,
PEUGEOT S.A., Université de Savoie, Université PARIS 9 DAUPHINE.

Bureau de l'AFIA

Marc SCHOENAUER, président
Serge DUPUY, secrétaire
Marc AYEL, trésorier

Comité :

Jean-Paul BAQUIAST, Gilles BISSON, Bertrand BRAUNSCHWEIG,
Alain CARDON, Jérôme EUZENAT, Christophe JACQUEMIN,
Eunika MERCIER-LAURENT, Sylvie PESTY, Jean-Paul SANSONNET,
Michèle SEBAG, Laurent SIKLÓSSY, Fabien TORRE,
Jean-Daniel ZUCKER.

Comité de rédaction

Jean-Paul SANSONNET
Rédacteur en chef
LIMSI — CNRS
B.P. 133, 91403 ORSAY Cedex
jps@limsi.fr

Isabelle ROBBA
Rédactrice adjointe
LIMSI — CNRS
B.P. 133, 91403 ORSAY Cedex
Isabelle.Robba@limsi.fr

Catherine BARRY-GRÉBOVAL
Rubrique « Présentation de laboratoires »
LaRIA, Equipe Ingénierie des
Connaissances,
Université de Picardie Jules Verne
5, rue du Moulin Neuf
80000 Amiens
barry@laria.u-picardie.fr

Brigitte GRAU
Rubriques « Sommaires des revues »
et « petites annonces »
LIMSI — CNRS
B.P. 133, 91403 ORSAY Cedex
grau@limsi.fr

Marc-Philippe HUGET
Rubrique « Thèses et habilitations »
Agent ART Group, Department of Computer
Science
University of Liverpool
LIVERPOOL L69 7ZF United Kingdom
M.P.Huget@csc.liv.ac.uk

Amedeo NAPOLI
Rubrique « livres »
LORIA
B.P. 239
F-54 506 Vandœuvre lès Nancy
Amedeo.Napoli@loria.fr

Sylvie PESTY
Rubriques
« Conférences et comptes rendus »
Laboratoire Leibniz-IMAG
46 Avenue Félix Viallet
F-38031 Grenoble Cedex
Sylvie.Pesty@imag.fr

Nicole TOURIGNY
Rubrique « IA au Québec »
Université Laval - Pavillon Adrien-Pouliot
Québec, Canada, G1K 7P4
tourigny@ift.ulaval.ca

Jean-Daniel ZUCKER
Rubrique « Débats »
LIP6, Université Paris 6
4, Place Jussieu, 75232 Paris Cedex 05
Jean-Daniel.Zucker@lip6.fr



Réa. Numérocopie – 01 46 31 64 53
ISSN 1273-1323
Dépôt légal juin 2002

Le « PortAI » un portail d'accès pour l'AFIA

Ce numéro de printemps du bulletin de l'AFIA est consacré essentiellement à Internet et à l'accès à l'information en ligne. C'est pourquoi nous présentons un dossier sur le « Web sémantique » coordonné par Chantal Reynaud (LRI), Jean Charlet (AP-HP) et Philippe Laublet (LaLICC) qui est tout à fait d'actualité.

Le Web Sémantique, vision prophétique de Tim Berners-Lee, est une frontière pour l'Ingénierie de la Connaissance et plus généralement pour l'Intelligence Artificielle. Au-delà des infrastructures de simple mise en ligne de l'information, c'est le *contenu* qui est en jeu et plus particulièrement *l'activité*, – activité professionnelle, mais aussi *activité collective des associations* comme par exemple l'AFIA qui depuis plus d'un an propose à la collectivité un site « le PortAI », entièrement rénové et entièrement dynamique :



Sur le site de l'AFIA, vous pouvez consulter les rubriques qui concernent l'animation de la communauté IA : tout d'abord la version électronique du bulletin (téléchargeable pour les adhérents AFIA en .pdf), les groupes de travail, les annonces et compte rendus de conférences mais aussi tous les services : annonces, annuaire, forums, etc.

Le bureau de l'AFIA souhaite rendre ce site encore plus interactif pour en faire un lieu vivant de rencontre entre les chercheurs de la communauté IA. Mais alors, en faire le lieu de passage central pour les activités de l'AFIA pose la question de « l'électronisation » du bulletin de l'AFIA, ou plutôt, puisqu'il est déjà présent sous forme électronique, de son découpage en rubriques éditées directement en ligne. Le bureau de l'AFIA réfléchit à ce transfert qui serait un changement majeur dans la vie de l'AFIA, peut être à partir de l'automne 2003. N'hésitez pas à nous faire part de votre opinion de vos suggestions ou commentaires à ce sujet : marc.schoenauer@inria.fr ou jps@limsi.fr.

Jean-Paul Sansonnet

Rédacteur en chef du Bulletin de l'AFIA

Nouvelles du Québec

Nicole Tourigny

Un laboratoire regroupant la recherche en intelligence artificielle du département d'informatique et de génie logiciel (<http://www.ift.ulaval.ca>) de l'Université Laval (<http://www.ulaval.ca>) a été récemment créé : le Laboratoire des Systèmes Intelligents (LSI).

Ce laboratoire compte 8 professeurs-chercheurs du département d'informatique et de génie logiciel de l'Université Laval, spécialisés en intelligence artificielle et qui travaillent avec une cinquantaine de personnes du département (stagiaires, étudiants des 3 cycles universitaires, professionnels de recherche). Des professeurs et des étudiants d'autres départements et d'autres établissements sont également impliqués dans ces travaux de recherche.

Le LSI est articulé autour d'équipes animées par un ou plusieurs professeurs. Chaque professeur est responsable d'obtenir du financement. Parmi les sources de financement du LSI, on trouve le Fonds québécois de la recherche sur la nature et les technologies (FQRNT), Valorisation-Recherche du Québec (VRQ), le Conseil canadien de Recherche en Sciences Naturelles et en Génie (CRSNG), le Conseil canadien de Recherche en Sciences Humaines (CRSH), les réseaux de centres d'excellence (NCE), dont le Réseau de centres d'excellence en géomatique (GEOIDE) et l'Institut de robotique et d'intelligence des systèmes (IRIS), Recherche et Développement de la Défense (RDDC) de Valcartier, Lockheed Martin Canada (LMC), le Consortium de foresterie FOR@C, le Réseau de Valorisation de l'Enseignement de l'Université Laval (RVE), le département d'informatique et de génie logiciel de l'Université Laval, la faculté des sciences et de génie de l'Université Laval.

Voici une brève description de chacune de ces équipes et des liens pour les contacter ou pour en savoir plus. On trouvera ces liens et des informations plus détaillées à partir du site du département d'informatique et de génie logiciel (<http://www.ift.ulaval.ca>) de l'Université Laval (<http://www.ulaval.ca>).

L'équipe **DAMAS** (Dialogue, Agents et Multiagents) (<http://www.damas.ift.ulaval.ca/>) est animée par le professeur Brahim Chaib-draa, Ph.D. (Chaib@ift.ulaval.ca).

Les thèmes de recherche de cette équipe sont les suivants : dialogues et communications entre agents (argumentation, dissonance, changements des attitudes, protocoles et jeux de dialogue, simulation) ; négociation entre agents (hyperjeux, cartes cognitives, enchères, coalitions, conception de mécanismes) ; environnements multiagents en temps réel (planification multiagent en temps réel, établissement d'horaires (*scheduling*) multiagent en temps réel ; apprentissage dans les environnements multiagents (apprentissage pour la coordination, la communication et l'apprentissage, planification multiagent) ; applications pour les techniques multiagents comme les chaînes logistiques, le commerce électronique, le *RobotCupRescue*, les systèmes de défense, l'automobile du 21^{ème} siècle, etc.

Parmi les projets en cours, on trouve ceux nommés ci-dessous.

Conception de mécanismes révélateurs en informatique. Dans les environnements distribués comme l'Internet, les protocoles gérant les interactions entre agents ont été conçus avec l'hypothèse que ces derniers obéissent à certaines règles, comme le principe de sin-

cerité. Or, dans bien des situations plausibles en pratique, les agents auraient tout avantage à ne pas être sincères. Aussi, de nouveaux protocoles à l'épreuve des manipulations comme ceux qui forcent les agents à révéler la vérité sur leur état doivent être développés afin d'éviter une dégradation de la performance des systèmes distribués.

Frégates Canadiennes. Ce projet vise à étudier les systèmes multiagents pour l'implémentation d'un système d'aide à la décision affecté à la gestion des ressources à bord d'une frégate canadienne de classe HALIFAX. Le but du système est de gérer toutes les ressources (armes, radars, systèmes électroniques, etc.) disponibles sur un navire de guerre, de type frégate, de manière à augmenter ses chances de survie lors d'attaques par des missiles aériens (ASM).

Chaînes logistiques. Ce projet vise à améliorer les techniques de gestion d'une chaîne logistique en étudiant par la simulation l'impact sur l'ensemble d'une telle chaîne de la décision prise par une seule entreprise. L'objectif est de proposer de nouveaux outils d'aide à la décision basés sur les nouvelles technologies de l'information.

Dialogue intelligent. Ce projet porte sur l'étude de la co-construction de dialogues dans les communications entre agents. Trois types de fondements théoriques sont à prendre en compte : les bases théoriques de l'informatique, les théories du langage naturel, les modèles d'agents. Le groupe Dialogue explore le domaine de la communication entre agents comme étant aussi bien de l'informatique avancée qu'un sous-ensemble des sciences cognitives. Le but premier est de prendre en compte des caractéristiques naturellement présentes dans les sociétés humaines.

Auto 21. Ce programme réunit plus de 200 chercheurs dans 28 universités situées partout au Canada. Il est organisé autour de six thèmes de recherche très variés : santé, sciences sociales, matériaux, fabrication assistée par ordinateur, carburants, émissions polluantes, systèmes intelligents, etc. Il y a 28 projets de recherche, chacun ayant un chef de projet qui coordonne au moins trois autres chercheurs provenant d'au moins trois institutions différentes dans au moins deux provinces canadiennes.

Collaborative Driving System. Ce projet implique des chercheurs de l'équipe DAMAS, de l'Université de Sherbrooke et de l'Université de Calgary. Il vise à contribuer à la conception d'un peloton de voitures. Dans un premier stade, ce peloton sera semi-autonome. Par la suite, le peloton deviendrait complètement autonome et une personne n'interviendrait qu'en cas de besoin. Le travail de l'équipe DAMAS s'inscrit dans une problématique de systèmes multiagents, en particulier pour la communication et la coordination entre voitures.

RobocupRescue. La *RoboCupRescue* est un championnat de simulation de gestion de catastrophes naturelles. Ce championnat se concentre sur les stratégies de planification et de coordination d'équipes de sauvetage. Il simule un tremblement de terre survenant dans une ville et pour lequel on vise à coordonner les différents agents (pompiers, ambulances, policiers, etc.). L'objectif est d'élaborer des algorithmes d'apprentissage multiagents qui pourraient minimiser les communications tout en améliorant la coordination.

L'Équipe d'**Informatique Cognitive et Géomatique** ([http:// www.ift.ulaval.ca/~moulin/](http://www.ift.ulaval.ca/~moulin/)) est animée par le professeur Bernard Moulin, Ph.D. (bernard.moulin@ift.ulaval.ca).

Cette équipe poursuit des recherches sur la représentation des connaissances, sur les méthodes d'analyse et de conception de systèmes à base d'agents logiciels et sur les applications de l'intelligence artificielle à la géomatique.

Ces recherches sont appliquées dans divers domaines dont le traitement du langage naturel (modélisation du contenu sémantique de textes, utilisation de graphes conceptuels, modélisation de conversations) ; la gestion des connaissances (leçons apprises, systèmes d'explications) ; la modélisation des interactions entre agents (modèles de conversation, argumentation); la représentation des connaissances spatiales (conception d'espaces géographiques, description d'itinéraires) ; la représentation des connaissances temporelles (dans les discours); les méthodes de conception de systèmes multiagents ; le développement d'applications à base

d'agents logiciels (ex. planification de trajet de navires dans le St-Laurent, système d'évacuation aéro-médicale à base d'agents mobiles et stationnaires, agents d'indexation et de recherche de documents sur le Web, agents logiciels évoluant dans un environnement virtuel 3D urbain) ; la simulation de l'utilisation d'espaces géographiques à base d'agents logiciels (comportements de foules en milieu urbain ou dans des parcs naturels, comportement de consommateurs dans des centres d'achat).

Plusieurs recherches explorent l'utilisation de techniques d'intelligence artificielle dans le domaine de la géomatique (Réseaux de neurones appliqués à la généralisation cartographique, approche par patterns, généralisation par l'exemple, simulation de l'utilisation d'espaces géographiques par des agents logiciels, interopérabilité de systèmes).

L'équipe du professeur Jean-Marie Beaulieu, Ph.D. (beaulieu@ift.ulaval.ca) effectue des travaux de recherche en analyse de l'image appliquée à la télédétection.

Le principal domaine de recherche est l'analyse automatique de l'image appliquée à la télédétection, et

plus particulièrement l'analyse des images radar (SAR-Synthetic Aperture Radar). Celles-ci sont fortement

bruitées (« speckle »), ce qui rend leur traitement très difficile. La recherche porte sur la détection des composantes de bas niveau et implique l'identification des lignes, des régions et des discontinuités. Cette équipe a développé une technique de regroupement hiérarchique par optimisation séquentielle pour la segmentation d'images. Cette technique utilise un critère d'étape. L'équipe a adopté ce critère aux caractéristiques des images radar. La valeur élevée du bruit dans les images

radar oblige à utiliser des critères de forme pour guider la segmentation. L'équipe a examiné le problème de la segmentation des images polarimétriques (SAR) et mis au point une technique très intéressante de filtrage des images SAR. La technique vise à réduire le bruit tout en préservant les discontinuités. Une technique de segmentation est utilisée. L'équipe exploite également la connaissance a priori de la distribution du signal.

L'équipe de recherche en **Intelligence Computationnelle** (<http://www.ift.ulaval.ca/~lci/>) est animée par le professeur Guy Mineau, Ph.D. (mineau@ift.ulaval.ca), également directeur du département d'informatique et de génie logiciel.

Cette équipe s'est donné pour mission d'explorer des avenues informatiques qui 1) proposent de doter d'autonomie tout système informatique ; 2) permettent de faire évoluer des systèmes informatiques de l'ère du traitement des données à l'ère du traitement (et de la gestion) des connaissances ; 3) favorisent les transferts technologiques qui en découlent.

Parmi les thèmes de recherche, on trouve les suivants : la modélisation des connaissances humaines et son utilisation par un système informatique, plus particulièrement les notations graphiques telles que les graphes conceptuels ; la théorie de l'apprentissage automatique (apprentissage effectué par l'ordinateur), incluant les techniques d'analyse, de regroupement conceptuel, d'abstraction et de résumé de données, telles les techniques de Knowledge Discovery et Data Mining ; la théorie des agents logiciels et leur structuration en société d'agents, communément appelée système multiagent ; l'utilisation des agents dans des systèmes d'aide à la décision et de support à la tâche.

Parmi les projets en cours, mentionnons les suivants.

Simplification de règles de classification par réécriture, une approche basée sur la création d'arbres alternatifs. Les arbres de décision sont devenus, avec les années, un type de classificateur fort populaire. Cependant, l'ensemble des règles de classification qu'ils représentent devient rapidement difficile à comprendre pour un être humain. L'équipe propose une méthode qui permet de développer des ensembles de règles alternatives plus courtes et moins complexes.

Simplification de règles de classification basée sur un mécanisme de factorisation. Ce projet a pour but de diminuer la complexité des représentations graphiques de règles de classification. Une structure compacte de représentation de règles, appelée EDAG (Exception Directed Acyclic Graph) est exploitée.

Appariement de graphes conceptuels et apprentissage automatique. Le but de ce projet est d'explorer diverses techniques d'apprentissage automatique pour pouvoir faire apprendre à un système les appariements de graphes qui auront le plus de chances de succès selon une fonction d'évaluation de la qualité des appariements proposés, et conséquemment, pour deux graphes donnés, proposer l'appariement le plus adéquat selon cette même fonction.

Description graphique de contraintes sémantiques sous UML. Ce projet vise à élaborer un formalisme graphique simple pour l'expression de contraintes lorsqu'on utilise le langage UML.

Modélisation intégrée des systèmes multiagents et de leur environnement. Ce projet propose d'ajouter deux nouveaux diagrammes aux méthodologies applicables aux Systèmes Multiagents (SMA) : le diagramme d'environnement et le diagramme d'agents. Le premier permettra de représenter l'évolution de l'environnement dans le temps tandis que le second représentera la structure des agents (rôles et tâches) ainsi que leurs relations (contrôle, communication et collaboration).

Un système de sélection automatique d'algorithmes de forage de données : une approche basée sur le méta-apprentissage. Un utilisateur naïf est souvent incapable de déterminer quel algorithme est le plus approprié pour les données à analyser (réseaux de neurones, générateurs d'arbres de décision, apprentissage à base d'instances, réseaux bayésiens, classificateurs conceptuels, etc.) Ce projet vise à réaliser un aiguilleur automatique de requête de forage de données.

Techniques d'indexation de patterns orientés objet dans un but de réutilisation. La réutilisation logicielle repose sur l'utilisation de connaissances ou d'objets d'un ou plusieurs systèmes existants pour développer de nouveaux systèmes. Les patrons (*patterns*)

supportent la réutilisation à un niveau d'abstraction élevé en favorisant la documentation, la compréhension et la réutilisation des éléments du système. Ce projet consiste à indexer les éléments des patrons pour faciliter leur recherche et leur réutilisation.

Le Data Mining distribué: filtrage de règles basée sur l'exploration de treillis de concepts complémentaires. Le regroupement sur un même ordinateur de données géographiquement distribuées pour y créer un

classificateur peut s'avérer très coûteux. Dans ce projet, l'équipe propose de construire plusieurs classificateurs, chacun bâti sur un petit ensemble de données. Ensuite, les règles constituant ces classificateurs sont regroupées puis filtrées, en se basant sur des mesures statistiques de leur validité, et ce afin de construire un méta-classificateur, qui donne une vue abstraite de tout l'ensemble de données et peut servir comme outil de prédiction pour toute nouvelle instance.

L'équipe **AMICA** (Acquisition et Modélisation Interactives de Connaissances pour l'Apprentissage) est animée par le professeur Ruddy Lelouche, Ph.D. (Ruddy.Lelouche@ift.ulaval.ca).

Les principaux thèmes de recherche de cette équipe portent sur l'informatique cognitive, les applications pédagogiques et les interfaces intelligentes et adaptatives.

Parmi les projets en cours, mentionnons les suivants.

Interfaces intelligentes dans les systèmes EIAO. Ce projet porte sur l'étude des caractéristiques que devrait posséder un bon système d'Enseignement Intelligent Assisté par Ordinateur (EIAO). On y aborde notamment les aspects utiles à la conception des interfaces personne-machine, tels l'apprentissage automatique, l'ergonomie cognitive, et le traitement de la langue naturelle.

TIRPA (Tuteur Intelligent pour la Résolution de Problèmes en Algorithmique). Le problème est d'amener le système à suivre l'étudiant au cours de sa démarche de conception de l'algorithme, sans toutefois le forcer à adopter une des stratégies de résolution du système. Le but de ce projet est de concevoir directement une interface usager plus robuste et conviviale, incluant diverses possibilités telles que: génération

automatique en langue naturelle de l'énoncé du problème et d'une instance de ce problème à partir des spécifications, diagnostics sélectifs adaptés à l'apprenant, feedback à deux temps, etc.

Système tuteur intelligent pour l'analyse économique. Ce projet consiste à concevoir et à développer un système tuteur intelligent pour l'analyse économique, dans le but d'étendre l'expertise en EIAO de l'équipe à la fois à un nouveau domaine d'apprentissage et à une nouvelle approche méthodologique pour les interventions tutorielles du système. Le tuteur interviendra essentiellement à la demande de l'apprenant, de quatre manières différentes selon que celui-ci désire approfondir un point particulier de la matière ou effectuer des exercices d'application, et selon que la connaissance impliquée porte sur le domaine lui-même ou sur la résolution de problèmes.

Use of abstraction levels in the design of intelligent tutoring systems. Dans ce projet, on propose d'utiliser l'abstraction comme paradigme unificateur dans la conception de systèmes éducatifs intelligents informatisés, notamment dans les systèmes tuteurs intelligents.

L'équipe **AWSA** (Agents du Web Sémantique et Applications) (<http://www.ift.ulaval.ca/~kone/index.htm>) est animée par le professeur Mamadou Koné, Ph.D (Kone.Mamadou@ift.ulaval.ca).

À court terme, l'équipe se concentre sur la nature, les propriétés et les composantes de la communication entre agents et propose trois objectifs complémentaires : construire une architecture pour l'interaction agent-service, redéfinir la coopération et la coordination entre agents, établir une plate-forme d'interaction personne-agent. Le but est d'apporter le mode de coopération flexible utilisé dans la recherche en agents

mobiles et dans les systèmes multiagents et de l'appliquer en sus des protocoles d'interaction rigides/mécaniques qui sous-tendent le Web Sémantique.

Les thèmes de recherche de cette équipe sont autant fondamentaux qu'appliqués. Les travaux s'articulent autour des axes suivants : le Web sémantique, les services Web, les agents mobiles, les systèmes multiagents, le commerce électronique.

L'équipe **ERICAÉ** (Équipe de Recherche en Ingénierie des ConnAissancEs) (ericae@ift.ulaval.ca) (<http://www.ift.ulaval.ca/~ericae/>) est animée par les professeurs Laurence Capus, Ph.D., (capus@ift.ulaval.ca) et Nicole Tourigny, Ph.D. (tourigny@ift.ulaval.ca).

L'objectif général est d'élaborer des méthodes, des modèles, des techniques et des outils en ingénierie des connaissances pour la modélisation, la conception et le développement de systèmes intelligents. L'un des principaux buts est d'aider à développer des systèmes à base de connaissances plus faciles à maintenir et à expliquer.

Les principaux thèmes de recherche sont la modélisation au niveau des connaissances, notamment pour les systèmes utilisant le raisonnement à partir de cas; la modélisation cognitive de tâches à apprendre ou à enseigner; les méthodes de développement; les explications; l'acquisition des connaissances; la représentation des connaissances, en particulier avec les graphes conceptuels et les objets; la réingénierie; la génération de description de systèmes dynamiques; la vérification et la validation des connaissances; les méthodes de résolution de problèmes; les interfaces personne-machine; la gestion des connaissances.

Les domaines d'applications privilégiés sont le résumé, les systèmes intelligents de transport, les systèmes intelligents d'aide à l'apprentissage humain, le Web sémantique.

Parmi les projets en cours, mentionnons ceux ci-dessous.

- **SPEAC.** Modélisation d'un Système de Présentation d'Exemples favorisant l'Auto-explication et la Collaboration. Ce système est destiné aux étudiants universitaires de 1er cycle pour l'apprentissage de l'intelligence artificielle. Il vise à favoriser l'apprentissage humain grâce à des activités permettant l'auto-explication d'exemples de résolutions de pro-

blèmes et à la collaboration entre apprenants pour construire une bonne explication.

- **Interfaces intelligentes.** Ce projet vise à élaborer des principes permettant la conception d'interfaces intelligentes pour les systèmes à base de connaissances, et plus particulièrement pour les systèmes destinés à l'apprentissage humain.
- **Web sémantique : l'amélioration des portails.** Ce projet vise à améliorer les portails grâce à une meilleure utilisation des connaissances. L'un des sous-projets porte sur la modélisation d'un portail de bibliothèques utilisant le Web sémantique dans le but de faire communiquer divers systèmes d'information grâce à des ponts ontologiques.
- **Modélisation de systèmes hybrides.** Ce projet porte sur la modélisation d'outils facilitant l'ingénierie des connaissances de systèmes hybrides, tant au niveau de la modélisation que de la conception. Un sous-projet porte sur la modélisation d'un système à base de connaissances pour la localisation d'avions disparus.
- **VCOH.** Modélisation d'un système pour aider à vérifier la cohérence dans les bases de connaissances selon une approche utilisant les réseaux neuronaux.
- **GARUCAS.** Modélisation d'un système d'aide pour apprendre à résumer au moyen d'exemples dans un environnement de raisonnement à partir de cas.
- **MOCASSIN.** Modélisation d'un atelier de rédaction de résumés utilisant diverses techniques de résumés automatiques dans un environnement de raisonnement à partir de cas. Un sous-projet porte sur la comparaison de graphes conceptuels.

Dans un prochain numéro, je présenterai une bibliographie des travaux effectués en intelligence artificielle au Québec.

Si vous avez des questions ou des commentaires, n'hésitez pas à me contacter.

Nicole Tourigny, Ph.D., professeur

Directrice LSI

Département d'informatique et de génie logiciel

Pavillon Adrien-Pouliot

Université Laval

Québec, CANADA

G1K 7P4

Téléphone : (1-418) 656-3432

Télécopieur : (1-418) 656-2324

tourigny@ift.ulaval.ca

<http://www.ift.ulaval.ca/~ericae/>

PRÉSENTATION DE LABORATOIRES

Présentations de laboratoires dans le bulletin de l'AFIA

LIFIA, Grenoble	Bulletin n°1	Institut autrichien de recherches en I.A.	Bulletin n°20
LRI, Orsay	Bulletin n°1	ENST Bretagne	Bulletin n°21
Service Systèmes Experts, Renault.....	Bulletin n°1	LIA - Université de Savoie	Bulletin n°22
CEDIAG,.....	Bulletin n°2	INRETS.....	Bulletin n°23
CERT, ONERA, Toulouse.....	Bulletin n°2	IRIN Nantes	Bulletin n°24
IRIT, Toulouse.....	Bulletin n°2	CRIN - INRIA Lorraine.....	Bulletin n°25
LAAS, Toulouse	Bulletin n°2	DIRO - Université de Montréal	Bulletin n°26
HEUDIASYC, UTC.....	Bulletin n°3	IRIT - Toulouse (1)	Bulletin n°28
IFP, Rueil Malmaison.....	Bulletin n°3	IRIT - Toulouse (2)	Bulletin n°29
DIAM, INSERM U194	Bulletin n°3	LAAS - Toulouse (1)	Bulletin n°30
Lab. Math. Info., Fac Médecine de Marseille..	Bulletin n°4	Sony CSL	Bulletin n°31
GMD, St. Augustin (RFA)	Bulletin n°4	LAAS - Toulouse (2)	Bulletin n°32
ONERA, Chatillon	Bulletin n°4	LIMSI - Département CHM	Bulletin n°33
KSL, Université de Stanford (USA)	Bulletin n°5	LAMSADE	Bulletin n°34
Dépt Applications de l'IA au CNET, Lannion.	Bulletin n°5	Institut autrichien de recherches en I.A.	Bulletin n°36
LAFORIA, Univ. Pierre et Marie Curie.....	Bulletin n°6	LIP6 – Université Pierre et Marie Curie.....	Bulletins n°37 & 38
L'institut FAW, ULM (RFA)	Bulletin n°6	GREYC – Université de Caen	Bulletin n°40
Institut IIIA, Compiègne	Bulletin n°6	LIFL – Université de Lille	Bulletin n°41
LAIR, OHIO State University (USA).....	Bulletin n°7	LRI (équipes IA et IASI)	Bulletin n°43
ARAMIHS, Labo mixte MATRA-CNRS,.....	Bulletin n°7	IMAG - Grenoble.....	Bulletin n°44
CEA, Service SERMA, Saclay	Bulletin n°8	PSI (Perception, Système, Information - Rouen).....	Bulletin n°45
Société ILOG.....	Bulletin n°8	INRIA – Sophia Antipolis	Bulletin n°46/47
LAIAC, Université de Caen	Bulletin n°9	LIH – Laboratoire d'Informatique du Havre	Bulletin n°46/47
Institut Français du Pétrole	Bulletin n°10	Tech-CICO – Université de Technologie de Troyes	Bulletin n°51
DFKI (Centre allemand de recherches en IA) .	Bulletin n°11	LIFO – Université d'Orléans – Équipe Contraintes et Apprentissage.....	Bulletin n°52
GRTC, Marseille	Bulletin n°11	LIIA – Ecole Nationale des Arts et Industries de Strasbourg	Bulletin n°52
Inst. d'Analyse des Systèmes, Ac. Russe	Bulletin n°12	LRL – Laboratoire de recherche sur le langage – Université Blaise Pascal Clermont 2	Bulletin n°53
Georges Mason Univ., Center for AI (USA)....	Bulletin n°13	MIG – Mathématique, Informatique et Génome – INRA.....	Bulletin n°53
IRISA, INRIA et Université de Rennes	Bulletin n°13	ESIEA Recherche, Laval et Paris	Bulletin 54
Société INGENIA	Bulletin n°14		
LIPN, Université de Paris Nord	Bulletin n°14		
Institut EURISCO	Bulletin n°15		
LRDC, Université de Pittsburgh (USA).....	Bulletin n°15		
Société ISOFT	Bulletin n°16		
Dépt. d'Info de l'Université d'Ottawa	Bulletin n°16		
Equipe CHM, Université du Colorado (USA) .	Bulletin n°17		
LIRMM, Montpellier	Bulletin n°19		

ESIEA Recherche

Directeur : Anne-Marie Kempf - mél : recherche@esiea.fr
38, rue des Docteurs Calmette et Guérin, 53000 LAVAL
et
9, rue Vésale, 75005 PARIS
Web : <http://esiea-recherche.esiea.fr>

Présentation

ESIEA Recherche est un centre de recherche privé (association loi 1901) créé en Septembre 1997. Ses équipes, à Laval et à Paris, sont constituées de 18 personnes qui travaillent à la constitution de nouveaux modèles pour la Découverte de Connaissances et à la mise au point des méthodes et outils associés. Véritablement ancrées dans la réalité des entreprises, y puisant des thèmes de recherche et y trouvant à la fois des champs de validation de leurs résultats et des opportunités de valorisation, elles développent une recherche fondamentale de haut niveau, sanctionnée par des publications nombreuses de grande qualité.

Les travaux de ESIEA Recherche portent sur l'apprentissage des Réseaux Bayésiens, l'Extraction Visuelle de Connaissances à partir de Données, les approches multi-agents pour la préparation et la fouille de données, l'analyse des données symboliques et les méthodes inductives pour le traitement automatique de la langue naturelle.

Les deux premiers axes sont les plus aboutis. Nous les décrivons en détail ci-après.

L'équipe « Visualisation en Extraction de Connaissances dans les Bases de Données »

Composition

François Poulet
Pierrick Chantrel
Lydia Boudjeloud
Thanh-Nghi Do
Edwige Fangseu-Badjio

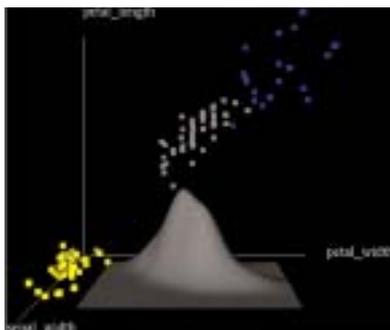
Résultats de recherche

Dans le processus de fouille de données, la visualisation n'intervient habituellement que lors de deux étapes particulières :

- l'une des toutes premières étapes pour « voir » les données brutes, ou la distribution de ces données,
- l'une des dernières étapes, celle qui permet de prendre connaissance des résultats obtenus par l'algorithme utilisé.

Entre ces deux étapes, il y a généralement exécution d'un algorithme automatique de fouille de données. Le but des recherches menées en « Visualisation et ECD » est de donner un rôle plus important à la visualisation dans le processus d'extraction de connaissances. Ceci peut être mené à bien de différentes manières :

- 1) en remplaçant l'algorithme automatique de fouille de données par un algorithme graphique interactif, on parle alors de fouille visuelle de données ou « Visual Data Mining »,
- 2) en faisant collaborer ensemble des outils graphiques et des outils automatiques,
- 3) en utilisant simultanément différentes représentations graphiques des mêmes données.



À l'heure actuelle, nos travaux de recherche s'effectuent essentiellement dans ces trois directions.

Fouille visuelle de données

La fouille visuelle de données ou Visual Data Mining peut être définie comme l'étape dans le processus d'extraction utilisant la visualisation comme canal de communication entre l'homme et la machine dans le but de produire de nouvelles connaissances. C'est l'utilisateur qui est au centre du processus et, contrairement à la plupart des autres approches utilisées à l'heure actuelle, l'utilisateur n'est pas ici un spécialiste en fouille de données, mais un spécialiste des données traitées.

Il s'agit d'une nouvelle approche (dont les premiers travaux remontent à 1999-2000) en fouille de données visant à impliquer de façon plus significative l'utilisateur dans le processus de fouille de données. Parmi les avantages de ce nouveau type d'approche, on peut citer :

- la possibilité d'utiliser les capacités de la perception humaine en reconnaissance des formes,
- l'utilisation de l'expertise de l'utilisateur dans le domaine des données pendant tout le processus de fouille,
- une confiance dans le modèle obtenu et une compréhension accrues puisque l'utilisateur a lui-même participé à sa création.

Plusieurs algorithmes interactifs de construction d'arbres de décision sont apparus dans les années 2000-2001 (cf. thèse de l'Université de Munich de M.Ankerst ou DTViz présenté à PAKDD'2001 par Han et Cercone). Ils sont basés sur une représentation initiale des données sous la forme de pixels (on parle de technique de « pixeli-

sation »), et permettent de créer de manière graphique interactive un arbre de décision. Les coupes de l'arbre sont des coupes unaires (c'est-à-dire ne faisant intervenir qu'un seul attribut ou variable). Graphiquement, cela signifie que la coupe est perpendiculaire aux axes. Nous avons fait un autre choix, celui de permettre des coupes « obliques en 2D », c'est-à-dire bivariées (faisant intervenir une combinaison linéaire de deux attributs) dans le but de réduire la taille (en nombre de nœuds) des arbres obtenus. Avec un tel choix, notre algorithme sera au pire (dans le cas où l'on n'effectue que des coupes unaires, perpendiculaires aux axes) équivalent aux deux précédents.

L'algorithme CIAD (Construction Interactive d'Arbres de Décision) que nous avons développé, est basé sur une représentation graphique des données initiales sous la forme d'un ensemble de projections 2D des données. Elles sont projetées suivant toutes les paires possibles d'attributs. À partir de cette représentation, l'utilisateur va interactivement construire son arbre de décision en dessinant (à la souris sur l'écran) les coupes successives de l'arbre, cherchant à mettre en évidence des zones pures. La première version de l'algorithme était 100% manuelle et imposait à l'utilisateur de tracer toutes les coupes correspondant à chaque nœud de l'arbre. Cette première version nous a permis de valider notre approche en comparant les résultats obtenus avec ceux des algorithmes automatiques. Le taux de précision est en général au moins aussi bon qu'avec les algorithmes d'induction d'arbres de décision automatiques (C4.5, CART, OC1, ...) et la taille des arbres obtenus est souvent nettement inférieure.

Une amélioration a alors été apportée à la version initiale de l'algorithme fournissant un mécanisme d'aide à l'utilisateur. Cette aide permet de calculer de manière automatique la meilleure coupe possible pour un nœud donné de l'arbre

en cours de construction. Pour ce faire, un première idée était d'utiliser un algorithme dérivé d'OC1 (un algorithme d'induction d'arbre de décision oblique), mais finalement l'idée retenue fut l'utilisation de Support Vector Machines (SVM). Cette famille d'algorithmes est apparue au milieu des années 1960 (en reconnaissance de formes) mais leur utilisation en fouille de données est beaucoup plus récente. Ils sont à l'heure actuelle parmi les plus performants des outils de fouille de données et sont de plus en plus utilisés dans des domaines nombreux et variés.

Avec cette nouvelle version appelée CIAD+, l'utilisateur a le choix entre un mode 100% interactif (manuel, où il trace lui-même le plan de coupe), un mode 100% automatique (l'arbre est construit de manière automatique comme les algorithmes usuels tels que C4.5 ou CART), et un mode mixte où l'utilisateur trace la séparatrice interactivement et la droite de séparation est automatiquement modifiée en LA meilleure droite de séparation entre les deux classes.

Pour résumer les résultats obtenus, nous pouvons dire que comparativement aux autres algorithmes interactifs, nous obtenons une précision au moins équivalente et la plupart du temps une taille d'arbre inférieure à précision égale. Cependant notre approche est pénalisée lorsque l'on doit traiter des ensembles de données de très grandes tailles. En effet, les projections 2D utilisées ne peuvent traiter qu'un nombre d'individus inférieurs à quelques milliers.

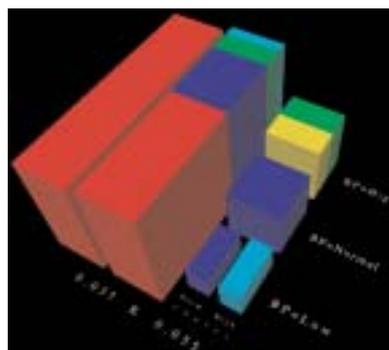
L'ensemble de ces travaux ne traite qu'un aspect du problème : si l'utilisateur d'un tel système n'est pas un spécialiste de la fouille ou analyse de données, mais un spécialiste des données, il faut lui permettre de passer avec succès les différentes étapes du processus, en partant des données et du problème initial à traiter jusqu'au résultat obtenu. Par exemple, à chacune de ces étapes, des choix doivent être faits, comme le choix

du meilleur type d'algorithme à utiliser en fonction du problème à résoudre et des données à traiter. Ce type de choix est habituellement fait par l'utilisateur (spécialiste en analyse de données). Ici il faut lui fournir un mécanisme d'aide pour lui faciliter les choix à faire pour arriver à une solution pertinente. Nous avons entamé des travaux dans cette voie.

Collaboration entre différents types d'outils

La description de CIAD+ est un premier exemple de collaboration entre outils graphiques et automatiques : un algorithme automatique dérivé des Support Vector Machines est utilisé pour aider l'utilisateur à construire de manière graphique, interactive un arbre de décision.

Un autre type de coopération utilisée est le post-traitement graphique d'un algorithme automatique. Ce type de post-traitement est destiné à améliorer la compréhensibilité des résultats des algorithmes automatiques utilisés. Il faut remarquer que ce critère de compréhension des résultats n'est absolument pas pris en compte lors de l'évaluation des performances des algorithmes à l'heure actuelle, alors qu'il est en général primordial pour le destinataire final de l'analyse. Des exemples de tels outils sont le « Tree Visualizer » et le « Decision Table Visualizer » qui étaient disponibles dans le logiciel MineSet de SGI. Nous avons appliqué ce post-traitement graphique à deux types d'algorithmes automatiques : les arbres de décision et les Support Vector Machines.



Les algorithmes d'induction d'arbres de décision donnent presque tous des résultats sous forme textuelle. Les résultats sont très faciles à comprendre pour un spécialiste de fouille ou analyse de données, mais ce n'est pas toujours le cas pour le spécialiste des données. Nous avons développé un outil de visualisation interactive des résultats fournis par un arbre de décision. Il s'agit d'une représentation graphique où chaque nœud de l'arbre est représenté par un cube dont la hauteur correspond au nombre d'individus contenus dans le sous-arbre correspondant, la couleur correspondant à la classe. Initialement, le cube représente la racine de l'arbre (et contient tous les individus). Par un clic de souris sur le cube en question, on descend d'un niveau dans l'arbre en effectuant une coupe suivant la valeur d'un attribut (représenté par exemple suivant l'axe des x). On obtient alors deux "sous-cubes".

Si l'on réitère le même procédé sur l'un des sous-cubes, on continue à descendre dans l'arbre en effectuant une autre coupe sur un autre attribut (représenté selon l'axe des y). On a donc un mécanisme permettant de parcourir graphiquement et interactivement les résultats obtenus par les algorithmes automatiques d'induction d'arbre de décision avec des mécanismes analogues au "drill-down" ou "roll-up" couramment utilisés en OLAP.

En ce qui concerne les SVM (Support Vector Machine), ils sont de plus en plus utilisés de nos jours et fournissent des résultats très intéressants dans bon nombre d'applications. Cependant, ils sont utilisés comme des boîtes noires. Les seuls résultats en général fournis par ces algorithmes sont l'équation de l'hyperplan de séparation en dimension m , (m étant le nombre d'attributs ou variables de l'ensemble de données traité) et un taux de bonne classification. À notre connaissance un seul article s'est intéressé en 2001 à l'interprétation des résultats fournis par les SVM. Cette

interprétation était basée sur une méthode de projections des vecteurs support par une méthode de "Grand Tour". Notre approche, quant à elle, consiste à visualiser l'hyperplan de séparation, ou plus précisément ses différentes projections en deux dimensions. Ce type d'information est approximatif (comme l'était la projection des vecteurs supports avec le Grand Tour et comme le sera toute représentation 2D d'un phénomène en nD). Par contre nous pensons que la projection en 2D de l'hyperplan de séparation est plus aisément interprétable que la notion de vecteur support qui impose une certaine connaissance du fonctionnement des SVM. Nous avons aussi étendu cette méthode pour visualiser les variations successives de l'hyperplan de séparation des deux classes (dans le cas d'utilisation d'un algorithme de SVM incrémental et décremental). Dans ce dernier cas, l'information est toujours approximative en ce qui concerne la position de l'hyperplan, mais la variation (ou non variation) de sa position au cours du temps est une information exacte. On peut ainsi mettre en évidence les attributs qui font varier la position de l'hyperplan et ceux qui n'ont aucun rôle sur cette modification temporelle. Il est ainsi aisé de détecter les attributs ou variables qui font varier la classification au cours du temps.

Ces deux exemples montrent l'intérêt qu'il y a à faire coopérer différentes méthodes (interactives et automatiques). Dans un cas, une méthode automatique permet d'améliorer les performances d'une méthode interactive et dans l'autre cas, on a le mécanisme inverse avec une méthode graphique qui permet d'améliorer la compréhension des résultats fournis par une méthode automatique. La collaboration entre les deux familles d'outils est bénéfique à la fois aux méthodes classiques et aux méthodes graphiques.

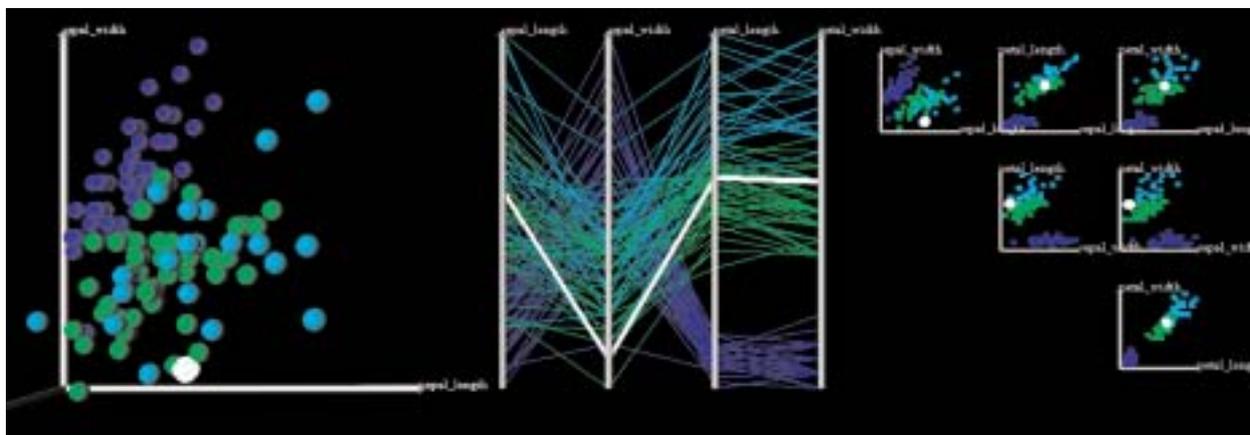
Différentes représentations simultanées et liées des mêmes données

Utiliser plusieurs représentations graphiques différentes des mêmes données peut aider à l'exploration, la compréhension de ces données. Plusieurs solutions ont déjà été étudiées, celle qui semble la plus intéressante dans la phase d'exploration des données est celle qui consiste à avoir dans une même fenêtre plusieurs représentations « en parallèle » des données (tirant ainsi avantage de chaque représentation). Ceci est encore plus vrai lorsque les vues sont liées entre elles, c'est à dire lorsque les actions effectuées dans une vue peuvent être automatiquement reportées sur les vues voisines. Ce type de mécanisme permet par exemple de sélectionner un individu particulier dans l'ensemble de données dans une vue, et que cette sélection apparaisse alors de manière automatique dans toutes les autres vues manipulant les mêmes données.

Nous avons développé un environnement graphique de fouille de données. Il regroupe une bonne quantité des algorithmes graphiques de visualisation (exploration) de données disponibles à l'heure actuelle. Etant donné que nous faisons aussi coopérer les différents types d'outils entre eux, nous lui avons ajouté des algorithmes automatiques de fouille de données tels que des arbres de décision, des règles d'association, etc.

Le fait d'avoir à disposition l'ensemble de ces outils simultanément dans un même environnement nous permet de plus d'éviter les habituels problèmes d'incompatibilité entre les formats de données lorsque l'on doit utiliser différents algorithmes.

Nous avons aussi testé l'utilisation réelle de la 3D lors de l'exploration des données (et non pas la projection en 2D d'objets 3D, ce qui est le cas dans la plupart des outils actuels). Il a été montré que le fait de visualiser les données en



3D permet d'augmenter la complexité de ces données de 60%, à compréhension égale pour l'utilisateur. Nous utilisons pour ce faire une visualisation stéréoscopique à l'aide de lunettes à cristaux liquides (disponibles de nos jours à faible coût sur la plupart des architectures) et un affichage alternant la vue de l'œil droit et la vue de l'œil gauche pour restituer la profondeur.

Ceci n'est qu'un premier pas vers un environnement utilisant des techniques de réalité virtuelle. La prochaine étape va être la réalisation d'une interface de navigation permettant à l'utilisateur de modifier les paramètres de la visualisation (zoom, rotation, translation, etc) à l'aide des mouvements de la main. Pour cela nous allons utiliser un gant de données et des capteurs de position 3D afin de repérer la position de la main de l'utilisateur et les mouvements de ses doigts pour les transformer en modifications du point de vue dans la scène visualisée afin d'augmenter le sentiment d'immersion et d'avoir un système de navigation naturel et intuitif permettant à l'utilisateur de se concentrer uniquement sur le résultat de la visualisation.

Publications :

F.Poulet, N.Do, « *Interactive Visualization Tools for Visual Data Mining* » in proc. of HCP'2003, Human Centered Processes'03, Luxembourg, May 2003.
 F.Poulet, N.Do, « *Mining Very Large Datasets with Support Vector Machine Algorithms* », in proc. of ICEIS'2003, 5th

Int.Conf. on Enterprise Information Systems, Angers, France, April 2003.
 N.Do, F.Poulet, « *Fouille de textes à l'aide de proximal Support Vector Machine* », in proc. of RIVF'2003, Recherche Informatique Vietnam Francophone, Hanoi, Feb.2003.
 F.Poulet, P.Kuntz, « *Visualisation et extraction des connaissances* », Atelier EGC'2003, 3e journées francophones d'Extraction et de Gestion des Connaissances, Lyon, Jan.2003.
 F.Poulet, « *Arbre de décision, clustering et SVM* », in proc. of SFC'2002, IX^e Rencontres de la Société Francophone de Classification, Toulouse, Sept.2002, pp.297-300.
 F.Poulet, N.Do, « *SVM parallélisé pour classifier un milliard de données* », in proc. of SFC'2002, IX^e Rencontres de la Société Francophone de Classification, Toulouse, Sept.2002, pp.301-304.
 F.Poulet, « *Cooperation between automatic algorithms, interactive algorithms and visualization tools for Visual Data Mining* », in proc. of ECML/PKDD 2nd Int. Workshop on Visual Data Mining, Helsinki, Aug.2002, pp.67-79.
 F.Poulet, « *FullView: A Visual Data Mining Environment* », in International Journal of Image and Graphics 2(1), 127-144, Jan. 2002.
 F.Poulet, « *La Visualisation en ECD* », Tutorial EGC'2002, 2e journées francophones d'Extraction et de Gestion des Connaissances, Montpellier, Jan.2002.
 F.Poulet, « *Construction interactive d'arbre de décision* », in proc. of SFC'2001, VIII^e Rencontres de la Société Francophone de Classification, Pointe à Pitre, Dec.2001, pp.297-300.
 F.Poulet, « *A Virtual Reality Environment for Visual Data Mining* », in proc. of VRIC'2001, 3rd Virtual Reality International Conference, Laval, May 2001.
 F.Poulet, « *CubeVis: Voir pour mieux comprendre* », in proc. of XXXIII^e Journées de Statistiques, Nantes, Mai 2001.
 F.Poulet, « *Visualisation des connaissances en ECD* », Tutorial EGC'2001, 1^{eres} journées francophones d'Extraction et de Gestion des

Connaissances, Nantes, Jan.2001.

F.Poulet, « *Improving Graphical Interactions and Result Comprehensibility in Data-Mining* », in proc. of International Symposium on Data Mining and Statistics, Augsburg, Germany, Nov.2000.

F.Poulet, « *Visualization in Data-Mining and Knowledge Discovery* » in Proc. of HCP'99, Human Centered Processes'99, Brest, Sept.1999.

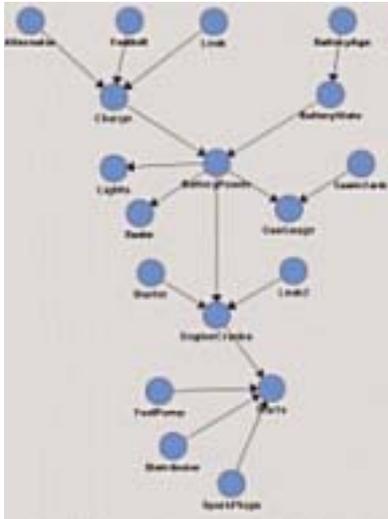
L'équipe « Réseaux Bayésiens »

Introduction

L'intérêt des réseaux Bayésiens pour l'ECD n'est plus à démontrer : ils offrent des solutions puissantes, dans un cadre théorique unitaire, à un grand nombre de tâches de fouille de données traditionnellement abordées par des techniques très hétérogènes : la classification, supervisée ou non, la découverte d'associations et la découverte de causalités. Ils permettent également une intégration naturelle des connaissances expertes. De plus, l'incertitude inhérente aux tâches d'ECD, généralement affectées par le bruit et l'indétermination, est traitée d'une manière rigoureuse dans ce cadre probabiliste.

Les travaux de recherche sur les réseaux Bayésiens ont été initiés au sein de ESIEA Recherche par Paul MUNTÉANU (chercheur) en 1998. Il a été rejoint par la suite par Lionel JOUFFE (chercheur), Denis CAU (ingénieur de

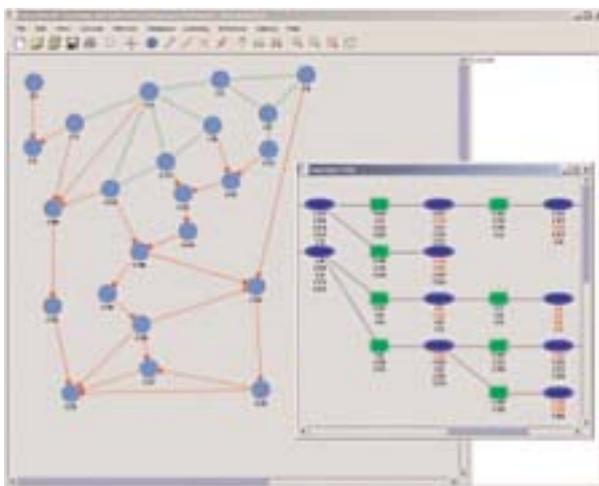
recherche) et Mohamed BENDOU (ingénieur de recherche, en fin de thèse de doctorat).



Résultats de recherche

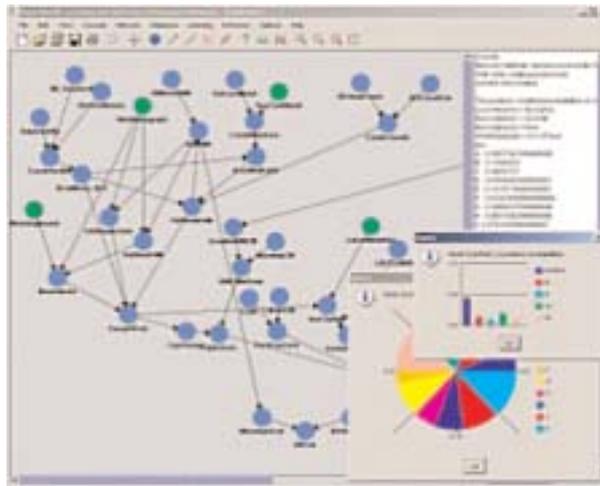
Nos travaux de recherche se sont focalisés essentiellement sur la conception de nouveaux algorithmes d'apprentissage de réseaux Bayésiens, plus précis, plus efficaces, plus robustes en présence du bruit et, donc, plus adaptés aux tâches pratiques d'ECD.

Un premier volet de ces recherches s'est attaqué au problème posé aux algorithmes d'apprentissage classiques par les optima locaux des fonctions d'évaluation, très nombreux dans l'espace des structures de réseaux Bayésiens et qui ont un impact négatif prononcé sur leurs performances.



Un premier volet de ces recherches s'est attaqué au problème posé aux algorithmes d'apprentissage classiques par les optima locaux des fonctions d'évaluation, très nombreux dans l'espace des structures de réseaux Bayésiens et qui ont un impact négatif prononcé sur leurs performances.

L'algorithme SG+ [1][3] contourne en partie ce problème en intégrant dans sa stratégie de recherche plus de sémantique propre aux structures de réseaux Bayésiens (modèles d'indépendances conditionnelles) que les fonctions d'évaluations classiques et en proposant un post-traitement qui continue l'affinement de la structure apprise dans l'espace des ordres topologiques des nœuds.



Partant de l'observation que la plupart des optima locaux dans l'espace des structures de réseaux Bayésiens sont directement liés à l'existence des classes d'équivalence (EQ) (ensembles de structures encodant les mêmes relations d'indépendance conditionnelle, représentées par des graphes semi-orientés), nous avons concentré

une partie importante de nos recherches sur l'élaboration d'une nouvelle famille d'algorithmes d'apprentissage : EQ [2][4], qui explorent directement l'espace des classes d'équivalence, ainsi que sur le développement d'une « boîte à outils » théoriques et algorithmiques pour l'analyse et le traitement des graphes semi-orientés [5][7]. Nous avons pu démontrer que les gains de précision significatifs apportés par ce type d'approche peuvent être obtenus tout en conservant des temps de calcul comparables à ceux des approches classiques et avons contribué ainsi au regain d'intérêt actuel pour l'apprentissage des classes d'équivalence de réseaux bayésiens (considéré pendant longtemps comme trop complexe par la communauté scientifique).

Dans ce cadre nous avons élaboré un Système Multi-Agents (SMA) pour la vérification de la consistance de certaines formes intermédiaires de graphes semi-orientés [7], appelés GSOI (Graphes Semi-Orientés Instanciables). La vérification de ce type de graphes est

nécessaire dans le processus de l'algorithme d'apprentissage des classes d'équivalence (EQ). Nous avons proposé une architecture d'agents réactifs dans laquelle nous nous sommes inspirés de notre approche conceptuelle de la détection des graphes semi-orientés instanciables qui consiste à détecter des sous-structures non-instanciables dans un graphe par propagation de messages. Chaque nœud est donc vu en tant qu'agent réactif, il communique et interagit avec d'autres agents. La propagation des messages dans le SMA est fondée sur la caractérisation des GSOI, les agents essayant de détecter à l'aide de la communication les sous-structures non-

instanciables. Pour ce faire, la propagation des messages est directement imposée par les règles d'orientation des arêtes des sous-structures du graphe en question.

Un autre concept théorique relativement peu utilisé dans l'apprentissage des réseaux Bayésiens est la couverture de Markov d'un nœud (ensemble des nœuds dont la connaissance de la valeur le rend indépendant de tous les autres nœuds du réseau). Nous nous sommes attachés à élaborer des méthodes qui cherchent à découvrir cette couverture de Markov (et, implicitement, le sous-réseau associé) « localement » (sans passer par la construction, somme toute assez coûteuse, du réseau Bayésien global). Utilisées en tant que pré-traitement, ces méthodes apportent un gain d'efficacité significatif à l'apprentissage des réseaux Bayésiens [6]. Elles sont particulièrement adaptées à l'apprentissage supervisé (caractérisation d'une variable « cible » à partir des autres variables).

Enfin, un troisième volet de nos recherches a été consacré à l'analyse des effets du bruit présent dans les données sur l'apprentissage des réseaux Bayésiens. Nous avons analysé et expliqué l'augmentation de la complexité des réseaux Bayésiens appris à partir de données bruitées et montré que, contrairement aux sur-spécialisations classiques affectant les autres classes de méthodes d'apprentissage, ce phénomène est justifié théoriquement et bénéfique pour le pouvoir prédictif des modèles appris [8][9].

Recherches en cours et perspectives

Au-delà de la poursuite de nos travaux sur l'apprentissage structurel des

réseaux Bayésiens, nous nous sommes orientés plus récemment vers l'étude des réseaux Bayésiens dynamiques, qui représentent explicitement le facteur temps, pour la modélisation de l'évolution temporelle des systèmes dynamiques complexes. Ces modèles ouvrent de nouvelles perspectives pour l'extraction des connaissances à partir de séries temporelles et des séquences et, couplés à des processus de décision Markoviens, sont de plus en plus utilisés pour la planification probabiliste (des actions de maintenance [10], par exemple).

Collaborations

Nous travaillons dans un domaine scientifique qui est, on doit le reconnaître, assez peu développé en France par rapport à son essor actuel sur le plan international. Une communauté de recherche française est cependant en train de se constituer autour des réseaux Bayésiens et nous essayons d'y apporter nos contributions (participation aux Journées Réseaux Bayésiens, aux manifestations organisées par le Collège Apprentissage, Fouille et Extraction de l'AFIA et par l'Action Spécifique STIC GaFoDonnées en lien avec les réseaux Bayésiens, etc.).

Le monde industriel semble également très ouvert et très intéressé par les réseaux Bayésiens. France Télécom R&D, PTOLMIX, Intelmark nous ont confié des études et des travaux de recherche et développement dans ce domaine.

Par ailleurs, une jeune société, BAYESIA, créée par essaimage de ESIEA Recherche, lauréate du Concours National de Création d'Entreprises Innovantes organisé par le Ministère de la Recherche, développe

des progiciels et des solutions métier s'appuyant sur les derniers résultats scientifiques dans ce domaine des Réseaux Bayésiens.

Publications :

- [1] L. Jouffe, P. Munteanu, Smart-Greedy+ : Apprentissage hybride de réseaux bayésiens, Colloque francophone sur l'apprentissage (CAP), St. Etienne, juin 2000.
- [2] P. Munteanu, D. Cau, Efficient Learning of Equivalence Classes of Bayesian Networks, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Lyon, septembre 2000.
- [3] L. Jouffe, P. Munteanu, New Search Strategies for Learning Bayesian Networks, Proceedings of Tenth International Symposium on Applied Stochastic Models, Data Analysis, Compiègne, juin 2001.
- [4] P. Munteanu, M. Bendou, The EQ Framework for Learning Equivalence Classes of Bayesian Networks, First IEEE International Conference on Data Mining (IEEE ICDM), San José, novembre 2001.
- [5] M. Bendou, P. Munteanu, Modèles graphiques semi-orientés pour l'apprentissage des réseaux bayésiens, Journées francophones d'Extraction et de Gestion des Connaissances (EGC), Montpellier, janvier 2002.
- [6] L. Jouffe, Nouvelle classe de méthodes d'apprentissage de réseaux bayésiens, Journées francophones d'Extraction et de Gestion des Connaissances (EGC), Montpellier, janvier 2002.
- [7] M. Bendou, MAS Approach for Detecting Instantiable PDAG, 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI), Orlando, juillet 2002.
- [8] M. Bendou, P. Munteanu, Analyse de l'effet du bruit dans les algorithmes d'apprentissage des réseaux Bayésiens, Journées francophones d'Extraction et de Gestion des Connaissances (EGC), Lyon, janvier 2003.
- [9] M. Bendou, P. Munteanu, Learning Bayesian Networks from Noisy Data, 5th International Conference on Enterprise Information Systems (ICEIS), Angers, avril 2003 (à paraître).
- [10] P. Weber, L. Jouffe, Reliability Modelling with Dynamic Bayesian Networks, 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS), Washington, juin 2003 (à paraître).

Dossier Web sémantique

Coordonneurs :

Chantal Reynaud (LRI), J. Charlet (AP-HP), Ph. Laublet (LaLICC)

Nota : cette introduction est un court extrait d'un article de synthèse « Sur quelques aspects du Web sémantique » fait pour les assises du GdR-I3 que l'on peut retrouver à <http://sis.univ-tln.fr/gdri3/>. Ce même article a été influencé par les travaux que nous pilotons au sein de l'action spécifique « Web sémantique » du département STIC du CNRS <<http://www.lalic.paris4.sorbonne.fr/stic/>>. Nous remercions les participants pour leurs différents apports.

L'expression Web sémantique, due à Tim Berners-Lee au sein du W3C, fait d'abord référence à la vision du Web de demain comme un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation, qualitativement supérieure, de grands volumes d'informations et de services variés. Espace virtuel, il devrait voir, à la différence de celui que nous connaissons aujourd'hui, les utilisateurs déchargés d'une bonne partie de leurs tâches de recherche, de construction et de combinaison des résultats, grâce aux capacités accrues des machines à accéder aux *contenus* des ressources et à effectuer des *raisonnements* sur ceux-ci.

Le Web sémantique, concrètement, est d'abord une *infrastructure* pour permettre l'utilisation de connaissances *formalisées* en plus du contenu informel actuel du Web, même si aucun consensus n'existe sur jusqu'où cette formalisation doit aller. Cette infrastructure doit permettre d'abord de localiser, d'identifier et de transformer des ressources de manière robuste et saine tout en renforçant l'esprit d'ouverture du Web avec sa diversité d'utilisateurs. Elle doit s'appuyer sur un certain niveau de consensus portant, par exemple, sur les langages de représentation ou sur les ontologies utilisés. Elle doit contribuer à assurer, le plus automatiquement possible, l'interopérabilité et les transformations entre les différents formalismes et les différentes ontologies. Elle doit

faciliter la mise en œuvre de calculs et de raisonnements complexes tout en offrant des garanties supérieures sur leur validité. Elle doit offrir des mécanismes de protection (droits d'accès, d'utilisation et de reproduction), ainsi que des mécanismes permettant de qualifier les connaissances afin d'augmenter le niveau de confiance des utilisateurs.

Mais restreindre le Web sémantique à cette infrastructure serait trop limitatif. Ce sont les applications développées sur celle-ci qui font et feront vivre cette vision et qui seront, d'une certaine manière, la preuve du concept. Bien sûr, de manière duale, le développement des outils, intégrant les standards du Web sémantique, doit permettre de réaliser plus facilement et à moindre coût des applications ou des services développés aujourd'hui de manière souvent ad-hoc.

Les recherches actuellement réalisées s'appuient sur un existant riche venant, par exemple, des recherches en représentation ou en ingénierie des connaissances. Mais leur utilisation et leur acceptation à l'échelle du (ou d'une partie du) Web posent de nouveaux problèmes et défis : changement d'échelle dû au contexte de déploiement, le Web et ses dérivés (intranet, extranet), nécessité d'un niveau élevé d'interopérabilité, ouverture, standardisation, diversités des usages, distribution bien sûr et aussi impossibilité d'assurer une cohérence globale. *Comme l'écrit, en substance, Tim Berners-Lee, le Web sémantique est ce que nous obtiendrons si nous réalisons le même processus de globalisation sur la représentation des connaissances que celui que le Web fit initialement sur l'hypertexte.*

Les propositions faites autour de l'infrastructure du Web sémantique doivent permettre aussi bien la réalisation d'outils généralistes avec des utilisateurs mal définis (un exemple pourrait être des moteurs de recherche prenant plus en compte le contenu sémantique de docu-

ments) que la réalisation d'applications pour des tâches plus complexes comme la gestion de connaissances au service des membres d'une entreprise. On soulignera, dans le premier cas, surtout l'utilisation de méta-données et dans le deuxième, la nécessité de systèmes d'intégration de données hétérogènes ou bien encore d'utilisation et de combinaison de services Web. Les langages proposés pour le Web sémantique sont au cœur de la démarche, même si l'infrastructure ne se réduit pas à ceux-ci. Les ontologies jouent un rôle important. Enfin, de nombreux travaux sont liés à des communautés de pratique, loin de la vision universaliste du Web sémantique.

Il est clair que la diversité des recherches relevant aujourd'hui du Web sémantique rend illusoire toute volonté d'exhaustivité. En pratique, ce dossier est l'occasion pour nous de fournir un panorama des recherches conduites à ce sujet en France.

**Groupe Gemo –
Université Paris-Sud,
CNRS et INRIA**

L.R.I./ Université Paris-Sud
Bâtiment 490
91405 Orsay cedex
<http://www.lri.fr/iasi>

Personne à contacter

Marie-Christine Rousset
Mél : mcr@lri.fr
Tél : 01 69 15 61 97
Fax : 01 69 15 65 86

Membres de l'équipe

Serge Abiteboul, Bernd Amann, Hélène Gagliardi, Gloria Giraldo, François Goasdoué, Hassen Kefi, Amar

Djalil Mesaour, Benjamin Nguyen, Nathalie Pernelle, Antonella Poggi, Chantal Reynaud, Marie-Christine Rousset, Brigitte Safar, Alexandre Termier, Véronique Ventos.

Thème général de l'équipe

Les travaux du groupe combinent des techniques d'Intelligence Artificielle et de Bases de Données pour traiter des problèmes nouveaux suscités par l'explosion de l'accès à un grand nombre de sources de données et de services, via le Web par exemple. Ces travaux concernent, plus particulièrement, l'intégration sémantique de données et de services à l'aide d'ontologies et ont conduit au développement d'outils méthodologiques et algorithmiques de construction de médiateurs à base d'ontologies. Un médiateur fournit une interface homogène entre un agent qui pose une requête et un ensemble des sources d'information distribuées et hétérogènes. L'objectif est de donner l'impression d'interroger un système centralisé et homogène. Un médiateur basé sur une ontologie comprend une ontologie (également appelée schéma global), décrivant le domaine d'application du système. Cette ontologie fournit un vocabulaire structuré qui sert de support à l'expression des requêtes et qui permet aussi de décrire le contenu des différentes sources accessibles. Ces descriptions, que l'on appelle « mappings » parce qu'elles mettent en correspondance le schéma global du médiateur et les schémas locaux propres à chacune des sources, jouent un rôle central. En effet, répondre à une requête posée à un médiateur nécessite de réécrire la requête en termes des descriptions des sources.

Projets représentatifs

PICSEL(<http://www.lri.fr/~picsel>) est le résultat d'une collaboration avec France-Telecom R&D. Ce projet a donné lieu au dépôt de deux logiciels : OntoClass, un outil de classification automatique de sources de données et de services décrits dans les termes d'une ontologie à base de classes et OntoQuery un outil de reformulation de requêtes conjonctives mettant en oeuvre

le calcul des réécritures dans un médiateur, les requêtes étant également exprimées dans les termes d'une ontologie à base de classes. Ces deux logiciels ont été utilisés au sein du projet européen sur le commerce électronique multilingue IST-MKBEEM. Dans le cadre du projet PICSEL, nous avons également travaillé sur l'utilisation d'ontologies pour établir un dialogue coopératif entre un médiateur et ses utilisateurs. Plus précisément, nous avons conçu un système d'affinement de requêtes qui n'obtiennent pas de réponses et nous avons travaillé sur l'évaluation de la similarité de deux requêtes, de façon à faire des choix lorsque le nombre de réponses à une requête est trop élevé.

Dans le cadre du projet PICSEL2, toujours en collaboration avec France Telecom R&D, nous étudions actuellement comment construire de façon semi-automatique une ontologie à base de classes à partir d'un ensemble de DTDs relatives à un même domaine d'application. Un prototype, OntoMedia, a été développé.

Xyleme est une action de recherche qui a regroupé plusieurs équipes de recherche françaises et une allemande. Dans Xyleme, le schéma global a une simple structure d'arbre. Il s'agit de réaliser une intégration de données à très grande échelle puisque l'objectif consiste à construire un entrepôt dynamique contenant toutes les données XML du Web. Les choix au niveau architecture et conception reposent sur des exigences en matière de performance, Xyleme devant, par exemple, pouvoir répondre en un temps raisonnable à de nombreuses requêtes concurrentes. Le modèle de données de Xyleme comprend un schéma global modélisé par un type d'arbre abstrait, correspondant à une vue sur un ensemble de types d'arbres spécialisés, chacun correspondant à un arbre de données réel, appelé type d'arbre concret, et « un mapping » entre le schéma global et les types d'arbres concrets. La simplicité de la relation de mise en correspondance (correspondance entre des chemins dans des arbres) facilite la génération de « mappings » automatique. Un prototype, SAMAG, a été développé. Il génère automatiquement des « mappings »

entre chemins de DTDs relatives à un domaine donné et une ontologie à base d'arbre portant sur ce même domaine.

Dans StyX, un langage de « mapping » simple mais expressif a été développé. Il permet de décrire des ressources XML en termes d'une ontologie globale. Un algorithme efficace de réécriture de requêtes en Xpath ou Xquery a par ailleurs été implémenté.

Travaux de l'équipe sur le Web sémantique

Les recherches que nous engageons actuellement viennent renforcer le travail effectué sur les ontologies en tant qu'outils de médiation pour le Web sémantique et ceci dans trois directions : la médiation entre ontologies, entre sources de données et entre le Web et ses utilisateurs.

1. Médiation entre ontologies

La nature décentralisée du Web rend inévitable le fait que différentes ontologies devront cohabiter, chacune d'elles ayant été choisie par une certaine communauté d'utilisateurs pour décrire des données ou des services. Ainsi, les ontologies seront également distribuées et on devra être capable de les échanger, de les fusionner, de les mettre en correspondance, de les étendre. Une ontologie est un vocabulaire structuré partagé par des communautés d'utilisateurs ou de praticiens qui ne sont ni ingénieurs cogniticiens ni logiciens. Dans le cadre du Web sémantique, on peut ainsi s'attendre à ce qu'elles soient relativement simples. Elles pourront, par exemple, correspondre à des arbres de termes ou à de taxinomies de classes décrites par leurs noms et leurs attributs. Le problème qui se pose consiste alors à établir une médiation entre les données, les services et les utilisateurs reposant sur des mises en correspondance sémantiques entre ontologies. La comparaison, la combinaison et l'intégration de données et de services décrits à l'aide d'ontologies ne pourront être effectifs que si nous sommes capables de définir des « mappings » complexes et de mettre en oeuvre des raisonnements sur ces « mappings ». Ainsi, le problème, qui

nous paraît être au cœur des recherches sur le Web sémantique et sur lequel l'équipe travaille, n'est pas tant le choix d'un langage de description des ontologies que le choix d'un langage pour représenter les « mappings » entre ontologies.

2. Médiation pair-à-pair entre sources de données et services

Les travaux réalisés jusqu'alors dans le domaine des systèmes médiateurs (tels que Tsimmis, Information Manifold, Infomaster, PICSEL, Momis, Xyleme et StyX) se situent dans le contexte d'une médiation centralisée entre sources de données distribuées et hétérogènes, reposant sur l'exploitant d'un unique schéma global. Dans le cadre du Web sémantique, l'intégration de sources d'information pourra s'appuyer sur de multiples systèmes de médiation. Certains pourront suivre une approche centralisée. D'autres pourront suivre une approche décentralisée consistant à considérer une coalition de serveurs d'information. Chaque serveur pourra indifféremment jouer le rôle de serveurs de données ou de médiateurs avec ses pairs en participant de manière collective et distribuée au traitement des requêtes utilisateurs. Les connexions entre systèmes de médiation donneront au Web toute sa puissance, autorisant la recherche de données dans des sources non directement connectées aux sources du serveur interrogées.

3. Médiation entre le Web et ses utilisateurs

Nous envisageons d'étendre nos travaux portant sur l'affinement de requêtes au contexte du Web sémantique. Notre objectif est d'utiliser les ontologies décrivant les données et services disponibles pour répondre à des requêtes de façon plus personnelle et coopérative. Les ontologies peuvent en effet être exploitées pour affiner des requêtes dont les réponses sont trop nombreuses. Il s'agit de proposer des requêtes plus précises sémantiquement proches des requêtes initiales des utilisateurs. Il peut également être utile d'exploiter des ontologies pour ordonner des réponses à des requêtes selon des critères sémantiques. Enfin, des interfaces de formulation de requêtes

personnalisées peuvent être conçues à partir de profils utilisateurs exploitant des ontologies et permettant, par exemple, de distinguer des concepts selon leur niveau d'abstraction.

Bibliographie

(<http://osage.inria.fr/verso/Gemo/PU/BLI>)

- [1] M.-C. Rousset, A. Bidault, C. Froidevaux, H. Gagliardi, F. Goasdoue, C. Reynaud, B. Safar, Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes : le projet PICSEL, Revue I3, Volume 2, Numéro 1, pages 9-59, 2002.
- [2] A. Bidault, C. Froidevaux, B. Safar. Repairing queries in a mediator approach. Proceedings of ECAI, 2000.
- [3] S. Abiteboul, S. Cluet, G. Ferran, M.-C. Rousset, The Xyleme project, Computer Networks, 39, pp. 225-238, 2002.
- [4] C. Delobel, C. Reynaud, M.-C. Rousset, J.-P. Sirot, D. Vodislav. Semantic Integration in Xyleme: a uniform tree-based approach. Data and Knowledge Engineering, 44(2), pp. 267-298, 2002.
- [5] I. Fundulaki, B. Amann, C. Beeri, M. Scholl, A.-M. Vercoustre. StyX: connecting the XML world to the world of semantics, Proceedings of EDBT Demo, 2002.

Axe D2C « Données, Documents, Connaissances »

LIRIS FRE 2672 – CNRS – INSA-Lyon
Univ. Lyon 1 – Univ. Lyon 2 – Centrale
Lyon
Bâtiment Nautibus - Université Claude
Bernard Lyon 1
43, boulevard du 11 novembre 1918 –
69622 Villeurbanne cedex
<http://liris.univ-lyon1.fr/>

Responsable

Alain Mille – 04 72 44 58 24
amille@bat710.univ-lyon1.fr

Liste des membres concernés

Jesus Arana (Doct.), Ahmed Arara (Doct), Salima Benbernou (MCF), Djamel Benslimane (PR), Sylvie Calabretto (MCF), Etienne Canaud

(Doct.), Pierre Antoine Champin (PostDoc), Mohand-Saïd Hacid (PR), Salima Hassas (MCF), Alain Mille (PR), Jean-Marie Pinon (PR), Yannick Prié (MCF), Catherine Roussey (MCF), Christine Solnon (MCF), Arnaud Stuber (Doct.)

Thème général de l'équipe

Le défi à relever par les systèmes informatiques omniprésents en tant que supports pour les activités extrêmement variées de la société est de « faire sens » pour les utilisateurs et les uns pour les autres. Les utilisateurs (ou agents systèmes) peuvent exploiter les services mis à disposition en les interprétant comme connaissances ou comportements intelligents dans le contexte de leur tâche courante. Nous travaillons à l'élaboration de propositions théoriques et pratiques pour relever ce défi, qui pourrait se résumer par la question : « Comment faire émerger des connaissances ou/et des comportements intelligents des systèmes informatiques ? ». Nous avons pour cibles communes la découverte, la gestion, l'exploitation et le partage des connaissances telles qu'elles s'expriment dans ce que nous appelons des « traces documentaires ». Une trace documentaire est un conteneur de connaissances qui peut prendre des formes très variées : les documents dans leur acception la plus habituelle du terme, en tant qu'artefacts construits, et en particulier les documents numériques ; les contenus des bases de données en tant que conteneurs de connaissances ; les traces d'interactions liées à l'usage du système informatique par l'utilisateur ; les traces d'interactions entre agents informatiques ; les traces d'usages captées dans les grandes bases de données (permettant de découvrir des profils) ; les codes complexes régissant des phénomènes d'évolution ou de comportement (génomés, séquences d'événements,...) ; etc.

Notre thématique générale se spécifie en sous-thèmes plus spécifiques, dont les trois premiers sont concernés par la problématique du Web sémantique.

1. Modélisation sémantique de documents (SC, JMP, CR ‡2002)
2. Cognition, expérience et agents situés (JA, PAC, SH, AM, YP, AS)

3. Représentation de connaissances, raisonnement et bases de données (AA, SB, DB, EC, MSH, CR, CS)
4. Découverte de connaissances

Description des travaux ou projets en lien avec le Web sémantique

1. Web sémantique et recherche d'informations

Nous avons développé un nouveau modèle, les graphes sémantiques, adapté aux documents multilingues et s'appuyant sur la notion de thésaurus sémantique. Contrairement au thésaurus multilingue traditionnel, le rôle du thésaurus sémantique n'est pas de limiter ou prescrire le vocabulaire d'un langage documentaire, mais de normaliser les connaissances pour définir la modélisation du domaine utilisé lors de l'indexation. Ces connaissances sont associées à des termes qui les référencent suivant un certain angle d'interprétation. Ce nouveau thésaurus se situe entre les ontologies, pour leur fonction de modélisation, et les bases de connaissances terminologiques pour leur nécessité d'associer un terme à son contexte. Nous avons développé un outil d'annotations sémantiques de pages Web permettant d'enrichir ces pages par le biais de connaissances pour pouvoir les retrouver sans tenir compte de leur langue d'écriture. L'outil SyDoM (Système Documentaire Multilingue) possède une composante gestion de terminologie permettant d'identifier l'interprétation des termes dans leur contexte et d'associer ainsi à un concept les termes de plusieurs langues. Nos perspectives de recherche sur le Web sémantique nous orientent vers l'interopérabilité des conceptualisations ou comment interroger des pages Web annotées par SyDoM lorsque les annotations ont été créées à l'aide de thésaurus sémantiques différents.

2. Modéliser les USages Et les Tâches pour Tracer (réutiliser et partager) l'Expérience sur le WEB : l'approche MUSETTE.

Le WEB nécessite la prise en compte de la complexité et de la diversité des situations d'usage des systèmes informatiques et pose la question de l'assistance à ces usages. Après avoir étudié et

réalisé des systèmes s'intéressant à la réutilisation de l'expérience dans le cadre de tâches relativement ciblées bien que générales comme la conception et la supervision de systèmes complexes, nous nous intéressons à des classes de tâches encore plus génériques comme l'apprentissage humain à distance ou l'indexation et la recherche d'information sur le Web. Si l'on pouvait imaginer de pouvoir modéliser convenablement ce qu'il est convenu d'appeler la « théorie du domaine » pour des applications ou des usages bien précis, ceci devient bien sûr très difficile pour des usages correspondant à des tâches ou des problèmes non complètement répertoriés a priori, ni identifiables précisément a posteriori. C'est donc la vraie question posée par le défi du « Web sémantique » pour lequel nous proposons une approche radicalement fondée sur la mise en contexte, et par conséquent exigeant la prise en compte des contextes à différents niveaux dans l'observation, l'indexation des usages et leur réutilisation à des fins d'assistance des utilisateurs. Nous construisons un cadre général de modélisation des traces d'utilisation à des fins de réutilisation en contexte baptisé MUSETTE (Modéliser les UsageS et les Tâches pour Tracer l'Expérience) exploité par plusieurs projets dans le domaine de l'assistance à l'apprentissage, l'assistance à la réutilisation et le partage d'expérience en annotations vidéos, l'assistance à la réutilisation et le partage d'expérience en conception collaborative multimédias, etc. La question de l'hétérogénéité sémantique intrinsèque au partage d'expériences d'utilisation est posée et des travaux sont menés sur une approche multi-agents facilitant dans un groupe l'émergence du sens issu des interactions liées à l'échange et la réutilisation d'expériences.

3. Web et Sémantiques Emergentes

Le besoin d'automatisation du processus de conception et de mise en œuvre des services Web rejoint les préoccupations à l'origine du Web sémantique, à savoir comment décrire formellement les connaissances de manière à les rendre exploitables par des machines. Un des premiers problèmes auquel nous nous intéressons de très

près est l'approximation de requêtes pour les services Web. Le cadre formel que nous développerons s'appuiera sur l'usage des logiques de description et des règles Datalog pour la description de services complexes, la résolution étendue et la substituabilité de contraintes pour réaliser des relaxations dans les requêtes, avec une emphase sur les aspects liés à la complexité des raisonnements dans ce cadre.

Un autre problème intéressant concerne l'interopérabilité des connaissances. Actuellement, la majeure partie des connaissances manipulées dans le Web sémantique est décrite dans des ontologies formelles. Or une ontologie décrit des connaissances valides uniquement pour un contexte de modélisation. Nos travaux sur l'interopérabilité des ontologies apportent une solution pour combiner dans une même ontologie plusieurs contextes de modélisation. Dans le domaine de la représentation formelle des connaissances pour le Web sémantique, notre but est de proposer des extensions au langage de représentation du contenu sémantique (tel que OIL ou RDFS). Ces extensions permettront de partager des connaissances entre plusieurs contextes de modélisation. Ainsi, ces connaissances adapteront automatiquement leur présentation et leur contenu en fonction du contexte de l'utilisateur.

Bibliographie

- [1] C. Roussey, S. Calabretto S., J.M Pinon. A new model of Conceptual Graph Adapted for Multilingual Information Retrieval Purposes. Proceedings of the 12th International Conference on Database and Expert Systems Applications DEXA'2001, Technical University of Munich (Germany), September 3-7 2001 Lecture Notes in Computer Science LNCS N°2113, pp. 92-101
- [2] P.-A. Champin, Y. Prié, A. Mille (2002) MUSETTE : un modèle pour réutiliser l'expérience sur le Web sémantique, Journées "Web sémantique" Action Spécifique STIC CNRS, oct. 2002, Paris.
- [3] M.S. Hacid, A. Léger, C. Rey and F. Toumani. Computing concept covers: a preliminary report.. International Workshop on Description Logics (DL'02). April 19 to April 21, 2002. Toulouse.
- [4] M.S. Hacid, A. Léger, C. Rey and F. Toumani. Dynamic discovery of e-services

in a Knowledge Representation and Reasoning context. 18èmes Journées Bases de Données Avancées 21-25 octobre 2002 Evry, France.

[5] D. Benslimane, A. Arara, C. Vangenot and K. Yetongnon, A Contextual Language approach for multirepresentation ontologies, to appear in International Conference on Intelligent Information Systems, June 2003.

Equipe ISRI - Indexation Sémantique et Recherche d'Information

ENST Bretagne
Département Intelligence Artificielle et
Sciences Cognitives
ZI de Kenevent
BP 832
29285 BREST CEDEX
Tél : 02 29 00 14 25
Fax : 02 29 00 10 30
<http://www-iasc.enst-bretagne.fr/>

Contact

Serge Garlatti
Tél : 02 29 00 14 53
Mél : Serge.Garlatti@enst-bretagne.fr

Membres de l'équipe concernés par le Web sémantique

Permanents : Yvon Kermarrec, Ioannis Kanellos, Philippe Picouet, Serge Garlatti

Doctorants : Benjamin Buffereau, Oussama Zein, Hussein Farhat.

Thèmes de l'équipe

Indexation sémantique des média (images plus particulièrement) à base de texte, texte, inter-texte et approches interprétatives.

Il est impossible de traiter automatiquement des images par leur contenu. Le plus souvent, on les indexe par un ensemble de mots clés, dont la pertinence reste fort subjective et liée seulement à une seule interprétation. Or, une image reçoit plusieurs interprétations. Et son indexation peut être faite sur plusieurs types d'attributs (notamment : plastique,

descriptif, interprétatif). Ce projet propose une méthodologie qui couple les processus d'indexation et de recherche d'information multimédia dans un espace pré-structuré par une analyse d'expert. L'architecture est pensée de manière anthropocentrée. L'utilisateur dont l'objectif est d'indexer massivement des images, utilise ces schémas. Il lui est même possible d'indexer des images par du texte (qu'il écrit ou récupère par ailleurs). Des modules de traitement textuel convertissent le texte en information susceptible en aval d'aider la recherche.

Documents virtuels adaptatifs/personnalisables, modélisation utilisateur, hypermédia adaptatifs, Composition sémantique de documents, Itinéraires

Un document virtuel personnalisable compose dynamiquement des documents à partir d'informations provenant de sources diverses, qui satisfont au mieux à des besoins utilisateurs. La composition est découpée en trois étapes, la sélection de l'information, son organisation et son assemblage en un document réel. Le document réel est le document qui est effectivement présenté à l'utilisateur. Pour composer un document réel, un DVP automatisé - ou semi-automatisé, notamment les processus d'organisation, de sélection des informations les plus pertinentes et les plus récentes par filtrage, ainsi que le processus de personnalisation en utilisant notamment un modèle utilisateur. un principe de spécification déclarative des processus de sélection, d'organisation et d'adaptation/personnalisation par un auteur, et d'autre part un principe de composition de documents réels (hypermédia), pour un lecteur particulier sur Internet ont été mis en oeuvre. Ces principes sont fondés sur les ontologies suivantes : une ontologie de l'application qui modélise notamment le savoir-faire de l'auteur, un modèle de domaine, un modèle utilisateur et un schéma de méta-données.

Découverte de Services sur Internet, annuaire (Trader), étude comportementale des services.

Projets en cours sur le thème du Web sémantique

KNOSSOS : Partenaire ODA-Pages Jaunes, Atlantide et ENST Bretagne

La société ODA-Pages Jaunes a développé un site à vocation à la fois culturelle et éducative concernant les oeuvres d'art (www.Louvre.edu). Souhaitant faire évoluer son site à un authentique service éducatif en ligne, elle est arrivée à se poser la question récurrente désormais, de caractérisation par le contenu des objets qu'elle dispose actuellement dans sa banque de données multimédia (plusieurs milliers d'images, essentiellement de tableaux d'art). Tout comme pour le projet SemIndex, cette forme de description est censée coupler le processus d'indexation à celui de la recherche dans une architecture anthropocentrée, cette dernière recevant une contrainte supplémentaire : de pouvoir intégrer plusieurs types d'utilisateur (profils), dans la mesure où l'outil final doit avoir une large gamme d'application éducative (allant de l'élève du primaire au professeur du secondaire) et pouvoir décliner différentes pratiques (de la simple information à la critique de spécialiste, en passant par la préparation d'un cours). Ici encore, je préconise la même approche de base en matière d'indexation et de recherche d'information que SemIndex.

CARISM : Projet GET : ENST Bretagne, ENST et Eurecom

Le projet vise à définir les caractéristiques essentielles des intergiciels ou « middlewares » dédiés aux systèmes mobiles. Par caractéristiques, nous entendons ici l'architecture logicielle et les services externes qui l'accompagnent. Pour satisfaire les contraintes matérielles très fortes de ces équipements, dans une première partie, il convient de structurer l'intergiciel afin qu'il offre des possibilités de configuration, de modularité et d'interopérabilité au sens large. Dans une deuxième partie, l'intergiciel doit également s'appuyer sur des services intelligents dont profite l'application finale. Par « service intelligent », nous entendons un service qui infère sur les requêtes de l'utilisateur pour éviter de solliciter inconsidérément le matériel prenant ainsi en compte la faible capacité en ressources de celui-ci. Un tel service intelli-

gent permettrait également de prendre en compte les préférences de l'utilisateur ainsi que les caractéristiques du support / voies de communication.

MESMUSES : Projet Européen

Dans le cadre du projet européen Mesmuses (<http://cweb.inria.fr/Projects/Mesmuses/index.html>), nous explorons l'intérêt d'une approche de type Web communautaire appliquée au domaine des musées scientifiques en ligne. Dans ce cadre bien identifié, notre réflexion se focalise sur le processus permettant de passer de ressources muséales préalablement indexées à la génération d'itinéraires, un objectif primordial pour les musées, autant pour valoriser leur fond documentaire et muséal, que pour s'ouvrir à de nouveaux publics. En tirant profit tant du contexte applicatif (indexation des ressources par leurs créateurs) que de la nature des ressources documentaires indexées (fragments de taille réduite et mono-thématiques), nous utilisons, du point de vue de la représentation des connaissances, la notion de cadre interprétatif comme support de l'indexation des ressources et de leur utilisation au sein d'itinéraires documentaires.

Projet ICCARS : collaboration avec Atlantide et le Télégramme. (<http://iccars.enst-bretagne.fr/>; <http://candle.enst-bretagne.fr/iccars/>)

Il s'intègre dans le réseau régional de la recherche en technologies de l'information et de la communication (R3TIC), Môle Armoricaïn de Recherche sur la Société de l'Information (Brest, Lannion, Rennes, Vannes) et dans le Programme de Recherche d'Intérêt Régional, presse régionale et technologie de l'information et de la communication.

Ce projet a pour but d'étudier un système d'aide à la création de dossiers thématiques adaptatifs/personnalisables publiables sur internet. Un dossier thématique est une synthèse faite par ou des journalistes sur un sujet particulier.

Projet CANDLE : Projet européen IST (IST-1999-11276). <http://www.candle.eu.org/>

Ce projet a pour objet l'utilisation

d'Internet pour améliorer la qualité et réduire les coûts d'enseignement en Europe. Il s'agit d'utiliser le Web, la technologie multimédia et de permettre la coopération entre les universités et l'industrie dans la création, le partage et la réutilisation de matériaux pédagogiques. D'un point de vue recherche, il s'agit de mettre en œuvre des outils de création auteur et un système de production de cours adaptés aux apprenants et à différentes approches pédagogiques. Le savoir-faire des auteurs est représenté à un niveau connaissance afin de le rendre partageable et réutilisable par les universités et les industries. Ce savoir-faire sera lié à l'approche pédagogique choisie.

KMP : Partenaires : Université de Sophia Antipolis /CNRS (LATAPSES, UMR 6563), INRIA Projet ACACIA, Association Telecom Valley, ENST, ENST Bretagne

L'objectif du projet est de construire une solution innovante de « *Knowledge Management* » partagé entre différents acteurs, au sein d'une même communauté de pratique. Cette solution repose sur l'analyse, la conception, la réalisation et l'évaluation d'un prototype de service Web de compétences, de type hypermédia adaptatif. Ce prototype sera une composante d'un portail web destiné à une communauté d'entreprises, d'institutionnels et d'organismes académiques intéressés par le domaine des télécommunications (communauté étudiée dans l'application : Telecom Valley).

Publications

- [1] B. Buffereau, P. Picouet, Des itinéraires pour accéder à l'information : un défi pour la représentation des connaissances, in Documents Virtuels Personnalisables 2002. 2002. Brest, 10-11 Juillet: ENST Bretagne.
- [2] Iksal, S. and S. Garlatti, Revisiting and Versioning in Virtual Special Reports, in Hypermedia: Openness, Structural Awareness and Adaptivity, Third Workshop on Adaptive Hypertext and Hypermedia, S. Reich, M.M. Tzagarakis, and P.D. Bra, Editors. 2001, Springer Verlag: Århus, Denmark. p. 264-279
- [3] Iksal, S. and S. Garlatti. Adaptive Special Reports for On-line NewsPapers. in Workshop Electronic Publishing, Adaptive Hypermedia (AH) 2002. 2002. Malaga, Espagne.

[4] I. Kanellos, T. Thlitis, A. Leger, « Indexation anthropocentrée d'images au moyen de textes : arguments théoriques et directions applicatives du projet SEMIN-DEX », In *Cognito*, 2000, N° 17, Hors série, pp. 33-44.

[5] I. Kanellos, « A propos de l'héritage critique du document électronique : multimodalité sémiotique, stratégies de lecture et intertextualité », CIDE'99 : deuxième colloque international sur le document électronique, Damas, 5-7 juillet 1999. Paris : *Europa*, 1999, pp. 97-109.

[6] Conférence DVP 2002 : Garlatti, S. and M. Crampes, eds. *Documents Virtuels Personnalisables 2002*. 2002, ENST Bretagne: Brest, 10-11 Juillet.

Groupes AMI (Architectures et Modèles de l'Interaction) et LIR (Langues, Information et Représentation) – LIMSI

LIMSI-CNRS,
Bâtiments 508 et 502bis,
Université de Paris-Sud,
BP133, 91403 ORSAY

Contacts

N. Sabouret, B. Habert et W. Turner
Mél : {nico,habert,turner}@limsi.fr

Participants

D. Béroule, B. Grau, B. Habert, M. Jardino, B. Turner, E. Folch, N. Hernandez, N. Sabouret, H. Zargayouna

Thèmes

Documents (indexation, structuration et classification), Pratiques Collectives distribuées, Agents conversationnels

Présentation

Le Web sémantique est une extension du Web actuel dans laquelle les documents sont munis d'une sémantique, pour que les systèmes informatiques (agents, services, etc.) puissent effectuer des traitements portant non seulement sur les contenus des documents mais aussi sur la signification de ces contenus. Les travaux que nous effectuons au

LIMSI portent sur la construction de tels documents munis d'une sémantique, sur l'élaboration de cartes sémantiques pour naviguer entre ces documents et sur l'étude de l'interaction, en bout de chaîne, entre les humains et ces « outils informationnels » fournis par le Web sémantique.

Si nous prenons en considération un ensemble de documents (éventuellement hétérogènes), nous pouvons annoter ces documents pour les munir d'une sémantique. Deux grandes approches sont étudiées au LIMSI pour l'indexation sémantique des documents :

- L'utilisation de techniques issues des communautés TAL et RI (structuration de documents textuels et repérage de descripteurs thématiques ou argumentatifs). Cette approche travaille à la fois pour décrire les contenus et pour faire ressortir la forme d'un document à partir de son « fond » ;
- L'utilisation d'une approche mixte, entre la RI et les langages de requêtes XML issus de la communauté BD, pour travailler à la fois sur la structure et le contenu, en lien avec l'ontologie du domaine.

Les documents annotés peuvent alors être utilisés pour effectuer des recherches sur des critères sémantiques, à partir de requêtes issues ou non de questions en langue naturelle.

Mais au delà, l'annotation sémantique des documents peut être utilisée pour construire de manière automatique des « cartes sémantiques » qui permettent de naviguer dans l'espace informationnel suivant des liens sémantiques non définis a priori. Deux approches distinctes mais complémentaires sont étudiées au LIMSI pour l'élaboration de ces cartes :

- L'apprentissage automatique de classes sémantiques à partir des documents annotés, à l'aide de méthodes statistiques. Les classes ainsi obtenues sont ensuite étiquetées manuellement pour permettre la génération des cartes, modélisées à l'aide des « Topic Maps ». L'interprétation des classes permet de matérialiser sur la carte, pour l'utilisateur, le sens construit de manière émergente.
- La reconnaissance de formes textuelles dans les documents qui permet, de manière incrémentale, de

construire des cartes actives. Dans ces cartes, les nœuds sont associés à des sens et reliés aux documents qui ont contribué à la définition de ce sens. Ils peuvent alors produire des explications en langue naturelle sur la nature de ces liens et leur élaboration.

Les outils de visualisation et de navigation ainsi produits constituent des « objets informationnels » qui vont permettre aux utilisateurs humains en interaction de se coordonner sur un sujet. Dans le cadre de l'étude des pratiques collectives distribuées (PCD), nous nous intéressons alors à l'observation des influences des objets informationnels sur cette coordination. Mais ces objets (cartes, interfaces...) ne sont pas seulement les catalyseurs (statiques) d'une organisation sociale, observable sur un historique : ils sont aussi actifs et donc caractérisés par leurs capacités d'interaction avec les utilisateurs (et éventuellement entre eux). Nous nous intéressons alors à la modélisation et au raisonnement sur ces capacités d'interaction dans les composants, afin de leur permettre d'interagir avec les utilisateurs non seulement à propos des données qu'ils manipulent (les documents) et de ce qu'ils ont fait (méorisé dans l'historique), mais aussi à propos de leur fonctionnalités (par exemple au niveau de la navigation).

Travaux effectués et en cours

Apprentissage dynamique de Cartes Conceptuelles

La pertinence d'un document est une mesure qui évolue au cours du temps. Il peut en effet s'avérer pertinent bien après son émission, lorsque le contenu d'autres textes sera venu alimenter le point de vue original qu'il avait exprimé. Pour tenir compte de ce constat lors de la conception d'un système de sélection de données textuelles, chaque nouvel article qui diffère des connaissances actuelles doit pouvoir donner naissance à une référence interne, au cas où des textes de contenu proche contribueraient ensuite à augmenter sa pertinence. En suivant cette approche, on adopte un principe de « Reconnaissance de Forme » avec apprentissage dynamique [1], pour lequel chaque texte à l'entrée du système est considéré comme une

forme à identifier par rapport à une base de références.

Le logiciel « QuatrAin » met en oeuvre cette stratégie, qui est issue du formalisme sous-jacent : les Réseaux à Propagation Guidée [2]. Les références internes sont des Cartes Conceptuelles évolutives qui possèdent une dimension « niveau d'activité ». En fonction des critères de pertinence choisis par l'utilisateur, les différentes cartes peuvent être classées par proximité décroissante avec un texte donné. Réciproquement, un ensemble de textes peut être classé par rapport à un réseau de concepts. Par ailleurs, le mécanisme de diffusion d'activité dont un tel réseau est le siège permet de mobiliser l'ensemble des concepts associés à un document, et d'en extraire sa sous-carte caractéristique.

Articuler traitements et annotations sémantiques

Dans un contexte de veille sociale, les données à annoter sont hétérogènes par leur origine, leur taille, leur statut, leur « genre ». Ce sont des « assemblages » de ces données qui sont soumis à des traitements diversifiés : acquisition non supervisée de classes sémantiques, analyses statistiques... Les annotations résultantes sont alors relatives aux assemblages constitués. Elles sont à leur tour le point de départ d'autres annotations (validation manuelle de propositions de relations sémantiques, par exemple). On doit alors disposer d'un modèle de traitement de données textuelles qui permette de « pister » non seulement les documents (ce qui a été proposé dans des projets comme TEI ou XCES) mais aussi les traitements et leurs ramifications [3,4].

Trier dans les mots « communs » convergences et divergences de sens

Un corpus assemble en général des (fragments de) textes d'auteurs différents ou produits à des dates différentes ou relevant de genres textuels différents. On peut alors le *partitionner* en fonction de ces distinctions (en les combinant éventuellement : textes de tel auteur à telle période). Un même mot n'est alors pas forcément employé avec le même sens selon les parties ainsi définies. C'est particulièrement le cas des corpus

relevant du domaine social et politique : les acteurs sociaux non seulement ne « s'entendent » pas sur la signification de tel ou tel mot mais cherchent à imposer leur acception des mots phares du débat qui les réunit et les oppose. Ces écarts de sens se manifestent par une variation des contextes où figure un mot d'une partie à l'autre. C'est par le biais des fluctuations ou au contraire des stabilités de ces associations <mot, contexte> qu'on peut progresser vers le repérage automatique des mots qui font consensus relatif et de ceux qui au contraire témoignent de divergences.

Ces deux directions sont développées au groupe LIR (Langues, Information, Représentations) dans le cadre du projet RNRT « Outiller les alliances » qui vise à permettre de repérer les convergences et divergences thématiques au sein de « forums citoyens ». Le standard Topic Maps permet d'annoter les documents en fonction de cartes sémantiques multiples voire divergentes.

De la sémantique pour structurer le document

Une autre composante du groupe LIR se positionne dans une perspective de Web sémantique adaptatif, dont l'objectif est de faciliter l'accès au contenu d'un document.

Nous partons de l'observation que la présentation d'un document électronique reprend souvent la linéarité d'un document papier et n'offre ainsi pas des modes d'exploration adaptés à un profil d'utilisation (résumé automatique, veille technologique, etc.). Les analyses et les annotations que nous produisons visent ainsi à formater ou reformater des documents.

L'approche que nous avons adoptée sous-tend une analyse essentiellement thématique des documents. Nous repérons des segments de texte thématiquement homogènes, leurs thèmes et les relations entre les segments (notamment à travers leurs thèmes). De plus, nous complétons cette analyse par le repérage d'expressions argumentatives [5].

Nos annotations renseignent donc sur différents aspects d'un document, que ce soit son contenu ou sa structure, et sont interprétées pour fournir différentes vues abstraites d'un même document

dans une perspective de visualisation et de navigation intra-document.

Indexation à partir des annotations

Les différentes annotations sur les documents constituent leurs structures logiques et peuvent servir à indexer ces derniers et permettre ainsi une recherche plus contextuelle et donc plus précise. Nous proposons une organisation d'index qui combine la structure au contenu sémantique. Cette indexation nous permet de répondre à des requêtes structurées et de présenter les résultats suivant un ordre de pertinence. Un des modèles classiques en RI (le modèle vectoriel de Salton) est ainsi adapté pour une vue bidimensionnelle du document qui représente les termes pertinents ainsi que les balises qui constituent la structure. Les fréquences des termes sont calculées par rapport à la balise (au contexte) et non au document et servent à mesurer le taux de spécificité du terme dans une balise donnée d'un document donné. Notre structure d'index est couplée également à une ontologie qui décrit les connaissances du domaine par des concepts et des rôles entre ces concepts. A long terme, nous comptons nous focaliser plus sur l'adaptation de l'ontologie par apprentissage aux différents profils d'utilisateurs qui peuvent exister sur le Web.

Accéder à la sémantique du fonctionnement

Les modèles actuels du Web sémantique permettent de munir les pages d'une représentation structurelle sémantisée. Mais les utilisateurs ordinaires n'accèdent pas directement à ces contenus : ils utilisent des composants actifs médiateurs (services, agent, etc.) qui sont caractérisés par un fonctionnement. Ils peuvent donc interagir non seulement à propos de l'information structurelle manipulée mais aussi à propos des actions de manipulation et de l'activité (présente ou passée) des composants. Dans ce contexte, le fonctionnement des composants actifs fait partie intégrante de l'information et devrait être représentée et manipulable en cours d'exécution pour expliquer aux utilisateurs ce que fait le composant. Contrairement aux ontologies de description de services (WSDL, WSMF, DAML-S...), l'ap-

proche que nous mettons en oeuvre dans le groupe AMI du LIMSI consiste à intégrer dans une même représentation la description des actions du composant et celle des données manipulées. Pour ce faire, nous proposons de programmer directement les composants en XML. Nous avons défini un schéma (<http://www.limsi.fr/Individu/nico/xml/vdl.xsd>) et un modèle d'activité [6,7] qui permettent d'utiliser XML comme langage de programmation des composants actifs. Ce modèle permet d'avoir accès en cours d'exécution à une description sémantisée du fonctionnement pour pouvoir répondre à des requêtes sur les actions des agents.

Bibliographie

- [1] D. Bérout, The Never-ending Learning, "Neural Computers, Vol 41, pages 219-230, R. Eckmiller and C. von Malsburg eds, Springer-Verlag, Berlin edition, 1988.
- [2] D. Bérout, Traitement connexionniste du langage, Histoire Epistémologie Langage, F. Rastier ed., 11-1, Juin 1989.
- [3] H. Folch, B. Habert, S. Lahlou. Navigable Topic Maps for overlaying multiple acquired classifications. MIT Press, Markup Languages: Theory & Practice, 2(3), pp. 269-280, 2000.
- [4] H. Folch. Articuler les classifications sémantiques induites d'un domaine. Thèse de doctorat, Université Paris XIII, 2002.
- [5] N. Hernandez et B. Grau, « Analyse Thématique du Discours : segmentation, structuration, description et représentation », CIDE'05, Hammamet, Tunisie, 20-23 octobre 2002.
- [6] N. Sabouret. Programmer des composants actifs dans le Web sémantique. Journées de l'Action Spécifique STIC "Web Sémantique", octobre 2002.
- [7] N. Sabouret. A model of requests about actions for active components in the semantic Web. In Proc. STarting AI Researchers Symposium (STAIRS) 2002, p.11-20, IOS Press, 2002.

Laboratoire ERIC - Equipe BDD

Université Lumière Lyon 2
5, avenue Pierre Mendès-France
69676 Bron Cedex

Responsable

Jérôme Darmont, 04 78 77 44 03

Mél : jerome.darmont@univ-lyon2.fr
http://bdd.univ-lyon2.fr

Membres de l'équipe concernés par le thème du Web sémantique

Kamel Aouiche (Doctorant), Fadila Bentayeb (MCF), Omar Boussaid (MCF), Jérôme Darmont (MCF), Sabine Rabaseda (MCF), Yamina Zellouf (ATER)

Thème général de l'équipe

Systèmes décisionnels, entrepôts de données

Description des travaux ou projets en lien avec le thème du Web sémantique

Intégration de données complexes dans un entrepôt de données : utilisation du Web comme source de données pour un entrepôt. Approche en 3 phases : (1) structuration des données complexes ; (2) modélisation multi-dimensionnelle des données complexes ; (3) analyse de données complexes (analyse en ligne ou fouille de données).

1. Problématique

La fin du XX^{ème} Siècle a vu le développement rapide de nouvelles technologies de communication basées sur la Toile et de nouvelles technologies d'aide à la décision. Les entreprises doivent relever de nouveaux défis en matière de commerce électronique ou mobile et sont en train de réorganiser fortement leurs systèmes d'information et leurs méthodes de gestion. Elles développent de nombreuses techniques dites d'intelligence économique afin de gérer leurs données et leurs connaissances. Ces nouveaux moyens leur permettent d'améliorer leur productivité et de mettre en œuvre une veille concurrentielle. Dans ce contexte, la Toile est devenue la principale source de données et il est maintenant indispensable de concevoir des systèmes d'information et d'aide à la décision basés sur la Toile. Notre travail se situe dans ce domaine. Nous avons notamment conçu une approche permettant de construire une Base de Données Décisionnelle (BDD) dont la source de données est la Toile.

Les technologies des entrepôts de données (data warehouses) et de l'analyse en ligne (OLAP — *On-Line Analytical Processing*) sont désormais matures pour des applications et des données classiques (numériques). Avec le développement d'Internet, la disponibilité de divers types de données (images, textes, sons, vidéos, données issues de bases de données...) s'est accrue. Ces données complexes sont très hétérogènes et peuvent être structurées ou non. La structuration et l'exploitation de grandes quantités de données complexes sont des tâches difficiles. Cependant, les concepts de l'entreposage de données demeurent valides pour des données complexes. Dans ce contexte, la Toile peut être considérée comme un système fournisseur de données (*web farming system*) à un entrepôt.

Les données issues de la Toile peuvent donc être stockées dans une BDD telle qu'un entrepôt de données afin d'être analysées en ligne ou grâce à des techniques de fouille de données (*data mining*). Cependant, ces données complexes doivent d'abord être structurées dans une base de données, puis intégrées dans l'architecture particulière d'un entrepôt (tables de faits, tables de dimension, magasins de données, cubes de données...). Or, l'approche classique d'entreposage de données n'est pas appropriée au traitement de données complexes, dont la modélisation multi-dimensionnelle est délicate et peut nécessiter l'introduction de nouveaux concepts. Les opérateurs OLAP classiques se révèlent en effet inadaptés ou inefficaces. L'administration de données complexes entreposées nécessite également des stratégies de rafraîchissement adaptées lorsque de nouvelles données doivent être prises en compte, ainsi que des politiques de réorganisation physique de la base qui dépendent de son usage, afin d'optimiser les performances.

2. Approche mise en œuvre

Afin d'appréhender la complexité de notre problématique, nous avons adopté une démarche en trois étapes : (1) structuration de données complexes issues de la Toile, (2) modélisation multidimen-

sionnelle et (3) analyse de données complexes. Jusqu'à présent, nos travaux de recherche ont principalement porté sur la première phase de ce processus et la structuration de données complexes en ODS (*Operational Data Storage*) relationnel, ainsi que sur la performance des entrepôts de données, qu'il sera particulièrement critique de garantir lorsque nous y stockerons des objets complexes. Nous commençons cependant à aborder les deux dernières phases

2.1. Structuration de données Web

Afin d'intégrer des données issues de la Toile dans un ODS, nous avons proposé un processus de modélisation « classique ». Tout d'abord, nous avons conçu un *modèle conceptuel* UML d'un objet complexe représentant une généralisation de tous les types de données complexes présents sur le Web et intéressants à entreposer. Ce modèle permet d'unifier ces données complexes hétérogènes au sein d'un cadre commun. De plus, il est important de noter que notre objectif n'est pas seulement de stocker les données, mais aussi de les préparer véritablement à l'analyse. Ceci représente une tâche plus complexe encore qu'un processus d'ETL (*Extracting, Transforming, and Loading*) classique.

Nous avons ensuite traduit notre modèle conceptuel UML en une définition de schéma XML (DTD ou XML-Schema) qui représente notre *modèle logique*. Finalement, nous avons instancié ce modèle logique en un *modèle physique* qui est un document XML. XML s'est révélé un format de choix pour à la fois stocker et décrire nos données, d'une part en raison de sa flexibilité et de son extensibilité, mais aussi parce que le schéma des données, implicitement inclus dans un document XML, représente en fait une partie des métadonnées du futur entrepôt. Les documents XML que nous obtenons sont finalement stockés dans une base de données relationnelle.

Afin de valider cette démarche, nous avons développé le Système Multi-Agents d'Intégration de Données Complexes (*SMAIDoC*), qui identifie les différentes tâches de ce processus

d'intégration de données complexes comme des services accomplis par des agents. L'architecture de SMAIDoC est évolutive et permet d'ajouter, modifier ou supprimer des services et de créer au besoin de nouveaux agents. Il est par exemple possible de donner à l'agent de collecte de données de SMAIDoC la capacité de communiquer avec des moteurs de recherche en ligne et d'exploiter les réponses de ces derniers. En plus du processus d'intégration des données complexes auquel il est dédié, SMAIDoC peut offrir, par le biais de nouveaux agents à créer, des services tels que la modélisation multidimensionnelle, l'analyse en ligne ou la fouille de données.

2.2. Modélisation multidimensionnelle de données complexes

Afin de modéliser de façon multidimensionnelle des données complexes, nos premiers travaux nous ont permis de définir un cadre méthodologique global qui se base d'une part sur les objectifs d'analyse et, d'autre part, sur la construction d'un référentiel de données. Ce référentiel doit lister de façon exhaustive les données nécessaires à la modélisation, décrire les caractéristiques de chacune des données en précisant son rôle dans le modèle à créer, aider dans le choix des éléments du modèle (descripteurs, indicateurs) conformes aux objectifs d'analyse et enfin vérifier la cohérence des données participant au modèle. Les données complexes sont décrites par des attributs globaux et possèdent également des descripteurs sémantiques. Ces derniers peuvent être obtenus par diverses techniques de fouille de données, de statistique, d'intelligence artificielle, de traitement d'images ou de signal. La sélection des données pertinentes dans un processus d'entreposage de données est une étape d'autant plus cruciale que les données sont complexes. Une exploration par fouille des données, par exemple, peut contribuer à l'identification de faits à analyser et peut permettre d'enrichir le référentiel par de nouveaux descripteurs sémantiques.

La spécificité des données complexes ne permet pas d'aborder facilement une

modélisation multidimensionnelle, mais nous considérons que les techniques de fouilles de données peuvent s'avérer utiles pour apporter une connaissance préalable de ces données. Ainsi, en plus de leur vocation d'analyse, ces techniques peuvent contribuer à la modélisation multidimensionnelle des données complexes, qui peut se révéler comme un contexte privilégié pour la construction de modèles prédictifs. L'exploration par des techniques de fouille de données (telles que les arbres de décision) de données complexes structurées dans une base de données relationnelle peut certainement aider à leur modélisation multidimensionnelle. Il s'agit ensuite d'implémenter ces modèles multidimensionnels pour construire des modèles prédictifs et de réfléchir à la mesure de performance de ces derniers.

2.3. Analyse de données complexes

L'intégration de données complexes dans une base de données relationnelle sous forme de documents XML nous permet d'envisager plusieurs voies pour les analyser. La première voie consiste à explorer directement les documents XML par des techniques de fouille de données. Ces dernières ont en effet fait leurs preuves dans le traitement du contenu de documents textuels et peuvent nous permettre d'exploiter les informations véhiculées par la structure des documents XML pour en extraire des connaissances. En revanche, la modélisation multidimensionnelle des données complexes est nécessaire pour aborder les autres voies d'analyse que nous envisageons. À partir de cubes de données complexes, il est possible d'appliquer des méthodes d'analyse en ligne et/ou de fouille de données. Mais l'utilisation conjointe des techniques d'analyse multidimensionnelle en ligne et de fouille de données sur des mêmes données fait émerger plusieurs problèmes.

- Comment associer fouille de données et analyse OLAP pour une même analyse afin d'exploiter leur complémentarité ? La vocation des opérateurs OLAP est en effet l'agrégation et la navigation dans le cube de données, alors que celle des techniques de

fouille de données est l'extraction des connaissances.

- Comment appliquer des algorithmes de fouille de données sur des données multidimensionnelles, sans perdre bénéfice de la multidimensionnalité des données ?
- Dans l'analyse multidimensionnelle, comment étendre les opérateurs OLAP à de nouveaux opérateurs basés sur des techniques de fouille de données, indispensables pour considérer la spécificité des données complexes ?

Courte bibliographie ciblée

- [1] F. Clerc, A. Duffoux, C. Rose, F. Bentayeb, O. Boussaid, « SMAIDoC : Un Système Multi-Agents pour l'Intégration des Données Complexes », Soumission à la Session Spéciale Entreposage et Fouille de Données des XXXV^{èmes} Journées de Statistique, Lyon, 2003.
- [2] J. Darmont, O. Boussaid, F. Bentayeb, S. Rabaseda, Y. Zellouf, « Web multiform data structuring for warehousing », *Multimedia Systems and Applications*, Vol. 22, Kluwer Academic Publishers, October 2002, pp. 179-194.
- [3] J. Darmont, O. Boussaid, F. Bentayeb, « Warehousing Web Data », 4th International Conference on Information Integration and Web-based Applications and Services (iiWAS 02), Bandung, Indonesia, September 2002, pp. 148-152.
- [4] S. Miniaoui, J. Darmont, O. Boussaid, « Web data modeling for integration in data warehouses », *First International Workshop on Multimedia Data and Document Engineering (MDDE 01)*, Lyon, France, July 2001, pp. 88-97.

France Telecom R&D – Site de Rennes

Programme Scientifique Traitement des Connaissances

France Telecom R&D
4, rue du Clos Courtel,
35512 Cesson
Tél : 02 99 12 42 23

Personne à contacter

Alain Léger
alain.leger@rd.francetelecom.com

Membres de l'équipe, concernés par le thème du Web sémantique

- TAL et Sémantique : Jérôme Vinesse, Gilles Prigent, Johannes Heinecke, Malek Boualem, Frédérique Arga
- Ontologie, Raisonnements et Web Services : Luigi Lancieri, Alexandre Deltheil, Alain Léger
- Recherche par le contenu : Henri Sanson, Françoise Chassaing, Christine Chardenon, Pascal Filoche
- Communautés d'intérêt : Michel Plu, Cécile Bothorel, Karine Chevalier, Layda Agosto
- KDD : Olivier Perrault, Fabrice Clérot, Valérie Beaudouin, Olivier Collin

Thèmes de l'équipe

- Représentation des Connaissances support d'algorithmes d'inférence « aux bonnes propriétés », Bibliothèque de service d'inférence générique, Bases de Connaissances Ontologiques
- Modèle sémantique, modèle du discours et modèle de traduction, Plateforme indépendante de la langue pour l'intégration rapide de nouvelles ressources linguistiques
- Apprentissage des profils d'internautes, Qualification anthropocentrée des contenus, Cybernavigation, Modèles de comportement collectif et métrique de coopération
- Web Mining, Text Mining, Fusion sémantique de données, Maintenance de Bases de Connaissances Ontologiques par apprentissage

Travaux ou projets en lien avec le thème du Web sémantique

IST-Mkbeem - Knowledge-Based Multilingual E-Commerce
<http://www.mkbeem.com>

IST-SmartEC - Knowledge-based E-Commerce transactions
<http://www.telecom.ntua.gr/smartec/>

IST-Ontoweb - Réseau d'excellence Web Sémantique
<http://www.ontoweb.org> <http://sig4.ago.fr>

Projet PICSEL 1 et 2 (LRI-FT) - Médiateur pour intégration d'informations hétérogènes
<http://www.lri.fr/~picse/>

Projet SemIndex (ENST Br-FT) - Accès par le contenu de banque de données Image
<http://www-iasc.enst-bretagne.fr/PROJECTS/SemIndex>

Projet Semantex (IRIT-FT) - Analyse sémantique de texte

Projet RNRT Cyrano, FT, INRIA, ECL (jusqu'en juin 03) ; Système de réplication actif personnalisé de flux audiovisuels sur Internet

Projet RNRT M-Univers (en préparation), Quadrille Ing., FT, INSA Lyon, UTC ; Gestion personnalisée de contenu sur réseau P2P multicast.

Bibliographie (site Web non public)

BDA'02, (Base de Données Avancées), « Dynamic discovery of e-services », M. Hacid, A. Léger, C. Rey, F. Toumani
 TALN 2002, workshop NLP techniques for speech analysis « Eurescom MUST project, Multimodal, multilingual information services for small mobile terminals », Nancy, France, 24-27 juin 2002. M. Boualem & partenaires du projet Eurescom P1104
 WWW2002 « On-Line hypermédia contents for learning and practicing foreign languages », mai 2002 Hawaïi, Michel PLU, Christine CHARDENON, Laurent MAU-PEU
 ECAI'2000, « Ontology Domain Modeling support for multi-lingual services in E-Commerce : MKBEEM », Berlin, août 2000, A. Léger (FTR&D), A. Lehtola (VTT), A. Gomez-Pérez (UPM)
 CIDE'2000, Indexation et recherche d'information imagée par le contenu : catégories de textes et catégories d'images dans le projet « SemIndex », Lyon, juillet 2000, Théodore Thliviitis, Ioannis Kanellos (ENSTBr) A.Léger (FTR&D)
 HCP'99 « Representation of content using natural language processing », J.Vinesse. Brest, septembre 1999.
 Workshop OntoWeb « Successful scenarios for ontology – based applications Vers 2 D22" Sardaigne juin 2002, A. Léger, Y. Bouillon, P. Ecoublet, M. Bryan, R. Dieng, A. Persidis, Y. Sure, A. Perez, Y. Ding
 ACL 2001, Workshop on Arabic Language Processing, « Hahoa Arabic Web Directory & Natural Language Processing for Arabic Information Retrieval » Toulouse, Juillet 2001, Malek Boualem (France Telecom R&D), Régina Sneifer (Wanadoo)
 e.work2001 – Proceedings of e.business 2001, pp 157-164, « MKBEEM, New Opportunities for Language Technology in Multilingual eMarkets » H. Fersoe, A. Jocelyne, (A. Léger, A. Lethola)
 CIDE2001, Le Web sémantique, Toulouse Alain Léger
 IDEAL02 - « Study of the Regularity of the Users' Internet Accesses » ; Nicolas Durand, Luigi Lancieri ; In proceedings of Third International Conference on Intelligent Data Engineering and Automated Learning , IDEAL'02 (IEEE, Springer, EPSCR,

UMIST);, Manchester, UK August 2002
 ICII01- « To exploit the collective intelligence thanks to the Co-operative replication » ; Luigi Lancieri, Nicolas Berthier Bonnel, Ludovic Stumme; In proceedings of International Conferences on Info-tech & Info-net -Beijing 2001.

IJCNN00 - « A connectionist approach for evaluating the complexity of interactions in the World Wide Web ». The case of News Groups ; Luigi Lancieri ; IEEE International Joint Conference on Neural Network 2000-(COMO - Italy)

ECAI 2002 - .A. Delteil, C. Faron (2002). *A Graph-Based Knowledge Representation Language*.

ICCS 2002 - A. Delteil, C. Faron, R. Dieng (2002). *Building Concept Lattices from RDF Graphs Annotating Web Resources*.

Projet ACACIA - Acquisition des Connaissances pour l'Assistance à la Conception par Interaction entre Agents

INRIA Sophia Antipolis
 2004 route des Lucioles BP 93
 06902 Sophia Antipolis
<http://www.inria.fr/acacia/acacia.html>

Personne à contacter
 Rose Dieng-Kuntz
 Rose.Dieng@sophia.inria.fr
 Tél : 04 92 38 78 10

Membres : Francis Avnaim, Olivier Corby, Alain Giboin, Laurent Alamarguy, Thanh-Le Bach, Tuan-Dung Cao, Khaled Khelif, Carolina Medina-Ramirez, Olivier Savoie
Collaborateurs : Catherine Faron-Zucker, Fabien Gandon.

Description des travaux

Le projet pluridisciplinaire ACACIA vise à développer des aides méthodologiques et logicielles pour la mémoire d'entreprise (ME), dans divers scénarios : mémoire de projet, veille technologique, cartographie de compétences, etc.

Par analogie avec le Web, nous étudions la ME matérialisée dans un « Web sémantique d'entreprise » (WSE), sous forme de ressources souvent documen-

taires, d'ontologies et d'annotations sémantiques, avec une éventuelle modélisation de multiples points de vue. Pour représenter ontologies ou annotations, nous utilisons le formalisme des graphes conceptuels (GC) et les langages du W3C (XML, RDF...).

1 La mémoire d'entreprise vue comme un WSE

La ME et le Web sont des paysages d'information hétérogènes et distribués, partageant le même problème de la recherche de ressources pertinentes. Un intranet / intraweb peut améliorer la diffusion d'informations voire de connaissances dans l'entreprise. Mais, comparée au Web ouvert, une ME a un contexte, une infrastructure et une portée délimités et mieux définis : l'organisation ; les problèmes de passage à l'échelle y sont moins cruciaux.

Les travaux sur le Web sémantique visent à rendre le contenu sémantique du Web interprétable par les machines. Notre approche repose sur la matérialisation de la ME sous forme d'un WSE, diffusé sur l'intranet et constitué d'ontologies, de ressources (i.e. des documents, des personnes, des services, des logiciels, des bases de données (BDs)) et d'annotations sémantiques sur ces ressources (i.e. sur le contenu des documents, sur les compétences des personnes, sur les caractéristiques des services / logiciels / BDs), ces annotations utilisant le vocabulaire conceptuel défini dans les ontologies [2]. Les principaux problèmes de recherche étudiés sont :

- Aux niveaux méthodologie, architecture et représentation, comment construire un WSE et ses composants inter-reliés (en particulier ontologies et annotations), et ce tant à partir de plusieurs experts qu'à partir de textes ?
- Comment diffuser et utiliser un WSE (recherche d'information (RI) vs dissémination pro-active (« push ») ?
- Comment évaluer et faire évoluer un WSE (cf. points de vue concepteur et utilisateur), en prenant en compte les liens entre ses composants (ressources, ontologies et annotations) ?

2 Construction d'ontologies et d'annotations

Ontologie d'entreprise

Notre méthode intègre (1) une approche descendante se basant sur une synthèse d'ontologies existantes, (2) une approche ascendante exploitant l'analyse (manuelle ou semi-automatique) d'entretiens ou de documents et (3) une approche centrifuge reposant sur la détermination, la généralisation et la spécialisation des concepts clés de l'entreprise. Nous avons ainsi construit l'ontologie O'CoMMA [4], évaluée ensuite selon la méthode des scénarios.

Construction d'ontologies à partir de textes

Suivant l'approche TIA, nous approfondissons la construction semi-automatique d'ontologies et d'annotations à partir de sources hétérogènes dans l'entreprise (personnes, documents textuels ou multimédia, BDs avec données textuelles), en utilisant au besoin des outils linguistiques. La méthode SAMOVAR [5] repose sur un enrichissement progressif semi-automatique des ontologies grâce à l'exploitation des résultats d'outils linguistiques tels que des extracteurs de candidats-termes appliqués au corpus textuel, et grâce à la proposition de règles heuristiques s'appuyant sur l'analyse des régularités de ces résultats. La représentation en RDF(S) des ontologies et des annotations permet ensuite d'exploiter le moteur CORESE pour la RI. Cette méthode, appliquée pour une mémoire de projet véhicule chez Renault, se généralise à la mémoire d'un projet de conception d'un système complexe.

Représentation formelle de l'ontologie et des annotations

Plusieurs langages permettraient de représenter les ontologies et les annotations :

- Des formalismes classiques de représentation des connaissances : le projet ESCRIRE, en coopération avec les équipes EXMO et ORPAILLEUR, a permis de comparer GC, représentation par objets et logiques de description (LD) pour représenter le contenu sémantique de documents à des fins de RI [Al-Hulou et al 2002].

- RDF Schema (RDFS) : pour pallier ses limitations en expressivité, nous en avons proposé une extension, DRDFS, qui permet d'exprimer des définitions explicites de concepts et de relations, et des connaissances contextuelles [3].
- De nouveaux langages : nous avons ainsi proposé le langage GDL, inspiré à la fois des LD et des GC [3].

Apprentissage d'ontologies pour un WSE

Nous avons proposé une méthode pour l'apprentissage incrémental des ontologies à partir d'annotations RDF. Cette méthode peut être utilisée pour générer des classes de documents à partir des annotations de ces documents [3].

Comparaison de multiples ontologies dans un WSE

Qu'il s'agisse d'un WSE interne à une organisation disposant de plusieurs ontologies métiers ou d'un WSE partagé entre plusieurs organisations, la manipulation de plusieurs ontologies nécessite de pouvoir les comparer, faire des correspondances entre elles voire les intégrer ou construire une ontologie multi-points de vue.

Outre nos algorithmes de comparaison entre ontologies représentées en GC [ECAI'98], nous en avons développé de nouveaux pour RDF(S), les avons implémentés en Java et testés sur les ontologies O'CoMMA et Aprobatom.

Fouille du Web pour l'intégration des sources externes dans un WSE

Outre les sources internes de l'entreprise, la construction d'un WSE peut reposer aussi sur des sources Web externes. Mais l'annotation manuelle de ces dernières étant peu réaliste, une aide méthodologique et logicielle pour la fouille du Web et pour la génération (semi)-automatique des annotations semble nécessaire. Nous étudions ainsi l'exploitation d'agents logiciels pour la fouille du Web et des techniques pour générer des annotations RDF sur des pages Web bien structurées (e.g. catalogues, rapports, articles scientifiques...).

3 Diffusion et utilisation d'un WSE

L'utilisateur souhaite accéder à des personnes ayant certaines compétences, retrouver des documents pertinents, découvrir des services utiles, publier dans le WSE pour partager des connaissances. RI, « push » et distribution du WSE sont donc des points essentiels à étudier.

Moteur de recherche sémantique CORESE

CORESE (COncEptual REsource Search Engine), notre moteur RDF(S) basé sur les GC, permet de charger une ontologie représentée en RDFS et des annotations RDF, de les traduire en GC puis, grâce à l'opérateur de projection, d'interroger la base d'annotations obtenue. CORESE exploite les propriétés des relations (symétrie, transitivité, etc.) ainsi que des règles d'inférence sur les graphes [1]. Utilisé dans les projets CoMMA, ESCRIRE, SAMOVAR, APROBATION et KMP, CORESE permet une RI « intelligente » dans un WSE.

Architecture multi-agents pour la gestion d'un WSE

L'hétérogénéité et la distribution des systèmes multi-agents (SMA) peuvent être une solution à l'hétérogénéité et à la distribution de la ME. Nous avons ainsi proposé :

- une architecture multi-agents pour la RI dans un WSE distribué ou pour le « push » vers les utilisateurs en fonction de leurs profils : agents dédiés aux ontologies, aux documents, aux utilisateurs, à la connexion entre agents.
- des algorithmes pour distribuer la base d'annotations et les requêtes d'un utilisateur entre les agents adéquats [4].

Ces travaux ont été appliqués dans le projet européen CoMMA qui a abouti à l'ontologie O'CoMMA et à un SMA pour la gestion d'un WSE pour les scénarios « insertion des nouveaux employés » et « veille technologique » [4].

4 Conclusions

Nous étudions certains problèmes du Web sémantique, mais notre contexte – la ME – nous permet des hypothèses simplificatrices : contexte de l'organisation, moindre hétérogénéité et échelle plus réduite que pour le Web ouvert, moindre volume des sources d'information pour guider la construction des ontologies et des annotations. Cependant, le scénario de la veille technologique oblige à affronter certains problèmes du Web ouvert. Nos travaux futurs vont s'étendre aux Webs sémantiques inter-entreprises ou inter-communautés.

Courte bibliographie

(<http://www.inria.fr/acacia/Publications>)

- [1] O. Corby, C. Faron, Corese: A Corporate Semantic Web Engine, Proc. of the WWW11 Int. Workshop on Real World RDF and Semantic Web Applications, Hawaii, May 2002, <http://paul.rutgers.edu/~kashyap/workshop.html>.
- [2] R. Dieng, O. Corby, F. Gandon A. Giboin, J. Golebiowska, N. Matta, M. Ribière, Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire pour le knowledge management, 2ème édition, Dunod, novembre 2001 (ISBN 2100053006).
- [3] A. Delteil. Représentation et apprentissage de concepts et d'ontologies pour le Web Sémantique. Thèse de Doctorat en Informatique, Université de Nice Sophia Antipolis (UNSA), 26 novembre 2002.
- [4] F. Gandon - Distributed Artificial Intelligence and Knowledge Management: ontologies and multi-agent systems for a corporate semantic web. Thèse de Doctorat en Informatique, UNSA, 7 novembre 2002.
- [5] J. Golebiowska. Exploitation des ontologies pour la mémoire d'un projet-véhicule : méthode et outil SAMOVAR. Thèse de Doctorat en Informatique, UNSA, 4 février 2002.

Équipe EXMO - Échanges de connaissance structurée médiatisés par ordinateur

INRIA Rhône-Alpes
655 avenue de l'Europe, 38330
Montbonnot Saint-Martin
<http://www.inrialpes.fr/exmo>

Contact

Jean-François Baget
Jean-Francois.Baget@inrialpes.fr

Membres : Jean-François Baget,
Jérôme Euzenat

Travaux de l'équipe

EXMO part du principe que, dans les futurs systèmes d'information, la connaissance formalisée sera échangée de façon routinière. Cette approche commence à se développer avec la généralisation de l'utilisation de langages d'échange standardisés (XML) lors de la communication en réseau. Dans le processus de communication, l'ordinateur peut introduire une plus-value à son rôle de médium et de mémoire en accomplissant des tâches comme le formatage, le filtrage, la catégorisation, le test de consistance ou la généralisation de la connaissance. Ces manipulations peuvent être vues comme des transformations. En contrepartie, les utilisateurs devront exiger une plus grande assurance sur les transformations opérées. EXMO a pour but le développement d'outils théoriques et logiciels pour aider à l'échange d'éléments de connaissance formalisée.

Notre angle d'attaque principal consiste à examiner les propriétés que doivent satisfaire les transformations appliquées aux représentations formalisées. Parmi ces propriétés, on trouve la préservation du contenu ou de la structure, la traçabilité des sources ou, au contraire, la confidentialité. On cherche à élaborer une « théorie générale des transformations » fondée sur les propriétés satisfaites par les transformations plutôt que sur les transformations elles-mêmes.

D'une part, nous nous attachons à montrer que cette problématique s'applique à un ensemble varié de contextes dépendant des langages, propriétés et transformations impliquées. Ainsi, nous étudions la préservation de l'information modélisée par des treillis ou la préservation de scénario de documents multimédia durant l'adaptation. D'autre part, nous nous intéressons plus particulièrement aux propriétés sémantiques lors

de la transformation de la connaissance d'un langage à un autre. La question essentielle est : les conséquences d'une représentation initiale sont-elles les transformées de celles de la représentation transformée ? Pour cela nous établissons les relations entre plusieurs types de propriétés sémantiques, connues ou non, en rapport avec celles-ci. Nous avons aussi étudié des cadres plus restreints (famille de langages, patrons) permettant d'utiliser ces relations et d'établir plus facilement les propriétés des transformations.

Enfin, sur un plus long terme, nous avons l'ambition d'explorer des propriétés appelées « sémiologiques », c'est-à-dire qui concernent l'interprétation par un utilisateur humain de la représentation communiquée, et sa préservation.

Du point de vue pratique, nous avons décrit la syntaxe et la sémantique de plusieurs langages de représentation de connaissance (TROEPS, logiques de descriptions) dans un format XML. Nous avons développé un langage capable de décrire des transformations composées et de les exécuter [5]. Notre but est maintenant de tirer parti des descriptions de la sémantique et des transformations pour vérifier des preuves de certaines propriétés des transformations avant de les utiliser. L'autre direction consiste à construire la preuve en fonction de celles des transformations initiales et des propriétés des constructeurs utilisés.

Les applications anticipées sont l'ingénierie des systèmes de transformations (où le système d'information est compris comme un flux de transformations) mais aussi l'infrastructure du « Web sémantique ». Le "Web sémantique" étant compris comme un Web de connaissance formalisée, il est patent que l'échange et l'incorporation de cette connaissance dans un contexte autre requerra des transformations dont on voudrait être capable d'établir les propriétés (Préservent-elles la consistance ? Les conséquences ?).

Dans ce cadre nous participons à différents groupes de travail (groupe WebOnt du W3C) ou de réflexion (réseau thématique Ontoweb, Action

spécifique "Web sémantique" du CNRS). Nous avons co-organisé en 2001 le semantic web working symposium à Stanford [1] et un strategic workshop EU-NSF sur le même thème à Sophia-Antipolis [4].

Au sein du groupe WebOnt du W3C nous participons à la conception du langage d'ontologie pour le Web OWL, successeur de DAML+OIL, lui-même fusion de DAML-ONT et de OIL. C'est un langage fondé sur RDF, fortement inspiré des logiques de descriptions dans une syntaxe objet exprimée en RDF/XML. Il est proposé en trois parfums: OWL-Lite est une version très simple d'un langage d'ontologies (il contient la conjonction, la spécialisation de classes et de propriétés, la restriction de domaine et de codomaine, l'égalité et l'inégalité

d'individus, les restrictions de cardinalité élémentaires, les relations inverses, symétriques et transitives), OWL-DL correspond à une logique de description expressive (il contient la négation, la disjonction, la définition de classes en extension, l'exclusion de classes, les contraintes de cardinalité générales), OWL-Full contient en outre tout RDF Schema et la possibilité de considérer une classe comme un individu.

Un découpage plus clair entre RDF et OWL aurait permis de développer d'une part des ontologies, et d'autre part des documents RDF dont les ressources seraient des classes ou des propriétés décrites dans un document OWL. Cette utilisation commune des deux langages, pourtant naturelle, n'a pour l'instant pas été étudiée. Même si la sémantique de ce langage RDF+OWL se définit immédiatement par les sémantiques des langages qui le composent, d'importants problèmes théoriques se posent : si des algorithmes sont connus pour raisonner sur des documents RDF (homomorphisme de graphe), et pour raisonner dans certains sous ensembles de OWL (travail réalisé pour les logiques de description), la juxtaposition RDF+OWL n'a pas été étudiée (des problèmes équivalents peuvent se retrouver en BD pour l'inclusion de requêtes, qui reste un travail en cours). Par ailleurs, les besoins différents des utilisateurs peuvent nécessiter un

découpage différent de celui retenu par le W3C. Ainsi, on pourrait préférer à OWL DL un langage dont la disjonction est exclue, la sémantique de la négation est intuitionniste et où les classes peuvent être considérées comme des instances. Un tel langage n'est pas défini actuellement. Pour en bénéficier, il aurait fallu développer une approche plus modulaire des langages du Web sémantique telle que nous l'avons proposée sur la base des travaux en logiques de description [2]. Nous avons pour ambition d'étudier les sous-ensembles de OWL, d'isoler les groupes de langages équivalents et de caractériser à la fois les limites de complexité et celles des algorithmes actuellement connus (jusqu'à où est-il possible de procéder par projection). Même le découpage de OWL (LITE, DL, FULL) pourrait être remis en cause par la complexité des sous-langages de RDF+OWL correspondant.

D'autre part, nous avons proposé une infrastructure de transformations sûres pour le Web sémantique [3] que nous élaborons à petit pas. Cette infrastructure est fondée sur l'idée de prouver les propriétés des transformations d'un langage à un autre et de publier les transformations, les propriétés et leurs preuves sur le Web. La construction d'une nouvelle transformation pourra alors se faire par composition de transformations préexistantes. L'obtention d'une preuve pourrait aussi pouvoir se faire par composition des preuves (ou des propriétés). Une de nos activités au sein du groupe WebOnt a été de développer une transformation XSLT du format XML de OWL vers le format RDF. Sachant que la sémantique d'OWL est définie dans son format XML et que celle-ci doit s'aligner sur un équivalent RDF, support de la sémantique de RDF, il est important de disposer d'une telle transformation. Il sera alors plus facile de montrer que l'ensemble des modèles d'un document OWL est un sous-ensemble de ceux du document RDF correspondant. Notre ambition est non seulement de produire cette transformation exécutable mais d'être capable de distribuer la preuve de la correction par rapport à la sémantique de RDF de manière à commencer à fonder le Web sémantique sur des transformations sûres.

Références

- [1] Isabel Cruz, Stefan Decker, Jérôme Euzenat, Deborah McGuinness (eds), *The emerging semantic web*, IOS press, Amsterdam (NL), 302p., 2002.
- [2] Jérôme Euzenat, Heiner Stuckenschmidt, *The 'family of languages' approach to semantic interoperability*, in: Borys Omelayenko, Michel Klein (eds), *Knowledge transformation for the semantic web*, IOS press, Amsterdam (NL), 2003, to appear.
- [3] Jérôme Euzenat, *An infrastructure for formally ensuring interoperability in a heterogeneous semantic web*, in [1], 2002.
- [4] Jérôme Euzenat (ed), *Research challenges and perspectives of the Semantic web*, Rapport de prospective EU-NSF, <http://www.ercim.org/EU-NSF/semweb.html>, 2002.
- [5] Jérôme Euzenat, Laurent Tardif, *XML transformation flow processing*, *Markup languages: theory and practice* 3(3):285-311, 2002.
- [6] Jérôme Euzenat, *Eight questions about semantic web annotations*, *IEEE Intelligent systems* 17(2):55-62, 2002.

Équipe Orpailleur – LORIA

LORIA - UMR 7503
BP 239
54506 Vandoeuvre-lès-Nancy Cedex
<http://www.loria.fr/equipes/orpailleur/>

Contact

Amedeo Napoli
napoli@loria.fr

Chercheurs impliqués

Rim Al Hulou, Mathieu d'Aquin, Jean Lieber, Emmanuel Nauer, Laszlo Sathmary, Yannick Toussaint.

Travaux

L'objectif de l'équipe Orpailleur au LORIA est de concevoir des « systèmes intelligents », capables d'exploiter des connaissances pour résoudre des problèmes dans un domaine donné. Parmi les problématiques étudiées par l'équipe figure le Web sémantique, et plus particulièrement les points suivants : la recherche d'informations combinée à la fouille de données, la gestion et le partage de connaissances sur le Web, l'exploitation du contenu des documents du Web par des systèmes intelligents.

1. L'accès intelligent à l'information sur le Web.

Le besoin en information est primordial dans de nombreux domaines, comme ceux de la recherche et de la veille scientifique et technique. Le volume des données relatives à un domaine est de plus en plus important, mais cette quantité croissante de données disponibles nécessite de mettre en oeuvre des moyens particuliers pour exploiter les données. Ainsi, la maîtrise de l'accès à l'information pertinente dans un fonds volumineux et hétérogène tel que le Web représente un enjeu majeur pour les consommateurs d'information (chercheurs, entreprises, etc.). Un objectif ici est de fournir aux chercheurs et aux spécialistes de l'information scientifique et technique un environnement dans lequel ils peuvent exploiter les données de leur domaine, pour des besoins de recherche d'information ou d'analyse du domaine.

Les moteurs de recherche sont débordés par l'explosion du Web et ne répondent plus aux tâches de recherche d'information. Une des préoccupations de l'équipe Orpailleur est de favoriser un accès intelligent aux données du Web en exploitant des connaissances relatives au domaine des données traitées. Dans ce cadre est proposée une approche qui combine l'exploitation de connaissances extraites par des techniques de fouille de données à un système de recherche d'information. Si l'exploitation de connaissances est classique pour favoriser la recherche d'information, l'originalité de l'approche introduite ici est d'utiliser un système de fouille de données qui fournit les moyens de manipuler les données structurées d'un domaine (références bibliographiques en particulier) pour faire émerger des connaissances sur le domaine, comme des réseaux d'auteurs, le vocabulaire employé par tel ou tel auteur, etc.

L'idée maîtresse est que la fouille de données et la recherche d'information sont deux approches complémentaires pour appréhender des données structurées ou non : la fouille de données permet de guider la recherche d'information à partir des connaissances extraites des données, et, inversement, la

recherche d'information permet de guider la fouille de données par l'exploitation des connaissances issues de la fouille de données elle-même.

La recherche d'information sur le Web passe aussi par la construction automatique de synthèses de textes pour raisonner sur un texte, un ensemble de textes ou de documents textuels. La fouille de données textuelles doit fournir à un utilisateur une vision synthétique et interprétable du contenu d'un ensemble de textes provenant du Web. La construction d'une synthèse ou résumé sémantique d'un ensemble de textes permet de situer un texte par rapport à d'autres textes, mais permet aussi d'expliquer en quoi le texte étudié s'en rapproche et en quoi il s'en différencie. Ainsi, les activités liées à la fouille de textes contribuent directement à l'élaboration du Web sémantique.

2. Les représentations par objets pour le Web sémantique.

La nécessité de contrôler les informations, documents ou données, par l'intermédiaire d'une sémantique, renvoie à la problématique de la représentation des connaissances en général, et des représentations par objets (RCO) en particulier. Alors que se déploient d'importants projets sur le Web sémantique, les RCO se font de plus en plus indispensables, en tant que supports pour la représentation des connaissances, pour la programmation et l'échange, et en tant qu'unité de déploiement modulaire de services. De plus, l'utilisation de XML sous toutes ses facettes comme passerelle entre documents et objets trouve dans l'idée de Web sémantique une justification naturelle.

La recherche et l'interrogation d'un site Web en s'appuyant sur le contenu des documents sont devenues des nécessités : les formalismes de représentation des connaissances sont des formalismes adéquats pour représenter ce contenu. La représentation du contenu d'un document permet de manipuler ce document pour faire de la recherche par spécialisation, par similitude, par analogie, etc. Une des lignes de recherche entreprise dans l'équipe Orpailleur consiste à

décrire les données ou documents en XML, puis à combiner les fonctionnalités de XML et celles des systèmes de RCO pour exploiter au mieux les données, à des fins de résolution de problèmes, de recherche d'information, de fouille de données, de navigation, etc. Une partie de ces travaux a été réalisée dans le cadre de l'ARC INRIA ESCRIRE, qui a consisté à comparer trois formalismes de représentations des connaissances pour la manipulation de documents par le contenu sur le Web.

L'ARC INRIA ESCRIRE (Embedded Structured Content Representation In REpositories) est une action de recherche coopérative entre trois groupes de recherche de l'INRIA : ACACIA (INRIA SOPHIA), EXMO (Grenoble), et Orpailleur (Nancy). L'objectif d'ESCRIRE est de comparer les trois formalismes de représentations de connaissances, graphes conceptuels, représentations des connaissances par objets et logiques de descriptions, du point de vue de la gestion de documents scientifiques et techniques, de la représentation du contenu des documents, et de leur manipulation, dans le cadre d'un système d'information intelligent. La comparaison de ces différents formalismes pour le traitement d'un jeu de documents a nécessité de mener une réflexion méthodologique sur le passage des textes à une représentation formelle en XML, indépendante de tout formalisme de représentation. Un ensemble de requêtes a été défini de manière coordonnée pour permettre de mener à bien l'évaluation comparée des trois formalismes.

L'équipe Orpailleur s'est intéressée principalement aux logiques de descriptions, et c'est le système RACER qui a été utilisé, pour son expressivité et pour ses possibilités d'intégration à Java. Pour mener à bien le travail sur ESCRIRE, il a fallu mettre au point une méthodologie de traduction dans une base de connaissances d'une ontologie du domaine des données et des documents décrits en XML. Plus précisément, les documents étudiés sont des résumés d'articles de biologie sur des interactions entre gènes. Les connaissances du domaine de la biologie sont représentées

par des concepts et des relations organisés en une ontologie. Les termes de cette ontologie servent à annoter les documents et leur contenu. Le système RACER est utilisé pour raisonner sur les données réifiées, et retrouver en particulier les documents qui satisfont une requête.

3. Systèmes de gestion des connaissances, Webs sémantiques locaux et intranets sémantiques.

Dans le cadre de la construction de systèmes de gestion de connaissances, que ce soit pour l'entreprise ou la médecine, deux des applications principales pour Orpailleur en matière de Web sémantique, il est nécessaire de mettre en place des serveurs de connaissances multi-dimensionnels sur les domaines traités. De tels serveurs doivent jouer le rôle de pivot entre les différents référentiels de connaissances disponibles comme des ontologies et des thésaurus par exemple, les gestionnaires de connaissances (qui peuvent être des programmes et/ou des êtres humains), et enfin les utilisateurs (spécialistes ou non) de ces connaissances.

Ici, la connaissance doit être considérée comme une ressource qui peut être retrouvée, exploitée, enrichie et étendue. Deux projets de recherche dans ce cadre sont actuellement en cours de développement dans l'équipe Orpailleur, et concernent plus particulièrement le domaine de la gestion des connaissances décisionnelles en oncologie et celui de la gestion d'une mémoire d'entreprise.

Bibliographie

- [1] R. Al-Hulou, O. Corby, R. Dieng-Kuntz, J. Euzenat, C. Ramirez, A. Napoli, and R. Troncy. Three knowledge representation formalisms for content-based manipulation of documents.
- [2] M. Cristani, editor, Workshop on Semantic Web SemWeb@KR-02, Toulouse, 2002 (CD Rom publication).
- [3] R. Al-Hulou and A. Napoli. Combining XML and DL for describing and querying documents. In I. Horrocks and S. Tessaris, editors, Proceedings of the International Workshop on Description Logics DL-2002, Toulouse, p. 140-147, 2002.
- [4] R. Al Hulou, A. Napoli, and E. Nauer. XML : un formalisme de représentation intermédiaire entre données semi-structu-

rées et représentations par objets. In C. Dony and H.A. Sahraoui, editors, Langages et Modèles à Objets (LMO'00), Montréal, p. 75-90. Hermès, Paris, 2000.

[5] S. Brachais, M. d'Aquin, J. Lieber, A. Napoli, Vers un Web sémantique en cancérologie, In A. Burgun and J. Charlet, editor, Journée « Web sémantique médical », Rennes, mars 2003.

[6] E. Nauer. Complémentarité entre fouille de données et recherche d'information dans le cadre d'analyses biliométriques. In 13^{ème} Congrès francophone AFRIF-AFIA de Reconnaissances des Formes et d'intelligence Artificielle, volume 3, p. 965-974, Angers, Janvier 2002.

Equipes Thème 3 « Interaction Autonomie Dialogue Coopération » – IRIT

IRIT

Université Paul Sabatier,
118 route de Narbonne,
31062 Toulouse Cedex 4

Equipe SMAC

Tél : 05 61 55 82 94
mél : gleizes@irit.fr
Contact : Marie-Pierre Gleizes
Autre membre concerné : André Machonin
Adresse Web : <http://www.irit.fr/SMAC>

Equipe MCE

Tél : 05 61 55 82 93
mél : aussenac@irit.fr
Contact : Nathalie Aussenac-Gilles
Autre membre concerné : Bernard Rothenburger
Adresse Web : http://www.irit.fr/SSI/ACTIVITES/EQ_SMI/nathalie.html

Thème général des équipes du thème 3 de l'IRIT

Les travaux de l'équipe SMAC depuis 1993, concernent la conception de systèmes complexes, multi-agents et plus particulièrement de systèmes adaptatifs et à fonctionnalité émergente. Les logiciels que nous concevons résolvent des problèmes pour lesquels il n'existe pas de modèle ou bien pour lesquels le modèle est mal connu. L'objectif de nos travaux est de fournir des théories et des méthodologies pour la conception de tels systèmes artificiels. Pour cela, nous

préconisons une méthode de conception ascendante.

L'approche multi-agent suppose que l'activité d'un système complexe adaptatif peut s'obtenir par l'agrégation de compétences plus simples en interaction dynamique. Nous avons développé des applications dans lesquelles un collectif est composé d'agents autonomes qui ont un comportement de sincérité et de serviabilité vis-à-vis d'autrui. Nous avons pu vérifier que lorsque les agents coopèrent réellement, l'activité collective est optimale. La coopération a pu ainsi devenir un critère de réorganisation interne du système lorsqu'il est plongé dans un environnement dynamique. Les créations/suppressions d'agents, les conflits, la gestion des croyances, l'intercompréhension sous-tendent cette activité. Comme des propriétés nouvelles peuvent apparaître par l'auto-organisation, l'émergence est aussi un de nos axes d'étude.

Les recherches de l'équipe MCE en matière d'ingénierie des connaissances au sein de l'équipe MCE ont d'abord pour sur le recueil et la modélisation de connaissances d'experts en s'appuyant sur une analyse de l'activité et selon une démarche ascendante (MACAO). Depuis 10 ans, l'équipe s'intéresse aussi à l'acquisition de connaissances à partir de textes selon des principes linguistiques ou statistiques, et à l'aide d'outils de traitement automatique de la langue. Les modèles de connaissances concernés sont alors des taxonomies, des terminologies et des ontologies, directement en lien avec les textes dont ils sont issus. Le fait d'aborder des applications de différents types (recherche d'information, consultation de documents, gestion d'archives de projets ou veille scientifique) a permis à l'équipe d'appréhender la diversité des méthodes à adopter.

Description des travaux ou projets en lien avec le thème du Web sémantique

Les avancées du Web Sémantique sont reconnues aujourd'hui comme relativement lentes par rapport à la dynamique et à la richesse potentielle du Web. Parmi les conditions de la réussite

de l'introduction de plus de moyens d'accéder au sens des informations, des documents disponibles sur le Web, le W3C met en avant la mise à disposition d'ontologies adaptées, partageables et exprimées dans des formats standards, à côté de l'existence de moteurs et de systèmes d'annotation capables d'utiliser et de manipuler ces ontologies. Même si on ne sait pas qui des ontologies ou des outils de recherche les utilisant devront être prêt en premier, le Web Sémantique ne sera pas une réalité tant qu'il sera coûteux de faire une ontologie ou d'annoter des pages (ses pages ?) avec des éléments d'ontologie.

Nos recherches portent sur les méthodes et outils de construction d'ontologies à partir de textes, en s'appuyant sur l'hypothèse que les textes sont de bons véhicules de connaissances, que l'usage des termes est révélateur de connaissances, et que des indicateurs statistiques ou linguistiques peuvent guider le repérage des connaissances dans un texte. Nous présentons ici nos travaux les plus significatifs en lien avec la construction d'ontologies de domaines, un des projets étant en lien avec les données du Web.

Projet TERMINAE en collaboration avec le LIPN

En collaboration avec le LIPN (Paris 13), et suite à notre expérience de développement de Géditerm, un éditeur d'ontologies, nous contribuons à la spécification d'un logiciel support à la modélisation d'ontologies et de terminologies à partir de textes, Terminae. Terminae a la particularité d'assurer une aide continue depuis l'étude des textes et l'exploration des résultats d'outils d'analyse de textes (Syntex, Linguae) jusqu'à la formalisation de connaissances en passant par la mise au point de fiches terminologiques et d'un modèle conceptuel non formel.

Outils de TAL pour la construction d'ontologies

En collaboration avec l'ERSS, laboratoire toulousain de linguistique, et le CEA de Cadarache, un système d'aide au repérage de relations conceptuelles à partir de leur expression en langue a été mis au point en s'appuyant sur une

approche par marqueur (thèse de P. Séguéla, 2000). Ce système, Caméléon, permet de mettre au point des marqueurs adaptés à un corpus donné, puis de les projeter et enfin d'enrichir un modèle conceptuel à partir des traces lexicales de relations trouvées dans les textes. Une nouvelle version s'appuie sur le concordancier Yakwa, développé par L. Tanguy, pour pouvoir définir des marqueurs plus puissants travaillant sur des textes étiquetés par un analyseur syntaxique. Nous expérimentons également les systèmes Syntex et Upery développé par D. Bourigault et C. Fabre. Tous ces outils permettent de scruter des textes en français ou en anglais.

Projet ABROSE : Agent Based Brokerage Services in electronic commerce (1998-2000)

Abrose est un projet ACTS européen. C'est un outil de commerce électronique de médiation de services fondée sur des agents. Une place de marché électronique est constituée d'une collection importante de fournisseurs et de clients qui ont des offres et des besoins très dynamiques. Chacun d'eux est représenté par un agent et l'approche par auto-organisation coopérative garantit une satisfaction individuelle et collective optimales alors qu'il n'existe jamais de connaissance globale de l'état du système. Chaque agent se construit et maintient une représentation du monde à l'aide d'un système multi-agent adaptatif. C'est ainsi que le système possède un certain nombre de fonctionnalités, notamment :

- La publicité ciblée : lorsqu'un fournisseur de services souhaite diffuser une offre, seuls les agents représentant des usagers dont les besoins sont conformes à celle-ci reçoivent l'information.
- La concurrence totale : un client reçoit du système la liste exhaustive des fournisseurs qui ont des produits répondant à la description de son besoin.
- L'apprentissage permanent : les agents représentants adaptent automatiquement leurs croyances tant sur les nouveaux produits des fournisseurs que sur les nouveaux besoins des usagers.

Projet DYNO

L'élaboration et la maintenance d'ontologie représentent un travail complexe et difficile jusqu'ici peu abordé. Nous souhaitons évaluer l'intérêt de systèmes multi-agent adaptatif pour la création et la maintenance automatique d'ontologies à partir de textes et, pour cela, concevoir un outil, DYNO (DYNAmical Ontology). Ce projet fait intervenir les équipes SMAC et MCE. Une première phase consiste en l'évaluation d'une ontologie réalisée à partir de l'outil SCIO (basé sur un système multi-agent adaptatif) dans le domaine de la chimie. Ensuite, la réalisation d'un outil interactif de transformation permettra d'obtenir l'ontologie écrite dans le langage standard OWL. L'ontologie obtenue pourra ainsi être comparée à l'ontologie élaborée à l'aide d'outil de traitement automatique des langues selon une approche supervisée par un cogniticien à partir des mêmes textes. A terme, nous envisageons d'enrichir les compétences des agents de capacité à traiter la syntaxe du texte et non pas seulement les relations de proximité entre termes.

Projet IndexWeb

En collaboration avec l'ERSS, nous avons évalué l'intérêt d'une analyse terminologique de pages de sites web et de documents d'entreprise pour améliorer l'aide au référencement proposée par le Système IndexWeb de la société Synapse Développement. Nous avons proposé une méthode s'appuyant sur la construction d'une ressource terminologique minimale à l'aide de logiciels d'analyse de textes (extracteur de termes Syntex, analyse distributionnelle Upery et recherche de patrons Yakwa). Les résultats ont finalement deux types d'utilisation possible pour un site donné : fournir un index hypertextuel du site, dont les entrées sont des termes (Groupes nominaux mais aussi verbes relatifs aux activités de l'entreprise et services fournis par le site) ; mise en évidence des termes saillants pour chaque page, qui peuvent être choisis comme méta-donnée « key-words » pour cette page.

Mesure de l'évolution des connaissances dans les textes

Lorsque l'on accède à de grande collection de textes (par exemple sur le Web), la démarche n'est pas seulement de découvrir des textes potentiellement pertinents. Souvent, on voudrait découvrir des tendances concernant les connaissances sur un domaine : quels sont les nouveaux domaines de connaissances qui sont apparus, quels sont ceux qui ont disparus, comment a évolué la prise en compte de tel ou tel domaine.

La méthode que nous proposons est de confronter une série, éventuellement temporelle, de textes sur un domaine donnée à un taxonomie de ce domaine. Un ensemble de métriques ont été définies pour prendre en compte la nature de cette confrontation. Ces métriques prennent en compte la *densité* de l'impact (combien de documents atteignent une catégories), sa *spécificité* (à quelle profondeur se trouve les catégories atteintes) et sa *concentration* (quelles sont les zones contiguës de la taxonomie qui ont été plus atteintes).

Cette méthode a été mise en œuvre dans le cadre d'application de suivis de projet ou de veille technologique avec de grands organismes scientifiques comme le CNES, notamment pour l'exploitation de ressources du Web.

Bibliographie

- [1] Georgé J-P., Gleizes M-P., Glize P., (2003) « Conception de systèmes adaptatifs à fonctionnalité émergente : la théorie des Amas », Revue d'IA (à paraître 2003)
- [2] Gleizes M-P., Glize P., (2002) « *ABROSE: Multi Agent Systems for Adaptive Brokerage* », Fourth International Bi-Conference Workshop on Agent-Oriented Information Systems (AOIS-2002), 27-28 May 2002, Toronto (Ontario, Canada) at CAiSE'02.
- [3] Aussenac-Gilles N., Seguela P., 2000 Les relations sémantiques : du *linguistique au formel*. Cahiers de grammaire, N° spécial sur la linguistique de corpus. A. Condamines (Ed.) Vol 25. Déc. 2000. Toulouse : Presse de l'UTM. pp 175-198.
- [4] B. Rothenburger, « A Differential Approach for Knowledge Management ». Proceedings of the Workshop: Machine Learning and Natural Language Processing for Ontology Engineering (ECAI02), Lyon, juillet 2002.
- [5] Szulman, S., Biébow B., Aussenac-Gilles N., 20002. Structuration de Terminologies à

l'aide d'outils d'analyse de textes avec TERMINAE. *Traitement Automatique de la Langue (TAL)*. Numéro spécial sur le Structuration de Terminologie. Eds A. Nazarenko, T. Hammon. Vol43, N°1; pp 103-128. 2002.

EQUIPE CONNAISSANCE, INFORMATION ET DONNÉE - IRIN

2, rue de la Houssinière BP 92208
44322 Nantes cedex 03
<http://www.sciences.univ-nantes.fr/irin/ThemeIC/>

Contact

Trichet Francky
trichet@irin.univ-nantes.fr

Membres impliqués dans la problématique du Web Sémantique

Cazalens Sylvie, Desmontils Emmanuel, Jacquin Christine, Lamarre Philippe, Leclere Michel, Trichet Francky

Travaux

Le projet scientifique de l'équipe « Connaissance, Information et Donnée » de l'IRIN., actuellement composée de deux thèmes intitulés respectivement « Extraction et Visualisation de Connaissances » et « Ingénierie des Connaissances », est d'apporter une contribution à la problématique de l'appropriation de connaissances par interprétation d'informations et/ou de données brutes. Les travaux menés actuellement portent sur la construction d'artefacts opérationnels de diverses natures (i.e., outil d'extraction à partir de données, outil de visualisation de règles d'association, outil de mesure de la qualité de règles d'association, outil de déploiement de systèmes multi-agents, outil d'annotation et d'indexation de documents textuels, etc.) et sur l'étude des problèmes théoriques sous-jacents à ces artefacts (i.e., algorithme de fouille, parcours de graphes et visualisation, représentation de connaissances et de croyances, définition de protocole pour des communautés d'agents, etc.). In fine, l'intégration de

ces travaux complémentaires contribuera, d'une part, au développement, pour différents modèles, de systèmes adaptatifs de fouille de données couplés à des systèmes de gestion des connaissances et, d'autre part, à l'émergence d'une véritable ingénierie du Web sémantique. Dans le cadre de cette dernière problématique, les travaux menés actuellement au sein de l'équipe visent à apporter des éléments de solutions aux trois axes de recherche suivants :

1. Connaissances distribuées et systèmes multi-agents

Définition d'infrastructures d'agents cognitifs pour le développement de services distribués.

2. Construction d'ontologies et intégration au sein d'un Système à Base de Connaissances

Formalisation et opérationnalisation d'une ontologie de domaine (concepts, relations et axiomes) à l'aide du modèle des Graphes Conceptuels et étude de l'adéquation avec les langages du Web sémantique.

3. Annotation et indexation de documents électroniques

Définition de méthodes d'indexation et d'annotation fondées sur l'exploitation d'ontologies pour la recherche d'information par le contenu.

D'un point de vue applicatif, ses travaux sont expérimentés dans le cadre de deux projets relevant du recrutement électronique, un domaine qui s'inscrit pleinement dans le cadre du Web sémantique : besoin d'explicitier les compétences sous-jacentes aux ressources électroniques liées à l'emploi (CV, bilan, offre d'emploi), besoin de gérer des documents massivement distribués, nécessité d'intégrer des informations provenant de sources hétérogènes, nécessité de partager des référentiels d'emploi/métier pour faciliter l'activité de rapprochement entre l'offre et la demande du marché de l'emploi, etc.

Bibliographie

[1] S. Cazalens, E. Desmontils, C. Jacquin, P. Lamarre (2002). « Sources d'informations et de connaissances : de la gestion locale à la

recherche distribuée », Revue L'Objet, 8 (4), Hermès.

[2] S. Cazalens and P. Lamarre (2001). « An organization of Internet Agents based on a hierarchy of information domains », Y. Demazeau and F.J. Garijo eds, Proceedings of « Modelling Autonomous Agents in a Multi-Agent World »: 12-27.

[3] E. Desmontils, C. Jacquin (2002). « Indexing a Web Site with a Terminology Oriented Ontology », The Emerging Semantic Web, I.F. Cruz, S. Decker, J. Euzenat and D. L. McGuinness Ed, pp. 181-197, IOS Press.

[4] Furst F., Leclère M., Trichet F. (2003). « Ontological Engineering and Mathematical Knowledge Management: A formalization of projective geometry », A paraître dans Annals of Mathematics and Artificial Intelligence. Kluwer Academic Publishers.

[5] Trichet F., Leclère M., Tixier B. (2002). « Capitalizing and Sharing Know and Know-How: an approach based on a Task/Method Knowledge-Based System », Handbook of Knowledge Management and Organizational Memories, pp. 31-40, Kluwer Academic Publishers.

avec les projets OSIRIS (Gestion unifiée des connaissances et des données) et ISIS (méthodologie de conception de systèmes d'information). Elle y travaille explicitement avec un travail sur la traduction d'ontologies [1], la thèse de G. Diallo sur la gestion des données structurées et faiblement structurées, et la réalisation en 2002 d'un éditeur d'ontologies, Ont'Osiris, qui est la première brique d'un environnement pour la conception d'ontologies. Les liens entre ces travaux sont représentés par le schéma ci-dessous :

Une ontologie peut être :

- (1) exploitée en base de connaissances, pour le classement d'instances (OSIRIS)
- (2) utilisée comme schéma initial dans la conception d'une base de données (ISIS)
- (3) Un schéma ISIS peut produire un schéma OSIRIS qui peut être exploité pour la gestion des données (SGBD) et comme base de connaissances (toute instance est systématiquement classée dans ses vues).

Equipe OSIRIS : Modélisation et Représentation des Données et des Connaissances

TIMC-IMAG

Faculté de Médecine - 38700 LA TRONCHE

<http://www-timc.imag.fr/osiris/>

Responsable

Michel SIMONET

Tel : 04 76 63 71 54

courriel : Michel.Simonet@imag.fr

Membres

Michel Simonet (chercheur CNRS), Ana Simonet (MCF), Patrick Palmer (MCF), Gayo Diallo (doctorant) Delphine Bernhard (DEA)

Thème général de l'équipe

Modélisation et représentation des données et des connaissances – Systèmes d'Information Intelligents – Ontologies.

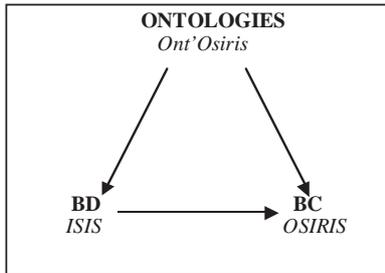
L'équipe travaille implicitement sur les ontologies depuis plusieurs années,

Travaux

La mise en œuvre du Web sémantique nécessite un certain nombre de travaux permettant la recherche d'information et le requêtage de données ainsi que l'interopérabilité entre sources de données hétérogènes. Les ontologies sont au cœur de ces processus : vocabulaire commun nécessaire à l'interopérabilité, thesaurus sémantique pour la recherche d'information, schéma conceptuel support des requêtes aux données. Le travail entrepris par G. Diallo pour la gestion unifiée des données structurées (les bases de données) et des données faiblement structurées (textes, pages Web), met en œuvre ces différentes composantes [4]. Il s'appuiera sur les réalisations de l'équipe : l'éditeur d'ontologies Ont'Osiris, le SGBD-BC OSIRIS pour l'exploitation des ontologies et la gestion des données, le système ISIS pour la rétro-ingénierie des bases de données ainsi que pour la définition des médiateurs sémantiques pour leur intégration. Ses travaux ont trouvé une première application dans le projet BC3 (Base de Connaissances Cœur-Cerveau), soutenu par la Région Rhône-Alpes, pour la

construction des relations anato-mofonctionnelles du cerveau [5].

L'éditeur d'ontologies **Ont'Osiris** a été conçu dans la perspective d'un environnement plus général pour la conception d'ontologies. D'une certaine manière, les systèmes ISIS et OSIRIS



possèdent déjà une interface pour la saisie d'ontologies. OSIRIS peut être considéré comme une logique de description avec un habillage objet (bien que les deux modèles fussent ignorés de ses concepteurs au moment de sa création). De ce fait il induit une structure de classes, que l'on peut assimiler à celle des frames, qui nous semble inappropriée pour le niveau de description propre à une ontologie. Nous avons touché la même question en ISIS où, pour la conception de bases de données, nous avons choisi de partir d'une description ontologique « plate », contrairement par exemple à une approche comme UML, qui induit des choix de type informatique qui sont préjudiciables à l'optimisation des bases de données qui seront produites à l'issue du processus de modélisation.

Ont'Osiris est implanté en JAVA. Il est réalisé dans une optique multilingue, à la fois au niveau de l'application (menu, messages, ...) et des termes associés aux concepts. L'application est aujourd'hui disponible en français, en anglais et en russe. Le chinois et l'arabe sont en projet. Le temps nécessaire pour l'introduction d'une nouvelle langue est celui de la traduction des libellés, contenus dans un fichier XML. Il est de l'ordre d'une journée.

L'accent a été mis sur l'import et l'export de données, qui est possible vers et à partir des formats RDF et DAML+OIL. L'ontologie elle-même est sauvegardée dans un format XML propriétaire. Une sortie HTML est égale-

ment possible, permettant de visualiser l'ontologie dans n'importe quel navigateur.

Ont'Osiris va être étendu en un environnement pour la conception d'ontologies, avec une aide pour l'acquisition à partir de textes. C'est le projet de thèse de D. Bernhard. Il ne s'agit pas seulement de mettre en œuvre des outils d'extraction terminologique, mais d'aider à la structuration des termes, en travaillant au niveau sémantique (cf travaux de Sabine Ploux à l'Institut des Sciences Cognitives sur les cliques de synonymes représentant des unités minimales de sens).

ISIS (Information System Initial Specification) désigne à la fois une méthodologie et un outil pour l'aide à la conception et à l'implantation de schémas de bases de données. Ces bases peuvent être de nouvelles bases ou des bases de données intégrant plusieurs bases existantes.

ISIS répond à un besoin qui n'est pas couvert par les méthodes et outils existants, qui concernent la conception de logiciel (e.g., UML) ou qui se limitent à la description d'une analyse conceptuelle de schéma de base de données (e.g., Merise) mais n'apportent pas d'aide à la conception elle-même. ISIS peut être considéré de manière autonome ou en complément de ces approches et des outils qui les implantent.

Un schéma ISIS consiste en une description « plate » des concepts de l'univers et de leurs interrelations. Cette description initiale est nommée « **diagramme ontologique** ». Ce niveau de description est suffisamment simple et compréhensible pour permettre une validation de la part des utilisateurs. Le diagramme ontologique est ensuite enrichi par les propriétés comportementales de l'application. Ces propriétés explicitent des critères qui interviennent implicitement dans les choix de modélisation. ISIS peut alors déduire les ensembles qui doivent être considérés comme des ensembles d'objets, des ensembles de valeurs ou des ensembles d'index. Selon le modèle cible choisi (relationnel, objet ou UML) ISIS propose les procédures de création d'une base (actuellement : Access, Oracle et Osiris), les requêtes

SQL d'interrogation et de mise à jour, les signatures des méthodes de mise à jour, des méthodes de calcul d'attributs et les triggers de mise à jour.

ISIS propose aussi des **simplifications** du schéma défini, c'est à dire la mise en évidence des sous-graphes dont la pertinence peut être discutée (soit ils ne sont pas nécessaires, soit la spécification est incomplète). Cet algorithme a été généralisé, en prenant en compte l'ontologie de représentation du modèle relationnel, afin de proposer le schéma objet maximal associé à un schéma relationnel désambiguïsé. Ainsi, plutôt que de travailler directement sur des tables du schéma physique, ce qui est tâche lourde et sujette à erreurs, le concepteur peut travailler sur un nombre réduit de concepts porteurs de sémantique.

Le projet ISIS faisant partie d'une action de valorisation, les travaux sont confidentiels. Voir cependant [2].

OSIRIS est un système de représentation et de gestion de données et de connaissances (SGBD-BC) orienté objet. Il est fondé sur le modèle des p-types [thèse A. Simonet, 1984], conçu à l'origine pour répondre aux besoins des bases de données, en particulier pour le partage des données entre différentes catégories d'utilisateurs. Dans ce modèle, concepteur et utilisateur ne travaillent qu'au niveau des schémas externes (les vues, dans l'architecture à trois niveaux de la norme ANSI-SPARC qui fonde les bases de données modernes) ; les niveaux des tables (dans le cas d'une implantation relationnelle) et des fichiers sont entièrement gérés par le système, de manière automatique et transparente à l'utilisateur comme au concepteur.

Le système de vues ainsi défini est très proche des concepts définis des Logiques de Description [3]. OSIRIS permet l'expression des propriétés logiques attachées aux classes d'objets, ici représentées par des vues, et le classement automatique d'instance dans les vues en fonction des propriétés de l'instance considérée. Osiris a été utilisé pour planter une base de connaissances en mycologie, et des travaux ont été entrepris pour développer une ontologie des attributs utilisés, en particulier les couleurs.

Bibliographie

- [1] H. Mihoubi, A. Simonet, M. Simonet, Towards a Declarative Approach for Re-using Domain Ontologies, Information Systems, Vol. 23, No 6, pp 365-381, 1998.
- [2] A. Simonet, M. Simonet, The ISIS Methodology for Object Database Conceptual Modelling, Poster E/R 99 : 18th Conference on Conceptual Modeling, Paris, 15-18 Nov. 1999.
- [3] M. Roger, A. Simonet, M. Simonet, A Description Logic-like Model for a Knowledge and Data Management System, DEXA 2000, Database and Expert System Applications, Greenwich (UK), 4-8 Sept. 2000.
- [4] G. Diallo, Ontologies pour la gestion des données structurées et faiblement structurées, Mémoire de DEA Systèmes d'Information, Grenoble, Juin 2001.
- [5] G. Diallo, D. Bernhard, B. Batrancourt, M. Simonet, Vers une Ontologie Anatomofonctionnelle du Cerveau, 1ères Journées Web Sémantique Médical, Rennes, Mars 2003.

**France Télécom R&D
Laboratoire D2I : Dialogues
et Intermédiations
Intelligentes
Direction des Interactions
Humaines**

France Télécom R&D DIH/D2I
2 avenue Pierre Marzin
22307 Lannion Cedex

Contacts : david.sadek@francetelecom.com ou philippe.bretier@francetelecom.com

Laboratoire

Le laboratoire D2I est composé de chercheurs, d'ingénieurs et de doctorants en informatique (Intelligence artificielle, architectures logicielles, agents intelligents), dialogue personne-machine, sciences cognitives, linguistique et psycho-ergonomie des interactions humains-agents.

Domaines d'activité du laboratoire

Développement d'outils méthodologiques et technologiques pour la mise en œuvre de dialogues personne-machine évolués et d'applications d'intermédiations intelligentes utilisateurs-services,

quels que soient les médias et modes d'interaction, ou les contextes linguistiques et socio-culturels cibles.

Les technologies développées au laboratoire Dialogues et Intermédiations Intelligentes (D2I) de France Télécom R&D, telles que la technologie ARTIMIS, mettent en œuvre des principes et mécanismes génériques de comportements intelligents (rationalité, communication, coopération, etc.) et de compréhension et génération de langages naturels et multimodaux/multimédia. La généralité de ces modèles et technologies se traduit par leur indépendance par rapport aux spécificités d'une application, d'une langue ou d'un média de communication particulier.

Parmi les domaines d'application des technologies développées par le laboratoire D2I, on peut citer celui de l'intermédiation, dans lequel des assistants intelligents, personnels ou publics, gèrent, à travers une interaction conviviale (dialogue en langage naturel, par exemple, à l'oral comme à l'écrit), la relation entre le client et des services de recherche ou de transaction, sur le Web notamment : identification des propositions qui répondent au mieux à la demande de l'utilisateur, construction de propositions pouvant combiner des informations issues de sources distinctes, proposition de solutions alternatives, etc.

Les domaines d'étude concernés sont la représentation des connaissances, la modélisation de comportements, la formalisation du raisonnement, l'inférence automatique, l'apprentissage des connaissances et des comportements, les modèles de dialogue personne-machine, la psycho-ergonomie cognitive pour le dialogue, la sémantique des interactions, les agents intelligents et le traitement automatique du langage naturel et de la multimodalité.

Travaux du laboratoire en rapport avec le Web sémantique

On retrouve dans les activités du laboratoire D2I les deux volets autour desquels s'articule l'objectif du **Web sémantique**, à savoir :

- permettre aux processus informatiques d'*accéder à la sémantique* des contenus des ressources du Web ;
- doter ces processus de capacités à *raisonner sur cette sémantique*.

Concernant le premier volet, les travaux du laboratoire portent plus particulièrement sur la sémantique des communications :

- inter-agents,
 - agents-ressources du Web (Web services, bases de connaissances, etc.),
 - humains-agents,
- ce dernier aspect étant une condition nécessaire à l'accroissement de l'utilisabilité du Web, notamment au travers d'interfaces permettant à l'utilisateur un accès pertinent et naturel aux contenus.

Web sémantique pour les interactions

Schématiquement, une action de communication d'un agent peut être analysée selon deux composantes :

- sa *fonction illocutoire*, qui caractérise sa nature même (telle que demander ou transmettre une information à son interlocuteur) ;
- son *contenu sémantique*, exprimé dans un langage de représentation des connaissances et faisant référence à la notion d'ontologie. C'est à ce niveau qu'apparaît, en particulier, la notion d'objets « métiers ».

Le consortium de standardisation FIPA (Foundation for Intelligent Physical Agents) a pour objectif de promouvoir le développement de spécifications de technologies génériques favorisant l'interopérabilité des applications à base d'agents logiciels. FIPA s'intéresse à ces deux composantes d'analyse de la communication inter-agent.

Outre les spécifications d'architecture pour l'interopérabilité des middlewares, FIPA a principalement produit les spécifications sémantiques du langage de communication inter-agent ACL (Agent communication language), du langage de contenu SL (Semantic Language) et d'un ensemble de protocoles génériques d'interaction. Les standards ACL et SL sont fondés sur le langage ARCOL (Artimis communication language) issu des travaux du laboratoire D2I. Des projets coopératifs tels que FACTS et

Agentcities, auxquels contribue D2I, montrent l'intérêt et la validité d'un tel standard.

A l'instar de FIPA, d'autres d'organismes (e.g. W3C, IEEE, DARPA, etc.), s'intéressent à la spécification d'ontologies. Cependant, jusqu'à présent, les propositions se sont concentrées principalement sur la définition de *syntaxes* standard, souvent à base d'XML. Or, si les syntaxes définies par les langages de représentation des connaissances proposés aujourd'hui, permettent aux agents de manipuler des concepts, elles ne garantissent pas le partage, la compréhension mutuelle et la découverte dynamique des concepts ou objets « métiers » manipulés.

Ce problème apparaît clairement lors de la construction de services d'intermédiation dans un environnement ouvert, où il est nécessaire de pouvoir composer de façon automatique, des services dont les capacités d'interaction (langages, ontologies, protocoles) sont décrites formellement. Ainsi, par exemple, étant donné deux agents indépendants, l'un fournissant des horaires d'avion, l'autre des horaires de train, un agent tiers jouant le rôle de médiateur devrait pouvoir intégrer ces services dynamiquement afin de fournir des voyages de bout en bout, combinant avion et train. Par la suite, l'agent médiateur pourra découvrir et intégrer un nouvel agent offrant des services de réservation de taxi, complétant ainsi automatiquement son offre de services.

Le laboratoire D2I étudie les langages de contenu tant du point de vue de leur expressivité que de leur capacité de découverte dynamique de nouveaux concepts ou objets « métiers ». L'objectif est d'identifier un langage de représentation des connaissances qui optimise ces deux aspects dans les contenus des actions du langage de communication ACL. Il est à noter que la prise en compte automatique de nouveaux concepts constitue l'un des principaux verrous technologiques à la mise en place d'un véritable **Web sémantique ouvert**.

Une autre problématique à laquelle

s'intéresse le laboratoire D2I est la représentation et la manipulation en contexte des concepts et objets d'une interaction, en particulier au sein des actions de communication. Cette problématique est comparable au traitement de la référence linguistique dans le domaine du dialogue et de la compréhension du langage naturel.

Web Sémantique pour les raisonnements

Raisonnement sur les connaissances, les concepts, les objets « métiers », et plus généralement, sur la sémantique des contenus du Web, requiert avant tout de disposer d'une représentation de ces éléments. Par conséquent, ce second volet des activités de recherche sur le Web Sémantique s'appuie en grande partie sur des techniques communes au volet relatif à l'accès à la sémantique des contenus. En particulier, les langages de représentation des connaissances sont souvent destinés à servir à la fois la composante interaction et raisonnement d'un agent.

Le laboratoire D2I étudie également les langages de représentation des connaissances du point de vue de leur capacité à exprimer et à mettre en œuvre des *inférences*. Les caractéristiques visées sont alors le potentiel inférentiel et son automatisation au sein de systèmes réels. Ce dernier point met l'accent sur des problématiques d'efficacité du raisonnement, de fonctionnement en « temps réel », ainsi que de montée en charge et de passage à l'échelle en environnement d'exploitation.

La technologie ARTIMIS d'*Agent Rationnel Dialoguant*, développée à D2I, comporte une unité rationnelle capable de raisonnements avancés sur des représentations d'attitudes cognitives (croyances, intentions, etc.) et sur des connaissances du domaine (dans ce dernier cas, ces raisonnements visent, par exemple, à affiner ou à élargir la recherche d'information). Cette unité met en œuvre un ensemble d'axiomes logiques génériques qui formalisent des principes de base du comportement rationnel, de la communication et de la coopération.

Bibliographie

<http://www.rd.francetelecom.fr/fr/technologies/ddm200210/dossier.php>

- [1] D. Sadek & H. Brouchoud : Agents intelligents. Réseaux – Internet, téléphone, multimédia – Convergences et complémentarités, ch. 10, pp. 343-374, Editeurs D. Hardy, G. Malléus, J.-N. Méreur eds., De Boeck Université, 2002, pp. 343-374, 2002.
- [2] FIPA Communicative Act Library Specification, <http://www.fipa.org/specs/fipa00037/>, 2000.
- [3] D. Sadek : Design considerations on dialogue systems: from theory to technology - the case of Artemis, Proceedings of the ESCA TR Workshop on Interactive Dialogue for Multimodal Systems (IDS'99), invited paper, Germany, 1999
- [4] D. Sadek, P. Bretier & F. Panaget : Artemis: Natural dialogue meets rational agency. Proceedings of IJCAI'97 (International Joint Conference on Artificial Intelligence), Nagoya, Japan, 1030-1035, 1997.
- [5] P. Bretier & D. Sadek : A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. Lecture notes in Artificial Intelligence: Intelligent Agents (III). Proceedings of ATAL'96 (European Conference on Artificial Intelligence workshop on Agent Theories, Architectures, and Languages (ATAL'96), Budapest, Hungary, 1996. J.P. Müller, M.J. Wooldridge & N.R. Jennings editors, pp. 189-203, 1997.

Tech-CICO - Université de Technologie de Troyes (UTT)

Laboratoire Technologie de la
Coopération pour l'Innovation et le
Changement Organisationnel (Tech-
CICO)

Université de Technologie de Troyes
(UTT)

12, rue de Marie Curie - BP 2060

10010 Troyes cedex

<http://tech-cico.utt.fr>

Contact

Manuel Zacklad

Tel: 03 25 71 76 95 / 80 13 -

fax: 03 25 71 76 98

Mél.: Manuel.Zacklad@utt.fr

Participants à l'axe « Web Sémantique » du laboratoire

Jean Caussanel, Christophe Doniat, Amalia Todirescu, Manuel Zacklad, Xavier Barbaud (Thèse), Jean-Pierre Cahier (Thèse), Jean-Philippe Leboeuf (Thèse), Xavier Pétard (Thèse).

Cadre général

Le cadre général est celui de l'étude des interactions cognitives dans l'ingénierie du Web sémantique. Si la problématique initiale du Web Sémantique (WS) est d'offrir à des agents logiciels la capacité d'exploiter le Web de manière « intelligente », cette capacité n'est possible qu'à condition que les domaines de connaissances aient été décrits par des acteurs humains, qu'ils soient des spécialistes du domaine lors de la constitution d'ontologies de référence ou des auteurs, voire lecteurs actifs, des documents, lors de leurs activités d'annotation à l'aide de méta-données (faisant ou non appel à des ontologies pré-existantes).

Par ailleurs, si un des objectifs du WS est de confier la recherche d'informations sur le Web à des agents automatiques, une partie non négligeable des corpus mis à disposition (catalogues, documents, plaquettes, fichiers...) est d'abord conçue pour des lecteurs humains. Il serait donc dommage que les efforts de structuration sémantique consentis pour les agents logiciels ne puissent pas également bénéficier à des humains engagés dans des activités de navigation.

Nous défendons dans notre programme l'idée selon laquelle la problématique de la construction de représentations sémantiques adaptées, comme celle de leur exploitation ultérieure par des lecteurs humains, doivent faire partie intégrante des recherches sur le WS, ce qui les inscrit d'ailleurs, sans ambiguïté, dans le champ de l'Ingénierie des Connaissances.

Notre préoccupation est aussi d'intégrer dans les recherches sur le WS l'ensemble des activités de conception initiale des représentations, de maintenance au fil de l'eau (au fur et à mesure que les connaissances évoluent) et d'évaluation de la pertinence des résultats des requêtes, autant d'activités

reposant sur des interactions cognitives dont la prise en considération est indispensable pour une industrialisation effective de services exploitant les potentialités du WS.

Nous plaçons ainsi les projets du laboratoire dans l'optique du « Web Cognitivement Sémantique », résumant toutes ces préoccupations.

Gestion coopérative des représentations sémantiques

Un premier projet vise la définition de protocole de coopération multi-experts dans une « place de marché à base de connaissance » via la norme Topic Maps. Une des conditions du fonctionnement du WS est en effet la définition de représentations émanant de plusieurs spécialistes de domaines. Qui plus est, elle peut être amenée à s'enrichir au fur et à mesure que des utilisateurs interagissent avec cette représentation.

Dans ce contexte, ce modèle (Knowledge Based MarketPlace ou KBM) permet de représenter un catalogue de ressources informationnelles selon différents points de vue et autorisant différentes formes de coopération, entre experts des domaines, entre utilisateurs finaux de la KBM, entre experts du domaine et utilisateurs finaux. Dans l'optique du Web Cognitivement Sémantique, les ressources sont ainsi considérées comme évoluant rapidement à l'instigation de multiples contributeurs non dotés de compétences poussées en formalisation. Si le modèle de départ s'inspire des « places de marché » sur internet, les domaines d'application ne relèvent pas uniquement du domaine marchand mais couvrent toutes les situations d'échange d'information qui sont médiatisées par un espace commun permettant de déposer les objets échangés.

Pour le modèle KBM nous nous appuyons sur les Topic Maps (TM). Ceux-ci sont un standard de notation qui permet de définir des vues multiples et concurrentes d'un ensemble d'information. Ce standard ISO dispose d'une représentation XML, la spécification XTM, que nous avons identifié comme l'un des plus adaptés pour la réalisation d'applications répondant aux critères du

Web Cognitivement Sémantique.

La réalisation d'une application dans un contexte industriel nous a conduit à formaliser davantage notre méthode de représentation et ses liens avec la norme Topic Maps, sous la forme d'un méta-modèle, HyperTopic. Une partie de nos réflexions portent aussi sur l'ergonomie de la navigation sémantique, en étudiant la relation entre le sémantisme d'une représentation et les performances navigationnelles.

Agents intelligents pour des moteurs de recherche sémantiques

Un autre projet consiste à enrichir l'architecture KBM en lui adjoignant des modules pour rechercher semi-automatiquement de nouvelles sources d'information. Si notre approche privilégie l'accès et la mise à jour par des acteurs humains plutôt que par des agents logiciels, nous sommes conscients du potentiel que les agents logiciels représentent pour assister les utilisateurs dans leurs interactions avec le Web. Mais cette formalisation est difficile à imposer comme un préalable à la conception de la plupart des applications. Elle doit plutôt résulter de la familiarisation progressive des utilisateurs et de l'équipe de conception avec le domaine abordé.

Pour développer cette amélioration, nous comptons utiliser certaines techniques d'analyse linguistique en reliant le formalisme des Topic Maps à celui des Logiques de Description. Par rapport à une approche statistique, qui a besoin d'un volume important de données pour construire une ontologie, nous privilégierons une approche qui utilise des techniques d'analyse robuste, et s'appuie également sur les mécanismes d'inférence proposés par le formalisme de la Logique Terminologique. Cette approche a l'avantage d'utiliser moins de ressources pour enrichir l'ontologie, comme le prouve la méthodologie DLIR développée dans les travaux de A. Todirascu.

Applications et perspectives

Les méthodes et outils que nous développons ainsi pour le Web Cognitivement Sémantique, sont confrontés à des applications sur trois

terrains : les projets d'innovation, les annotations en conception et les systèmes collaboratifs pour les réseaux de soin.

Dans le premier domaine, le développement d'un système de gestion des projets d'innovation a été assuré en 2002 pour France Telecom ; il s'est agi de concevoir une place de marché de l'innovation basée sur les Topic Maps et le modèle Hypertopic, pour le partage des résultats de la R&D dans l'ensemble de ce Groupe. Dans le second domaine, le développement d'un système de gestion des annotations collectives a concerné un ensemble de projets de conception industrielle distribués (projet MEDIAPRO). Enfin, le troisième terrain d'expérimentation concerne le développement d'un système collaboratif pour les réseaux de soin pour personnes âgées en Champagne-Ardenne.

Ces terrains vont permettre de pousser l'analyse et la modélisation des interactions coopératives dans le sens d'une *web cognitivement et socialement sémantique*. Après avoir surtout travaillé sur la construction d'ontologies flexibles et multipoints de vue pour représenter les connaissances du domaine, nous souhaitons aujourd'hui développer des modèles de coordination à travers la proposition d'ontologies génériques permettant de représenter les organisations et les rôles puis les processus dans lesquels les acteurs sont engagés pour, par exemple, comparer des offres, lancer une transaction, assurer une veille conjointe sur un ensemble de thèmes ou co-construire des classifications.

Bibliographie

- [1] Cahier J.-P., Zacklad M., « Approche cognitive de la co-construction d'un catalogue multipoints de vue et multi-experts », Actes de la Conférence Extraction et Gestion des Connaissances EGC 2002, Montpellier du 21 au 23 janvier 2002, p.221-226, Hermès.
- [2] Caussanel J., Cahier J.-P., Zacklad M., Charlet J., « Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique ? », Conférence Ingénierie des Connaissances IC2002, Rouen Mai 2002.
- [3] Cahier J.-P., Zacklad M., « Towards a Knowledge-Based Marketplace model (KBM) for cooperation between agents » in Blay-Fornarino M., Pinna-Dery A., Schmidt K., Zarate P. (Eds), Cooperative Systems

Design, A challenge of the mobility age, Actes de la conférence COOP'2002, St Raphael, 4 - 7 juin 2002, IOS, Volume 74 p. 226-238.

[4] Zacklad M., Caussanel J., Cahier J.-P., « Proposition d'un méta-modèle basé sur les Topic Map pour la structuration et la recherche d'information », journées scientifiques Web sémantique - CNRE-AS2W, Paris-10 et 11 octobre 2002.

Projet Langue et Dialogue - UMR 7503 - LORIA

Campus Sciences,
Bâtiment LORIA BP 239
543506 Vandoeuvre-lès-Nancy
www.loria.fr/equipes/led/index.html

Personne à contacter

Marie-Dominique Devignes
Tél : 03 83 59 20 65
Mél : devignes@loria.fr

Membres de l'équipe concernés par le thème du Web sémantique

Marie-Dominique Devignes (CR CNRS), Malika Smaïl (MCF, Université Henri Poincaré), Nacer Boudjlida (Professeur, Université Henri Poincaré, projet ECOO)

Thème général de l'équipe Langue et Dialogue

L'objectif du projet est de définir des modèles et des techniques permettant de mettre en oeuvre à court, moyen ou long terme des systèmes de dialogue homme machine finalisés reposant sur une forte composante langagière. Nous nous limitons au DHM (Dialogue Homme-machine) finalisé, c'est-à-dire dédié à la commande, à l'apprentissage ou à la recherche d'informations dans un domaine clairement spécifié par exemple en génomique ; cela suppose que l'on peut expliciter complètement l'action - ou plus généralement l'intention - visée par l'utilisateur, de manière à ce qu'elle puisse être réalisée par le système.

Description des travaux en lien avec le thème du dossier

Site web spécifique (en construction) : <http://bioinfo.loria.fr>

Il s'agit d'assister le biologiste lors de la collecte de données sur le génome, en vue d'une meilleure exploitation de la masse d'informations disponibles sur le Web dans les domaines biologiques, pharmaceutiques ou médicaux. Deux champs d'applications ont été explorés : la cartographie des gènes humains en relation avec les maladies génétiques et la recherche dans le génome humain de gènes candidats pour des pathologies complexes (obésité, arthrose...). Cette activité constitue l'une des opérations du PRST Intelligence Logicielle de Lorraine, thème : Bioinformatique et Applications à la génomique.

Peu de systèmes aujourd'hui prennent en compte les problèmes liés à la collecte des données biologiques à partir de sources hétérogènes et réparties (multiplicité des ressources, hétérogénéité et variabilité des formats, mises à jour inégales, redondance des nomenclatures, insuffisance de la documentation sur les données...) ou à leur interprétation (recherche d'informations complémentaires, mise en oeuvre d'outils d'analyse, intégration de l'ensemble des résultats...).

Dans chaque domaine d'application, des partenaires biologistes (CNRS FRE2571 et UHP UPRES 3446) ont été amenés à modéliser le scénario de recherche d'information, jusqu'à ce qu'il soit compréhensible et exploitable par les informaticiens. Ce travail sert de base à la conception de systèmes de recherche d'information dédié à chaque problème biologique posé. Cette phase de conception a été suivie du développement d'une application interactive, qui a fait l'objet d'une expérimentation par les partenaires biologistes. La seconde application nous a amenés à construire une application générique capable de mettre en oeuvre tout scénario décrit selon un modèle général ; ce modèle étant défini sous forme d'une DTD XML. Selon ce modèle, un scénario est composé d'une suite d'étapes ; une étape est caractérisée par des données

d'entrée, l'adresse de la ressource interrogée, la syntaxe de la requête et les données de sortie exprimées sous forme d'expressions régulières caractérisant le passage recherché dans le document retourné par la source. Les données collectées lors de la mise en œuvre d'un scénario particulier sont structurées selon une DTD calquée sur celle d'un scénario.

Ces premières études nous ont permis de nous familiariser avec les caractéristiques des bases de données biologiques utilisées en génomique mais aussi d'aborder le problème de la collecte de données à partir de sources identifiées et répondant à un problème connu. Il est à présent opportun d'aller plus loin en abordant le problème de l'identification des sources pertinentes et de la définition d'un scénario de collecte à partir d'une question biologique quelconque. Pour cela, nous sommes amenés à généraliser notre approche par l'étude et la définition d'une structure de médiation conviviale pour l'inventaire et la navigation à travers les sources de données génomiques hétérogènes et accessibles via le Web. Un aspect intéressant du problème concerne l'intégration de données homologues retournées par plusieurs sources avec gestion des contradictions et des redondances. Cette intégration devra en outre exploiter la qualité/fiabilité relative des données et des sources dont elles sont issues afin de trier ces données et, dans la mesure du possible, aider à l'authentification de certaines données.

Une série de logiciels a été mise au point (bientôt sur bioinfo.loria.fr) :

- Xmap_INTERACTIVE : Assistant pour l'interrogation des sources génomiques, le filtrage des résultats et l'intégration des données collectées. Vise à répondre au problème de la co-localisation sur le génome humain des nouveaux gènes découverts par le séquençage avec les pathologies orphelines (dont on ne connaît pas le gène responsable).
- Xmap_AUTO : Version automatisée de Xmap_INTERACTIVE.
- Xmap_DB : Base des données collectées par Xmap_AUTO (gérée sous le SGBD PostgreSQL); import/export des données au format XML.

- Xmap_SHOW : Outil pour la visualisation des résultats de session Xmap. Convertit le document de session XML en document SVG.
- Xprom : Application basée sur la modélisation générique d'un scénario de recherche d'information et permettant l'exécution de ce scénario. Pour chaque source, les fonctionnalités suivantes sont mises en œuvre successivement : formulation des requêtes, soumission des requêtes, filtrage des données retournées, stockage dans un document XML.

Bibliographie

- [1] Devignes, Marie-Dominique et Smaïl, Malika et Boudjlida, Nacer. *Collecte de données biologiques à partir de sources multiples et hétérogènes. Vers une structure de médiation conviviale et orientée source*. In Journées scientifiques sur le Web sémantique. (Paris, France). 2002. 5 p.
- [2] Devignes, Marie-Dominique et Schaaff, André et Smaïl, Malika. *Collecte et intégration de données biologiques hétérogènes sur le Web : application dans le domaine de la cartographie du génome humain*. Ingénierie des systèmes d'information. 2002. vol 7. n° 1-2. pp.45-61.
- [3] Boudjlida, Nacer and Smaïl-Tabbone, Malika and Devignes, Marie-Dominique. *Services for a Genomics Open Distributed Environment*. In XEWA'2000: IEEE Workshop on XML-Enabled Wide Area Search in Bioinformatics. (League City, Texas). 2000. 2 p.
- [4] Devignes, Marie-Dominique and Schaaff, André and Smaïl, Malika. *Querying Heterogeneous Databases : a user oriented system for collecting and structuring genome information*. In Intelligent Systems for Molecular Biology 1999 (ISMB'99). (Heidelberg, Germany). Poster. 1999. pp.30.

Mission de recherche STIM (AP-HP)

STIM/DPA/DSI/AP-HP
CHU Pitié-Salpêtrière,
91, boulevard de l'Hôpital,
75634 PARIS CEDEX 13.

Contact

01 45 83 67 28
Courriel générique :
diam@biomath.jussieu.fr
URL : <http://www.biomath.jussieu.fr/>

Membres de l'équipe intervenant sur le thème Web sémantique

Jacques Bouaud, Jean Charlet, Brigitte Séroussi, Pierre Zweigenbaum, {jb,jc,bs,pz}@biomath.jussieu.fr

Doctorants : Yun-Chuang Chiao, Natalia Grabar, Pierre Jacquemart, Véronique Malaisé

Objectifs et cadre général de l'équipe

Les activités de la mission de recherche en sciences et technologies de l'information médicale (STIM) reposent historiquement sur des travaux de recherche fondamentale et appliquée avec des objectifs de conception, développement, et réalisation de méthodes et d'outils pour (i) l'aide à la décision médicale, (ii) faciliter l'accès aux connaissances médicales dans un objectif de recherche d'informations, que cette recherche mette en œuvre des techniques de traitement du langage naturel ou que ces connaissances soient représentées dans des formats semi-structurés au sein de documents électroniques, et (iii) la prise en compte des informations médicales (dossier patient, guides de bonnes pratiques, etc.) dans un contexte de gestion des connaissances. Issus de la communauté scientifique de l'intelligence artificielle et engagés dans des projets finalisés mettant en œuvre l'évaluation en pratique des outils d'informatique médicale développés pour servir ces trois objectifs, les membres du STIM sont actuellement impliqués dans des réflexions qui placent le texte, le document et plus généralement les approches documentaires au centre du paradigme d'accès à l'information médicale. Ce choix se justifie parce qu'il permet une meilleure prise en compte du contexte, condition indispensable à l'appropriation et à l'usage effectif par les professionnels de santé de systèmes informatiques à base de connaissances. Le Web sémantique prolonge cette approche documentaire, en particulier et pour nous autour des problématiques d'interopérabilité (services Web) et d'indexation.

Interopérabilité en médecine et services Web

L'interopérabilité en médecine est un

problème toujours pas vraiment résolu : la plupart des logiciels du monde médical développent des formats propriétaires qui les rendent incompatibles et ininteropérables les uns par rapport aux autres. Des propositions du consortium américain *Health Level 7 (HL7)* convergent avec des propositions européennes du CEN TC251 autour de la norme EHRCOM. Pourtant, l'un comme l'autre représentent un saut qualitatif très élevé pour l'industrie du logiciel et une rupture technologique qui risque d'être difficile à absorber. Au sein d'ÉDISANTE, organisme de normalisation des échanges en santé, un groupe de travail propose de réutiliser les efforts du e-commerce et maintenant du Web sémantique pour améliorer l'interopérabilité en santé. Ce travail se développe au sein d'un projet « EDI données cliniques », labellisé par le RNTS. Les premiers résultats obtenus sont une proposition d'enveloppe de messagerie, normant l'envoi et les échanges entre praticiens en laissant libre le contenu, qu'il soit désordonné ou qu'une sémantique de l'échange soit présente – e.g. l'envoi d'une image et de son compte rendu associé. Cette proposition s'appuie en particulier sur les normes ebXML et SOAP (*Simple Object Access Protocol*) en leur associant des caractéristiques propres au milieu médical comme l'envoi d'information d'un praticien vers un autre au sujet d'un patient et permettant la gestion du contenu à l'arrivée [1].

Une telle approche ouvre la voie vers l'*interopérabilité syntaxique*. L'*interopérabilité sémantique*, plus ambitieuse, s'intéresse au contenu des informations échangées. Elle passe par des approches ontologiques seules à même de représenter les contenus échangés. C'est sur le développement de ces ontologies que nous travaillons dans le but de fournir un support conceptuel aux thésaurus médicaux [2,3].

Indexation conceptuelle du Web médical

La sémantisation du Web médical suppose d'associer aux sites Web des descriptions conceptuelles. La majeure partie de ces sites se composent de documents textuels : il s'agit alors d'aller

vers une indexation conceptuelle de ces documents, qui dépasse l'indexation directe en texte intégral effectuée par les moteurs de recherche classiques du Web.

Une particularité du domaine médical est l'existence de nombreux produits terminologiques (thesaurus, classifications, etc.). L'ULMS (Unified Medical Language System [4]) regroupe et interconnecte en 2003 cent de ces terminologies, actuellement essentiellement en anglais, et comprend plus de 800 000 concepts. En attendant l'arrivée de ressources plus ontologiques ou une ontologisation plus forte de l'UMLS, une indexation par les concepts de l'UMLS constitue un moyen terme réaliste. Pour la rendre opératoire pour des documents en français, il reste à renforcer la part de français dans l'UMLS. C'est le thème de deux projets auxquels participe l'équipe.

UMLF (ACI Technologies pour la Santé, 2002-2004, coordonné par P. Zweigenbaum [5]) vise à mettre à disposition un lexique du français médical incluant des variantes morphologiques (flexion et dérivation). VUMeF est une proposition de projet RNTS (coordonnée par S. Darmoni, CHU de Rouen) qui a pour objectif de renforcer la part de terminologies en français dans l'UMLS.

Références

- [1] Charlet J, Cordonnier E & Gibaud B. Interopérabilité en médecine : quand le contenu interroge le contenant et l'organisation. *Revue Information – Interaction – Intelligence*, 2 (2). 2002
- [2] Charlet J. L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. Habilitation à diriger des recherches, Université Paris 6, décembre 2002
- [3] Le Moigno S, Charlet J, Bourigault D, Degoulet P & Jaulent MC. Terminology extraction from text to build an ontology in surgical intensive care. In : Actes du AMIA Annual Symposium 2002, San Antonio, Texas. November, 9-13, 2002.
- [4] McCray AT et Nelson SJ. The semantics of the UMLS knowledge sources. *Methods of Information in Medicine* 1995 ; 34 (1/2).
- [5] Zweigenbaum P, Baud R, Burgun A, et al. Towards a unified medical lexicon for French. In : Le Beux P et Baud R, eds, Actes Medical Informatics Europe, Amsterdam. IOS Press, 2003. À paraître.

**L@STICS,
Faculté de médecine de
Rouen - PSI FRE CNRS 2645 -
Université de Rouen
& INSA de Rouen**

L@STICS Laboratoire des Sciences & Technologies de l'Information et de la Communication en Santé, Faculté de médecine de Rouen
22, boulevard Gambetta 76183 Rouen Cedex
<http://www.chu-rouen.fr/L@STICS>

Contact

Stéfan Darmoni
Tel: 02.32.88.88.29
Fax: 02.32.88.88.32
Stefan.Darmoni@chu-rouen.fr

Participants à l'axe « Web Sémantique » du laboratoire

Benoit Thirion, Jean-Philippe Leroy, Magaly Douyère, Lina Soualmia (Thèse 2A), Aurélie Névéal (Thèse 1A)

Travaux

1. Introduction

Etant donnée la quantité d'information disponible sur le Web, celui-ci doit faire face aux problèmes d'exhaustivité et de précision en recherche d'information. Le catalogue CISMef [1] a été développé afin de faciliter l'accès à l'information de santé disponible sur l'Internet selon trois axes (médecine factuelle, ressources pédagogiques, ressources pour les patients et leur famille). La problématique d'aujourd'hui se veut aussi être une recherche d'information « intelligente » dans l'infrastructure du Web sémantique, une extension du Web actuel qui permettrait de rendre interprétable le contenu des ressources par les hommes mais aussi par les machines. Un principe basique du Web sémantique est de décrire les ressources d'information à l'aide de marqueurs exploitables par différents logiciels, par exemple des métadonnées au format RDF [2], pour guider une recherche.

2. Structure du catalogue CISMef

La terminologie CISMef : La terminologie CISMef est composée des mots clés et des qualificatifs du MeSH. Un ensemble de métatermes (spécialités

médicales) et une hiérarchie de types de ressources ont été rajoutés afin d'optimiser la recherche d'information dans le catalogue par une expansion automatique des requêtes dans les arborescences de mots clés, de qualificatifs et de types de ressources, mais également pour permettre une vision plus globale concernant une spécialité, ceci n'étant pas possible au niveau du MeSH.

Les métadonnées CISMéF : La recherche d'information est la première utilité des métadonnées [2]. Elle est aussi utile pour l'interopérabilité (testée avec succès sur la plate-forme d'e-learning du Campus Virtuel d'Archimède). Chaque ressource est décrite par onze des quinze éléments du Dublin Core ainsi que huit autres éléments spécifiques à CISMéF. Pour les ressources pédagogiques : onze éléments de la catégorie « Educational » de l'IEEE 1484 ; pour les ressources de médecine factuelle : les champs définis par CISMéF *indication du niveau de preuve* et la *méthode* utilisée ; pour qualifier la qualité de l'information de santé, le langage HIDDEL du projet européen MedCIRCLE (CISMéF devient tiers de confiance explicite) [3]. Le format de ces métadonnées est passé du langage HTML en 1995, au langage XML en septembre 2000 et depuis décembre 2002 à RDF. RDF est à la base du Web Sémantique pour la représentation des métadonnées dans un contexte de recherche d'information : il améliore les performances des moteurs de recherche par mot clé et la recherche par navigation.

3. Recherche d'Information : thèse L.Soualmia

Nous évaluons à une échelle réelle, en terme de recherche d'information dans le catalogue CISMéF, les améliorations possibles de trois approches issues de domaines différents ainsi que leur complémentarité. *Traitement Automatique du Langage Naturel* : Le MeSH ne comportant ni variations morpho-syntaxiques ni termes synonymes (ceux du MeSH ne correspondent pas aux termes en usage courant par les cybercitoyens), nous proposons de construire une base de connaissances morphologiques de la terminologie et d'étudier les logs (pour déterminer les termes synonymes) pour

une réécriture automatique des requêtes. Ce module s'inscrit dans le projet UMLF 1 (équivalent du Specialist Lexicon de l'UMLS). Bases de Données : Nous proposons de découvrir de nouvelles règles d'association entre termes (à partir de la base de données CISMéF) pour une expansion automatique des requêtes. *Intelligence Artificielle* : Nous proposons de modéliser sous forme de règles sémantiques une partie du réseau sémantique de l'UMLS pour permettre un raisonnement sur le contenu des ressources et d'effectuer des inférences grâce à des outils comme TRIPLE.

4. Indexation Automatique : thèse A.Névéol

A l'heure actuelle, la recherche de nouvelles ressources à intégrer dans CISMéF et l'indexation de ces ressources sont des tâches effectuées manuellement par l'équipe de documentalistes de CISMéF. Nous proposons d'implémenter un système de veille et d'indexation semi-automatique qui, supervisé par les documentalistes, permettrait d'élargir la couverture du catalogue CISMéF tout en maintenant la qualité de l'indexation. Nous nous appuyons pour cela sur des méthodes relevant de plusieurs domaines: *Traitement Automatique du Langage Naturel* : Utilisation d'automates dictionnaires pour la reconnaissance des mots clés et qualificatifs MeSH et de leurs synonymes. *Apprentissage* : Détermination de règles d'association à partir du corpus indexé des 11.000 ressources CISMéF à notre disposition.

5. Web Services

L'équipe CISMéF a développé un Web Service permettant l'intégration des métadonnées extraites de CISMéF par une requête quelconque afin d'être intégré avec n'importe quelle application (site Web, dossier électronique du patient, méta-catalogue).

6. Perspectives

Dans les années à venir, l'équipe souhaite travailler sur la navigation sémantique, plus riche que la navigation hiérarchique actuelle, ainsi que sur la création d'un méta-catalogue en partenariat avec le LERTIM de Marseille. Ces perspectives s'inscrivent

au sein du consortium VUMéF, récemment labellisé par le Ministère de la Recherche.

Bibliographie

- [1] LF. Soualmia, A. Névéol, B. Dahamna, M. Douyère, JP. Leroy, B. Thirion, SJ. Darmoni. CISMéF dans l'infrastructure du Web sémantique. Première Journée du Web sémantique médical, Rennes, mars 2003.
- [2] B. Thirion, G. Loosli, M. Douyère, SJ. Darmoni. Metadata element set in a Quality-Controlled Subject Gateway: a step to an health semantic Web. MIE 2003 (in press).
- [3] MA. Mayer, SJ. DARMONI, M. Fiene, C. Köhler, T. Roth-Berghofer, G. Eysenbach. MedCIRCLE - Modeling a Collaboration for Internet Rating, Certification, Labeling and Evaluation of Health Information on the Semantic World-Wide-Web. MIE 2003 (in press).
- [4] SJ. Darmoni, B. Thirion, JP. Leroy, M. Douyère, B. Lacoste, C. Godard, I. Rigolle, M. Brisou, S. Videau, E. Goupy, J. Piot, M. Quéré, S. Ouazir, H. Abdulrab H. A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the Internet. Medical Informatics & The Internet in Medicine 2001; 26(3):165-178.

LIM - Laboratoire d'Informatique Médicale

Faculté de Médecine - Université de Rennes I
Avenue du Prt Léon Bernard
35043 Rennes
<http://www.med.univ-rennes1.fr>

Tel : 02 99 28 42 15

Fax : 02 99 28 41 60

Membres de l'équipe concernés par le thème du Web sémantique

Membres permanents : Anita Burgun, Christine Golbreich, Pierre Le Beux, Franck Le Duff

Ingénieurs : Denis Delamarre, Nicolas Garcelon

Doctorants : Marc Cuggia, Gwenaëlle Marquet, Vincent Mary, Fleur Mougin

Thème général de l'équipe

« La modélisation symbolique et représentation des connaissances en

médecine », thème divisé en cinq thématiques :

- Extraction des connaissances,
- Construction d'ontologie,
- Représentation symbolique de l'image,
- Conception de méthodes et d'outils pour l'indexation et la recherche d'information,
- Intégration des modèles.

Description des travaux ou projets

Notre équipe mène depuis longtemps une activité de recherche sur la représentation des connaissances en médecine et des ontologies (1). En partant des bases de connaissances pour l'aide à la décision médicale et la pédagogie, nous avons développé des outils d'accès à cette connaissance notamment pour le projet UMVF (Université Médicale Virtuelle Francophone), qui est piloté par P. Le Beux. Dans ce contexte, il apparaît important de renforcer cet axe de recherche sur les outils d'indexation automatique dans le cadre du Web Sémantique (2), et d'étendre cette approche aux outils d'interrogation et de manipulation de ressources médicales et tout particulièrement de génomique (3).

1. Ontologies et modèles symboliques

La construction d'ontologies en médecine est un point central pour un Web Sémantique Médical. Les méthodes sur lesquelles nous nous appuyons associent une approche formelle et une approche « à partir des données » [1]. L'approche formelle apporte les principes permettant de construire des ontologies de haut niveau bien formées et améliore incontestablement les représentations conceptuelles « intuitives » ou pragmatiques d'un domaine. Par exemple, elle permet de distinguer clairement les rôles (par exemple donneur d'organe, transporteur de fer) des autres concepts (par exemple personne, protéine) et de représenter l'anatomie en utilisant un modèle théorique des relations de proximité spatiale. Elles permettent de bâtir des systèmes logiquement cohérents à partir desquels il est possible de raisonner. Par exemple, nous collaborons à un projet de représentation de l'anatomie cérébrale avec l'UPRES EA 2232 (B. Gibaud et O.

Dameron), l'objectif étant d'associer un modèle symbolique aux images du cortex. Ce modèle comprend un modèle général et un modèle latéralisé et un travail théorique sur les dépendances entre relations méréologiques est en cours. Mais la taille du domaine biomédical incite à compléter cette approche par une méthode plus pragmatique « à partir des données » ou bottom-up permet d'acquérir des ontologies à partir de l'analyse des corpus de textes, des terminologies existantes ou des données stockées dans les entrepôts. Par exemple, l'UMLS (Unified Medical Language System) est créé par mapping lexical de sources terminologiques existantes. Dans le cadre de notre participation au projet Medical Ontology Research (MOR) de la National Library of Medicine (NLM) aux États-Unis, nous avons collaboré au développement et à l'évaluation d'outils linguistiques pour assurer la cohérence des ontologies médicales. Par exemple, des techniques linguistiques sont développées pour vérifier les liens taxonomiques entre un terme et un terme modifié (par exemple arthrite et arthrite aiguë) et pour incorporer à une ontologie préexistante des termes modifiés acquis à partir d'un corpus de textes. Un travail de recherche avec T.C. Rindfleisch et O. Bodenreider [2] a permis de proposer une méthode basée sur les modificateurs antonymes pour analyser la cohérence et la structure d'ontologies médicales existantes, le principe étant que deux termes x_1 et x_2 ne différant que par des modificateurs antonymes doivent être deux concepts frères opposés dans la structure taxonomique. Cette méthode a été appliquée à l'UMLS. Par ailleurs, des méthodes visant à enrichir les terminologies à partir de corpus de documents semi-structurés et non structurés sont en cours de développement et d'évaluation.

2. Accès à l'information médicale pédagogique

L'accès à l'information est un enjeu majeur pour les projets de type Université Virtuelle ou les sites à l'intention des patients et du grand public. Ceci est particulièrement nécessaire dans le domaine biomédical qui dispose d'un vocabulaire très étendu et qui incorpore

la plupart des termes et langages scientifiques d'autres disciplines et également une grande partie du langage courant. Dans le cadre des projets UMVF et du campus numérique d'imagerie médicale, nous avons décrit un standard des ressources pédagogiques en XML au travers de notices de description de ces ressources [3]. En complément de ce travail, une notice d'indexation plus détaillée utilisant une extension du Dublin Core a été définie à partir de la norme IEEE 1484. Dans le cas de système d'indexation de type CISMef¹, les notices sont remplies par des documentalistes humains et il est bien évident que l'augmentation du volume des ressources ne permettra plus de les indexer toutes, même en se limitant au français. L'annotation automatique est incontournable dans ce cadre. C'est au moyen des technologies du Web Sémantique que nous pourrions réaliser ce projet. Le standard des ressources de l'UMVF utilisera RDF(S) qui est plus adapté pour la description des métadonnées et dont les spécifications incluent une ontologie « légère », peu contrainte sur le plan formel, pour supporter les échanges et le partage de connaissances sur le Web. De plus, nous disposons de Nomindex [4], un outil d'indexation conçu et développé dans notre laboratoire, qui permet de reconnaître les concepts dans les textes médicaux, ce qui impose de résoudre les problèmes posés par le traitement automatique du langage naturel. Il nous permettrait pour ce projet d'extraire automatiquement les informations sémantiques des ressources documentaires et pédagogiques de l'UMVF.

3. Conception d'un modèle Web sémantique appliqué à la génomique fonctionnelle

La recherche dans le domaine de la génomique fonctionnelle nécessite d'accéder à une multitude de bases de données et de connaissances accessibles via le Web, mais hétérogènes dans leur structure et leur terminologie. Aujourd'hui, le problème qui est posé est celui d'une collecte manuelle d'informations, qui ne peut être que partielle ou incroyablement fastidieuse, étant donné le nombre de sources disponibles et leur diversité, ver-

1. <http://www.chu-rouen.fr/cismef/>

sus une collecte d'informations automatisée « intelligente ». L'intérêt du Web sémantique en génomique fonctionnelle est d'apporter suffisamment de renseignements sur les ressources au moyen des métadonnées, de décrire leur contenu de manière à la fois formelle et signifiante via des ontologies de domaine, de telle sorte que des agents intelligents de recherche sur le Web puissent sélectionner les informations pertinentes pour une question donnée.

Notre but est de réaliser une plateforme de connaissances multi-sources en génomique fonctionnelle. Pour cela, nous avons développé le système BioMeKe [5] (BioMedical Knowledge Extraction réalisé au LIM) qui intègre, à ce jour, des ontologies (l'Unified Medical Language System®, Gene-Ontology™ et GO Annotation) et d'autres bases de connaissance (HUGO, MeSH Supplementary Terms). Cet outil sera combiné à un entrepôt de données constitué à l'U522 de l'INSERM pour l'analyse du transcriptome hépatique (GEDAW).

Notre méthode se basera alors sur une approche de type médiation et sur l'ontologie de domaine de BioMeKe. Elle permettra de faire l'interface entre le biologiste et les sources à interroger où sont stockées les données. Au moyen d'un tel système, les biologistes pourront effectuer une requête unique et obtenir un résultat global calculé de façon transparente par combinaison des réponses des différentes sources de données. Le biologiste aura ainsi l'illusion d'interroger un système homogène et centralisé.

Bibliographie

<http://www.med.univ-rennes1.fr/cgi-bin/limpublic/limpublic.pl?action=annuaire>

- [1] Burgun, A., et al., *Issues in the design of medical ontologies used for knowledge sharing*. J Med Syst, 2001. 25(2): p. 95-108.
- [2] Bodenreider, O., Burgun A., Rindflesch T.C., *Assessing the consistency of a biomedical terminology through lexical knowledge*. Int J Med Inf, 2002. 67(1-3): p. 85-95.
- [3] Mouglin F., Cuggia M., Le Beux P., *Development of an indexing search engine for the UMVF : Proposal for an indexing method based on Dublin Core and XML*. Accepté à Medical Informatics Europe 2003.
- [4] Pouliquen B., Delamarre D., Le Beux P.,

Indexation de textes médicaux par extraction de concepts, et ses utilisations. 6th International Conference on the Statistical Analysis of Textual Data, JADT'2002. St. Malo, France, March 13-15 2002.

[5] Marquet G., Burgun A., Moussouni F., Guérin E., Loreal O., *An integrative approach of biomedical knowledge via ontologies for liver transcriptome analysis*. Workshop on Ontology in biology, Heidelberg, Nov 7-8 2002

LSI - Laboratoire des Systèmes Intelligents – Université de Laval

Le Laboratoire des Systèmes Intelligents (LSI) du Département d'Informatique et de Génie Logiciel de l'Université Laval, dirigé par le professeur Nicole Tourigny, a été constitué récemment en réunissant des équipes de recherche existantes, qui pourront ainsi allier leurs expertises pour travailler sur des projets communs. Ces projets gravitent autour de plusieurs axes de recherche, en particulier sur l'Intelligence Artificielle et le Web sémantique.

Personne à contacter

Dr Koné Mamadou Tadiou, professeur au LSI, coordonnateur du thème
Département d'informatique et de génie logiciel
Local 3908, Pavillon Adrien-Pouliot
Université Laval Sainte-Foy,
Canada G1K 7P4
Téléphone : (418) 656-2131, poste 4984
Télécopie : (418) 656-2324
Courriel :
Kone.Mamadou@ift.ulaval.ca

Membres du laboratoire concernés par le thème du Web sémantique

- 1) Nicole Tourigny et Laurence Capus : Équipe de Recherche en Ingénierie des ConnAissancEs (ERICAE), résumé, ontologies, graphes conceptuels.
- 2) Guy Mineau : Équipe Intelligence Computationnelle (IC), théorie des graphes conceptuels, apprentissage automatique.

- 3) Bernard Moulin : Équipe Informatique Cognitive et Géomatique.
- 4) Mamadou Tadiou Koné: Web Sémantique, Services Web et Langage de Communication des Agents.
- 5) Brahim Chaib-Draa : Équipe Distributed Agents and Multi-Agent Systems (DAMAS).
- 6) Ruddy Lelouche : Équipe Acquisition et Modélisation Interactives de Connaissances pour l'Apprentissage (AMICA), Interfaces intelligentes, Environnements d'apprentissage.

Thème général du laboratoire

Le LSI s'intéresse à l'ingénierie des connaissances, la théorie des graphes conceptuels, les apprentissages humain et automatique, les interfaces intelligentes, les systèmes multi-agents et langages de communication des agents et la géomatique.

Description des travaux

Au sein du Laboratoire des Systèmes Intelligents, plusieurs thèmes de recherche complémentaires impliquant le Web sémantique et les Services Web existent.

Ainsi, l'ERICAE propose dans le cadre du Web sémantique une approche de résumé automatique de pages Web basée sur l'extraction de phrases afin d'évaluer et d'améliorer les résumés courants. De plus, un nouveau projet sur l'utilisation de ponts ontologiques vient d'être mis en place pour permettre la fusion de plusieurs applications Web d'une part, et pour rendre accessibles des données, des informations ou des connaissances sans avoir à modifier les applications existantes d'autre part. Ainsi, l'accessibilité devient possible à partir de n'importe laquelle de ces applications.

De son côté, pour résoudre le problème épineux de l'indexation pour les moteurs de recherche actuels, l'équipe Informatique Cognitive et Géomatique a conçu un système capable de mettre en correspondance les structures conceptuelles extraites de la requête de l'utilisateur et celles des documents trouvés

par un moteur de recherche général comme Google. Ainsi ce système présente une nouvelle architecture pour des expériences plus étendues dans la manipulation des structures conceptuelles extraites de documents en langage naturel. De plus, les applications géo-référencées avec la multitude de PDA et de téléphones cellulaires en circulation seront les premières bénéficiaires des résultats de cette recherche.

Pour sa part, l'équipe d'Intelligence Computationnelle s'intéresse à la théorie de l'apprentissage automatique par les agents logiciels avec le regroupement conceptuel, l'abstraction et le résumé de données (techniques de Knowledge Discovery et Data Mining). Dans ce contexte, les serveurs de connaissances seront accessibles sur le Web grâce à la modélisation des agents logiciels par graphes conceptuels.

Quant à elle, l'équipe AMICA compte mettre en commun les travaux des deux équipes précédentes avec les techniques et approches utilisés dans les systèmes éducatifs en vue de s'ouvrir davantage à l'éducation mobile et géographiquement située. Cela consiste naturellement à utiliser les outils du Web sémantique à des fins spécifiques d'apprentissage, pour adapter les interactions offertes à un apprenant à ses besoins spécifiques. Il s'agit donc d'abord d'utiliser adéquatement les connaissances du système de formation (ou des sites en tenant lieu) et du modèle de l'apprenant stockés sur le Web : présentation des connaissances, articulation des raisonnements et des simulations aux fins de l'apprentissage. Mais il s'agit aussi d'adapter les interfaces et interactions offertes à l'apprenant en déplacement aux possibilités de l'équipement dont il dispose ponctuellement: interactions multimodales, multilingues, prise en compte des contraintes imposées par la surface d'affichage.

Enfin, dans le contexte du Web sémantique, le coordonnateur du thème projette, dans son programme de recherche à long terme, de créer un nouveau paradigme pour les interactions, la coopération et la coordination entre les agents logiciels. Dans sa phase initiale

en cours, ce programme se penche sur la nature, les propriétés et les composantes de la communication des agents et poursuit trois objectifs complémentaires:

- construire une architecture pour l'interaction entre agent et service utilisant des outils du Web sémantique;
- redéfinir la coopération et la coordination entre agents en tant que représentants de services Web;
- établir une plate-forme d'interaction entre l'humain et l'agent logiciel reposant sur le profil du premier et une ontologie descriptive du second.

D'un côté, l'objectif de ce programme est de mettre à profit les modes flexibles de coopération et de coordination utilisés dans la recherche sur les systèmes multi-agents plutôt que les protocoles d'interaction rigides qui sous-tendent le Web sémantique. D'un autre côté, il s'agit de déterminer jusqu'à quel niveau un agent doit contenir des informations sur son propriétaire pour être un substitut valable. Cette recherche en tant que valeur ajoutée pour le projet bien connu de Gouvernement en Ligne (eGovernment) du Canada, vise à créer un accès fluide et intégré aux informations sur les services fédéraux.

Bibliographie

- [1] M.-A. Nolin, N. Tourigny et L. Capus. Portail de bibliothèques utilisant le Web sémantique, *Congrès de l'Association Francophone pour le Savoir (ACFAS)*, Rimouski (Québec), Canada, , 19-23 mai 2003 (soumis).
- [2] D. Nadeau et N. Tourigny. Évaluation d'un système pour le résumé automatique de documents électroniques, *Advances in Artificial Intelligence, AI2001-14th Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, 7-9 juin 2001, LNAI 2056, Springer-Verlag, 277-286.
- [3] S. Nicolas, G. W. Mineau, B. Moulin. Extracting conceptual structures from English texts using a lexical ontology and grammatical parser, G. Angelova, D. Corbett, U. Priss (eds.), *Supplementary Proceedings, 10th International Conference on Conceptual Structures ICCS'02*, Borovets, Bulgaria, 2002, pp.15-28.
- [4] Kone Tadiou, Mamadou, Shimazu A., and Nakajima T. The State of the Art in Agent Communication Languages, Knowledge and Information Systems Journal, 2(3), *Springer (Berlin, Germany)*, 2000, pp. 1-26.

« Cognition & Usages » - Laboratoire FRE 2627 Paris

Université Paris VIII-St Denis,
2 rue de la Liberté
F93526 St Denis cedex 02
<http://www.ipt.univ-paris8.fr/~psycog/>
Contact : Charles Tijus,
Tel :33 1 49 40 64 79,
tijus@univ-paris8.fr

Laboratoire d'Informatique de Paris 6
(LIP6), UMR-CNRS 7606
SYSDEF (Systèmes d'Aide à la
Décision et à la Formation)
Université Paris 6, CNRS
8 rue du Capitaine Scott
75015 Paris
<http://www-poleia.lip6.fr/~brezil>
Contact : Patrick Brézillon,
Tel : 33 1 44 27 70 08,
Patrick.Brezillon@lip6.fr

Membres

Charles Tijus, Patrick Yeu, Sébastien Poitrenaud, Elizabeth Hamilton, Christine Leproux, Patrick Brézillon, Arnaud Santolini.

Thèmes de l'équipe

Catégorisation, Contexte, Sémantique, Apprentissage

Description des travaux ou projets en lien avec le thème du dossier Web sémantique

Formaliser le contexte d'utilisation (ou d'apprentissage) à partir des graphes contextuels (Brézillon, 2002) et de la catégorisation contextuelle (Tijus, 2001) sous la forme d'ontologies qui servent également à l'indexation comme méta-données pour le Web sémantique.

Modèles sémantiques pour le Web : accès à la signification pédagogique

Le W3C a mis au point un dispositif très cohérent composé d'un langage personnalisable, le XML (eXtensible Markup Language) et la possibilité d'ajouter des données sur la donnée avec le RDF (Resource Description Framework), un identificateur unique (URI : Uniform Resource Identifier) et

des Ontologies. Cet ensemble permet d'établir des classes d'appartenance de générer des hiérarchies et d'articuler l'ensemble avec les ontologies. Des applications fonctionnent déjà, par exemple Protégé pour des ontologies personnelles, ou encore Wordnet à partir des travaux de Miller & Fellbaum (1995). D'autres sont en développement à partir de LSA (Latent Semantic Analysis, Landauer & Dumais, 1997 ; Kintsch, 2001).

Notons que, fondamentalement, le fonctionnement d'un Web Sémantique (WS) est possible parce que les traitements, les opérations et le raisonnement sont basés sur la catégorisation. C'est en effet la catégorisation qui permet la réalisation des classes et des hiérarchies. La catégorisation désigne (i) la tâche qui consiste à répartir un ensemble d'objets en un certain nombre de groupes, ou selon des dimensions prédéfinies, (ii) le processus de création de nouvelles catégories pour des objets inconnus, le processus d'acquisition des catégories qui peuvent être lexicalisées et (iii) le processus d'affectation catégorielle, ceci dès la reconnaissance d'un objet: « quand on reconnaît un objet, on est en train de le catégoriser ».

Chez l'homme, les traitements cognitifs de compréhension et de raisonnement sont basés sur ces activités de catégorisation (cf. Piaget), comme le sont les traitements liés à l'utilisation. En logique et en mathématiques, les calculs sont basés sur la catégorisation (cf. l'universalité du treillis de Galois). D'où la possibilité d'un SW dans la mesure où la structure fondamentale repose sur les hiérarchies, avec la possibilité de définir une donnée de deux manières : ce qu'elle est (attributs et classes incluant) et comment elle a été obtenues (modes de productions des attributs et des classes incluant).

En termes de processus et d'opérations mentales, les travaux sur la catégorisation sont très nombreux, et ce dans les différents champs de la psychologie, depuis les travaux piagétiens jusqu'aux catégories ad hoc de Barsalou.

En termes de représentation des connaissances, les travaux de Rosch sur les catégories sémantiques et de Keil sur les ontologies se poursuivent en psychologie du développement, mais se retrou-

vent surtout développés dans le domaine de la philosophie de l'esprit et en intelligence artificielle, plus récemment dans les recherches sur le Web sémantique.

Un SW basé sur les opérations logiques des classes, - et dans les cas d'incertitude, avec des prises de décision fondées sur les probabilités (théories bayésiennes) ou sur les possibilités (logique floue)-, et sur des ontologies préfixées (WordNet, Sensus) ou dynamiques (Protégé), permet de fournir aux données un statut de connaissances. Parce que ce qui définit une connaissance, c'est sa mise en relation avec d'autres données. La constitution d'une connaissance par la mise en relation peut ainsi être très diverse : ontologie, partage d'attributs avec le treillis de Galois, nombre d'occurrences conjointes comme pour LSA, etc. Il s'agit toujours d'un découpage en catégories de l'ensemble des données à partir desquelles des traitements sont possibles, pour fournir par exemple une réponse à une requête, voire même pour évaluer la cohérence d'un texte.

Outre la catégorisation, la prise en compte du contexte nous semble primordiale pour l'accès à la signification, que ce soit en termes de concepts, de relations ou des deux. Une idée (W3C, communication personnelle) est la prise en compte du contexte de l'utilisation, mais sans qu'on sache encore comment s'y prendre. Une proposition que nous avons déjà faite est de représenter le contexte d'utilisation par un réseau de catégories contextuelles, une ontologie individuelle située. Notre proposition présente est qu'il n'est pas suffisant pour un SW performant de s'attacher seulement au contexte de conception (relations entre données, voire même aux relations entre traitement et données) : le SW doit prendre en considération l'utilisateur, en l'occurrence, l'élève, voire l'enseignant, et son contexte, et surtout le contexte d'utilisation. C'est l'objet de nos recherches en considérant deux modèles, celui des graphes contextuels et celui de la catégorisation contextuelle. Pour soumettre à l'épreuve, d'une manière opérationnelle, l'idée que la sémantique est basée sur le contexte, nous sommes en train de réaliser un modèle informatique de l'interaction homme-machine sous la forme

d'une maquette interactive d'accès aux bases de connaissances. L'interaction comprend trois phases : une phase de contextualisation de la requête initiale qui consiste à concaténer les branches « utilisateur » et « tâche » au contenu de la requête. La seconde phase consiste à retrouver le domaine concerné par la requête et son ontologie puis à réaliser le réseau de catégories qui a la forme d'un treillis de Galois à partir de l'ontologie et du contenu de la requête reformulée. La troisième phase consiste à dialoguer avec l'utilisateur en parcourant le treillis du général au spécifique.

Bibliographie

Brézillon P., <http://www-poleia.lip6.fr/~brezil/Pages2/Publications/index.html>

Tijus C., <http://www.ipt.univ-paris8.fr/~psycog/>

[1] Brézillon P, Cavalcanti M., Naveiro R. & Pomerol J-Ch (2000) SART: An intelligent assistant for subway control. *Pesquisa Operacional, Brazilian Operations Research Society*, 20(2): 247-268.

[2] Brézillon P. (2002) Expliciter le contexte dans les objets communicants. In: G. Poulain, G. Privat, C. Kintzig (Eds.): *Les Objets Communicants*. France Telecom Editions. (to appear).

[3] Pudelko, B., Hamilton, E., Legros, D., & Tijus, C.A. (1999). How Context contributes to Metaphor Understanding. In J.G. Carbonell & J. Siekmann (eds), *Lectures Notes in Artificial Intelligence, vol. 1688, Modeling and Using Context*, (pp. 511-514). New-York: Springer

[4] Tijus, C. A. (2001). Contextual Categorization and Cognitive Phenomena, in V. Akman, P. Bouquet, R. Thomason, & R. A. Young, *Modeling and Using Context*. Berlin: Springer-Verlag, pp. 316-329

[5] Tijus, C., & Yeu, P. (2002). Web semantic models : accessing users meanings. *First Human Factors & the WEB meeting*, Nice, 16 et 17 mars 2002 (CDRom).

Groupe « Sûreté et Cognition » - INCOD – LISIS Marseille

Le groupe de recherche « Sûreté et Cognition » est une composante de l'équipe de recherche Ingénierie des Connaissances Distribuées (INCOD) du

Laboratoire des Sciences de l'Information et des Systèmes (LSIS – UMR n° 6168)

Avenue Escadrille Normandie-Niemen
13397 Marseille cedex 20
www.lsis.org

Membres

- Permanents : Chouraqui Eugène (Professeur) eugene.chouraqui@lsis.org 04 91 05 60 12, Faucher Colette (Maître de Conférences) colette.faucher@lsis.org 04 91 05 60 58, Mercantini Jean-Marc (Maître de Conférences) jean-marc.mercantini@lsis.org 04 91 05 60 15, Santoni Charles (Maître de Conférences) charles.santoni@lsis.org 04 91 05 60 14
- Doctorants : Koné Séga, Sadoc Sophie, Guerrero Claudia, Scaico Alexandre
- Post-Docs : Loschmann Ronald (ATER)

Personne à contacter

E. Chouraqui

Thème général du groupe

Les travaux de recherche, que nous conduisons actuellement se répartissent selon deux axes interdépendants : le premier, qui adopte une démarche déductive, porte sur l'étude cognitive et formelle des raisonnements pouvant être mobilisés lors de la résolution de problèmes relatifs à la sûreté des organisations (causalité, classification, analogie, raisonnement à base de cas, formalisation des écarts, ...); le second, qui adopte une démarche inductive porte sur l'étude de la sûreté des organisations mobilisant les sciences cognitives, en relation avec des domaines expérimentaux (pilotage d'avions, conduite automobile, réaction des opérateurs confrontés à des situations d'urgence, sécurité de sites industriels urbains, ..).

Les personnes du groupe, concernées par le thème du Web sémantique sont les quatre permanents ainsi qu'un doctorant en première année de thèse. Le regroupement des compétences autour de ce thème a été déclenché par la réalisation d'un projet industriel (projet n°1) et se poursuit aujourd'hui par le projet n° 3 (voir ci-dessous). Nous abordons le

thème du Web sémantique depuis octobre 2001. Nous préparons actuellement nos premières publications.

Description des projets en lien avec le thème du Web sémantique

Projet n°1 : Etude d'un portail dans le domaine de la gestion du trafic aérien

Les travaux que nous avons réalisés s'inscrivent dans le cadre d'un projet de conception et de réalisation d'un « portail veille ATM » (Air Trafic Management) mis à la disposition des usagers intéressés par les informations et les connaissances relevant de cette thématique. Ce projet a été développé en collaboration avec la société TRANS-SIM (Aix-en-Provence, Les Miles). Vue la diversité du domaine, nous avons retenu comme définition partielle de l'ATM, *la gestion du transport des passagers pour les voyages de tourisme et d'affaires*.

Un tel portail doit donc en premier lieu jouer le rôle d'une interface « intelligente » entre ses usagers et les sites du réseau Internet dont les contenus sont pertinents eu égard à ce domaine de connaissance. L'expression première et obligée de cette intelligence réside dans la nécessité d'élaborer un vocabulaire spécialisé qui soit partagé par tous les usagers du portail et à partir duquel les sites seront accessibles. Pour ce faire, nous avons construit une ontologie d'application relevant de la gestion des passagers pour les voyages de tourisme et d'affaires, sous thème circonscrit à un corpus homogène des sites WEB proposés par les aéroports français.

La démarche utilisée est fondée sur une « extraction manuelle » de termes pertinents à propos desquels nous avons défini et avons étudié leurs relations sémantiques (synonymie, polysémies, relation de générique à spécifique ...). Cette analyse sémantique a permis de fournir un ensemble organisé de termes désambiguïsés à partir duquel l'ontologie a été construite. Cette ontologie est organisée sous la forme d'un graphe borné supérieurement dont la racine engendre les deux classes d'activité des usagers au sein d'une zone aéroportuaire : les activités d'accès et les activités de service.

Par ailleurs nous avons conçu et déve-

loppé une maquette du portail dont la conception repose sur la mise en relation des termes de l'ontologie avec les pages des sites « Web » concernés. Cette maquette a permis en particulier de valider l'ontologie. Ce projet a fait l'objet d'un DEA.

Projet n°2 : Filtrage des réponses fournies par les moteurs de recherche en vue de contribuer à éliminer le bruit.

L'un des problèmes majeurs qui a été et qui est encore reproché aux moteurs de recherche est la quantité importante de résultats peu ou pas pertinents qu'ils retournent à l'utilisateur. L'application de filtres sur ces résultats, de manière à éliminer ceux qui ne sont pas pertinents, est un des moyens d'aide à l'utilisateur.

Ce projet a précisément pour objet de concevoir et de réaliser un filtre des réponses fournies par un moteur de recherche en vue d'éliminer le bruit. Cette étude a permis d'élaborer, grâce à l'analyse syntaxico-sémantique des résultats aux questions posées par les utilisateurs, les critères de sélection permettant d'une part de rejeter les documents non pertinents et d'autre part de classer les documents retenus. Ces critères sont les suivants : synonymes, repérage des syntagmes nominaux, noms propres, extension de domaines de pays, respect de l'ordre des termes dans les réponses et enfin liste de mots vides.

Un prototype a été implémenté en Java et a été exploité dans le cadre des réponses fournies par le moteur Google, le domaine de connaissance choisi étant l'ATM pour lequel nous disposons d'une ontologie (voir projet N°1). Ce prototype a permis de valider la méthode de filtrage que nous avons élaborée à travers plusieurs dizaines de questions / réponses. Cette validation a mis en évidence la nécessité de prendre en compte le problème de la polysémie qui avait été volontairement écarté pour des questions de temps de réponses. Ce projet a fait l'objet de la soutenance d'un DEA.

Projet n°3 : Elaboration d'un portail coopératif relatif à la sécurité des organisations, fondé sur la construction et la mise à jour automatique d'ontologies.

Ces travaux de recherche sont la

continuité de ceux qui ont été développés au cours du projet n°1 et font l'objet d'une thèse de doctorat. L'objectif de ce projet est la conception et la réalisation d'un « système portail interactif » dans le domaine de la sécurité des organisations, mis à la disposition des acteurs relevant de ce domaine. Ce portail devant fournir le plus rapidement possible les connaissances pertinentes contribuant à la résolution des problèmes de sécurité auxquels sont confrontés ces acteurs, et ce à partir d'un point d'entrée unique et bien identifié sur le Web. Les usagers d'un tel système sont tous les responsables (les acteurs) qui ont en charge la gestion de la sécurité des organisations à tous les niveaux : administratif, réglementaire, scientifique, intervention, formation, prévention, ...

Ces travaux se situent à la frontière de trois disciplines : l'informatique, les sciences cognitives et la sécurité des organisations. Les apports essentiels de ce travail de recherche se situent au niveau des quatre points suivants :

1. La construction, l'intégration et la mise à jour automatique des ontologies à partir des connaissances disponibles sur l'internet.
2. La conception d'une interface intelligente permettant d'accéder rapidement au savoir et savoir-faire disponible sur le réseau internet.
3. La contribution à l'élaboration d'une théorie dans le domaine de la sécurité.
4. L'aide à la résolution de problèmes auxquels sont confrontés les experts du domaine.

La démarche adoptée pour conduire ce projet est la suivante :

1. Etude d'un inventaire raisonné des travaux relevant de cette thématique dans les trois domaines qui la concerne.
2. Accès et analyse des sites Web dont les contenus relèvent de la sécurité des organisations.
3. A l'issue de cette analyse, élaboration d'un langage terminologique
4. Elaboration de critères de sélection d'un domaine de compétence particulier, fondé sur l'état de l'art et le langage terminologique.
5. Construction d'une ontologie relative

à ce domaine de compétence et établissement des fondements d'une théorie relative à ce domaine de compétence.

6. Elaboration d'une méthode automatique d'aide à la construction et à l'intégration d'ontologies fondée sur le point précédent.
7. Elaboration d'un portail thématique coopératif intégrant la méthode de construction et le comportement cognitif des utilisateurs.
8. Développement d'une maquette et validation.

L'état d'avancement de ce projet concerne les trois premiers points de notre démarche.

Bibliographie

- [1] Faucher C. (2001) Approximate knowledge representation and management in a frame-based language : the system CAIN, *International Journal of Computer Research, Published by NOVA Publisher's New York*.
- [2] Mercantini J.M., Loschmann R., Chouraqui E. (2000) A safety Previsional analysis method of an Urban Industrial Site, *Scientific Journal of the Finnish Institute of Occupational Health, Série : People and work. Safety in the modern Society* - (Research Report 33) , pp. 105-109.
- [3] Caussanel J., Chouraqui E. (1999) Information et Connaissances : quelles implications pour les projets de Capitalisation des Connaissances , *Revue Document numérique - Gestion des documents et Gestion des connaissances*. 3(numéro spécial 3-4) ; pp. 101-119.
- [4] Mercantini J.M., Chouraqui E., Bélanger N. (1999) Etude d'un système d'aide au diagnostic des accidents de la circulation routière. In *IC'99 Ingénierie des Connaissances 99*, Palaiseau, France.
- [5] Ch. Santoni, E. Furtado, Ph. François. (1995) *Adaptive human computer interfaces for supervision systems*. HCI International'95, 6th International Conference on Human-Computer Interaction, Yokohama, Japon, 9-14 juillet 1995, Edité par Y. Anzai, K. Ogawa et H. Mori, « Symbiosis of Human and Interfact : Future Computing and Design for Human-Computer Interaction ». ELSEVIER Science, Vol.1, pp 1077-1082, ISBN : 0 444 81795 6

DoC - Document et Connaissance

UMR CNRS 6599 Heudiasyc
BP 20529
60205 Compiègne cedex
www.hds.utc.fr/~ptrigano/doc

Personne à contacter

Philippe Trigano
philippe.trigano@utc.fr
Tel : 03 44 23 45 02

Membres de l'équipe concernés par le thème du Web sémantique

Abel Marie-Hélène, Bachimont Bruno, Barthès Jean-Paul, Fontaine Dominique, Lenne Dominique, Morizet Pierre, Moulin Claude, Trigano Philippe

Thème général de l'équipe

L'étude des Communications entre l'Homme et la machine nécessite des méthodes de recherche différentes de celles mises en œuvre dans les autres équipes du laboratoire. La relation entre l'utilisateur et l'ordinateur est en effet prépondérante, l'expérimentation prend nécessairement le pas sur la modélisation, les réalisations concrètes de prototypes deviennent essentielles. Cette différence dans les méthodologies de recherche est fondamentale.

- L'étude d'un système intégrant l'Homme rend difficile toute étude abstraite du système complexe. Seules des modélisations partielles sont envisageables, et la validation de l'ensemble nécessite des expérimentations, qui, elles-mêmes, imposent des réalisations : systèmes d'apprentissage assisté (EIAH et e-Learning), systèmes multi-agents, etc.
- L'étude de systèmes complexes impliquant l'Homme relève de l'informatique dite "symbolique". Celle-ci a pour but d'interagir et de s'adapter aux raisonnements humains, en représentant et en manipulant des connaissances.

Un thème autour des documents et des connaissances fait écho à différentes tendances que l'on voit émerger dans les technologies de l'information. On peut citer :

- Le Web sémantique : vocable qui désigne le projet de publier en ligne des documents munis de la représentation formelle de leur signification pour permettre à des applications de s'en saisir et d'y appliquer les traitements adéquats.
- L'indexation multimédia : le numérique permettant aujourd'hui de réunir sur un même support différents média, il devient nécessaire de pouvoir les analyser et d'en croiser les index.
- La médiation intelligente : les réseaux facilitent le travail à distance, mais posent avec acuité les problèmes de l'interculturalité. Comment travailler ensemble quand on ne se connaît pas ? Les agents permettent de reformuler les échanges de manière à renforcer leur acceptabilité culturelle des deux côtés de la communication.

Description des travaux ou projets en lien avec le thème du Web sémantique

Le projet MEMORAe (soutenu par le pôle NTE de la région Picardie), responsable M.H. Abel.

Avec l'émergence des Technologies de l'Information et de la Communication (TIC), une nouvelle forme de formation, ou plus exactement un nouveau mode d'apprentissage, est apparu. Souvent appelé "e-learning", ce mode est basé sur l'accès à des formations en ligne, interactives et parfois personnalisées, diffusées par l'intermédiaire d'un réseau (Internet ou intranet) ou d'un autre média électronique. Cet accès permet de développer les compétences des apprenants, tout en rendant le processus d'apprentissage indépendant du temps et du lieu.

Dans le cadre du projet MEMORAe (MEMOire ORganisationnelle Appliquée au e-learning) [ABE, 02a][ABE, 02b] nous nous intéressons à la problématique du contenu mis à disposition des apprenants. Nous avons choisi de découper et de représenter le contenu pédagogique d'une formation au moyen de grains de connaissance. La gestion de ce contenu pose des problèmes de structuration, d'organisation, d'acquisition, de maintenance, de diffusion des informations et des connais-

sances. Nous proposons de traiter cette problématique au moyen d'une mémoire organisationnelle qui sera mise à la disposition des utilisateurs (apprenants, enseignants, etc.). Le partage de cette mémoire repose, entre autres, sur des mécanismes de consultation de son contenu via l'utilisation d'Internet. Les grains de connaissance représentent les notions à appréhender et servent d'index pour accéder aux ressources (livres, sites Web, etc.) qui traitent de ces notions. De ce fait on se retrouve dans une problématique Web Sémantique.

Le projet MEMORAe est donc situé au carrefour de trois domaines de recherche : Ingénierie des Connaissances, Ingénierie Éducative et Web Sémantique. Il s'agit tout d'abord d'effectuer un travail pédagogique afin de découper le contenu de la formation en grains de connaissance et d'articuler ces derniers les uns par rapport aux autres (domaine de l'ingénierie éducative) ; ensuite, le choix d'organiser et de gérer un tel contenu au moyen d'une mémoire organisationnelle basée sur des ontologies concerne le domaine de l'ingénierie des connaissances. Enfin, le choix de la norme ISO Topic Maps [IEC, 99] pour structurer les informations afin de présenter l'accès aux ressources de la mémoire organisationnelle mise à disposition des apprenants concerne le domaine du Web Sémantique. Cette norme permet de distinguer un niveau ressources et un niveau de représentation sémantique décrivant, indexant, ces ressources.

Bibliographie

- [1] Abel M.-H., Lenne D., Cissé O. (2002) « E-learning and Knowledge Management: The MEMORAe project » Proceedings of E-Learn 2002, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Montreal, Canada, octobre 2002.
- [2] Abel M.-H., Lenne D., Cissé O. (2002) « e-learning et Web sémantique : le projet MEMORAe » In actes électroniques des journées scientifiques web sémantique, Ivry sur Seine, 10-11 octobre 2002.
- [3] International Organisation for Standardization (ISO), International Electrotechnical Commission (IEC), Topic Map, International Standard ISO/IEC 13250, 19 avril 1999.

Groupe des Ecoles des Télécommunications

Equipe constituée de chercheurs de Paris (ENSTINFRES/III) et de Brest (ENST-Bretagne/IASC)
url : <http://www.get-telecom.fr/>

Personne à contacter

Jean-Marc Saglio
Département Informatique et Réseaux
Ecole Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris cedex 13
<http://www.infres.enst.fr>

Membres de l'équipe concernés par le thème du Web sémantique

Talel Abdessalem, mél : talel.abdessalem@enst.fr, tél : 01 45 81 77 79
<http://www.infres.enst.fr/~talel/>
Philippe Picouet, mél : Philippe.Picouet@enst-bretagne.fr, tél : 02 29 00 13 95
<http://www-iasc-perso.enst-bretagne.fr/~picouet/> et Benjamin Buffereau (doctorant)
Jean-Marc Saglio, mél : saglio@enst.fr, tél : 0 1 45 81 80 62
<http://www.infres.enst.fr/~saglio> et Tuan Anh Ta (doctorant)

Thème général de l'équipe

Notre approche du Web sémantique

Depuis plus de deux ans nous avons adopté, pour nos recherches sur le Web sémantique, l'approche « Web communautaire » (C-Web) en pensant particulièrement soit à des institutions expertes en conservation, classement et présentation au public de ressources numérisées, comme les *bibliothèques*, *médiathèques*, musées et autres centres d'exposition publique, soit à des communautés rapprochant professionnels et amateurs par des échanges mutuels, comme les *clubs et sociétés scientifiques ou culturelles*. Cette approche fait l'hypothèse que la construction et la diffusion d'une représentation commune des connaissances - une « ontologie » - est dans l'intérêt général de la communauté.

Dans ce contexte nous nous sommes particulièrement intéressés à la production, par les professionnels ou même par

les amateurs, de *présentations* faisant non seulement une sélection dans les ressources référencées par la communauté mais les organisant comme des *itinéraires de visite guidée* ou des *plans recommandés de lecture*, que nous avons appelées « parcours ».

Etudes déjà réalisées et conclusions provisoires

Pour définir - et recommander - des « parcours » de présentation de ressources, on a montré dans [1] l'utilité des notions d'*itinéraire* (ou « ligne »), de « station » (étape n sur la ligne l , pouvant être aussi étape n' sur la ligne l' , etc..) faisant fonction de cadre/boîte de présentation (« showbox ») pour des ressources regroupées par « propos » (ou « topique »).

En effet avec ces seules notions un concepteur peut :

- grouper préalablement des ressources dont les indexations se recouvrent en leur reconnaissant un "propos" commun,
 - les (re)présenter dans une boîte étiquetée par le nom (titre) de ce "propos",
 - construire des parcours comme structures de présentation (narratives) avec des boîtes préexistantes ou à créer (propos encore vide).
- Ce genre de structure de présentation laisse le lecteur libre d'une navigation
- en réseau, avec changement d'*itinéraire* (de « ligne ») sur les nœuds (les « stations ») offrant des « correspondances » (ayant même « propos »)
 - à plusieurs niveaux selon des règles (que l'on précisera) de décomposition/recomposition des nœuds du réseau (« showroom »/« showbox » et propos/sous-propos).

A la suite de l'article [2] on admet l'utilité des notions de « points de vue » ou de « représentations » multiples pour une unique ontologie que l'on a, pour la simplicité, ramenées à une notion unique de « cadre interprétatif » :

- dans un ensemble I de cadres interprétatifs possibles sur l'ontologie générale $\langle \text{Objets}, \text{Propriétés}, \text{Valeurs} \rangle$, on choisit, pour chaque présentation (« parcours »), un cadre interprétatif i pour sélectionner certains objets o^j et certaines propriétés p^k .

Avec cette notion, l'indexation des ressources de la communauté (documents numériques accessibles par leur URI) consiste à leur associer soit un couple $\langle i, o^j \rangle$ (dans le cas d'une *indexation synthétique*) soit plusieurs couples (dans le cas d'une *indexation analytique*).

Description des travaux en cours ou en projet

Objectifs

On veut s'interroger à présent sur les relations entre la possibilité de décomposition des nœuds de parcours et la possibilité de fragmentation des ressources composites, dans le but d'aider l'écriture multi-niveaux (permettant des « zoom in »/« zoom out » lors de la lecture) de boîtes de présentations (« showboxes ») de ressources, étiquetées par des propos eux-mêmes inscriptibles dans des hiérarchies (propos/sous-propos) parce que se rapportant à des objets et des cadres interprétatifs hiérarchisés.

Dans l'univers très vaste des ressources hypertextuelles, nous nous intéresserons restrictivement aux ressources manifestant une unité et une structure documentaire classiques : pages ordonnées et parties emboîtées. Dans ces documents un fragment, identifiable par son URI, est soit une page, soit une image attachée à une page (ressource « détachable »), soit une partie de page située entre une ancre et l'ancre suivante (ressource « découppable »).

Une indexation particulière des fragments est réalisée par la « table des matières » (TdM) qui est un arbre à profondeur variable dont tous les nœuds sont ordonnés, et étiquetés par des (sous-)titres. Elle guide la lecture classique, séquentielle mais hiérarchique car le lecteur peut choisir son « niveau de survol » pour chaque (sous-)partie. Une lecture réalise toujours une « coupe » de cet arbre. Mais l'auteur peut aussi recommander d'autres itinéraires en indiquant une suite de nœuds de l'arbre qui ne suit pas l'ordre de leur numérotation dans la TdM.

Nous faisons l'hypothèse que la TdM, « surchargée » d'autres itinéraires de lecture recommandée, est un guidage concret, à travers des fragments de documents, « piloté » par des itinéraires

abstrait, à travers des couples $\langle i, o^j \rangle$ ou des tuples $\langle i, o^i1, o^j2, o^k3, \dots \rangle$ dont les (sous-)titres sont, sinon des indicateurs exacts, du moins des « suggesteurs » (ayant le même rôle que la légende sous une figure).

Définitions

Nous distinguons les classes *ressource* (uniquement des *documents*), *objet*, *propos* (syn. de *topique*), *station* (syn. de *showbox*) et nous les associons par les fonctions suivantes (dont les noms requièrent peut être des synonymes plus acceptables) :

- *indexation* ($\langle i, o^j \rangle$) = { *ressource* } -- multivaluée
- *sujet* (*propos*) = $\langle i, o^j \rangle$ -- monovaluée
- *illustration* (*propos*) = { *ressource* } - - multivaluée
- *présentation* (*station*) = *propos* -- monovaluée
- *titre* (*station*) = $\langle i, o^j \rangle$ -- monovaluée
- *décoration* (*station*) = { *ressource* } - - multivaluée

qui permettent des compositions transitives:

- *titre* (*station*) = *sujet* (*présentation* (*station*))
- *illustration* (*propos*) = *indexation* (*sujet* (*propos*)) -- en général suivi d'un choix manuel
- *décoration* (*station*) = *illustration* (*présentation* (*station*))

Au niveau d'un document les nœuds d'une TdM sont des stations dont les décorations sont des fragments du document.

On peut alors dire qu'un document est « totalement indexé » si tous les fragments en décoration des nœuds de sa TdM font l'objet d'au moins une indexation, seulement "bien indexé" si cela est vrai seulement de nœuds formant une « coupe » de la TdM.

Problématique

On se demande comment l'indexation dont peut faire l'objet le fragment décorant un nœud de la TdM peut se propager aux nœuds pères d'une part, aux nœuds fils d'autre part. Cela pose le problème de composition/décomposition

d'objets oij si on ne change pas de cadre interprétatif i, et le problème de savoir comment synthétiser/détailler les cadres interprétatifs. Par exemple si des nœuds frères sont indexés par le même objet mais dans trois cadres interprétatifs différents, il est clair que le nœud père doit être indexé par le même objet mais dans un cadre interprétatif plus "large". Peut-on parler de généralisation/spécialisation pour les cadres interprétatifs ?

Programme de travail Recherche

Il conviendra d'approfondir et de comparer les concepts de « cadre interprétatif », de « point de vue » et de « représentation ». Le dernier véhicule les concepts de dimensions spatiales ou porteuses d'autres sémantiques, de thèmes et d'échelle, que les cartographes s'efforcent de bien définir et formaliser.

Les outils de manipulation de documents par leurs TdM relèvent plus de l'ingénierie linguistique que de l'ingé-

nerie des (méta-)données semi-structurées. Il sera sans doute nécessaire d'explorer les ponts possibles entre les deux.

Enfin il semble difficile de rester dans des réflexions sur la manipulation formelle des structures des trois domaines - ontologie, parcours, documents - sans proposer/prototyper des interfaces pour les utilisateurs leur présentant clairement les diverses sémantiques des navigations à travers les hiérarchies.

Développements et tests

Un certains nombre d'outils pour manipuler les documents et leurs TdM sont à développer: indexeur de fragment, extracteur/expliciteur de TdM implicite, indexeur de TdM. Il faudrait montrer, au moins sur un exemple, ce que l'on entend par spécialisation de cadres interprétatifs.

Bibliographie

[1] Philippe Picouet, Jean-Marc Saglio, « Définition de parcours sur un Web

Communautaire », Rapport projet GET, Paris, mars 2002.

<http://www.infres.enst.fr/~bdtest/webdb/cr/e-parcoursRPT.html>

[2] Benjamin Buffereau, Philippe Picouet, « Des itinéraires pour accéder à l'information : un défi pour la représentation des connaissances », DVP2002, Brest, juillet 2002.

http://iasc.enst-bretagne.fr/DVP2002/Articles/Buffereau_DVP2002.pdf

[3] Jean-Marc Saglio, Tuan Anh Ta, « Génération de parcours recommandés dans un web communautaire », Journées scientifiques CNRS Web Sémantique, Paris, octobre 2002. <http://www.lalic.paris4.sorbonne.fr/stic/octobre/octobre1/Ta.pdf>

Projets présentant des similitudes

- MESMUSES, <http://cweb.inria.fr/Projects/Mesmuses/index.html>
- HyperMuseum, <http://www.archimuse.com/mw2001/papers/stuer/stuer.html>

Recension de l'ouvrage « A new kind of science »

de Stephen Wolfram, éditeur Wolfram Media Inc., 1280 p., 2002

Antoine CORNUÉJOLS

Difficile de parler de cet ouvrage sans commencer par en évoquer l'aspect. D'abord le poids, 2,5 kg. Autant qu'un portable, que le manuel de Mathematica du même auteur ou que le livre « Gravitation », livre mythique, à juste titre, sur la théorie de la relativité. Deux fois plus lourd que le « Newton » de Westfall. Ensuite l'intérieur : 1280 pages dont 349 de notes en point 8 qui forment en fait l'essentiel du livre, environ 1000 figures en niveaux de gris et en haute résolution, un index de 15000 termes, ... et pas de section de références. Selon un commentateur, la raison en serait qu'avec une seule page de plus, l'ouvrage s'effondrerait en un minuscule trou noir. Expérience à essayer dans son jardin, mais d'après « Gravitation », le risque est faible. En fait, Stephen Wolfram estime que tout compte fait peu de chose est vraiment à sauver, tout au moins à savoir, de la science d'avant lui, et comme ses notes sont sensées former une synthèse, forcément la meilleure, des différentes disciplines discutées dans l'ouvrage, il n'est pas utile de connaître les références antérieures.

« Gravitation » de Misner, Thorne et Wheeler (1973), livre de même poids rappelons-le, présente un paradigme : la physique comme émanation de la géométrie, expose merveilleusement, selon plusieurs perspectives, les outils conceptuels et techniques associés : tenseurs, invariances, etc., et dresse un bilan des résultats et des confrontations avec l'expérience. Formidablement pédagogique et très plaisant à lire et à étudier, il s'agit d'un ouvrage de référence sur un paysage scientifique déjà bien balisé.

Qu'en est-il de « A new kind of science » ?

D'abord, c'est l'œuvre d'un auteur, d'un chercheur solitaire qui croit avoir trouvé la pierre philosophale et qui ne veut courir aucun risque que l'histoire oublie de faire justice de cette éminente contribution. C'est donc l'exposé d'une thèse, mais qui au lieu de la placer dans un contexte, de reconnaître et saluer les « épaules de géant » qui en forment le socle, choisit au contraire de

faire comme si toutes les idées exposées étaient dues, d'une manière ou d'une autre, à l'auteur qui seul en a saisi la portée.

Stephen Wolfram est physicien de formation, mais très vite il a abandonné les arcanes de la physique quantique pour se consacrer à l'étude des automates cellulaires avec l'intime conviction que ces systèmes, apparemment les plus simples, recèlent les clés de la compréhension de l'univers sous toutes ses manifestations. Des plus fondamentales : particules et forces, aux plus philosophiques : la nature du libre-arbitre par exemple, en passant par l'explication de processus cognitifs comme la perception ou l'analogie. Qu'est-ce qui peut l'autoriser à une telle stupéfiante extrapolation ?

Rappelons ce qu'est un automate cellulaire. Il s'agit d'un système dynamique généralement déterministe consistant en un réseau de machines à états finis identiques, ou cellules, qui changent de couleur d'un instant au suivant en fonction d'une règle de mise à jour. Cette règle est appliquée simultanément (en général) à toutes les cellules à des temps discrets. Une cellule particulière calcule sa nouvelle couleur en fonction de sa couleur actuelle et des couleurs de ses voisines. Il existe de nombreuses définitions de ces voisinages, ainsi que de règles de mises à jour. L'un des exemples les plus célèbres est celui du jeu de la vie mis au point par John Conway en 1970. Les cellules y sont placées sur une grille plane à maille carrée et calculent leur état : vivante ou morte (d'où le nom du jeu), en fonction de leur huit plus proches voisines selon une règle très simple. Si les cellules voisines vivantes sont exactement au nombre de deux, alors la cellule maintient son état courant. Si le nombre de voisines vivantes est de trois, la cellule sera vivante à la génération suivante quelque soit son état actuel. Dans tous les autres cas, la cellule sera morte à la génération suivante.

Ce qui est surprenant c'est qu'une structure et des

règles aussi simples donnent naissance à des évolutions très complexes. Le jeu de la vie étant devenu un programme très courant, fréquemment utilisé comme nuit d'écran, tout le monde est devenu familier des formes qui naissent, se métamorphosent, se déplacent et parfois meurent sur l'écran (voir les ouvrages « The recursive universe » de William Poundstone, 1985 et « Laws of the game » de Manfred Eigen et Ruthild Winkler, 1981, remplis d'illustrations et de détails sur ce jeu. Voir aussi le site de Paul Rendell et celui de David Griffeth psoup.math.wisc.edu). Ce que l'on sait peut-être moins est qu'il a été démontré que ce jeu peut simuler une machine de Turing et a donc la puissance d'une machine universelle. On profitera de cette occasion pour citer l'extraordinaire livre de Berkelamp, Conway et Guy « Winning ways for your mathematical plays », dans lequel, entre autre mille fascinantes facettes, on trouve une esquisse de cette preuve.

Wolfram, dès le début des années quatre-vingt, a étudié des automates encore plus simples puisqu'ils sont définis sur un réseau linéaire dans lequel le voisinage d'une cellule comporte seulement ses deux plus proches voisines. Cela signifie que l'état d'une cellule est déterminé par l'état de trois cellules (dont elle-même) à la génération précédente. Il y a donc $2^3=8$ contextes possibles et $2^8=256$ règles d'évolution envisageables. L'intérêt de ce type d'automates linéaires est qu'il est aisé d'analyser leur évolution en superposant ligne après ligne sur une même figure les générations successives des cellules. C'est ce qu'a réalisé Wolfram de manière systématique et ce qu'il a obtenu ainsi a transformé sa vie et devrait, pense-t-il, révolutionner la science. L'observation des évolutions des différents automates (partant tous du même état initial) suggère en effet de distinguer plusieurs classes de comportements. Selon Wolfram, il y aurait la classe 1 de ceux qui évoluent vers un attracteur et dont les cellules conservent leur état après un certain nombre de générations, la classe 2 de ceux qui évoluent vers un processus périodique, la classe 3 de ceux qui exhibent des « trajectoires » qui ont un caractère aléatoire et dont l'apparence indique une transmission d'information à longue distance entre les cellules, et enfin la classe 4 de ceux qui seraient à la

frontière entre les comportements des classes 2 et 3 avec des régions apparemment aléatoires jouxtant des régions localement structurées. Deux automates en particulier ont fasciné Wolfram, correspondant aux règles 30 et 110 (voir dans le livre la manière astucieuse qu'utilise Wolfram pour décrire ses règles). La première correspondrait à la classe 3 et serait une sorte de génératrice de hasard (on s'amuse d'ailleurs d'apprendre qu'elle est effectivement utilisée pour cela dans Mathematica). Aucun test connu ne permettant d'y distinguer des régularités. La seconde, la règle 110, étant plus extraordinaire encore car, comme le jeu de la vie mais d'une définition plus simple, elle serait équivalente à une machine de Turing. Le lecteur de « A new kind of science » pourra admirer, grâce à des figures superbement bien choisies et rendues, l'extraordinaire bestiaire de comportements possibles.

Wolfram est donc stupéfait que partant de systèmes déterministes aussi simples une telle richesse de comportements soit possible et puisse inclure quelque chose ayant une puissance de calcul universelle. D'autres avant lui l'avaient vu, citons Emile Post en 1943 avec les fameuses « règles de production », Stan Ulam et John von Neuman, inventeurs des premiers automates auto-reproductibles, John Conway bien sûr, et, en France, Françoise Fogelman, Yves Robert, Maurice Tchuente et Henri Atlan étudiant la dynamique de réseaux booléens au début des années quatre-vingt. Cependant Wolfram n'hésite pas à y discerner deux conséquences qu'il élève au rang de lois fondamentales.

Premièrement, des règles discrètes extrêmement simples peuvent produire des comportements et des structures aussi complexes qu'il est possible d'atteindre. C'est l'essence de ce qu'il appelle « le principe d'équivalence computationnelle ». Il n'est donc pas besoin d'expliquer le complexe par le complexe. Notons au passage que Wolfram assimile ici le fait que le simple puisse produire du complexe de tout ordre pour laisser entendre que tout complexe doit nécessairement être produit par du simple.

Deuxièmement, les règles simples produisant du complexe sont très abondantes plutôt qu'exceptionnelles comme on aurait pu le penser.

Pour Wolfram, cela amène à penser que l'univers entier n'est que la manifestation du fonctionnement d'un automate très simple (au maximum quatre lignes de code en Mathematica a-t-il conjecturé dans une interview). Cela le conduit aussi à traquer les automates sous-jacents dans tout un ensemble de phénomènes naturels, cognitifs et culturels. Après avoir montré dans les premiers chapitres l'extraordinaire richesse du monde des automates (Wolfram en explore une énorme variété au-delà des automates linéaires qui servent de base à sa thèse), c'est ainsi qu'il consacre les chapitres 7 à 10 à l'exploration des formes naturelles (que l'on trouve dans la nature, tels les arrangements de couleurs sur les coquillages), les lois fondamentales de la physique et les processus de la perception et de l'« analyse ».

Pour ma part, j'ai trouvé la partie consacrée aux phénomènes cognitifs faible, mal informée et la moins intéressante. Il est visible que, même après quelques coups de téléphones nocturnes à Terry Sejnowski, Wolfram ignore en grande partie ce que sont les sciences cognitives et l'intelligence artificielle. Il en est essentiellement resté à l'image de l'IA des années 70 avec les grands programmes, tels ARCH, AM, ACT* ou SOAR, qui se décrivaient au niveau d'une architecture cognitive et à celle d'une IA désormais entièrement fondée sur l'étude des réseaux connexionnistes. Les modélisations de la perception, essentiellement visuelle, qu'il propose sont attrayantes, mais pas plus que beaucoup d'autres. Il parle également d'apprentissage artificiel sans en utiliser le nom en assimilant apprentissage et compression de données, ce qui d'ailleurs n'est pas très éloigné de la perspective actuelle. Cependant il ignore tout des résultats obtenus sur l'analyse théorique de l'apprentissage durant ces dernières décennies. Il semble en particulier penser qu'il peut exister un système universel et optimal de recherche du meilleur modèle sous-jacent à des données. Mais cela est interdit par le no-free-lunch theorem. Globalement, il ne voit pas que l'apprentissage est un « problème inverse » dans lequel il s'agit de chercher un modèle qui ait pu produire les données. Or les données seules ne peuvent suffire à déterminer le modèle optimal, il faut ajouter des

connaissances a priori extérieures aux données. Pour en terminer sur ce chapitre, Wolfram affirme que l'intelligence est avant tout liée à l'utilisation efficace et à bon escient d'une énorme mémoire. C'est intéressant, même si d'autres avant lui, comme Roger Schank, l'avaient affirmé avec plus de force encore. Malheureusement il en reste essentiellement là, sans donner la clé permettant cette utilisation si puissante de la mémoire. D'un certain côté, je me demande si, curieusement, la contribution la plus féconde à l'étude de l'IA ne serait pas à chercher dans le chapitre 9 sur la physique fondamentale dans lequel il décrit des réseaux causaux pour rendre compte de la structure de l'espace et des forces entre particules. Même si d'éminents physiciens ont montré l'insuffisance de ce modèle en ce qui concerne la physique (voir par exemple la critique de Scott Aaronson dans *Quantum Information and Computation*, vol.1, 2001), il n'est pas impossible que l'on puisse détourner ce modèle pour l'étude de certains phénomènes de transmission de l'information dans les systèmes cognitifs.

Reste à discuter ce qu'entend Wolfram quand il parle de « nouvelle science » et en quoi cela pourrait concerner par exemple les chercheurs en IA. Wolfram considère que la science traditionnelle est avant tout occupée à chercher des normes, à décrire les contraintes que respectent les systèmes naturels. De ce fait, elle passerait à côté de la vraie explication des phénomènes qu'il identifie évidemment à l'identification de l'automate sous-jacent. Selon lui, il faudrait donc arrêter de chercher des modèles globaux dont la complexité est forcément du même ordre que celle du phénomène étudié pour se tourner vers la recherche des règles des systèmes dynamiques sous-jacents qui sont d'après les deux lois fondamentales de Wolfram nécessairement simples. Par ailleurs, Wolfram propose – non, je me trompe, Wolfram ne propose jamais, il affirme – que pour juger de la complexité d'un processus on se repose essentiellement sur notre capacité d'analyse visuelle puisqu'en effet, toujours selon lui, les techniques d'analyse sont en nombre limité, essentiellement équivalentes dans leur pouvoir d'analyse et que notre système visuel est sans doute le meilleur.

Je vois au moins trois problèmes à cette position. Le premier est que c'est une chose d'affirmer qu'il existe une règle simple sous-jacente et une autre de la trouver. On se retrouve à nouveau face à un problème inverse redoutable. Wolfram n'est pas le seul à s'être fait le chantre de l'émergence de la complexité à partir du simple. Pour le moment, personne ne sait comment s'y prendre de manière non empirique. Wolfram lui-même ne recule pas, dans son livre, à une exploration systématique de toutes les règles possibles dans certains cas (il a ainsi exploré le nombre hallucinant de 4 294 967 296 règles pour les automates mobiles, chap.3). Mais cette approche est évidemment vouée à l'échec dans le cas général. Le second tient à la nature de ce que l'on tient pour une explication ou ce que signifie comprendre. Est-ce que connaître, à supposer qu'il existe, l'automate sous-jacent à un phénomène signifie qu'on le comprend ? Certes d'un certain côté on est ainsi capable de le reproduire, mais on n'est plus capable de faire des prédictions dans les cas complexes intéressants, il manque les normes et les contraintes justement (voir sur le sujet de la nature de l'explication, les livres « Prédire n'est pas expliquer » de René Thom et « L'aveuglante proximité du réel » de Michel Bitbol, tous deux dans la collection Champs Flammarion). De plus, et c'est le troisième problème de la nouvelle science, Wolfram insiste sur le caractère justement imprévisible des comportements des automates intéressants. Il faut donc attendre et observer, mais combien de temps ? Et puis pour quoi faire à la fin ? Vérifier que l'on a reproduit l'univers ou le cerveau ? On retrouve d'ailleurs là un débat qui a agité les sciences cognitives il y a une dizaine d'années lorsque l'émergence des réseaux de neurones sapait l'approche traditionnelle fondée sur l'idée de représentation et de raisonnement. Décrire en termes de lois élémentaires (au niveau neuronal) peut-il remplacer ou à tout le moins rendre inutile une description au niveau mental ou psychologique ? Il y a là certainement un débat qui touche crucialement l'étude de la vie artificielle (nous examinerons dans une prochaine rubrique le livre « Vie artificielle. Où la biologie rencontre l'informatique » de Jean-Philippe Rennard, Vuibert Informatique, 2002).

Pour finir, comme beaucoup d'autres commenta-

teurs (chercher ANKOS (A New Kind Of Science) sur google), je ne discerne pas vraiment d'idées nouvelles dans le livre de Wolfram, ni de révolution scientifique en germe. Je ne pense pas non plus que ce livre doive provoquer l'intérêt des chercheurs en IA toutes affaires cessantes. Cependant, c'est un livre riche. D'abord parce qu'il réexamine de nombreux problèmes en essayant d'en rendre compte avec des automates, or cela est souvent très éclairant, presque comme une renaissance. Par exemple les discussions sur le codage, sur la compression de données, sur la nature de la complexité ne sont pas toujours parfaitement satisfaisantes sur le plan académique mais elles font réfléchir à neuf. Les illustrations étayaient souvent magistralement ces discussions. Et si on prend le recul nécessaire pour ne pas être dupe de ces « démonstrations par l'œil », on ne peut qu'admirer l'ingéniosité qu'elles recèlent. Ensuite, le livre vaudrait sa lecture rien que pour les notes de fin d'ouvrage qui reprennent des grandes disciplines et, à côté de synthèses souvent brillantes, même si Wolframocentriques, fourmillent de détails sur des concepts, des techniques, des algorithmes. Également, pour ceux qui comme moi sont des utilisateurs heureux de Mathematica, on y trouve des centaines d'exemples de petits programmes magnifiques. Finalement, « A new kind of science » est un très beau livre, superbement réalisé. Cela le met à part et en fait presque un objet d'art.

À la fin de leur célèbre article sur la structure de l'ADN en 1953, Watson et Crick en trois lignes « It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material » annonçaient la révolution de la biologie du génome. Wolfram a lui choisi de marteler son message en 1280 pages et 2,5 kg. L'avenir dira si sa portée sera inversement proportionnelle ou non à cette insistance.

Je tiens à signaler aux lecteurs motivés des ouvrages très intéressants sur des sujets proches : « Cellular automata machines » de Toffoli et Margolus, MIT Press, 1987, et surtout « The origins of order. Self-organization and selection in evolution » (1993) et « Investigations » (2000) de Stuart Kauffman qui valent vraiment le temps passé à les étudier.

Compte rendu de la conférence

EGC 2003

Journées francophones d'Extraction et de Gestion des Connaissances

Lyon les 22, 23 et 24 Janvier 2003

<http://www.univ-lyon3.fr/EGC2003>

Danielle Boulanger, Présidente du comité d'organisation
Comité d'organisation : équipe MODEME, IAE, Université Jean Moulin, Lyon

La troisième édition des Journées Francophones d'Extraction et de Gestion des Connaissances (EGC 2003) qui se sont déroulées à l'Université Jean Moulin à Lyon a rassemblé 170 chercheurs travaillant dans les domaines suivants : bases de données, recherche d'information, représentation des connaissances, apprentissage, interaction homme-machine et statistiques.

Trois chercheurs de très haut niveau ont accepté de nous présenter une conférence invitée. Le 23 au matin, nous avons écouté Grégory Piatetsky-Shapiro – de KDnuggets, Boston – qui est le fondateur du domaine de la fouille de données. L'après-midi Attilio Giordana – professeur à l'Université du Piémont Oriental – a exposé ses derniers résultats : il a participé à de multiples avancées de l'apprentissage automatique. Le 24 au matin, en concomitance avec la session industrielle, Madame Marie-Paule Lefranc – professeur à l'université de Montpellier et appartenant à l'Institut Universitaire de France – qui a créé et qui exploite une base d'immunogénétique moléculaire nous a fait partagé son expérience relative à tous les aspects du maintien et de la fouille de données.

Ont été retenus quarante sept articles sur une centaine de soumissions. Les auteurs d'articles et les participants venaient d'universités et d'organisations francophones (Québec, Suisse, Belgique, Tunisie et France) mais aussi des Etats-Unis, d'Italie et d'Espagne.

Les journées, précédées par quatre tutoriels et quatre ateliers, ont été 'denses' avec plusieurs sessions en parallèle. La session « industrielle » a démontré l'intérêt que portent les entreprises au « Knowledge Management ».

Deux prix ont été remis :

Celui de la Fondation Vedioirbis, distinguant le meilleur papier de recherche des Journées dont le titre est « Exploration de données multimedia par réalité virtuelle » co-écrit par Gilles Venturini de l'Ecole polytechnique de Tours et Christiane Guinot du CERIES.

Celui de l'AFIA valant une inscription à la plateforme AFIA se déroulant à Laval en juillet 2003, distinguant le meilleur papier jeune chercheur : « Découverte des règles associatives non redondantes : application aux corpus textuels » par Saïd Ben Yahia et Georges Mineau de l'Université de Laval (Québec).

Les actes du congrès, publiés chez l'éditeur scientifique Hermès, font l'objet de trois numéros de la Revue des Sciences et Technologies de l'Information, série « Recherche en Intelligence Artificielle – Extraction des Connaissances et Apprentissage » : EGC 2003, RSTI série RIA-ECA – Volume 17 – N° 1-2-3 / 2003, 552 pages.

Si ce congrès a pu se dérouler dans d'excellentes conditions c'est grâce au soutien de l'Université Jean Moulin de l'IAE et des différents sponsors. Nous remercions l'AFIA, l'association Inforsid, le GDR I3, l'Association Française de Statistiques, la Société Française de Classification, le Crédit Mutuel Enseignant, la MAIF et la fondation Vedioirbis.

Danielle Boulanger, Présidente du comité d'organisation EGC'2003, équipe MODEME, IAE, Université Jean Moulin, Lyon.

Calendrier des conférences AFIA

JUILLET 2003			
1 - 4	Plate-forme AFIA	Laval	<p>3 conférences</p> <ul style="list-style-type: none"> * CAp 2003 : Conférence d'apprentissage * IC 2003 : Ingénierie des Connaissances * RJCIA 2003 : Rencontre Jeunes Chercheurs en IA <p>5 ateliers :</p> <ul style="list-style-type: none"> * Usage des Treillis de Galois pour l'Intelligence Artificielle * Raisonnement à partir de cas * Apprentissage machine et bio-informatique * Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel. * Risque et Intelligence Artificielle <p>Des sessions posters et démonstrations</p>
NOVEMBRE 2003			
26 - 27	LFA	Tours	<p>Rencontres Francophones sur la Logique Floue et ses Applications</p> <p>http://www.rfai.li.univ-tours.fr/lfa03/</p>
27 - 29	JFSMA	Hammamet	<p>Journées Francophones sur les Systèmes Multi-Agents</p> <p>http://www.cck.rnu.tn/JFSMA/</p>
DECEMBRE 2003			
3 - 4	CITE	Troyes	<p>Coopération, Innovation, Technologie</p> <p>http://tech-cico.utt.fr/cite2003/</p>
JANVIER 2004			
28 - 30	RFIA	Toulouse	<p>Reconnaissance des Formes et Intelligence Artificielle</p> <p>http://www.laas.fr/rfia2004</p>

Présentation des Ateliers, Démonstrations et Posters

Plate-Forme AFIA 2003

Laval du 1er au 4 Juillet 2003

<http://www.afia.polytechnique.fr/plateforme-2003/>

Président du programme : **Gilles Bisson** Leibniz-IMAG, Grenoble
Présidente du comité d'organisation : **Anne-Marie Kempf** ESIEA Recherche, Laval

La plate-forme de l'AFIA qui se déroulera du 1er juillet au 4 juillet à Laval, est accueillie par le centre de Recherche du groupe ESIEA (<http://www.esiea.fr/>).

Outre les trois conférences **CAp**, **IC** et **RJClA**, la plate-forme accueillera le vendredi 4 juillet un ensemble de 5 ateliers (workshops) autour de l'IA ou utilisant ses approches :

- **A1 : Usage des Treillis de Galois pour l'Intelligence Artificielle**

Responsables : Engelbert Mephu Nguifo (CRIL, Univ. Artois), Michel Liquière (LIRMM, Montpellier, France) et Petko Valtchev (DIRO, Univ. Montréal)

- **A2 : Raisonnement à partir de cas**

Responsable : Jean Lieber, (LORIA)

- **A3 : Apprentissage machine et bio-informatique**

Responsables : Jean-Philippe Vert (Ecole des Mines) et Florence d'Alche-Buc (LIP6).

- **A4 : Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel.**

Responsables : Claire Nédellec (INRA-MIG) et Pierre Zweigenbaum (STIM-DIAM)

- **A5 : Risque et Intelligence Artificielle**

Responsable : Sophie Levionnois (ESIEA Recherche)

Toujours dans le cadre de la plate-forme 2003, le jeudi 3 juillet se déroulera une session commune « posters » et « démonstrations » ouverte à tous les participants. Elle a pour but de permettre à chacun de présenter librement des travaux originaux en IA et de faire des démonstrations de logiciels. En outre, chaque intervenant peut décrire ses travaux en une page, l'ensemble des contributions étant regroupé dans un fascicule qui sera distribué à chacun des participants de la plate-forme.

Présentation de la conférence

LFA03

*Rencontres Francophones sur la Logique Floue et ses Applications
Tours les 26 et 27 novembre 2003*

<http://www.rfai.li.univ-tours.fr/lfa03/>

Carl Frélicot, Laboratoire d'Informatique, Image, Interaction (L3i) - Université de La Rochelle

Les rencontres sur la logique floue et ses applications sont l'occasion pour les chercheurs de la communauté francophone de se réunir afin de faire le point sur les développements actuels de la discipline.

Depuis plusieurs années, la manifestation s'est élargie, au delà des ensembles flous et de la théorie des possibilités, à d'autres théories de l'incertain. L'édition 2003 de la manifestation s'inscrit dans cette volonté d'évolution et d'ouverture.

Un nombre croissant de domaines est concerné par l'exploitation de ces différentes théories. Si d'un point de vue historique la commande floue a joué un rôle précurseur, elle est aujourd'hui rejointe par bon nombre de disciplines telles que la recherche opérationnelle et l'aide à la décision, le mesurage et l'évaluation subjective, l'analyse de données, la classification et la reconnaissance de formes, le traitement et l'interprétation des images, les bases de données et les systèmes d'informations, l'intelligence artificielle, la fusion d'informations, la robotique, mais aussi l'économie, la gestion, la psychologie, l'ergonomie ...

Les organisateurs de ces rencontres souhaitent pouvoir offrir aux participants de LFA03 un panorama aussi complet que possible des avancées réalisées au sein de la communauté francophone universitaire mais aussi industrielle. Toute contribution théorique, méthodologique ou applicative est la bienvenue ; les contributions jeunes chercheurs sont accueillies avec un intérêt tout particulier.

SOUTIENS

- * Association Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF)
- * European Society for Fuzzy Logic And Technology (EUSFLAT)
- * Groupement de Recherche en Information, Signal, Image et viSion (ISIS)
- * Groupement de Recherche en Information, Interaction, Intelligence (I3)
- * Conseil Général d'Indre-et-Loire (cg37), Ville de Tours (Tours),
- * Ecole Polytechnique de l'Université de Tours (PolytechTours)
- * Association Française d'Intelligence Artificielle (AFIA)

DATES IMPORTANTES

- * Date limite de soumission des articles : 2 mai 2003
- * Notification d'acceptation : 18 juillet 2003
- * Date limite de réception des versions finales : 19 septembre 2003

RENSEIGNEMENTS

Carl Frélicot, Lab. d'Informatique, Image, Interaction (L3i), Université de La Rochelle
Tél. : (+33) (0)5 46 45 82 34
Fax : (+33) (0)5 46 45 82 42
lfa03@univ-tours.fr
<http://www.rfai.li.univ-tours.fr/lfa03>

Présentation de la conférence

JFSMA 2003

*Journées Francophones sur les Systèmes Multi-Agents
Hammamet, Tunisie les 27, 28 et 29 Novembre 2003*

<http://www.cck.rnu.tn/JFSMA/>

Présidente du comité scientifique : **Jean-Pierre Briot**, LIP6, Paris
Président du comité d'organisation : **Khaled Ghédira** (ENSI, Tunisie)

Les Journées Francophones sur les Systèmes Multi-Agents (JFSMA) constituent un forum d'échange privilégié entre chercheurs provenant de plusieurs communautés scientifiques telles que les systèmes multi-agents, l'intelligence artificielle, la vie artificielle, les sciences humaines, les systèmes distribués et le génie logiciel. Ces journées (jusqu'à l'an dernier appelées Journées Francophones en Intelligence Artificielle Distribuée et Systèmes Multi-Agents - JFIADSMA) se sont tenues respectivement à Toulouse (1993), Grenoble (1994), Chambéry (1995), Port-Camargue (1996), Nice (1997), Nancy (1998), L'Ile de la Réunion (1999), Saint-Etienne (2000), Montréal (2001) et Lille (2002). Pour leur onzième édition, les JFSMA'03 se tiendront à Hammamet, en Tunisie du 27 au 29 novembre 2003.

Une des motivations principales des systèmes multi-agents est d'offrir un cadre pour la conception et la réalisation de systèmes informatiques coopératifs fortement décentralisés (ex : supervision, commerce électronique, domotique...). La recherche dans le domaine des systèmes multi-agents (SMA) a déjà eu un très grand impact et de nombreux succès. Un des enjeux qui se pose maintenant à elle est le passage à des systèmes cette fois à très grande échelle (ex : commerce électronique, services Web, jeux en réseau, GRID, informatique amorphe, transports, informatique citoyenne, simulation d'éco-systèmes, robotique sociale...), avec les problèmes afférents (passage à l'échelle, hétérogénéité, robustesse, adaptation dynamique, facteurs humains...). L'enjeu du passage à l'échelle nécessite de pousser plus loin recherches conceptuelles, théoriques, expérimentales et également de favoriser une coopération accrue avec d'autres communautés (génie logiciel, objets et composants distribués, systèmes répartis, algorithmique répartie, ingénierie des connaissances, informatique sociale...) qui abordent également cet enjeu.

L'enjeu du passage à l'échelle se pose à différents stades de la construction d'applications multi-agents, tels que :

- niveau conception - méthodes systématiques de conception, modèles d'adaptation, modélisation, lien avec des propositions telles que la « Model Driven Architecture » de l'OMG.
- niveau construction - langages d'agents, architectures à base de composants et d'objets, frameworks, intergiciels (middleware).
- niveau déploiement et exécution - répartition, gestion des ressources, adaptation (aux contraintes, à l'environnement...), résistance aux pannes, sécurité, mobilité, déploiement en milieu industriel.
- niveau social - interaction entre agents artificiels et humains, acceptation par les utilisateurs, adaptation aux utilisateurs, apprentissage.

Pour cette édition des Journées, ont été sollicitées en priorité des contributions originales abordant l'enjeu du passage à l'échelle, mais également des contributions originales sur les thématiques traditionnelles des systèmes multi-agents. Les contributions peuvent proposer des avancées conceptuelles, formelles ou logicielles, ou elles peuvent être analytiques (sur les problèmes, enjeux...) ou encore synthétiques (par exemple des retours sur expérimentations).

Ces Journées, comme les précédentes, seront placées sous le signe de l'échange, de la convivialité et de l'ouverture vers d'autres disciplines, mais aussi vers les entreprises et les organismes de recherche privés.

Les Journées JFSMA se tiennent sous l'égide du Collège « Systèmes Multi-Agents » de l'AFIA.

DATES IMPORTANTES

Date limite de soumission : 9 mai 2003
Notification aux auteurs : 5 juillet 2003
Réception des versions définitives : 25 septembre 2003

Appel à communications de la conférence

CITE 2003

*Coopération, Innovation, Technologie
Troyes les 3 et 4 Décembre 2003*

<http://tech-cico.utt.fr/cite2003/>

Président du Comité de Programme : **Manuel ZACKLAD** (Tech-CICO, Troyes)

Recherches interdisciplinaires, sciences de l'ingénieur et sciences humaines, pour le développement des technologies de la coopération et de la connaissance dans l'innovation technique, managériale, sociétale.

Il ne fait aujourd'hui plus de doute que la mise en place de processus d'innovation et de changement en continu, enjeu économique et social crucial, repose sur la capacité à gérer efficacement les activités intellectuelles organisées, ou, ce qui revient au même, à faciliter la prise en charge, par les acteurs eux-mêmes, de leurs activités coopératives sur les plans sociaux et cognitifs. L'hypothèse majeure du congrès CITE'2003 est que les formes de cognition collective dépendent profondément des technologies de la coopération et de la connaissance mises en œuvre.

Ces technologies résultent d'une association intime entre, d'une part, un ensemble d'artefacts matériels mis à la disposition du collectif pour se réunir, mémoriser les connaissances et transmettre à distance des informations via différents média, et d'autre part, un ensemble de conventions collectives et de méthodes d'organisation relevant de savoir-faire aussi diversifiés que peuvent l'être l'usage des règles de politesses, l'utilisation de méthodes de travail de groupe ou, plus abstraitement, l'invocation à bon escient des valeurs du collectif pour régler un différend.

Le développement des technologies de la coopération et de la connaissance confronte les chercheurs à plusieurs questions de recherche fondamentale relevant des sciences cognitives, des sciences humaines et sociales et des sciences de l'ingénieur. L'outillage des activités coopératives et de la production collective de connaissances qui les accompagnent agit en effet comme un révélateur des phénomènes de cognition distribuée, des processus d'organisation et de gestion de l'action collective et de la dynamique des relations et

des identités mises en scène dans les activités communicationnelles. Par ailleurs, elle pose aux sciences de l'ingénieur des problèmes de modélisation ardues d'activités heuristiques et émergentes auxquelles les systèmes d'information doivent en permanence s'adapter.

CONTRIBUTIONS ATTENDUES

L'importance de cette problématique nous semble nécessiter la rencontre régulière de chercheurs de disciplines diverses assurant le croisement de thématiques plurielles afin que s'amorce une synergie mutuellement profitable. Les contributions attendues, à caractère théorique ou issues de recherches empiriques, doivent traiter de la problématique de la conception des technologies de la coopération et de la connaissance. Elles peuvent mettre l'accent sur deux dimensions :

- * La description des technologies de la coopération et de la connaissance innovantes relevant des systèmes d'information, mais aussi de méthodes d'organisation, de gestion des connaissances et de management.
- * La présentation de recherches portant sur la compréhension et l'évaluation des effets des technologies sur un plan cognitif, social et économique à l'échelle individuelle et collective.

Enfin, une journée particulière pourra être consacrée à la rencontre avec les porteurs d'innovations industrielles d'origine diverse : industrie de production de biens matériels, éditeurs de logiciel, entreprises de services (dont cabinets de consultants, industries culturelles, etc...).

DISCIPLINES

Sciences de l'ingénieur : ingénierie des connaissances, gestion des connaissances, systèmes d'informa-

tion, CSCW (groupware, workflow, architectures distribuées...), interaction et interface homme-machine, génie industriel, génie mécanique, productique...

Sciences humaines : communication, pragma-linguistique, sciences de gestion (gestion de l'innovation, stratégie...), économie (économie industrielle, service...), sociologie des organisations, psychosociologie, sciences cognitives (ergonomie cognitive, anthropologie cognitive...).

DATES IMPORTANTES

Date limite de soumission : Juillet, 2003

Notification d'acceptation : 22 Septembre, 2003

Date limite d'envoi de la version finale : 20 Octobre, 2003

Appel à communications de la conférence

RFIA 2004

**14^e Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle
Toulouse les 28-30 Janvier 2004**

<http://www.laas.fr/rfia2004/>
rfia2004@laas.fr

Président du comité de programme : **Jean Charlet**

Co-présidents du comité d'organisation :

Michel DEVY (LAAS-CNRS, Toulouse) et **Philippe JOLY** (IRIT, Toulouse)

RFIA 2004, quatorzième édition du congrès francophone de Reconnaissance des formes et d'Intelligence artificielle, sera organisé par le LAAS-CNRS et l'IRIT et se tiendra du 28 au 30 Janvier 2004 au Centre de Congrès Pierre BAUDIS de Toulouse.

Ce congrès, à la tradition maintenant parfaitement établie, est un rendez-vous attendu des scientifiques issus des domaines de la Reconnaissance des formes et de l'Intelligence artificielle sous l'égide de l'AFRIF <<http://www.afrif.asso.fr/>> et de l'AFIA <<http://www.afia-france.org/>>. Il constitue une opportunité unique de rassemblement et d'échanges fructueux entre ces deux communautés et, depuis la première édition en 1977, RFIA s'est forgé une réputation de tout premier plan.

Nos deux disciplines ont une longue tradition de collaboration autour de problèmes impliquant de combiner analyses, interprétations et raisonnements dans des environnements toujours plus complexes. Cette année, cette collaboration sera particulièrement mise en avant dans les domaines de la bio-informatique et de la

robotique : la bio-informatique sollicite fortement et conjointement ces deux disciplines, notamment pour l'annotation des génomes et la recherche des gènes et des signaux de régulation, pour la prédiction de la structure conformationnelle des protéines et des ARN, et pour l'analyse des données issues de puces à ADN. La robotique, quant à elle, est un domaine de recherche essentiellement pluridisciplinaire dont la problématique porte sur la conception et l'étude de fonctions de perception, de décision et d'action, et sur l'intégration cohérente de ces fonctions en une machine physique. La Reconnaissance des formes et l'Intelligence artificielle y trouvent naturellement un terrain applicatif de prédilection.

RFIA 2004 doit aussi permettre de mettre en avant des synergies entre des approches maintenant classiques à RFIA dans les domaines de la conception et la mise en œuvre de capteurs, du traitement du signal sonore et de l'image, de la reconnaissance et l'interprétation des formes, de la vision artificielle, du raisonnement, de la représentation des connaissances, de l'ingénierie des connaissances, de l'apprentissage, de la linguistique,

avec des applications dans le domaine des documents (analyse, interprétation, fouille de texte, indexation, ...), du Web sémantique (ontologies, méta-données, ...), des bases de données (analyse vidéo, recherche d'images, entrepôts, BD déductives...), de la robotique (mobile, manufacturière, ...), de l'imagerie médicale (analyse d'images, reconstruction volumique, ...), des interfaces homme-machine (capture de geste, d'émotions, d'expressions, ...), de la surveillance (télé-détection, vidéo surveillance, ...), des Environnements informatiques pour l'apprentissage humain (EIAH), de l'Informatique médicale, de la mémoire d'entreprise ou des multi-agents...

Dans cette optique, l'essor grandissant des sciences du traitement de l'information et de la communication est un atout indéniable. Plus que jamais, il apparaît nécessaire de mettre en exergue les spécificités et les complémentarités de ces deux disciplines et d'encourager leur enrichissement mutuel. L'évolution technologique actuelle offre des opportunités nouvelles d'échanges fondés sur la parole, l'image et le texte. La genèse d'outils performants d'analyse, d'interprétation, de visualisation et de combinaison de ces vecteurs est nécessaire en tenant compte de l'appropriation cognitive des systèmes mis en oeuvre par les utilisateurs.

Le congrès RFIA 2004 doit être l'occasion de relever le défi. Nous vous invitons à répondre nombreux à cet appel à communications. Seront bienvenues les soumissions décrivant des recherches effectuées en Reconnaissance des formes ou en Intelligence artificielle, et tout particulièrement celles qui se situent à la frontière entre ces deux domaines. Les recherches présentées pourront être de natures diverses : travaux méthodologiques, théoriques ou expérimentaux ; développement d'algorithmes, d'outils, de systèmes ou d'applications ; évaluation des systèmes, de leur appropriation cognitive par les utilisateurs ou de leur insertion dans l'environnement technique et organisationnel. Elles doivent correspondre à des travaux originaux et les thématiques pourront être choisies dans la liste (non limitative) ci-dessous. Le congrès comportera des sessions orales et des sessions posters, ainsi que des tutoriels et des tables rondes. Des démonstrations seront également organisées, qui donneront lieu à l'édition d'une plaquette spécifique : les chercheurs intéressés sont priés de se mettre en contact avec les organisateurs.

THÈMES

- * Algorithmes génétiques / évolutionnaires
- * Algorithmique de l'IA ou de la RF
- * Analyse de documents
- * Analyse de scènes
- * Apprentissage symbolique et/ou numérique
- * Architectures et intégrations de systèmes
- * Bio-informatique
- * Capteurs de vision (calibration géométrique et photométrique, modélisation de la fonction de transfert, de la couleur, ...)
- * Caractérisation et Evaluation de performances
- * Classification
- * Compression
- * Construction d'ontologies (épistémologie, liens linguistique-conceptuel, fusions, ...)
- * Contraintes (CSP, SAT, ordonnancement, ...)
- * Coopération personne-système (modélisation de l'utilisateur, aide à l'apprenant, aide au travail coopératif, dialogue, ...)
- * Diagnostic
- * Démonstration automatique
- * Documents électroniques, documents virtuels (organisation, navigation, adaptation, ...)
- * Environnements informatiques pour l'apprentissage humain
- * Extraction des connaissances
- * Formalisation des raisonnements et de la cognition
- * Fouille de données
- * Fusion (de capteurs, données, informations, décisions, ...)
- * Indexation et recherche d'information
- * Ingénierie des connaissances
- * Interaction homme-machine
- * Modèles conceptuels
- * Modélisations des images (photométrie, colorimétrie, approches probabilistes, approches variationnelles, ...)
- * Planification
- * Problèmes inverses
- * Raisonnement à partir de cas
- * Raisonnements en environnement physique, évolutif ou incertain
- * Réalité virtuelle et augmentée
- * Reconnaissance et synthèse de la parole (vérification du locuteur, transcription textuelle, ...)
- * Recherche d'informations (indexation documents multimédia, ...)

CONFÉRENCES

- * Reconnaissance et interprétation de l'écrit
- * Reconnaissance et interprétation du geste humain
- * Représentation des connaissances
- * Représentation, reconnaissance et interprétation des formes et des images
- * Réseaux neuronaux et approches connexionnistes
- * Résolution de problèmes et programmation des jeux
- * Robotique (couplage perception/action, asservissement visuel, ...)
- * Synthèse d'image (simulation, animation, visualisation,...)
- * Systèmes d'aide à la décision
- * Systèmes multi-agents
- * Texture (modélisation, estimation, synthèse, ...)
- * Traitement d'image
- * Traitement des langues naturelles
- * Vision dynamique et active
- * Vision 3D et géométrie (calibration, mise en correspondance, reconstruction, ...)
- * Web sémantique

MODALITÉS DE SOUMISSION

Les auteurs sont invités à soumettre des contributions non publiées par ailleurs en langue française, mais pouvant toutefois être soumises simultanément à un congrès non francophone.

L'appel à communications détaillé et la procédure de soumission sont disponibles sur le serveur du congrès où les auteurs seront tenus de déclarer leur intention de soumission à partir du 15 mai 2003 : <http://www.laas.fr/rfia2004/>

Les textes complets doivent ensuite être adressés dans leur format définitif (entre 8 pages minimum et 10 maximum) avant le 20 juin 2003. Les articles acceptés devront être présentés lors du congrès sous forme de communication orale ou de poster et figureront dans les actes.

DATES IMPORTANTES

- * 15 mai 2003 : Déclaration de soumission sur le serveur du congrès
- * 20 juin 2003 : Date limite de réception des communications
- * 30 septembre 2003 : Notification aux auteurs
- * 30 octobre 2003 : Réception des versions définitives des articles
- * 26-27 janvier 2004: Tutoriels et Ateliers
- * 28-30 janvier 2004: Conférence RFIA 2004

SOMMAIRES DES REVUES

N'hésitez pas à envoyer un message à Brigitte Grau (grau@limsi.fr) pour lui indiquer toute suggestion permettant d'améliorer cette rubrique. Les revues figurant régulièrement au sommaire mais n'ayant pas de nouveau numéro apparaissent seulement avec leur nom et leur adresse WEB.

REVUE D'INTELLIGENCE ARTIFICIELLE

<http://ria.e-revues.com/>

REVUE D'INTELLIGENCE ARTIFICIELLE V17 N°1-3 2003

- cf. titres des articles sur le Web

ELECTRONIC TRANSACTIONS ON ARTIFICIAL INTELLIGENCE (ETAI)

Free publication and interactive reviewing on the internet
<http://www.ida.liu.se/ext/etai/>

ARTIFICIAL INTELLIGENCE

<http://www.elsevier.nl/inca/publications/store/5/0/5/6/0/1/>

ARTIFICIAL INTELLIGENCE V144 N°1-2 MARCH 2003

- Knowledge, action, and the frame problem, *R.B. SCHERL, H.J. LEVESQUE*
- A relevance restriction strategy for automated deduction, *D.A. PLAISTED, A. YAHYA*
- Robustness of regional matching scheme over global matching scheme [rapid communication], *L. CHEN, N. TOKUDA*
- Qualitative simulation of temporal concurrent processes using Time Interval Petri Nets, *V. BULITKO, D.C. WILKINS*
- A logic programming approach to knowledge-state planning, II: The DLV^K system, *T. EITER, W. FABER, N. LEONE, G. PFEIFER, A. POLLERES*
- Lessons from a failure: Generating tailored smoking cessation letters, *E. REITER, R. ROBERTSON, L.M. OSMAN*
- Developing reproducible and comprehensible computational models [rapid communication], *P.C.R. LANE, F. GOBET*
- Evaluating the effect of semi-normality on the expressiveness of defaults [rapid communication], *T. JANHUNEN*
- On market-inspired approaches to propositional satisfiability, *W.E. WALSH, M. YOKOO, K. HIRAYAMA, M.P. WELLMAN*

ARTIFICIAL INTELLIGENCE V145 N°1-2 APRIL 2003

- Multiobjective heuristic state-space planning, *I. REFANIDIS, I. VLAHAVAS*
- Region Connection Calculus: Its models and composition table, *S. LI, M. YING*
- Backjumping for Quantified Boolean Logic satisfiability, *E. GIUNCHIGLIA, M. NARIZZANO, A. TACHELLA*
- Establishing motion correspondence using extended temporal scope [rapid communication], *C.J. VEENMAN, M.J.T. REINDERS, E. BACKER*
- Towards an analytic framework for analysing the computation time of evolutionary algorithms, *J. HE, X. YAO*
- Multiple agent-based autonomy for satellite constellations, *T. SCHETTER, M. CAMPBELL, D. SURKA*
- Diagram processing: Computing with diagrams, *M. ANDERSON, R. MCCARTNEY*
- AI-A multiple book review, *U. FURBACH*
- BOB: Improved winner determination in combinatorial auctions and generalizations, *T. SANDHOLM, S. SURI*

ARTIFICIAL INTELLIGENCE V146 N°1 MAY 2003

- Controlling the learning process of real-time heuristic search, *M. SHIMBO, T. ISHIDA*
- Hybrid backtracking bounded by tree-decomposition of constraint networks, *P. JEGOU, C. TERRIOUX*
- Object-based visual attention for computer vision, *Y. SUN, R. FISHER*

AI MAGAZINE

<http://www.aaai.org/Magazine/magazine.html>

AI MAGAZINE V24 N°1 SPRING 2003

- Intelligent Control of a Water-Recovery System: Three Years in the Trenches, *P. BONASSO, D. KORTENKAMP, C. THRONESBERRY*
- The AAI-2002 Mobile Robot Competition and Exhibition, *H. A. YANCO, T. BALCH*
- The AAI-2002 Robot Rescue, *J. L. CASPER, M. J. MICHERE*
- The Robot Host Competition at the AAI-2002 Mobile Robot Competition, *D. A. GUSTAFSON, F. MICHAUD*
- The AAI-2002 Robot Challenge, *B. KUIPERS, A. STROUPE*
- The AAI-2002 Robot Exhibition, *I. HORSWILL*
- The 2002 Trading Agent Competition: An Overview of Agent Strategies, *A. GREENWALD*
- TAC-03--A Supply-Chain Trading Competition, *N. SADEH, R. ARUNACHALAM, J. ERIKSSON, N. FINNE., S. JANSON*

SOMMAIRES DES REVUES

COGNITIVE SCIENCE

<http://www.elsevier.com/inca/publications/store/6/2/0/1/9/4/>

COGNITIVE SCIENCE V27 N°2 MARCH 2003

- Modeling developmental transitions on the balance scale task, *H. VAN RIJN, M. VAN SOMEREN, H. VAN DER MAAS*
- Sequential sampling models of human text classification, *M.D. LEE, E.Y. CORLETT*
- Navigating joint projects with dialogue, *A. BANGERTER, H.H. CLARK*
- Intrinsic cognitive models, *J.A. WASKAN*
- Area activation: a computational model of saccadic selectivity in visual search [rapid communication], *M. POMPLUN, E.M. REINGOLD, J. SHEN*
- Better limited systematicity in hand than structural descriptions in the bush: A reply to Hummel, *S. EDELMAN, N. INTRATOR*
- Lexical effects on compensation for coarticulation: the ghost of Christmas past [rapid communication], *J.S. MAGNUSON, B. MCMURRAY, M.K. TANENHAUS, R.N. ASLIN*
- "Effective systematicity" in, "effective systematicity" out: a reply to Edelman and Intrator (2003), *J.E. HUMMEL*
- ERP evidence for task modulations on face perceptual processing at different spatial scales [rapid communication], *V. GOFFAUX, B. JEMEL, C. JACQUES, B. ROSSION, P.G. SCHYNS*

COGNITIVE SCIENCE V27 N°3 MAY-JUNE 2003

- Modeling of memory and perception, *R. M. SHIFFRIN*
- A Bayesian approach to the evolution of perceptual and cognitive systems, *W. S. GEISLER, R. L. DIEHL*
- Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression, *D. E. HUBER, R. C. O'REILLY*
- Spacing and repetition effects in human memory: application of the SAM model, *J. G. RAAIJMAKERS*
- Inferring causal networks from observations and interventions, *M. STEYVERS, J. B. TENENBAUM, E.-J. WAGENMAKERS*
- Data-driven approaches to information access, *S. DUMAIS*
- Controlled & automatic processing: behavior, theory, and biological mechanisms, *W. SCHNEIDER, J. M. CHEIN*

APPLIED ARTIFICIAL INTELLIGENCE

<http://www.tandf.co.uk/journals/tf/08839514.html>

APPLIED ARTIFICIAL INTELLIGENCE V17 N°2 FEBRUARY 2003

- Evolution Of The Safe Path For Ship Navigation, *X. ZENG*

- Analysis Of Induction Motor Fault Diagnosis With Fuzzy Neural Network, *M. ALEXANDRU*
- Temporal Constraint Satisfaction In Matrix Method, *C. ZHANG, S. ZHANG, Z. ZHANG*
- Adaptive Behaviorin Cellular Automata Using Rough Set Theory, *Y. HASSAN, E. TAZAKI*

APPLIED ARTIFICIAL INTELLIGENCE V17 N°3 MARCH 2003

- Introduction Concept Lattice-Based Theory, Methods And Tools For Knowledge Discovery In Databases: Applications, *E. MEPHU NGUIFO, V. DUQUENNE, M. LIQUIERE*
- Multi-Level Knowledge Discovery From Rule Bases, *D. RICHARDS, U. MALIK*
- The Cube Lattice Model And Its Applications, *L. CHAUDRON, N. MAILLE, M. BOYER*
- Structuration Of Phenotypes And Genotypes Through Galois Lattices And Implications, *V. DUQUENNE, C. CHABERT, A. CHERFOUH, et al.*
- Document Retrieval For E-Mail Search And Discovery Using Formal Concept Analysis, *R. COLE, P. EKLUND, G. STUMME*
- Conceptual Knowledge Discovery--A Human-Centered Approach, *J. HERETH CORREIA, G. STUMME, R. WILLE, et al.*

APPLIED ARTIFICIAL INTELLIGENCE V17 N°4 APRIL 2003

- Automatically Constructing Multi-Relationship Fuzzy Concept Networks For Document Retrieval, *Y.-J. HORNG, S.-M. CHEN, C.-H. LEE*
- Prediction And Modeling Of The Rainfall-Runoff Transformation Of A Typical Urban Basin Using Ann And Gp, *J. DORADO, J. R. RABUNAL, A. PAZOS, et al.*
- Neural Network Based Biomass And Growth Rate Estimation Aimed To Control Of A Chemostat Microbial Cultivation, *S. POPOVA, P. KOPRINKOVA, T. PATARINSKA*
- Neural Networks In Production Scheduling: Intelligent Solutions And Future Promises, *K. METAXIOTIS, J. PSARRAS*

COMPUTATIONAL INTELLIGENCE

<http://www.blackwellpublishers.co.uk/asp/journal.asp?ref=08247935&src=cts>

MINDS AND MACHINES

JOURNAL FOR ARTIFICIAL INTELLIGENCE PHILOSOPHY AND COGNITIVE SCIENCE

<http://www.wkap.nl/prod/j/0924-6495>

SOMMAIRES DES REVUES

MINDS AND MACHINES V13 N°2 MARCH 2003

- The Self as an Embedded Agent, *C. DOBBYN, S. STUART*
- How to Psychoanalyze a Robot: Unconscious Cognition and the Evolution of Intentionality, *D. LEVY*
- Self-Deception and Emotional Coherence, *B. SAHDRA, P. THAGARD*
- The Genealogy of the Moral Modules, *J. BOLENDER*
- A Reductio of Kripke-Wittgenstein's Objections to Dispositionalism about Meaning, *J. HOHWY*
- vExplanation by Computer Simulation in Cognitive Science, *J. FERNÁNDEZ*
- The Chinese Room Argument Reconsidered: Essentialism, Indeterminacy, and Strong AI, *J. C. WAKEFIELD*
- Book Review
- The Mind Doesn't Work That Way, Cambridge, MA: Bradford Books/MIT Press, 2000, Jerry Fodor, by *Y. WILKS*
- The Computer and the Brain, 2nd edition, John von Neumann, by *G. PICCININI*

ARTIFICIAL INTELLIGENCE REVIEW

<http://www.kluweronline.com/issn/0269-2821/contents>

AI REVIEW V19 N°2 MARCH 2003

- Sexual Selection for Genetic Algorithms, *K. SONG GOH, A. LIM, B. RODRIGUES*
- Knowledge Discovery from Decision Tables by the Use of Multiple-Valued Logic, *K.J. ADAMS, D.A. BELL, L.P. MAGUIRE, J. MCGREGOR*
- Application of an Expert System on the Fracture Mechanics of Concrete, *A. KAZAZ*

AI REVIEW V19 N°3 APRIL 2003

- Information Compression by Multiple Alignment, Unification and Search as a Unifying Principle in Computing and Cognition, *J.G. WOLFF*
- A Tabu-Based Exploratory Evolutionary Algorithm for Multiobjective Optimization, *K.C. TAN, E.F. KHOR, T.H. LEE, Y.J. YANG*

AI REVIEW V19 N°4 MAY 2003

- Interactive Visual User Interfaces: A Survey, *F. MURTAGH, T. TASKAYA, P. CONTRERAS, J. MOTHE, K. ENGLMEIER*
- A Taxonomy of Recommender Agents on the Internet, *M. MONTANER, B. LOPEZ, J. LLUIS DE LA ROSA*

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

<http://www.academicpress.com/ijhcs>

IJHCS V58 N°3 MARCH 2003

- Factors affecting the adoption and diffusion of XML and Web services standards for E-business systems, *M. CHEN*
- An empirical study of the effects of interactivity on web user attitude, *H.-H. TEO, L.-B. OH, C. LIU, K.-K. WEI*
- Comparing smart card adoption in Singapore and Australian universities, *C. HUI MIN LEE, Y. WING CHENG, A. DEPIK-KERE*
- Effects of content representation and readers' prior knowledge on the comprehension of hypertext, *H. POTELE, J.-F. ROUET*

IJHCS V58 N°5 MAY 2003

- Introduction: design and evaluation of notification user interfaces, *D.S. MCCRICKARD, M. CZERWINSKI, L. BARTRAM*
- Moticons: - detection, distraction and task, *L. BARTRAM, C. WARE, T. CALVERT*
- Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems, *D.S. MCCRICKARD, R. CATRAMBONE, C.M. CHEWAR, J.T. STASKO*
- Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal, *J.G. TRAF-TON, E.M. ALTMANN, D.P. BROCK, F.E. MINTZ*
- Notification and awareness: synchronizing task-oriented collaborative activity, *J.M. CARROLL, D.C. NEALE, P.L. ISENHOUR, M.B. ROSSON, D.S. MCCRICKARD*

COMPUTATIONAL LINGUISTICS

<http://mitpress.mit.edu/catalog/item/default.asp?sid=8563C099-9701-4DD2-85C8-8F3502E9C8AE&tttype=4&tid=10>

COMPUTATIONAL LINGUISTICS V28, N°4 - SPECIAL ISSUE ON SUMMARIZATION

- Introduction to the Special Issue on Summarization, *D. R. RADEV, E. HOVY, K. MCKEOWN*
- Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status, *S. TEUFEL, M. MOENS*
- Automatic Summarization of Open-Domain Multiparty Dialogues in Diverse Genres, *K. ZECHNER*
- Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization, *H. G. SILBER, K. F. MCCOY*
- Generating Indicative-Informative Summaries with SumUM, *H. SAGGION, G. LAPALME*
- vUsing Hidden Markov Modeling to Decompose Human-Written Summaries, *H. JING*
- Human Variation and Lexical Choice, *E. REITER, S. SRIPADA*

Book review

- Early Years in Machine Translation: Memoirs and Biographies of Pioneers edited by W. John Hutchins, *W. J. PLATH*
- Patterns of Text: In Honour of Michael Hoey edited by Mike Scott and Geoff Thompson, *G. HIRST*
- Word Frequency Distributions by R. Harald Baayen, *G. SAMPSON*
- Data-Driven Techniques in Speech Synthesis edited by R. I. Dampier, *T. DUTOIT*
- Empirical Linguistics by Geoffrey Sampson, S. P. ABNEY
- Computational Nonlinear Morphology with Emphasis on Semitic Languages by George Anton Kiraz, *M. WALTHER*

COMPUTATIONAL LINGUISTICS V29, N°1 - MARCH 2003

- Optimization Models of Sound Systems Using Genetic Algorithms, *J. KE, M. OGURA, W. S-Y WANG*
- A Systematic Comparison of Various Statistical Alignment Models, *F. J. OCH, H. NEY*
- Graph-Based Generation of Referring Expressions, *E. KRAHMER, S. VAN ERK, A. VERLEG*
- A Machine Learning Approach to Modeling Scope Preferences, *D. HIGGINS, J. M. SADOCK*
- Word Reordering and a Dynamic Programming Beam Search Algorithm for Statistical Machine Translation, *C. TILLMANN, H. NEY*
- Weighted Deductive Parsing and Knuth's Algorithm, *M-J. NEDERHOF*

Book review

- The Cambridge Grammar of the English Language by Rodney Huddleston and Geoffrey K. Pullum, *C. BREW*

NATURAL LANGUAGE ENGINEERING

<http://www.journals.cambridge.org/bin/bladerunner?REQUNIQ=976614197&REQSESS=5004848&116000REQEVENT=&REQINT1=5&REQSTR1=NLE&REQAUTH=0>

USER MODELING AND USER-ADAPTED INTERACTION

<http://www.wkap.nl/jrnltoc.htm/0924-1868>

USER MODELING AND USER-ADAPTED INTERACTION V12 N°4 NOVEMBER 2002

- Hybrid Recommender Systems: Survey and Experiments, *R. BURKE*
- Using Bayesian Networks to Manage Uncertainty in

Student Modeling, *C. CONATI, A. GERTNER, K. VANLEHN*

Book Review

- User Interfaces for All: Concepts, Methods and Tools, *B. A. NARDI*

COMPUTER SPEECH AND LANGUAGE

http://www.sciencedirect.com/science?_ob=JournalURL&_isn=08852308&_auth=y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=be00614a6a8826664cf3538182118628

COMPUTER SPEECH & LANGUAGE V17 N°2-3 APRIL-JULY 2003

- Introduction to the special issue on new computational paradigms for acoustic modeling in speech recognition, *M. J. RUSSELL, J. A. BILMES*
- Non-parametric probability estimation for HMM-based automatic speech recognition, *F. LEFÈVRE*
- A probabilistic framework for segment-based speech recognition, *J. R. GLASS*
- Modelling the uncertainty in recovering articulation from acoustics, *K. RICHMOND, S. KING, P. TAYLOR*
- Bayesian network structures and inference techniques for automatic speech recognition, *G. ZWEIG*
- Robust speech recognition and feature extraction using HMM2, *K. WEBER, S. IKBAL, S. BENGIO, H. BOURLARD*
- Buried Markov models: a graphical-modeling approach to automatic speech recognition, *J. A. BILMES*
- Parameter reduction schemes for loosely coupled HMMs, *H. J. NOCK, M. OSTENDORF*
- Dynamic Bayesian networks for multi-band automatic speech recognition, *K. DAOUDI, D. FOHR, C. ANTOINE*
- vSpeech recognition with unknown partial feature corruption – a review of the union model, *J. MING, F. J. SMITH*
- Special Issue on World Sense Disambiguation

MACHINE LEARNING

<http://www.wkap.nl/jrnltoc.htm/0885-6125>

MACHINE LEARNING V51 N°2 MAY 2003

- Variance and Bias for General Loss Functions, *G. M. JAMES*
- Learning from Different Teachers, *D. ANGLUIN, M. KRIJS*
- Microchoice Bounds and Self Bounding Learning Algorithms, *J. LANGFORD, A. BLUM*
- Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy, *L. I. KUNCHEVA, C. J. WHITAKER*

MACHINE LEARNING V51 N°3 JUNE 2003

- Guest Editor's Introduction, *C. GENTILE*
- Boosting and Hard-Core Set Construction, *A. R. KLIVANS, R. A. SERVEDIO*
- Potential-Based Algorithms in On-Line Prediction and Game Theory, *N. CESA-BIANCHI, G. LUGOSI*
- Estimating the Optimal Margins of Embeddings in Euclidean Half Spaces, *J. FORSTER, N. SCHMITT, H. U. SIMON, T. SUTTROP*
- A Theoretical Analysis of Query Selection for Collaborative Filtering, *S. DASGUPTA, W. SUN LEE, P. M. LONG*

NEURAL NETWORKS

<http://www.elsevier.com/inca/publications/store/8/4/1/>

NEURAL NETWORKS V16 N°2 MARCH 2003

- Neural Networks Letter - A complex-valued nonlinear neural adaptive filter with a gradient adaptive amplitude of the activation function, *A.I. HANNA, D.P. MANDIC*
- Self-organizing continuous attractor networks and motor function, *S.M. STRINGER, E.T. ROLLS, T.P. TRAPPENBERG, I.E.T. DE ARAUJO*
- Zonal function network frames on the sphere, *H.N. MHASKAR, F.J. NARCOWICH, J.D. WARD*
- On motion detection through a multi-layer neural network architecture, *A. FERNANDEZ-CABALLERO, J. MIRA, M.A. FERNANDEZ, A.E. DELGADO*
- Dual extended Kalman filtering in recurrent neural networks, *C.-S. LEUNG, L.-W. CHAN*
- Kalman filters improve LSTM network performance in problems unsolvable by traditional recurrent nets, *J.A. PEREZ-ORTIZ, F.A. GERS, D. ECK, J. SCHMIDHUBER*
- Intelligent optimal control with dynamic neural networks, *Y. BECERIKLI, A.F. KONAR, T. SAMAD*
- Clustering ensembles of neural network models, *B. BAKKER, T. HESKES*
- Neural network based tracking control of a flexible macro-micro manipulator system, *X.P. CHENG, R.V. PATEL*
- Book review
- Graphical Models: Foundations of Neural Computation - Eds. Michael I. Jordan and Terrence J Sejnowski, The MIT Press, 2001, *G.R. BRADSKI*
- Introduction to neural and cognitive modeling (second edition) - D.S. Levine, Lawrence Erlbaum Associates, Hillsdale, NJ, 2000, *H. OGMEN*
- 2004 Special Issue - VISION AND BRAIN

NEURAL NETWORKS V16 N°3-4 APRIL 2003

- Measuring solar magnetic fields with artificial neural networks, *H. SOCAS-NAVARRO*

- Application of the kernel method to the inverse geosounding problem, *H. HIDALGO, S. SOSA LEON, E. GOMEZ-TREVINO*
- Using self-organizing maps to identify potential halo white dwarfs, *E. GARCA-BERRO, S. TORRES, J. ISERN*
- Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting, *D. BARATTA, G. CICIONI, F. MASULLI, L. STUDER*
- Improved system for object detection and star/galaxy classification via local subspace analysis, *Z.-Y. LIU, K.-C. CHIU, L. XU*
- Evolving neural networks to identify bent-double galaxies in the FIRST survey, *E. CANTU-PAZ, C. KAMATH*
- A feature extraction unsupervised neural network for an environmental data set, *G. ACCIANI, E. CHIARANTONI, G. FORNARELLI, S. VERGURA*
- Astronomical image segmentation by self-organizing neural networks and wavelets, *J. NUNEZ, J. LLACER*
- Neural networks in astronomy, *R. TAGLIAFERRI, G. LONGO, L. MILANO, F. ACERNESE, F. BARONE, A. CIARAMELLA, R.D. ROSA, C. DONALEK, A. ELEUTERI, G. RAICONI, S. SESSA, A. STAIANO, A. VOLPICELLI*
- Some neural network applications in environmental sciences. Part II: advancing computational efficiency of environmental numerical models, *V.M. KRASNOPOLSKY, F. CHEVALLIER*
- Overview of independent component analysis technique with an application to synthetic aperture radar (SAR) imagery processing, *S. FIORI*
- Some neural network applications in environmental sciences. Part I: forward and inverse problems in geophysical remote measurements, *V.M. KRASNOPOLSKY, H. SCHILLER*
- Independent component analysis for artefact separation in astrophysical images, *M. FUNARO, E. OJA, H. VALPOLA*
- Ground-based telescope pointing and tracking optimization using a neural controller, *D. MANCINI, M. BRESCIA, P. SCHIPANI*
- Outlier detection in scatterometer data: neural network approaches, *R.J. BULLEN, D. CORNFORD, I.T. NABNEY*
- Fuzzy neural identification and forecasting techniques to process experimental urban air pollution data, *F.C. MORABITO, M. VERSACI*
- Introduction: Neural networks for analysis of complex scientific data: astronomy and geosciences, *R. TAGLIAFERRI, G. LONGO, B. D'ARGENIO, A. INCORONATO*
- Neural maps in remote sensing image analysis, *T. VILLMANN, E. MERENYI, B. HAMMER*
- Source separation in astrophysical maps using independent factor analysis, *E.E. KURUOGLU, L. BEDINI, M.T. PARATORE, E. SALERNO, A. TONAZZINI*

SOMMAIRES DES REVUES

ARTIFICIAL INTELLIGENCE IN MEDICINE

<http://www.elsevier.com/inca/publications/store/5/0/5/6/2/7/>

AI IN MEDICINE V27 N°2 FEBRUARY 2003

- Characterization of medical time series using fuzzy similarity-based fractal dimensions, *M. SARKAR, T.-Y. LEONG*
- Building a case-based diet recommendation system without a knowledge engineer, *A.S. KHAN, A. HOFFMANN*
- Decision support tools for clinical trial design, *S. MODGIL, P. HAMMOND*
- Self-organizing map for cluster analysis of a breast cancer database, *M.K. MARKEY, J.Y. LO, G.D. TOURASSI, C.E. FLOYD*
- Evolutionary computing for knowledge discovery in medical diagnosis, *K.C. TAN, Q. YU, C.M. HENG, T.H. LEE*
- Erratum to "Data mining a diabetic data warehouse" - [Artif.Intell. Med. 2002;26:37-54], *J.L. BREAUULT, C.R. GOODALL, P.J. FOS*

Book review

- Clinical Applications of Artificial Neural Networks – Richard Dybowski, Vanya Gant (Eds.), Cambridge University Press, Cambridge, UK, New York, Oakleigh, Madrid, Cape Town, 2001, *W. WIEGERINCK*

AI IN MEDICINE V27 N°3 MARCH 2003

- Software agents in health care, *A. MORENO, C. GARBAY*
- Anthropropic agency: a multiagent system for physiological processes, *F. AMIGONI, M. DINI, N. GATTI, M. SOMALVICO*
- A multi-agent intelligent environment for medical knowledge, *R.M. VICARI, C.D. FLORES, A.M. SILVESTRE, L.J. SEIXAS, M. LADEIRA, H. COELHO*
- Component-based mediation services for the integration of medical applications, *Y. XU, D. SAUQUET, P. DEGOULET, M.-C. JAULENT*

- Formalizing an electronic institution for the distribution of human tissues, *J. VAZQUEZ-SALCEDA, J.A. PADGET, U. CORTES, A. LOPEZ-NAVIDAD, F. CABALLERO*
- A multi-agent system approach for monitoring the prescription of restricted use antibiotics, *L. GODO, J. PUYOL-GRUART, J. SABATER, V. TORRA, P. BARRUFET, X. FABREGAS*
- Automated monitoring of medical protocols: a secure and distributed architecture, *T. ALSINET, C. ANSOTEGUI, R. BEJAR, C. FERNANDEZ, F. MANYA*

Book review

- Workflow Management-Models, Methods and Systems - Wil van der Aalst, Kees van Hee, MIT Press, Cambridge, MA/London, England, 2002, *S. QUAGLINI*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

<http://www.elsevier.com/inca/publications/store/5/0/5/7/8/7/>

IJAR V33 N°1 APRIL 2003

- A new understanding of subjective probability and its generalization to lower and upper prevision, *G. SHAFER, P.R. GILLET, R.B. SCHERL*
- Fuzzy active noise modeling and control, *J.M. SOUSA, C.A. SILVA, J.M.G. SA DA COSTA*
- Implicitly preserving semantics during incremental knowledge base acquisition under uncertainty, *E. SANTOS, E.S. SANTOS, S.E. SHIMONY*
- Extreme points of credal sets generated by 2-alternating capacities, *E. MIRANDA, I. COUSO, P. GIL*

Contrôle et Coordination Orientés Multi-Agent

Olivier Boissier

Habilitation à diriger des recherches de l'Ecole des Mines de Saint-Etienne. Soutenue le mercredi 12 février 2003 à l'EMSE

Résumé

Les recherches présentées dans cette Habilitation abordent deux préoccupations principales des Systèmes Multi-Agents : l'autonomie des agents et le comportement global cohérent du système qualifié notamment par les termes de coopération, de collaboration ou de compétition. Ces deux préoccupations sont abordées dans un contexte de systèmes décentralisés et ouverts où, de facto, aucune vision globale cohérente du système ne peut exister.

Afin de définir et de mettre en oeuvre des modèles et des outils répondant à ces préoccupations, deux directions principales ont été suivies : Coordination Orientée Multi-Agent et Contrôle Orienté Multi-Agent.

Nous avons ainsi élaboré des modèles d'Agent, d'Interaction (ACL, Protocoles, Coordination de négociations) et d'Organisation (Structure Organisationnelle, Dépendances, Réorganisation) pour mettre en place un comportement global cohérent au sein du système. Ces modèles ainsi que les différentes applications développées ont constitué un cadre d'étude de la Coordination Orientée Multi-Agent par combinaison de modèles d'(A)gent, d'(E)nvironnement, d'(I)nteraction et d'(O)rganisation. L'étude du Contrôle Orienté Multi-Agent s'est concrétisée par l'élaboration de différentes architectures d'agent intégrant ces modèles

A, E, I et O étant capables d'exhiber une autonomie vis à vis de l'environnement, des autres agents ou des organisations. Ces différentes réalisations ont également servi de cadre à l'étude du difficile problème de la confrontation entre contrôle de l'agent et coordination du système.

Olivier BOISSIER
Systemes Multi-Agents
SIMMO - ENSM.SE
158 cours Fauriel
42023 Saint-Etienne Cedex 02
Mél : Olivier.Boissier@emse.fr
<http://www.emse.fr/~boissier>

Un modèle d'analyseur syntaxique robuste fondé sur la modularité et la lexicalisation de ses grammaires

Núria Gala Pavia

Thèse de l'Université Paris 11, soutenue le 21 Mars 2003 au Limsi, effectuée à Xerox et au Limsi

Résumé

Ce travail s'inscrit dans le domaine du Traitement Automatique de la Langue (TAL) et plus spécialement celui de l'analyse syntaxique robuste (robust parsing).

La notion d'analyse syntaxique robuste s'est répandue dans les années 90 à la suite d'une effervescence générale pour les techniques informatiques d'analyse de corpus. Dans cette optique, l'analyse syntaxique consiste à associer automatiquement à la chaîne découpée en unités, une représentation des groupements structurels et/ou des relations fonctionnelles existant entre ces unités. En général, les traitements syntaxiques ne sont pas un but en soi : l'analyse syntaxique est ainsi vouée à la création d'outils permettant l'extraction d'information d'ordre linguistique qui pourra être exploitée par la suite par d'autres applications (par exemple, pour l'extraction d'information, pour la désambiguïsation sémantique ou pour les systèmes de question-réponse).

Placés dans un paradigme lié à la recherche d'information, les analyseurs de cette approche se heurtent à la difficulté de maintenir un équilibre entre la finesse de la description linguistique et l'efficacité de l'analyseur, ainsi qu'entre cette finesse descriptive et leur adéquation empirique. En effet, la volonté de traiter des grandes quantités de texte tout venant est parfois au détriment d'une analyse globalement précise. C'est-à-dire qu'on retrouve des phénomènes généralement bien modélisés par la plupart des analyseurs (marquage de syntagmes de base -syntagmes nominaux, verbaux, etc.-, extraction de la dépendance sujet et objet, etc.) mais un bon nombre d'autres structures ne sont pas prises en compte par leurs grammaires, ce qui fait diminuer considérablement leurs performances. Par exemple, nous avons créé des grammaires spécialisées pour le traitement de structures mettant en jeu des paramètres typodispositionnels (listes et titres).

Ces aspects nous ont amenée à réfléchir sur un modèle d'analyseur robuste qui puisse traiter avec précision une grande variété de corpus et phénomènes, ceci tout en conservant une architecture efficace pour le traitement de grands volumes de données.

Nous avons proposé (décrit et implémenté) un modèle

d'analyseur robuste qui maîtrise la variété des phénomènes présents dans des corpus tout venant (quel que soit leur genre ou leur domaine) et qui garantit la qualité des analyses produites. Pour ce faire, l'idée de base a été le traitement de l'analyse en deux grandes étapes, selon les caractéristiques de chaque unité en entrée. A la différence de la plupart des analyseurs existants, toute phrase (au sens large, incluant titres et listes) est "évaluée" de façon à estimer les types de phénomènes présents, pour être traités par la suite par une grammaire particulière. L'analyseur est ainsi composé de différents modules grammaticaux garantissant un traitement spécifique de chaque phénomène.

De plus, nous nous sommes proposée de modifier les caractéristiques du système initial (système symbolique, écriture manuelle des règles) par l'ajout d'un mécanisme d'apprentissage automatique à partir d'un grand corpus, dans le but de lexicaliser la grammaire de base et améliorer ainsi l'extraction de certaines dépendances. En effet, les systèmes actuels d'analyse syntaxique sont fondés sur les parties de discours et n'exploitent que très peu d'informations d'ordre lexical, telles que les rections ou les fréquences de cooccurrences des mots. Cette omission d'informations de type lexical dans une première phase de l'analyse permet d'assurer la robustesse, l'efficacité et la large couverture des analyseurs, mais elle est péna-

lisante si l'on veut une analyse plus riche et plus précise. Dans notre modèle, l'incorporation d'informations d'ordre lexical, puisées automatiquement dans de grandes collections de textes comme le WWW, permet d'enrichir la grammaire et de traiter de façon pondérée l'extraction de dépendances liées au rattachement prépositionnel. Dans notre système, nous avons donc combiné une description structurale riche (produite par l'analyseur initial) avec des informations lexicales et statistiques obtenues à partir du Web.

Le résultat de notre travail est un modèle d'analyseur à large couverture, robuste et modulaire, avec la particularité de donner des indices de fiabilité sur l'ensemble de structures et phénomènes analysés. Devant la difficulté d'obtenir des résultats toujours précis pour des corpus hétérogènes, notre choix a été de donner des informations sur la fiabilité de l'analyse d'une phrase traitée par l'analyseur. Ainsi, selon que cette analyse a été calculée par une grammaire ou autre, le résultat pourra être utilisé différemment dans d'autres applications.

Mots clés : traitement automatique de la langue, analyse syntaxique robuste à large couverture, spécialisation et lexicalisation des grammaires, apprentissage non-supervisé pour la désambiguïsation structurelle.

Dossier « Systèmes d'information, connaissances et IA »

A paraître dans le bulletin numéro 55, fin juin 2003

Coordonnateurs :

Danielle Boulanger db@sunlyon3.univ-lyon3.fr,
Eunika Mercier-Laurent eml@wanadoo.fr, Manuel Zaklad manuel.zaklad@utt.fr

Objectifs

Depuis l'invention de l'ordinateur les technologies de l'électronique et celles de la programmation ont évolué. Dans notre vision, celle de l'Innovation à partir des connaissances (Knowledge Innovation[®]), l'ordinateur connecté traite des connaissances et travaille en symbiose avec l'utilisateur, il est son assistant intelligent. Il fait partie d'un système d'information de l'organisation étendue et permet d'apprendre en permanence des échanges dans ces réseaux. Ce dossier a pour l'objectif de faire le point sur les approches et les techniques de l'IA symbolique et numérique pouvant moderniser l'informatique dans les entreprises et les organisations ainsi que garantir la sécurité du système et le paiement sécurisé. Selon cette vision, les utilisateurs doivent pouvoir déléguer à l'ordinateur des tâches pour lesquelles il est meilleur que l'humain, comme les suivantes :

- la mémorisation ;
- l'indexation et la recherche efficace dans une grande quantité d'informations et de connaissances ;
- la découverte des connaissances ;
- l'association floue de l'offre et demande ;
- l'organisation et le traitement du retour d'expériences ;
- la prise en compte d'un grand nombre de paramètres pour la prise de décision
- ...

Toutes les approches et les techniques de l'IA sont concernées.

Qui peut soumettre ?

Toute personne et équipe concernée par le sujet : recherche public ou privée, industriels et organisations, impliqués dans des actions d'acquisition, modélisation, collectivisation et traitement des connaissances, ayant pour objectif de faciliter la vie de l'utilisateur. Les spécialistes de l'architecture d'ordinateurs et ceux de sciences humaines sont tout particulièrement invités.

Comment soumettre ?

Les contributions devront présenter succinctement l'équipe et les principaux travaux concernant le thème IA et Systèmes d'Information : concepts, techniques, méthodes, outils, expériences, applications mises en œuvre. Elles seront organisées selon le schéma suivant :

- Identification de l'équipe: noms, adresse, site Web, personne à contacter (téléphone, mél) ;
- Membres de l'équipe concernés par le thème ;
- Contexte global de recherche/application ;
- Description des travaux ou projets en lien avec le thème du dossier ;
- Courte bibliographie ciblée (5 références maximum) et adresse d'un site Web où l'on peut trouver l'ensemble des références et articles.

Présentation

Les contributions feront **1400** mots au maximum et devront être envoyées par courrier électronique au format Word ou RTF. Aucune mise en forme particulière ne doit être faite sur les textes autre que gras, italique et la taille des caractères. Si vous êtes concernés par cet appel, merci d'envoyer votre contribution :

***** *avant le 5 juin 2003* *****

à l'adresse suivante : eml@wanadoo.fr.

Pour tout renseignement, s'adresser par courrier électronique, à l'un des coordonnateurs.

Adhésion individuelle et abonnement Demande Renouvellement

Nom : Prénom :
 Affiliation :
 Adresse postale :
 N° de téléphone : N° de télécopie :
 Adresse électronique :
 Activités (à titre professionnel / à titre privé (*rayez la mention inutile*)) :

	Consultation du bulletin sur WEB (pour une personne)	Envoi du bulletin papier + un accès pour consultation du bulletin sur WEB
<input type="checkbox"/> Adhésion simple :	30 Euros	60 Euros
<input type="checkbox"/> Adhésion étudiant (sur justificatif) :	15 Euros	30 Euros
<input type="checkbox"/> Adhésion de soutien	Sans objet	90 Euros
<input type="checkbox"/> Abonnement au bulletin sans adhésion	Sans objet	55 Euros

- Adhésion au collège *IAD-SMA* : ajouter 7,5 Euros pour les étudiants, 15 Euros pour les autres.
 Adhésion au collège *Cafe (Apprentissage)* : gratuit.

Adhésion Personne morale Demande Renouvellement

Organisme :
Adresse postale commune aux bénéficiaires couverts par cette adhésion :
 Nom et prénom du représentant : Fonction :
 Mél : Tél : Fax :
 Adresse postale :

Le tarif d'adhésion comprend une partie fixe et une partie par bénéficiaire
Coordonnées des bénéficiaires (10 maximum) :

NOM, prénom	Mél.	Tél.	Fax

	Tarif de base fixe :	Tarif par bénéficiaire :
<input type="checkbox"/> Laboratoires universitaires	100 Euros	30 Euros
<input type="checkbox"/> Personnes morales non universitaires	300 Euros	30 Euros
<input type="checkbox"/> Adhésion de soutien	600 Euros	Sans objet

- j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA.
 j'accepte que les renseignements ci-dessus soient transmis à l'ECCA pour constituer un fichier européen.

Veillez trouver un règlement (à l'ordre de l'AFIA) de Euros

Trésorier AFIA : Marc AYEL, LIA-Université de Savoie, 73376 Le Bourget du Lac cedex.

Mode d'adhésion :

De préférence, en ligne via le site Internet de l'AFIA : <http://www.afia-france.fr>

A défaut, cette page doit être envoyée au trésorier.

Modes de paiement :

- 1) par chèque, à l'ordre de l'AFIA, envoyé au trésorier.
- 2) par bon de commande administratif, à l'ordre de l'AFIA, envoyé au trésorier.
- 3) Par virement bancaire sur le compte de l'AFIA : Société Générale, Résidence du Val de Seine, 78430 LOUVECIENNES. Code banque 30003, code guichet 01902, numéro de compte 00037283856 clef RIB 3.

TVA non applicable, article 293B du CGI

Les dossiers du *Bulletin de l'AFIA*

Web sémantique	Bulletin n°54
L'IA dans le RNTL	Bulletin n°53
IA et diagnostic.....	Bulletin n°52
Temps, espace et évolutif	Bulletin n°51
Equipes d'IA en France	Bulletin n°49/50
IA et Médecine	Bulletin n°48
Fouille de données	Bulletin n°46/47
IA et document	Bulletin n°44
IA et connexionisme	Bulletin n°43
IA et Vie Artificielle.....	Bulletin n°42
IA et CHM	Bulletin n°41
IA et EIAH.....	Bulletin n°40
Plates-formes multi-agents	Bulletin n°39
IA et WEB.....	Bulletin n°38
Mémoires d'entreprises.....	Bulletin n°36
IA et logique	Bulletin n°35
Ingénierie des connaissances.....	Bulletin n°34
IA et Télécommunications.....	Bulletin n°33
IA et Terminologie	Bulletin n°32
Décision et IA	Bulletin n°31
Raisonnement IA et Image.....	Bulletin n°30
Raisonnement temporel et spatial	Bulletin n°29
Systèmes Multi-agents	Bulletin n°28
IA et robotique.....	Bulletin n°27
I.A . et biologie moléculaire	Bulletin n°26
A. et droit	Bulletin n°25
I.A. et fusion de données	Bulletin n°24
I.A. et musique	Bulletin n°23
Apprentissage	Bulletin n°22
Les explications dans les SBC.....	Bulletin n°20
Pétrole-Chimie.....	Bulletin n°19
Le raisonnement à partir de cas	Bulletin n°18
I.A. et temps-réel.....	Bulletin n°17
Planification et action.....	Bulletin n°16
Traitement automatique des langues	Bulletin n°15
I.A. et médecine.....	Bulletin n°14
Diagnostic à base de modèles	Bulletin n°13
Validation des SBC	Bulletin n°12
Le connexionisme	Bulletin n°11
I.A. et jeux	Bulletin n°10
E.I.A.O.	Bulletin n°9
I.A. et gestion	Bulletin n°8
Conception et I.A.	Bulletin n°7
Intelligence artificielle distribuée.....	Bulletin n°6
Acquisition des Connaissances.....	Bulletin n°5
IA et ordonnancement.....	Bulletin n°4

SOMMAIRE DU BULLETIN N° 54

La vie de l'A.F.I.A.	4
L'IA au Québec	6
Présentation de Laboratoires	9
Dossier IA et RNTL	25
Conférences	43
Sommaires des Revues	49
Résumés Habilitations et Thèses	55
Appels à dossiers	61

CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Eté</i>
Réception des contributions: 15 décembre	Réception des contributions: 15 juin
Sortie le 31 janvier	Sortie 31 juillet

<i>Printemps</i>	<i>Automne</i>
Réception des contributions: 15 mars	Réception des contributions: 15 septembre
Sortie le 30 avril	Sortie le 31 octobre