

BULLETIN

DE

L'AFIA



AVRIL 2006

N° 61

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 :

Michèle SEBAG

L.R.I., Bât. 490,

Université Paris-Sud,

91405 ORSAY

Tel : +33 (0)169 15 76 02

Fax : +33 (0)1 69 15 65 86

Michele.Sebag@lri.fr

<http://www.lri.fr/~sebag>

Serveur WEB

<http://afia.lri.fr>

Adhésions, Liens avec les adhérents

Jérémie MARY

LRI, Bâtiment 490

Université Paris-Sud

91405 ORSAY

Mél. : Jeremie.Mary@lri.fr

Contributions au bulletin

Jérôme EUZENAT

Philippe MORIGNOT

voir ci-contre

Parrainage de manifestations

Marc-Philippe HUGET

Voir ci-contre

Membres d'honneur

Marie-Odile Cordier (1999), Jean-Paul Haton (1999), Jacques Pitrat (1999),

Jean-Marc David (2000), Daniel Kayser (2000), Claude Vogel (2000),

Henri Farreny (2001), Alain Colmerauer (2002), Jean-Louis Laurière (2002),

Gérard Sabah (2003), Jean-Claude Latombe (2004), Yves Kodratoff (2004),

Malik Ghallab (2005).

Personnes morales adhérentes à l'AFIA

ADIT, AI*IA, CEA, 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

Michèle SEBAG, président

Serge DUPUY, secrétaire

Jérémie MARY, trésorier

Comité :

Emmanuel d'ANGELO, Jean-Paul BAQUIAST, Jacques BLANC-TALON,

Gilles BISSON, Laurent CANDILLIER, Alain CARDON, Amal EL FALLAH,

Jérôme EUZENAT, Olivier FRANCOIS, Emmanuel GUERE,

Marc-Philippe HUGET, Christophe JACQUEMIN, Nicolas LACHICHE,

Vincent LEMAIRE, Patrick MARTY, Eunika MERCIER-LAURENT,

Philippe MORIGNOT, Jean-Denis MULLER, Gérald PETITJEAN,

Sylvie PESTY, Michèle SEBAG, Marc SCHOENAUER, Olivier TEYTAUD,

Fabien TORRE, Samuel WIECZOREK, Jean-Daniel ZUCKER.

Comité de rédaction

Jérôme EUZENAT

Co-rédacteur en chef

INRIA Rhône-Alpes

655, avenue de l'Europe

Montbonnot St-Martin,

38334 Saint-Ismier

jerome.euzenat@inrialpes.fr

Philippe MORIGNOT

Co-rédacteur en chef

AXLOG Ingénierie

19-21, rue du 8 mai 1945,

94110 ARCUEIL

philippe.morignot@axlog.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

« Conférences et comptes rendus »

ESIA-LISTIC

Université de Savoie

B.P. 806

74016 Annecy cedex

Marc-philippe.Huget@univ-savoie.fr

Amedeo NAPOLI

Rubrique « livres »

LORIA

B.P. 239

F-54 506 Vandœuvre lès Nancy

Amedeo.Napoli@loria.fr

Nicole TOURIGNY

Rubrique « IA au Québec »

Université Laval - Pavillon Adrien-Pouliot

Québec, Canada, G1K 7P4

tourigny@ift.ulaval.ca



Réa. Numérocopie – 01 46 31 64 53

ISSN 1273-1323

Dépôt légal à parution

Surfez intelligent !

Le dossier de ce Bulletin est consacré à l'ingénierie du Web, hors Web sémantique (pour cela, voir le numéro 54 du Bulletin). Il est coordonné par Gilles Venturini (LI, Université François-Rabelais de Tours) et Chantal Reynaud ENSC (Université Paris-Sud XI, CNRS (LRI) & INRIA (Futurs)).

Ce Bulletin reproduit une interview de Bradley Efron (Stanford) dans le domaine des statistiques, et sur la genèse de ses méthodes du bootstrap et du jackknife.

Amedeo Napoli (LORIA) nous offre la recension du livre de G. S. Boolos et R. C. Jeffrey sur la computabilité & logique.

Marc Boulle (France Télécom R&D) nous propose un compte-rendu de la quatrième conférence internationale sur l'apprentissage et la fouille de données (MLDM'05), tenue en juillet dernier à Leipzig. Sébastien Laborie et Jérôme Pierson (INRIA Rhône-Alpes) rendent compte de la conférence sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'06), tenue fin janvier 2006 à Tours. Enfin, Olivier Teytaud (Université Paris-Sud) relate les 24e Journées d'Arithmétiques Faibles, tenues fin mai 2005 à Fontainebleau.

Ensuite le Comité Stratégie (composé de Serge Dupuy, Jean-Paul Baquiast et Jean-Denis Muller) a achevé ses réflexions sur les termes en lesquels il fallait penser l'I.A., ce qui fera l'objet d'une communication dans un prochain Bulletin. Enfin, toujours dans l'optique d'une plus grande synergie entre les internautes et notre site web (<http://www.afia-france.org>), un Comité Web (dirigé par Marc-Philippe Huet) finit d'effectuer une refonte complète de ce site (PortAI), pour une plus grande convivialité et ergonomie. Nous prévoyons toujours de mettre la table des matières des Bulletins en ligne sur ce site (et non plus seulement l'éditorial), avec éventuellement la liste des laboratoires et dossiers déjà présentés et toujours d'autres surprises encore.

Nous renouvelons toujours notre appel à contribution : si vous, cher lecteur, avez un sujet que vous aimeriez faire partager à la communauté française d'I.A. (chants d'amour ou coups de gueule, ou plus sérieusement recension de livres que vous avez aimés, compte-rendus de conférences que vous avez aimées, etc), n'hésitez pas à nous faire parvenir vos textes, nous serons ravis de les publier.

Philippe MORIGNOT et Jérôme EUZENAT

Rédacteurs en chef du Bulletin de l'AFIA

Nous publions ici les résumés des deux articles qui ont reçu le prix de l'AFIA aux Rencontres Jeunes Chercheurs en Intelligence Artificielle de 2005 (RJCIA'05).

Un cadre théorique et pratique commun aux formalismes SAT et CSP n-aires.

Lionel Paris, Belaïd Benhamou, Pierre Siegel
Laboratoire des Sciences de l'Information et des Systèmes (LSIS),
Équipe Inférence Contraintes et Application (InCA),
CMI, Université de Provence,
Technopôle de Château Gombert,
39, rue Joliot Curie, 13453 Marseille Cedex 13
<http://www.lsis.org> et <http://www.cmi.univ-mrs.fr>

{Lionel.Paris, Belaïd.Benhamou, Pierre.Siegel}@cmi.univ-mrs.fr

L'approche « contraintes » est un des moyens utilisés pour formaliser et résoudre des problèmes en Intelligence Artificielle. Les CSPs binaires à domaines finis discrets est un des formalismes les plus utilisés dans ce cadre. On ne perd pas de généralité en théorie, car le problème de décision de consistance de CSP binaires est NP-complet. Mais, cette restriction aux contraintes binaires rencontre ses limites pour la formalisation et la résolution de problèmes réels. La plupart de ces problèmes s'expriment naturellement sous forme de contraintes n-aires quelconques et il est souvent difficile de leur trouver une représentation sous forme de CSP binaires. Le cadre général est plus expressif, mais les outils de résolution de CSPs n-aires sont moins développés et moins efficaces que ceux des CSP binaires.

Pour traiter les CSP n-aires, deux solutions ont été proposées.

- Transformer les CSPs n-aires en CSPs binaires, et utiliser les outils de CSPs binaires;
- Généraliser les techniques de résolution de CSPs binaires aux CSPs n-aires.

La deuxième approche est plus prometteuse, mais n'a pas encore donné de résultats satisfaisants.

Une troisième voie originale que nous voulons suivre consiste à coder toute instance CSP n-aires en un ensemble de clauses propositionnelles généralisées et d'adapter les techniques classiques de SAT pour les résoudre.

Dans ce papier, nous étudions une généralisation de la forme CNF qui permet de coder efficacement les CSP n-aires sous une forme booléenne sans accroissement de la taille du problème. Plus précisément, nous montrons que la complexité en espace de ce codage est identique à celle de la représentation sous forme de CSP. Nous introduisons une nouvelle règle d'inférence, dont la saturation équivaut à l'application de la consistance d'arc (en temps linéaire) pour les CSP n-aires exprimés dans ce codage booléen.

Deux méthodes énumératives pour la résolution de problèmes ainsi exprimés sont proposées : la première (équivalente à MAC dans les CSP) maintient la propriété de consistance d'arc à chaque nœud de l'arbre de recherche, tandis que la seconde (équivalente à FC) ne maintient qu'une forme partielle de consistance d'arc à chaque nœud. Ces deux méthodes sont expérimentées et comparées sur des instances du problème de Ramsey, ainsi que sur des CSP d'arités 3/4 générés aléatoirement. Des résultats satisfaisants sont obtenus.

Contrôle dynamique d'agents autonomes

Caroline Chopinaud

Thales Systèmes Aéroportés – 2 avenue Gay Lussac 78851 Elancourt Cedex
LIP6 – Laboratoire informatique de Paris 6 – 8 rue du Capitaine Scott 75015 Paris
caroline.chopinaud@fr.thalesgroup.com

Nous nous plaçons dans le cadre d'applications critiques, telles que les systèmes embarqués et les systèmes temps réels, que nous souhaitons concevoir à l'aide des systèmes multiagents (SMA), en particulier dans l'objectif de réduire la complexité de conception de telles applications. Dans un tel contexte, il est indispensable de pouvoir avoir confiance dans le comportement du SMA. Or l'émergence du comportement du système due à la mise en relation des agents le constituant, peut engendrer l'apparition de comportements incompatibles avec la bonne exécution du système. De même, l'autonomie que l'on accorde aux agents, c'est-à-dire la capacité à prendre seul leurs décisions, accroît les risques de voir apparaître, ce que nous appellerons par la suite, des **comportements indésirables**.

L'utilisation des méthodes classiques de validation, telles que le model-checking, les tests ou la démonstration automatique, n'est pas suffisante si l'on souhaite garantir que, une fois mis en condition réelle d'exécution, un SMA ne générera pas de comportements indésirables. Nous proposons donc une approche complémentaire d'auto-surveillance et d'autorégulation permettant aux agents de contrôler leur propre comportement. Ce contrôle d'agent se divise en trois étapes :

- 1- La surveillance du comportement d'un agent ou d'un groupe d'agents.
- 2- La détection de l'apparition des comportements indésirables résultant d'un agent ou d'un groupe d'agents.
- 3- La régulation du comportement du ou des agents incriminés.

Pour la surveillance du comportement des agents nous nous sommes intéressés au monitoring software. La surveillance d'un programme se fait par instrumentation, en insérant des sondes logicielles dans le code pour détecter certains événements. Cette instrumentation peut être faite manuellement mais cela s'avère

assez complexe : l'insertion des sondes dans un programme prend du temps, est difficile et sujette à l'erreur. Nous proposons donc d'effectuer cette instrumentation de façon automatique à partir de la spécification de liens entre des concepts représentant l'application et le code des agents.

Pour détecter les comportements indésirables nous utilisons des lois. Ces dernières définissent des comportements ou des situations désirées ou redoutées au sein d'un agent ou d'un groupe d'agents. Les lois peuvent être assimilées aux exigences primordiales au bon fonctionnement du système. Ainsi, détecter un comportement indésirable revient à détecter la transgression d'une loi.

La régulation quant-à-elle sera effectuée par les agents eux-mêmes à partir de leur capacité de raisonnement. Lorsqu'une violation de loi est détectée, l'agent est prévenu de cette transgression, puis un comportement de régulation est suivi par l'agent dans la perspective de se soustraire au comportement indésirable ayant entraîné la transgression de la loi.

Pour mettre en place ce contrôle au sein du système, nous proposons de générer automatiquement des agents capables de contrôler leur propre comportement. L'application du contrôle se fait alors de la façon suivante :

- Description de l'application. Le modèle d'agent et le système sont décrits à l'aide de concepts. Nous fournissons un ensemble de concepts de base qui pourra être étendu à l'aide de sous-concepts et d'instances pour affiner la description.
- Description des lois. Nous fournissons un langage de description utilisant les opérateurs déontiques d'interdiction et d'obligation. Les lois sont définies à l'aide de ce langage et portent sur les concepts utilisés pour décrire l'application.

- Génération automatique des agents auto-contrôlés. Le point central de cette étape consiste à générer automatiquement, à partir d'une loi, le réseau de Petri la représentant. Ce réseau sera utilisé pour vérifier que le comportement du ou des agents concernés respecte la loi en utilisant la technique de l'observateur de M. Diaz (1979), c'est-à-dire en exécutant en parallèle les agents et les réseaux de Petri, représentant les lois qui leur sont associées, pour détecter des incohérences.

Les agents générés ont une architecture introspective leur permettant de surveiller leur comportement et

de détecter les transgressions des lois. Cette architecture est divisée en deux parties : une partie comportement, regroupant le comportement réel de l'agent instrumenté dans le but d'en effectuer la surveillance, ainsi que les stratégies de régulation fournies par le développeur ; une partie contrôle incluant les réseaux de Petri représentant les lois auxquelles l'agent est soumis et se chargeant de détecter les transgressions de lois. Ce mécanisme de génération sera implanté dans un framework, SCAAR, fournissant de plus les outils nécessaires à la description de l'application et des lois.

Nouvelles du Québec

Nicole Tourigny

Voici les résumés des thèses soutenues depuis juin 2005 au département d'informatique et de génie logiciel (<http://www.ift.ulaval.ca>) de l'Université Laval (<http://www.ulaval.ca>). Ces travaux ont été réalisés dans le cadre du Laboratoire des Systèmes Intelligents (LSI). Ces résumés reprennent ceux utilisés pour l'annonce de la soutenance par l'auteur concerné. LSI est composé de plusieurs équipes de professeurs chercheurs, animées par un ou plusieurs professeurs : l'équipe AWSA (Agents du Web Sémantique et Applications) par Dr Mamadou Koné (Kone.Mamadou@ift.ulaval.ca), l'équipe DAMAS (Dialogue, Agents et Multiagents) Dr Brahim Chaib-draa (Chaib@ift.ulaval.ca), l'équipe ERICAE (Équipe de Recherche en Ingénierie des ConnAissancEs (ERICAE) par Dr Laurence Capus (capus@ift.ulaval.ca) et Dr Nicole Tourigny (Nicole.Tourigny@ift.ulaval.ca), l'équipe ICG (Informatique Cognitive et Géomatique) par Dr Bernard Moulin (Bernard.Moulin@ift.ulaval.ca), l'équipe de recherche en intelligence computationnelle par Dr Guy Mineau (Guy.Mineau@ift.ulaval.ca) et l'équipe de traitement de l'image par Dr Jean-Marie Beaulieu.

Titre : **Aspects cognitifs des dialogues entre agents artificiels : l'approche par la cohérence cognitive**

Diplôme: Ph.D.

Auteur : Philippe PASQUIER

Soutenance : 30 juin 2005

Groupe : LSI-DAMAS

Examineurs :

Brahim Chaib-draa, Ph.D. (directeur de thèse), Université Laval.

Jean Caelen, Ph.D. (examineur), CNRS, CLIPS – GEOD, Grenoble, France. Luc Lamontagne, Ph.D. (examineur), Université Laval.

Jean-Paul Sansonnet, Ph.D. (examineur), LMSI-AMI, Paris, France

Résumé :

Les cadres interactionnels actuels pour les communications entre agents cognitifs (protocoles, stratégies de conversation, jeux de dialogue, ...) garantissent la cohérence structurelle des conversations tenues. Pourtant, ce n'est pas tant l'habileté des agents à structurer leurs conversations qui nous intéresse que leurs aptitudes à tenir de manière autonome des conversations utiles quant à leurs objectifs individuels et collectifs. En particulier, les aspects cognitifs de la pragmatique des dialogues entre agents n'ont pas été étudiés dans le cadre des nouvelles approches reposant sur les engagements sociaux. Pour traiter cette problématique, nous avons défini et implanté un modèle complet de la

communication entre agents qui couvre les quatre dimensions classiques de la communication langagière : syntaxe, structure, sémantique et pragmatique. Les contributions principales de cette thèse sont les suivantes : 1. un modèle de l'engagement social flexible et de son respect. Ce modèle capture les aspects sémantiques de la communication ; 2. le langage de communication agent DIAGAL, reposant sur les jeux de dialogue, qui couvre les dimensions syntaxique et structurelle des conversations ; 3. un modèle motivationnel de type homéostatique qui fonde les aspects cognitifs de la pragmatique des dialogues entre agents ; 4. une validation informatique du modèle dans le contexte des agents BDI (Beliefs, Desire and Intentions) implantée grâce à une plate-forme commerciale de développement multiagent appelée JACKÔ.

Titre : **Prise de décision distribuée et coordination de tâches dans des environnements multiagents dynamiques et incertains**

Diplôme: Ph.D.

Auteur : Sébastien PAQUET

Soutenance : 19 décembre 2005

Groupe : LSI-DAMAS

Examineurs :

Brahim Chaib-draa, Ph.D. (directeur de thèse), Université Laval.

François Laviolette, Ph.D. (examineur), Université Laval.

Roland P. Malhamé, Ph.D. (examineur), École Polytechnique de Montréal.

Joëlle Pineau, Ph.D. (examinatrice), McGill University.

Résumé :

La prise de décision dans l'incertain et la coordination entre agents sont au cœur des systèmes multiagents. Dans ce type de systèmes, les agents doivent percevoir leur environnement et prendre des décisions de manière à se coordonner les uns avec les autres. Lorsque l'environnement est partiellement observable, les agents doivent être en mesure de gérer cette incertitude pour prendre les décisions les plus éclairées possible. Dans cette thèse, nous considérons des environnements multiagents coopératifs complexes (dynamiques, incertains et en temps-réel). Pour ce type d'environnements, nous proposons une approche de prise de décision dans un environnement incertain permettant une coordination flexible entre agents. Notre approche est bâtie autour d'un algorithme de résolution en ligne utilisant les processus de décision de Markov partiellement observables (POMDPs). Dans les environnements précédemment évoqués, les tâches que doivent accomplir les agents sont généralement fort complexes. Il est alors très difficile pour les agents de déterminer le nombre de ressources nécessaires à l'accomplissement de chacune de leurs tâches. Pour résoudre ce problème, nous proposons un algorithme permettant aux agents d'apprendre le nombre de ressources nécessaires à l'accomplissement de leurs tâches selon les caractéristiques de ces mêmes tâches. Dans le même ordre d'idée, nous proposons aussi une méthode d'ordonnement permettant d'ordonner les différentes tâches des agents de manière à maximiser le nombre de tâches pouvant être accomplies dans un temps limité. Toutes ces approches ont pour but de permettre la coordination d'agents pour l'accomplissement efficace de tâches complexes dans un environnement multiagent partiellement observable, dynamique et incertain. De telles approches ont démontré leur efficacité au moyen de tests effectués dans l'environnement de simulation de la RoboCupRescue.

Titre : **A Multi-Agent System for On-The-Fly Web Map Generation and Spatial Conflict Resolution**

Diplôme: Ph.D.

Auteur : Nafaâ JABEUR

Soutenance : 9 janvier 2006

Groupe : LSI-ICG

Examineurs :

Bernard Moulin, Ph.D. (directeur de thèse), Université Laval.

Yvan Bédard, Ph.D. (codirecteur de thèse), Dép. des sciences géomatiques, Université Laval.

Jean-Marie Beaulieu, Ph.D. (examineur), Université Laval.

Christophe Claramunt, Ph.D. (examineur), Naval Academy Research Institute, France.

Sylvain Giroux, Ph.D. (examineur), Université de Sherbrooke.

Résumé :

Internet est devenu un moyen de diffusion de l'information géographique par excellence. Il offre de plus en plus de services cartographiques accessibles par des milliers d'internautes à travers le monde. Cependant, la qualité de ces services doit être améliorée, principalement en matière de personnalisation. A cette fin, il est important que la carte générée corresponde autant que possible aux besoins, aux préférences et au contexte de l'utilisateur. Ce but peut être atteint en appliquant les transformations appropriées, en temps réel, aux objets de l'espace à chaque cycle de génération de la carte. L'un des défis majeurs de cette génération de la carte à la volée est la résolution des conflits spatiaux qui apparaissent entre les objets, essentiellement à cause de l'espace réduit des écrans d'affichage. Dans cette thèse, nous proposons une nouvelle approche basée sur la mise en œuvre d'un système multiagent pour la génération à la volée des cartes et la résolution des conflits spatiaux. Cette approche est basée sur l'utilisation de la représentation multiple et la généralisation cartographique. Elle résout les conflits spatiaux et génère les cartes demandées selon une stratégie innovatrice : la génération progressive des cartes par couches d'intérêt. Chaque couche d'intérêt contient tous les objets ayant le même degré d'importance pour l'utilisateur. Ce contenu est déterminé à la volée au début du processus de génération de la carte demandée. Notre approche multiagent génère et transfère cette carte suivant un mode parallèle. En effet, une fois une couche d'intérêt générée, elle est transmise à l'utilisateur. Dans le but de résoudre les conflits spatiaux, et par la même occasion générer la carte demandée, nous affectons un agent logiciel à chaque objet de

l'espace. Les agents entrent ensuite en compétition pour l'occupation de l'espace disponible. Cette compétition est basée sur un ensemble de priorités qui correspondent aux différents degrés d'importance des objets pour l'utilisateur. Durant la résolution des conflits, les agents prennent en considération les besoins et les préférences de l'utilisateur afin d'améliorer la personnalisation de la carte. Ils améliorent la lisibilité des objets importants et utilisent des symboles qui pourraient aider l'utilisateur à mieux comprendre l'espace géographique. Le processus de génération de la carte peut être interrompu en tout temps par l'utilisateur lorsque les données déjà transmises répondent à ses besoins. Dans ce cas, son temps d'attente est réduit, étant donné qu'il n'a pas à attendre la génération du reste de la carte. Afin d'illustrer notre approche, nous l'appliquons au contexte de la cartographie sur le web ainsi qu'au contexte de la cartographie mobile. Dans ces deux contextes, nous catégorisons nos données, qui concernent la ville de Québec, en quatre couches d'intérêt contenant les objets explicitement demandés par l'utilisateur, les objets repères, le réseau routier et les objets ordinaires qui n'ont aucune importance particulière pour l'utilisateur. Notre système multiagent vise à résoudre certains problèmes liés à la génération à la volée des cartes web. Ces problèmes sont les suivants : 1. Comment adapter le contenu des cartes, à la volée, aux besoins des utilisateurs ? 2. Comment résoudre les conflits spatiaux de manière à améliorer la lisibilité de la carte tout en prenant en considération les besoins de l'utilisateur ? 3. Comment accélérer la génération et le transfert des données aux utilisateurs ? Les principales contributions de cette thèse sont : 1. la résolution des conflits spatiaux en utilisant les systèmes multiagents, la généralisation cartographique et la représentation multiple ; 2. la génération des cartes dans un contexte web et dans un contexte mobile, à la volée, en utilisant les systèmes multiagents, la généralisation cartographique et la représentation multiple ; 3. l'adaptation des contenus des cartes, en temps réel, aux besoins de l'utilisateur à la source (durant la première génération de la carte) ; 4. une nouvelle modélisation de l'espace géographique basée sur une architecture multicouche du système multiagent ; 5. une approche de génération progressive des cartes basée sur les couches d'intérêt ; 6. la génération et le transfert, en parallèle, des cartes aux utilisateurs, dans les contextes web et mobiles.

Titre : **Developping 2D and 3D multiagent**

geosimulation, a method and its application: The case of shopping behavior geosimulation in Square One mall (Toronto)

Diplôme: Ph.D.
Auteur : Walid ALI
Soutenance : 10 janvier 2006
Groupe : LSI-ICG

Examineurs :

Bernard Moulin, Ph.D. (directeur de thèse), Université Laval.
Alexis Drogoul, Ph.D. (examineur), UR GEODES 079, IRD, France.
Christophe Claramunt, Ph.D. (examineur), Naval Academy Research Institute, France
Jean Pierre Muller, Ph.D. (examineur), CIRAD TA 60/15, France.

Résumé :

Dans cette thèse, nous proposons une méthode générique de développement d'applications de géosimulation, en 2D et 3D, de divers phénomènes ou comportements complexes (exp. comportements humains) dans des environnements géographiques. Notre travail vise à résoudre quelques problèmes dans le domaine de la simulation informatique et, plus particulièrement, dans le domaine de la simulation multiagent. Les principaux problèmes que nous visons à résoudre dans cette thèse sont : 1. l'absence de méthodes génériques de développement de simulations multiagents de phénomènes ou de comportements dans des environnements géographiques ; 2. le manque de techniques de collecte et d'analyse des données spatiales et non-spatiales: (i) données en entrée de la géosimulation multiagent (qui sont utilisées pour alimenter la simulation) ou (ii) données en sortie de la géosimulation (qui sont générées par cette simulation); 3. l'absence d'un prototype de géosimulation qui peut être à la fois 'réaliste' et 'utile' pour simuler le comportement du magasinage des êtres humains dans un environnement géoréférencé représentant un centre commercial. L'idée principale de notre thèse consiste en : 1. la création d'une méthode générique de développement de géosimulations multiagents, en 2D et 3D, des phénomènes complexes (impliquant par exemple des êtres humains) dans des environnements géographiques ; 2. l'application de cette méthode en utilisant le comportement de magasinage dans un centre commercial comme cas d'illustration. Les principales contributions de cette thèse sont : 1. une

nouvelle méthode de développement d'applications de géosimulation multiagent, en 2D et 3D, des phénomènes complexes (tels que ceux qui impliquent des comportements humains) dans des environnements géographiques ; 2. une application de la méthode proposée en utilisant le comportement du magasinage dans un centre commercial comme cas d'illustration. Le cas de test qui a servi pour développer le prototype de simulation est le centre commercial Square One (Toronto). Ce prototype 'réaliste' et 'utile' est intitulé Mall_MAGS ; 3. des techniques et des outils pour : (i) collecter des données spatiales et non-spatiales qui servent à alimenter des géosimulations ; (ii) saisir, simultanément, des données spatiales et non-spatiales qui vont alimenter des géosimulations ; (iii) collecter des données spatiales et non-spatiales en provenance de la géosimulation en utilisant le paradigme d'agents, ainsi qu'un outil d'analyse de ces données ; 4. un couplage des techniques d'analyse et d'exploration de données OLAP (On Line Analytical Processing)/SOLAP (Spatial On Line Analytical Processing) et de notre prototype de géosimulation du comportement du magasinage des êtres humains dans un centre commercial. Ce couplage sert à analyser et explorer les données générées par ce prototype.

Titre : La géosimulation orientée agent : un support pour la planification dans le monde réel

Diplôme: Ph.D.

Auteur : Nabil SAHLI

Soutenance : 11 janvier 2006

Groupe : LSI-ICG

Examineurs :

Bernard Moulin, Ph.D, (directeur de thèse), Université Laval.

Alexis Drogoul, Ph.D. (examineur), UR GEODES 079, IRD, France.

Driss Kettani, Ph.D. (examineur), School of Science and Engineering, Maroc.

Jean-Pierre Muller, Ph.D. (examineur), CIRAD TA 60/15, France

Résumé :

La planification devient complexe quand il s'agit de gérer des situations incertaines. Prédire de façon précise est une tâche fastidieuse pour les planificateurs

humains. Le problème est encore plus complexe lorsque viennent s'ajouter des contraintes spatiales. Par exemple, lors d'un feu de forêt, des bulldozers doivent construire une ligne d'arrêt pour arrêter le feu. Ils doivent alors tenir compte non seulement de l'avancée des feux mais aussi des caractéristiques du terrain afin de pouvoir avancer plus facilement. Nous proposons une approche de géosimulation basée sur les agents et qui a pour but d'assister la planification dans un espace réel, à large échelle géographique et surtout à forte composante spatiale. L'approche consiste à établir un parallélisme entre l'Environnement Réel (par exemple une forêt incendiée) et un Environnement de Simulation basée sur des données SIG. Cet environnement virtuel utilise des agents logiciels sensibles à l'espace et dotés de capacités cognitives avancées pour planifier et coordonner les interventions sur le terrain. Étant donné que les plans générés dans le cas étudié (feux de forêts) sont essentiellement des chemins, nous proposons également une approche basée sur la géosimulation orientée agent pour résoudre des problèmes particuliers de recherche de chemins. Finalement, pour appliquer notre approche au cas des feux de forêts, nous avons utilisé MAGS comme plate-forme de géosimulation et Prometheus comme simulateur du feu.

Pour plus d'informations, vous pouvez contacter les directeurs de thèse. Vous trouverez leurs adresses électroniques sur le site du département d'informatique et de génie logiciel (<http://www.ift.ulaval.ca>) de l'Université Laval. Il me fera également plaisir de répondre à vos questions. Vos commentaires sont toujours appréciés.

Nicole Tourigny, Ph.D., professeure titulaire

LSI-ERICAE

Département d'informatique et de génie logiciel

Faculté des sciences et de génie

Pavillon Adrien-Pouliot

Université Laval

Québec, CANADA

G1K 7P4

Nicole.Tourigny@ift.ulaval.ca

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

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

Ingénierie du Web et Intelligence Artificielle

Coordonnateurs :

*Gilles Venturini (LI, Université François-Rabelais de Tours, gilles.venturini@univ-tours.fr),
Chantal Reynaud (Université Paris-Sud XI, CNRS (LRI) & INRIA (Futurs), cr@lri.fr)*

Introduction

Ce dossier fait suite au dossier publié en 2003 sur le Web sémantique, qui avait été coordonné par Chantal Reynaud (LRI), Jean Charlet (AP-HP) et Philippe Laublet (LaLICC). Il s'agit de recenser, de manière complémentaire, les équipes francophones travaillant en Intelligence Artificielle (I.A.) et dans les autres domaines du Web, qui ne sont pas directement concernées par le Web sémantique. Nous avons inclus initialement les thèmes suivants :

- la fouille du Web, qui traite à la fois de thèmes liés aux moteurs de recherche et à la fouille de textes, du contenu du Web et des usages du Web,
- les données et services distribués, tels que l'automatisation du raisonnement pour répondre à des requêtes, découvrir ou composer des services, les approches et architectures d'intégration (médiation, entrepôt de données, architecture pair-à-pair), les agents mobiles et la sécurité, utilisant, bien entendu, l'I.A.,
- les documents et contenus multimédia pour le Web ainsi que leur accès et leur utilisabilité. Ce thème concerne, par exemple, les hypertextes pour le Web (génération, analyse, utilisation), l'art numérique sur le Web, les documents électroniques et multimédia et les différentes modalités de leur accès sur le Web,
- la personnalisation et l'adaptation des applications Web avec les interfaces et sites Web adaptatifs, l'analyse d'audience (faisant partie également de la fouille du Web) et de profils utilisateurs,
- la visualisation d'information au travers du Web et la visualisation 3D pour le Web,
- enfin, toutes les applications potentielles de l'I.A. au Web : le commerce électronique, la veille stratégique, la

gestion des connaissances, le travail coopératif (entre individus ou organisations), le calcul scientifique, etc.

Les équipes présentées dans ce dossier sont les suivantes :

- Equipe ADN, Division R&D, Technologies, Laboratoire Easy, France Télécom
- Groupe APIS, équipe bases de données et systèmes d'information, LIMOS, Université Blaise Pascal – Clermont II
- Equipe Bases de Données, LICP, Université de Cergy-Pontoise
- Pôle de recherche DataLedge, Division R&D, France Telecom
- Equipe ERECD, LARODEC, IHEC de Carthage
- Equipe IASI/GEMO, LRI (Université Paris-Sud, CNRS) et UR Futurs de l'INRIA
- Equipe Informatique – INA P-G/INRA-Mét@risk
- Laboratoire LIP6, Equipes ACASA, Apprentissage statistique et connexionnisme, Bases de Données, LOFTI, Systèmes Multi-Agents, VirtualSysBio, Université Paris VI
- Laboratoire LIPN, Institut Galilée, Université Paris XIII
- Équipe Orpailleur, LORIA
- Laboratoire Paragraphe, Université Paris VIII
- Equipe ROI, LAMIH, Université de Valenciennes
- Equipe RTIC, LI, Université François-Rabelais de Tours
- Equipe SMAC, IRIT, Université Paul Sabatier
- Thème Systèmes de Bases de Données, Laboratoire PRISM, Université de Versailles et Saint-Quentin-en-Yvelines

Les thèmes principalement traités par les équipes participant à ce dossier sont :

Equipe ADN (Agent intelligent et Dialogues Naturels), Division R&D, Technologies, Laboratoire Easy, France Télécom	Fouille du web, moteurs de recherche, langage de communication inter-agents, services multi-supports sur mobiles/études d'ergonomie et d'usages
Groupe APIS, équipe bases de données et systèmes d'information, LI MOS, Université Blaise Pascal – Clermont II	Services web (modèle de protocoles, compatibilité/substitution), systèmes de médiation
Equipe Bases de Données, LICP, Université de Cergy-Pontoise	Médiation distribuée, Adaptation du contenu pour mobiles
Pôle de recherche DataLedge, France Telecom Division RD	Fouille de données
Equipe ERECD, LARODEC, IHEC de Carthage	Moteurs de recherche, Requêtes temporelles
Equipe IASI/GEMO, LRI (Université Paris-Sud, CNRS) et UR Futurs de l'INRIA	Extraction de connaissances, approches et architectures d'intégration, exploration ciblée du Web

Equipe Informatique – INA P-G/INRA-Mét@risk	Extraction de connaissances, intégration de données
Laboratoire LIP6, Equipes ACASA, Apprentissage statistique et connexionnisme, BD, LOFTI, Systèmes Multi-Agents, VirtualSysBio	Accès à la connaissance via des media électroniques, recherche et extraction d'information, modélisation sémantique de ressources P2P, ordonnancement de résultats de requêtes, comportement utilisateur, agents mobiles cognitifs, visualisation/navigation
Laboratoire LIPN, Institut Galilée, Université Paris 13	Analyse/prédiction du comportement utilisateur, extraction d'information, services web/planification d'actions
Équipe Orpailleur (LORIA)	Web sémantique, représentation de connaissances, méta-donnée, extraction des connaissances, recherche d'information
Laboratoire Paragraphe, Université Paris VIII	Hypermedia, hypertextes, moteur de recherche, travail collaboratif
Equipe ROI, LA-MIH, Université de Valenciennes	Adaptation de services transactionnels (personnalisation), services et données distribués.

Equipe RTIC, LI, Université François Rabelais de Tours	Fouille du web, Veille stratégique
Equipe SMAC, IRIT, Université Paul Sabatier	Multi-agents, recherche de services adaptés, aide à la navigation, fouille de textes et des usages
Thème Systèmes de bases de Données, PRISM, Université de Versailles et Saint-Quentin-en-Yvelines	Fouille de données, médiation pair-à-pair, entrepôt orienté services

Gilles Venturini, Chantal Reynaud

Equipe ADN (Agent intelligent et Dialogues Naturels)

France Télécom, Division R&D,
TECH/EASY/ADN
2 avenue Pierre Marzin
22307 Lannion Cedex

Contact : Philippe Bretier
philippe.bretier@francetelecom.com

L'équipe ADN 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.

Thèmes de l'équipe

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 travaux de l'équipe ADN portent plus particulièrement sur la sémantique

des communications :

- inter-agents,
- agents-ressources du web (web services, bases de connaissances, etc.), et
- 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.

Les technologies développées dans l'équipe ADN, 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 l'équipe ADN, 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. Dans ce contexte, l'équipe ADN s'intéresse plus particulièrement à l'identification des propositions qui répondent au mieux à la demande de l'utilisateur, à la construction de propositions pouvant combiner des informations issues de sources distinctes, à la 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 l'Ingénierie du Web et Intelligence Artificielle

Un sous-ensemble des activités de l'équipe ADN est relatif aux thèmes suivants de l'Ingénierie du Web et Intelligence Artificielle :

- Web mining : moteurs de recherche, recherche d'information et extraction de connaissance.
- Données et services distribués : automatisation du raisonnement (pour répondre à des requêtes).
- Documents et multimédia : diversité des modalités d'accès.

Moteurs de recherche thématiques dialoguants, extraction de connaissance

L'équipe ADN s'intéresse à la mise en œuvre de moteurs de recherche alternatifs par rapport aux moteurs classiques qui, à chaque requête, renvoient une liste (souvent très longue) d'URL. L'objectif recherché est de fournir une liste réduite d'URL pertinentes grâce à un dialogue aidant l'internaute à affiner ou clarifier sa requête ou suggérant des propositions alternatives.

L'équipe ADN dispose actuellement de technologies permettant de réaliser des moteurs de recherche dialoguants thématiques. Ces solutions ont été illustrées, en particulier, sur :

- un portail bancaire et
- une application de type *Virtual Buddy* accessible depuis le Messenger Wanadoo (<http://www.wanadoo.fr>; rubrique *Le Messenger*), avec qui l'internaute peut dialoguer en français pour obtenir des rubriques d'aides sur l'utilisation du Messenger.

Une autre problématique à laquelle s'intéresse l'équipe ADN est l'extraction de connaissance depuis de grandes bases documentaires et, en particulier, le Web. L'objectif est d'extraire des informations sémantiques (par exemple, sur une personne, un événement ou un objet) afin de les exprimer dans un langage de représentation des connaissances (voir le paragraphe suivant) ; ce langage devant être manipulable par un agent artificiel, avec qui des utilisateurs pourront dialoguer.

Automatisation du raisonnement

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.

L'équipe ADN étudie 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 par l'équipe ADN, 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 (voir ci-dessous).

Les langages de représentations des connaissances étudiés par l'équipe ADN sont destinés à servir à la fois au raisonnement sur ces connaissances mais aussi à l'interaction entre l'agent artificiel et l'utilisateur. 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.

Le consortium de standardisation FIPA (Foundation for Intelligent Physical Agents, <http://www.fipa.org>) 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 de l'équipe ADN. Des projets coopératifs tels que FACTS, Agentcities et JADE, auxquels contribue ADN, ont pour objectif la promotion d'un tel standard.

L'équipe ADN 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.

Une autre problématique à laquelle s'intéresse l'équipe ADN 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.

Diversité des modalités d'accès

Aujourd'hui, de nouvelles applications ou services apparaissent, dans lesquelles des supports multiples et coordonnés sont utilisés dans des contextes ou des environnements géographiques variés pour accomplir une tâche. Citons par exemple les services de messagerie ou de e-banking accessibles depuis un PC, une télévision interactive, un PDA ou un téléphone fixe ou mobile. Ces services permettent une grande mobilité et profitent des spécificités de chacun des supports utilisés. Par contre, ces services peuvent engendrer deux types de problèmes d'usage

spécifiques. Le premier a trait à la découverte d'un service sur un nouveau support, dont les propriétés peuvent induire des modifications en termes de fonctionnalités accessibles, de styles d'interaction et d'apparences. Le second problème a trait à la discontinuité possible d'une action commencée sur un support et achevée sur un autre avec un intervalle de temps variable entre les deux. Faciliter les transitions entre supports s'avère nécessaire pour aider l'utilisateur à s'adapter aussi rapidement que possible à chaque nouveau support et assurer une continuité du service.

L'équipe ADN s'intéresse aux méthodes et techniques permettant de développer rapidement des services multisupport tout en limitant les coûts supplémentaires par rapport à un service monosupport. Ces travaux sont couplés à des études d'ergonomie et d'usage visant à garantir une simplicité d'utilisation.

Bibliographie

- [1] Les agents rationnels dialoguants, Dossier France Télécom. <http://www.rd.francetelecom.fr/fr/technologies/ddm/200210/dossier.php>
- [2] 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.
- [3] 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.
- [4] 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.
- [5] C. Denis, L. Karsenty & F. Panaget, Favoriser les transitions avec des services multisupports, Colloque sur la mobilité, Nancy, 2002.

Groupe APIS (Architectures et Protocoles d'Intégration orientés Services)

Equipe Bases de données et Systèmes d'information, LIMOS

LIMOS, Université Blaise Pascal - Clermont II, Complexe scientifique des Céseaux, F - 63177 AUBIERE cedex, France <http://bdlimos.isima.fr/apis/>

Contact : Farouk Toumani
mél : ftoumani@isima.fr
Tél. : +33 (0)4 73 40 50 28
Fax : +33 (0)4 73 40 50 01

Membres de l'équipe concernés par le thème :

Michel Schneider (Pr.), Marinette Bouet (MC), Patrick Kellert (MC), Farouk Toumani (MC), Hélène Jaudoin (ATER, Doctorante), Julien Ponge (Doctorant), Christophe Rey (ATER).

Thème général de l'équipe

Le groupe APIS (Architectures et Protocoles d'Intégration orientés Services) fait partie de l'équipe bases de données et systèmes d'information du LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes, UMR 6158 du CNRS). L'activité scientifique de cette équipe couvre divers domaines tels que la fouille de données, les systèmes d'information géographique, l'interopérabilité de sources hétérogènes, la recherche d'information multimédia et l'intégration d'applications intra ou inter-entreprises dans le contexte des services web et du web sémantique.

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

Les travaux du groupe APIS portent sur les problèmes d'intégration d'applications intra ou inter-entreprises

dans le contexte des web services et du web sémantique. La problématique d'intégration constitue aujourd'hui un des enjeux majeurs de l'informatique d'entreprise. Les problèmes sous-jacents d'interopérabilité de données et de programmes sont connus pour être difficiles à résoudre compte tenu de la nature hétérogène, distribuée, dynamique et autonome des applications existantes. Le travail de recherche du groupe APIS s'inscrit dans cette perspective, en se focalisant particulièrement sur les problèmes de modélisation et d'automatisation décrits ci-dessous.

Approche dirigée par les modèles pour la conception, le développement et la gestion des services (en collaboration avec l'Université du New South Wales, Sydney, Australie et les laboratoires de Hewlett-Packard, Palo Alto, Etats-Unis).

L'objectif de ce projet de recherche est de développer une approche d'ingénierie générative, par laquelle tout ou partie d'un service web est généré et administré à partir de modèles. Deux résultats importants ont été obtenus dans le cadre de ce projet :

Un modèle enrichi de protocoles de services.

Nous nous sommes intéressés à la modélisation du comportement externe d'un service (appelé aussi protocole d'un service). Après une analyse de nombreux sites de commerce électronique (e.g., amazon.com, expedia.com, ford crédits services, etc.), nous avons proposé un modèle enrichi de protocoles de service permettant non seulement de décrire la chorégraphie des opérations supportées par un service mais également les contraintes temporelles et transactionnelles liées aux différentes opérations. Nous avons aussi étudié comment des squelettes de services décrits dans le langage BPEL¹ peuvent être générés automatiquement à partir du modèle proposé.

1. Business Process Execution Language (<http://www-106.ibm.com/developerworks/library/ws-bpel>).

Analyse de la compatibilité et de la substitution entre services.

Nous nous intéressons ici à l'analyse des protocoles de services selon les deux dimensions suivantes : (i) Compatibilité : il s'agit d'identifier dans quelles conditions deux services peuvent interagir correctement, (ii) Substitution: il s'agit d'identifier dans quelles conditions un service peut remplacer un autre service de manière transparente à l'utilisateur. Nous nous sommes intéressés dans un premier temps à des protocoles de services simples qui décrivent uniquement la chorégraphie des opérations supportées par un service. Les contributions précises concernant ce point sont les suivantes :

- Formalisation d'un modèle de protocole simple et étude de différents types de sémantiques adaptées au cas de protocoles déterministes et non-déterministes.
- Définition d'opérateurs, comme, par exemple, l'intersection ou la différence, permettant l'analyse et la manipulation de protocoles de services. Un étude théorique a été menée pour définir de manière précise la sémantique de ces opérateurs, analyser leur complexité et proposer des algorithmes permettant de les implémenter.
- Définition de plusieurs classes de compatibilité/substitution entre services et formalisation de ces classes à l'aide des opérateurs définis précédemment.

L'extension de ce cadre formel pour la prise en compte des contraintes temporelles fait l'objet d'une thèse de doctorat qui est réalisée en cotutelle avec l'University du New South Wales de Sydney.

Automatisation de la découverte des services web sémantiques (projet européen IST MKBEEM, collaboration avec France Telecom R&D et le LIRIS, Université Lyon I, France).

Nous nous sommes intéressés ici au problème de la découverte dynamique des services web sémantiques. Nous avons proposé de formaliser le problème de la découverte dynamique de services comme étant une nouvelle instance du cadre général de la réécriture de concepts en utilisant une terminologie.

Pour cela, un nouveau type de raisonnement, appelé découverte des meilleures couvertures, a été étudié en s'appuyant sur un cadre formel basé sur les logiques de description. Le problème de la découverte des meilleures couvertures peut s'énoncer de la manière suivante : « étant donné un concept Q (une requête) et une terminologie T de concepts (les services), le problème consiste à rechercher les conjonctions de concepts de T (les ensembles de services) qui sémantiquement se rapprochent le plus de Q ». La notion de proximité sémantique entre concepts est formellement définie en s'appuyant sur l'opérateur de différence sémantique entre concepts des logiques de description. Elle consiste à minimiser les différences entre une requête Q et ses réécritures potentielles, maximisant ainsi l'information commune entre elles. La découverte des meilleures couvertures d'une requête Q revient alors à identifier les ensembles de services de T qui vérifient cette propriété. L'intérêt de notre approche réside dans l'utilisation d'un critère de réécriture qui est plus souple que les relations de subsomption ou d'équivalence utilisées habituellement dans les approches de réécritures existantes comme, par exemple, les approches de réécriture de requêtes en utilisant des vues développées dans le contexte des systèmes de médiation. Nous nous intéressons actuellement à l'extension de cette approche pour permettre la composition (semi-)automatique de services web sémantiques.

Réécriture de requêtes en présence de contraintes de valeurs (en liaison avec le Cemagref, Clermont-Ferrand).

Un système de médiation (médiateur/wrapper) permet d'interroger via une interface uniforme, un ensemble de sources de données hétérogènes et distribuées. Nous nous intéressons, dans le cadre d'un tel système, au problème de réécriture de requêtes en utilisant les vues lorsque les requêtes et/ou les vues peuvent contenir des contraintes de valeurs. En effet, le contexte applicatif de nos travaux, lié au domaine du développement durable, a montré que les vues qui décrivent les sources de données

ont souvent la même description conceptuelle et qu'elles se distinguent uniquement par des contraintes de valeurs, c'est-à-dire les valeurs autorisées pour un attribut donné comme les numéros de commune ou les identifiants de fermes. Les contributions de ce thème sont les suivantes. Nous avons formalisé ce problème dans le cadre des logiques de description où le constructeur OneOf est utilisé pour représenter ces contraintes dans la description des requêtes et des vues. Puis afin de calculer effectivement les réécritures ainsi que de permettre le passage à l'échelle de l'algorithme de réécriture, nous avons montré comment notre problème peut être ramené à un cadre formel de Découverte de Connaissance dans les Bases de Données (DCBD). L'utilisation de ce cadre formel nous permet à la fois de bénéficier d'une solution capable de traiter un grand nombre de vues, mais aussi d'utiliser et d'adapter des algorithmes existants dans l'implémentation d'une des étapes les plus coûteuses de notre approche.

Médiation sémantique de données

Un médiateur est un système capable d'interroger des sources de données hétérogènes en offrant à l'utilisateur une vision uniforme des données. Un médiateur doit en particulier résoudre les problèmes d'hétérogénéité sémantique. La solution traditionnelle consiste à installer un schéma global pour l'ensemble des sources et à décrire les correspondances entre les schémas locaux et le schéma global. Cette solution présente une certaine lourdeur lorsque les schémas locaux évoluent ou lorsque un nouveau schéma local est introduit. Une autre solution consiste à s'appuyer sur une ontologie pour rechercher automatiquement les correspondances qui peuvent exister entre la requête de l'utilisateur et les schémas locaux. La requête est généralement approximative car l'utilisateur formule en s'appuyant sur une connaissance implicite des sources et du domaine. De même les correspondances peuvent être approximatives en raison des ambiguïtés qui peuvent survenir.

L'objectif de ce thème de recherche est de concevoir et de mettre en œuvre un médiateur répondant à ces principes et d'évaluer ses performances. On se

place plus particulièrement dans le cas où les données proviennent de sources XML ou de bases de données relationnelles.

Bibliographie

- [1] Boualem Benatallah, Fabio Casati, Farouk Toumani. Representing, Analysing and Managing Web Service Protocols, Data Knowl. Eng. Journal (DKE), (à paraître).
- [2] Boualem Benatallah, Mohand-Said Hacid, Alain Leger, Christophe Rey, Farouk Toumani: On automating Web services discovery. VLDB J. 14(1): 84-96 (2005).
- [3] Michel Schneider, "Dealing with semantic annotation for manipulating and mapping relational and XML sources", IADIS International Conference on Applied Computing, Algarve, Portugal, 22-25 February 2005, 8 pages.
- [4] Boualem Benatallah, Fabio Casati, Farouk Toumani: Web Service Conversation Modeling: A Cornerstone for E-Business Automation. IEEE Internet Computing 8(1): 46-54 (2004).
- [5] Hélène Jaudoin, Michel Schneider, Frédéric Vigier: Querying Data Sources in a SuperPeer Data Management System. FQAS 2004: 54-67.

Références bibliographiques du groupe APIS :
<http://bdlimos.isima.fr/apis/publis>

Equipe Bases de Données – LICP

Laboratoire de recherche en Informatique de Cergy-Pontoise (EA 2175)
 Université de Cergy-Pontoise
 U.F.R. des Sciences et techniques
 2, avenue Adolphe Chauvin
 BP. 222, Pontoise
 95302 Cergy-Pontoise Cedex

Tel. : 01 34 25 66 07

Fax : 34 25 66 30

Web : <http://www.u-cergy.fr/rech/labo/equipes/licp>

Contact : Dominique LAURENT

Membres de l'équipe concernés par le thème

Dominique LAURENT (PR), Tuyêt Trâm DANG NGOC (MC), Virginie

SANS (doctorante)

Thème général de l'équipe

L'équipe Bases de Données s'intéresse principalement aux thèmes suivants : Entrepôts de données, Extraction de connaissances dans les bases de données, Données semi-structurées (XML)

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

En collaboration avec le laboratoire PRiSM de l'université de Versailles-Saint-Quentin (sur le projet ACI Masse de données)

Les sources d'informations sur l'Internet sont nombreuses et très diversifiées. Afin de les présenter sous la forme d'une vue uniforme pour permettre leur interrogation, des architectures de médiation sont en général utilisées. Une architecture de médiation s'appuie sur le concept médiateur/adaptateur. Les adaptateurs se chargent de l'hétérogénéité des sources en faisant office de traducteurs vers un modèle commun. Le médiateur se charge de la distributivité des données en intégrant ces données. Le modèle commun que nous utilisons est XML et le langage permettant de l'interroger est XQuery. Nos études liées à XML se font dans les directions suivantes :

- La conception d'architectures de médiation optimisée dans un environnement pair-à-pair
- La définition d'une structure (nommée TGV) permettant de modéliser des requêtes XQuery afin de les optimiser et d'en tirer un plan d'exécution. Cette structure TGV doit tenir compte - des capacités fonctionnelles des sources ainsi que de la distribution des données.
- La définition d'une algèbre (nommée XAlgèbre) pour la construction d'un plan d'exécution pour la résolution de requête XQuery sur des données distribuées.
- La définition et le calcul de modèles de coût par rapport à une requête XQuery.

En cours de soumission pour l'appel à projet CNRS jeunes chercheurs

Il n'est plus besoin de présenter le succès que représente le web tant en tant

que source d'information que d'échanges de services et de support de commerce électronique. Parallèlement, les outils mobiles tels que les téléphones portables et les assistants personnels sont devenus des outils incontournables de la société moderne. Or si le web est parfaitement adapté aux ordinateurs de bureau et aux stations de travail, il en est tout autre quant à son accès depuis des périphériques mobiles, les problèmes récurrents étant l'affichage (taille de l'écran, rafraîchissement, couleurs, précision), le débit de connexion et l'interaction avec l'utilisateur (stylet, roulettes, touches limitées).

Des solutions comme le wap et l'i-mode existent mais cela nécessite de la part des serveurs web de prévoir un accès de la part des périphériques mobiles afin d'adapter le contenu des sites aux fortes contraintes des périphériques mobiles. De plus, cette solution ne prend pas en compte les spécificités de chacun des périphériques.

Tout récemment, le W3C, à l'origine des standards et protocoles régissant le web, a lancé une réflexion sur le web mobile (Web mobile Initiative - WMI). WMI vise à permettre (entre autres), aux constructeurs de définir les contraintes de leur périphériques suivant une norme encore à spécifier. Cela démontre un intérêt tout particulier et soutenu de la part des deux mondes (web et mobiles) concernant leur unification.

Néanmoins, il subsiste le problème de permettre aux périphériques mobiles d'accéder aux sites web tels qu'ils existent actuellement. L'inertie du web étant, il n'est pas envisageable de concevoir à court et moyen terme que tous les sites web offrent leur contenu adaptés aux périphériques demandeurs.

Notre objectif est donc de créer un intermédiaire (proxy) permettant aux périphériques mobiles d'accéder au web de façon transparente et personnalisée, sans pour autant modifier quoique ce soit sur les sites concernés.

Notre proxy, appelé PADAWAN (Proxy for All Devices Accessing the Web And Neighborhood), intégrera les nombreuses fonctionnalités d'un proxy

classique (intermédiaire et cache de données), mais exploitera aussi la nature semi-structurée du web. Le cache ne sera donc plus simplement un « simple stockage » de pages web par date, mais sera une vue matérialisée des sites mis au format XML. Une architecture de médiation, des adaptateurs spécifiques et un gestionnaire de vues permettra l'intégration des données du web.

Le contenu des pages ainsi récupérées ou les vues demandées seront alors formatés suivant les contraintes du périphérique demandeur, du profil utilisateur et de sa localisation.

En plus d'être un projet porteur en phase avec les besoins des utilisateurs actuels, notre projet aura le mérite de mener de front les thèmes de recherches suivants :

- La mise à jour des vues XML matérialisées : ce thème de recherche est complexe, il s'agit non seulement d'intégrer des sources autonomes telles que les pages web, et d'assigner des identifiants aux différentes parties des documents choisis, mais il s'agit en plus de repérer les mises à jour des pages et de pouvoir les reporter efficacement sur la vue matérialisée.
- La génération de requêtes XQuery en fonction de contraintes : suivant les contraintes données par le client (contraintes physiques du périphérique mobile, profil utilisateur), la vue demandée devra être adaptée. De plus, en intégrant la notion de localisation du mobile, il s'agira d'intégrer éventuellement dans la vue, des notions de sources volatiles et/ou dépendant de la localisation du client.
- Dans le voisinage du mobile, peut se trouver d'autres mobiles ou bornes, on pourra alors éventuellement les utiliser également comme source de données externes à intégrer, et donc savoir comment router et intégrer ces nouvelles sources vers notre proxy.
- Enfin, le langage de définition de contraintes est en début de réflexion au W3C (WMI, DDG), nous utiliserons donc un modèle extensible permettant d'intégrer la normalisation lorsque celle-ci sera disponible. Et rien ne nous empêche de participer à cette normalisation.

Bibliographie

- [1] Dang-Ngoc T.T., Gardarin G., Conception et Evaluation de XQuery dans une architecture de médiation 'Tout-XML', Revue ISI (Integration de systèmes d'information) : Numéro spécial Bases de Données Semi-structurées, (Hermès), volume 8, numéro 5-6, pp 11-25, 2003.
- [2] Dang-Ngoc T.T., Gardarin G., Travers N., Tree Graph View: On Efficient Evaluation of XQuery in an XML Mediator. 20èmes Journées Bases de Données Avancées (BDA), 2004.
- [3] Dang-Ngoc T.T., Sans V. Laurent D., Classifying XML Materialized Views for their Maintainance on Distributed Web Sources, Extraction et gestion des connaissances, EGC 2004. In Revue des Nouvelles Technologies et de l'Information, RNTI, Cepaduès Editions 2005.
- [4] Dang-Ngoc T.T., Sans V. Laurent D., On the Maintenance of Materialized XML Views, ISIP 2005, Lyon, France, 2005.
- [5] Jamard C, Travers N., Dang-Ngoc T.T., Gardarin : XLive - Prototypage d'un médiateur Xquery. Logiciel.

Pôle de Recherche DataLedge France Telecom R&D

Traitement et Gestion des Connaissances et des Données
France Telecom Division R&D Rennes
4 rue du Clos Courtel BP 91226
35512 Cesson Sévigné

Contact : Alain Léger
Tél : 02 99 12 42 23
Mél : alain.leger@francetelecom.com

Membres de l'équipe concernés par le thème du dossier

Alain Léger (Web sémantique, Web services), Gilles Prigent (Traitement automatique des langues naturelles), Vincent Lemaire (Knowledge Discovery from Data and Machine Learning), Luigi Lancieri (Interactions humaines et Intelligence collective)

Thèmes de l'équipe

Le pôle de recherche DataLedge, composé d'une quarantaine de chercheurs, constitue un des 16 pôles de recherche de la division R&D de France

Télécom, qui lui permettent :

- de développer et de maintenir des compétences scientifiques sur des thématiques stratégiques en anticipation des besoins du Groupe,
- de développer et de protéger des techniques directement utilisables pour les réseaux et services du Groupe,
- d'identifier les évolutions et ruptures technologiques et en évaluer les conséquences sur les systèmes et services télécoms,
- de tisser des partenariats scientifiques et techniques pour accélérer la mise en oeuvre des innovations.

L'activité de Recherche du Pôle scientifique DataLedge est au carrefour des communautés scientifiques et techniques de l'Intelligence Artificielle, des sciences cognitives et de l'informatique Avancée. Les facultés de « haut niveau » considérées sont : représenter, comprendre, apprendre, expliquer, communiquer et résoudre. C'est à ces tâches de "haut niveau" que DataLedge tente d'apporter des réponses Innovantes - au plus près des ruptures technologiques ou au meilleur de l'état de l'art des résultats les plus avancés - aux besoins spécifiques du groupe France Télécom. DataLedge a donc centré ses recherches Scientifiques et Technologiques sur le domaine du traitement des connaissances (IA) appliquées aux technologies de l'Internet (Web). Il est ainsi organisé autour de trois champs scientifiques clés, à savoir la représentation des Connaissances et des Raisonnements, la Qualification des Connaissances, le traitement Automatique des Langues naturelles, et l'émergence de connaissances à partir des données et de l'apprentissage. Les travaux portent sur la maîtrise du sens (Web sémantique), le traitement des langues naturelles et la recherche de sens dans les gigantesques silos d'information hétérogènes mais stratégiques.

Le pôle se concentre en particulier sur les sujets de Knowledge Discovery from Data et Machine Learning. Cet axe s'intéresse au développement de méthodes et de systèmes permettant d'organiser, de prévoir, d'interpréter des données. Ces données peuvent être structurées, comporter de l'imprécision et être de taille importante. De telles

données s'accumulent de plus en plus dans des bases de données réparties sur un réseau et suscitent les questions suivantes auxquelles le pôle cherche à répondre : comment représenter des données et des connaissances pour pouvoir extraire des régularités ? Comment organiser les observations en classes ou concepts facilement interprétables ? Comment trouver des fonctions permettant de décider de l'appartenance d'une observation à des classes connues à priori ? Comment traiter efficacement de gros ensembles de données ? Comment traiter des flux de données en temps réel (« à la volée ») ?

Par ailleurs, un autre axe de recherche concerne l'exploitation des phénomènes d'intelligences collective la gestion automatisée des connaissances. L'intelligence collective exprime que la capacité d'un groupe d'individu est supérieure à la somme des capacités individuelles. Le réseau Internet et un milieu naturel privilégié favorisant l'émergence et l'expression de cette forme d'intelligence. La principale difficulté en ce qui concerne la gestion des connaissances est que l'intelligence collective empreinte de complexité et difficile à modéliser. De plus à l'image des forces naturelles, ce phénomène évolue librement sans que l'on puisse encore clairement envisager d'en canaliser ou ré-exploiter ces effets.

Bibliographie

- [1] Wiley Book, G Stamou et Stefano Kollias Ed., Chapitre sur « Semantic Web Applications », Alain Léger (FTR&D), Pramila Mullan (FT R&D SF), Shishir Garg (FT R&D SF), Jean Charlet (STIM), 2005.
- [2] Vincent Lemaire, Fabrice Clérot, The many faces of Kohonen Map, Chapter 1 of Classification and Clustering for Knowledge Discovery, Series: Studies in Computational Intelligence, Vol. 4 Halgamuge, Saman K.; Wang, Lipo (Eds.) 2005, Approx. 300 p., Hardcover; ISBN: 3-540-26073-0 - will be published in July 2005.
- [3] Emilie Guimier de Neef et Jean Véronis, Le traitement des nouvelles formes de communication écrite, Dans le traité « Compré-

hension des langues et interaction » (Gérard Sabah ed.), à paraître en 2005.

- [4] Michel Plu, Layda Agosto, Laurence Vignollet, Jean-Charles Marty, A Contact Recommender System for a Mediated Social Media. ICEIS (5): 107-114, 2004.
- [5] L.Lancieri, Interactions humaines dans les réseaux, Edition Hermes ISBN : 2-7462-1108-4, Avril 2005.

Equipe ERECD - Laboratoire LARODEC

IHEC de Carthage
2016 Carthage-Présidence
<http://www.cck.rnu.tn/larodec>

Contact : Rim Faiz
Mél : Rim.Faiz@ihec.rnu.tn
Tél. : +216 71 775 948

Membres de l'équipe

Rim Faiz, Nizar Khayati, Khaled Mellouli, Sami Mahfoudhi, Najeh Naffakhi.

Thèmes de l'équipe

Fouille de textes, Extraction de connaissances, Modélisation et Représentation des données et des connaissances, Fouille de données, Apprentissage et Classification.

Travaux de l'équipe

Analyse du contexte temporel pour améliorer la recherche d'informations événementielles sur le web

Les requêtes permettant de répondre aux demandes d'information des utilisateurs deviennent très complexes et l'extraction des données les plus pertinentes devient de plus en plus difficile dès que les sources de données sont hétérogènes et nombreuses. Il est impératif de considérer la sémantique de ces données et de l'exploiter le plus possible afin d'améliorer la recherche sur le web. L'intérêt de la consultation de ce média est lié à l'efficacité des moteurs de recherche d'informations. De ce fait, les moteurs de recherche doivent prendre en compte la sémantique des requêtes. En plus les méthodes de traitement d'informations fondées sur les statistiques ne sont plus suffisantes pour

répondre aux besoins des utilisateurs pour manipuler (rechercher, traduire, résumer) les informations sur le Web. Un constat tend à s'imposer : introduire « plus de sémantique » pour la recherche d'informations pertinentes issues des textes.

Lorsqu'un utilisateur lance sa requête, il serait souhaitable de trouver précisément ce qu'il cherche, c'est à dire trouver « l'information pertinente », sans qu'il soit submergé avec un volume de réponses non maîtrisables et ingérables. Dans ce contexte, nous proposons une approche, à l'instar de celles visant à améliorer les performances des moteurs de recherche, tels que l'introduction de la notion de contexte, l'analyse des pages web ou aussi la création de moteurs de recherche spécifiques dans un domaine donné.

Dans notre travail, nous traitons un aspect particulier : la prise en compte du contexte temporel. En effet, dans le but d'améliorer la précision et permettre une recherche plus contextuelle, nous proposons une méthode basée sur l'analyse du contexte temporel d'une requête afin d'obtenir des informations événementielles pertinentes.

La question qui se pose est comment découvrir les informations événementielles et transformer les collections de données en connaissances intelligibles, utiles et intéressantes dans le contexte temporel où l'on se trouve.

Nous avons constaté que, d'une manière générale, les requêtes recherchant un ou plusieurs événements ayant lieu à une date donnée ou pendant une période de temps bien déterminée ne permettent pas d'obtenir les résultats attendus. Par exemple : *Les découvertes scientifiques depuis 1940*. Dans cet exemple de requête, l'utilisateur veut chercher les découvertes scientifiques depuis 1940 jusqu'à aujourd'hui et non pas pour l'année 1940 uniquement, il s'agit ici de traiter un intervalle de temps. En effet, un moteur de recherche classique effectue la recherche uniquement sur le terme « 1940 » et non sur la période de temps en question.

D'où l'idée de la reformulation de la requête lancée par l'utilisateur, en basant la recherche sur l'expression introduite par l'utilisateur ainsi qu'une combinaison de mots synonymes aux termes de la requête initiale. Le traitement de la requête se fait surtout au niveau contextuel. Le système que nous proposons doit repérer les expressions temporelles dans les requêtes et les séparer de l'expression de l'événement recherché par ailleurs. Ainsi le système doit pouvoir comprendre le contexte temporel de la requête donc le doter d'une certaine intelligence (pour s'approcher du raisonnement humain) en plus d'une analyse sémantique (pour la compréhension de la requête).

Pour le repérage des expressions temporelles, nous utilisons notre méthode de filtrage automatique d'information temporelle que nous avons développée dans des travaux antérieurs. L'extraction de l'information temporelle à partir de la requête se fait par l'identification de marqueurs temporels (*depuis, au cours, avant, jusqu'à ...*) ou par la présence de date explicite dans la requête.

Ensuite, pour l'interprétation de ces expressions, nous utilisons notre représentation du temps à partir de la notion d'intervalle. Cette représentation est basée sur les dates de début et de fin des événements. Nous distinguons idéalement deux types d'événements :

- L'événement ponctuel ou instantané (Evi) : c'est un événement qui est considéré comme instantané par rapport au grain temporel et sa date de début est égale à sa date de fin ; $Deb(Evi) = Fin(Evi)$. Ce type d'événement est exprimé dans la requête, par exemple, par l'expression : « en 2001 ».
- L'événement duratif (Evd) : c'est celui un événement qui se déroule dans le temps sans interruption par rapport au grain temporel ; $Deb(Evd) < Fin(Evd)$ d'où l'intervalle de temps : $[Deb(Evd), Fin(Evd)]$. Ce type d'événement est exprimé dans la requête, par exemple, par l'expression : « entre 1990 et 2001 » ou l'expression : « depuis 1980 ».

Il est à noter que le grain temporel

sur lequel nous travaillons est l'année et nous envisagerons plus tard de travailler sur d'autres grains.

Dans le contexte de notre travail nous avons besoin de représenter, sous forme d'intervalle, l'information temporelle contenue dans la requête pour pouvoir l'utiliser comme une information complémentaire à la requête.

Ainsi, nous appliquons les règles d'interprétation pour déterminer, de manière explicite, l'intervalle de temps. Exemple : « Si la requête contient « depuis » + année_début Alors intervalle de temps = [année_début , année_en_cours] ». Donc, si le document contient l'événement recherché ayant lieu pendant l'intervalle de temps généré, alors nous le considérons comme pertinent.

Pour mieux cerner le contexte de la requête introduite par l'utilisateur, nous avons également envisagé l'extension de la requête et ce par l'ajout de termes synonymes à l'événement recherché. Exemple : pour le terme « attentat », nous utilisons les synonymes : « attaque, explosion, crime, etc. ».

Nous avons validé notre travail par le développement du système SORTWeb (Système d'Optimisation de Requêtes Temporelles sur le web) qui permet d'améliorer l'efficacité et la précision de la recherche d'informations événementielles sur le web et ce par le biais d'une reformulation automatique des requêtes afin d'obtenir des résultats pertinents et répondant aux attentes des utilisateurs.

Modélisation et classification d'images à partir du web

Nous envisageons d'étendre nos travaux portant sur le traitement d'informations recueillies lors des recherches sur le web au contexte des informations multimédias.

Nous nous sommes intéressés à la problématique de la classification automatique d'images à partir du web. La classification correspond à une tâche de traitement d'information, à laquelle est associée une forte sémantique cognitive. Il en résulte une très grande difficulté

pour proposer des systèmes performants de classification et de recherche pertinente à partir de grandes bases d'images. Un problème important de la production automatique de règles de classification concerne la durée de génération de ces règles.

Dans ce contexte, le but de notre travail est de proposer une nouvelle méthode de classification basée sur l'arbre de décision en utilisant une structure de donnée appelée P-tree (arbre de Peano Count) et ce en vue d'accélérer le processus de classification et de recherche dans de grandes bases d'images. Pour valider notre méthode, nous avons développé un système nommé C.I.A.D.P-tree qui a été implémenté et confronté à une application réelle dans le domaine du traitement d'images à partir du web.

Bibliographie

<http://www.cck.rnu.tn/laroddec/pages/Recherche.htm>

- [1] Faïz R., Mahfoudhi S. et Mellouli K., Web services as a tool for information systems effectiveness. *Symposium on Applications and the Internet. IEEE Computer Society*, M. Papazoglou and K. Yamazaki (Editors), Los Alamitos, California, pp. 362-365, 2005.
- [2] Faïz R., Naffakhi N. et Mellouli K., Apprentissage supervisé pour la classification des images basé sur la structure P-tree. *Revue des Nouvelles Technologies de l'Information*, D.A. Zighed et G. Venturini (Editeurs), Cépadues-Éditions, Vol.2, pp.469-480, 2005.
- [3] Faïz R., Khayati N. et Mellouli K., Optimisation des requêtes temporelles sur le web. *Conférence sur l'Extraction et Gestion des Connaissances*, Clermont Ferrand, France, 20-23 janvier, 2004.
- [4] Faïz R. et Biskri, I., Hybrid approach for the assistance in the events extraction in great textual data bases. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Tunisie, 6-9 octobre, 2002.
- [5] Faïz R., Exev: extracting events from news reports. *Actes des Journées Internationales d'Analyse statistiques des Données Textuelles (JADT 2002)*, A. Morin et P. Sébillot (Editeurs), Vol. 1, France, pp. 257-264, 2002.

Equipe IASI/GEMO, LRI
(Université Paris-Sud, CNRS)
et UR Futurs de l'INRIA

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

Contact : Chantal Reynaud
Mél : cr@lri.fr
Tel : 01 69 15 66 45
Fax : 01 69 15 65 86

Membres de l'équipe IASI/GEMO

P. Adjiman, P. Chatalic, H. Gagliardi, G. Giraldo, F. Goasdoué, H. Kefi, A. -D. Mezaour, N. Pernelle, C. Reynaud, M. -C. Rousset, F. Saïs, B. Safar, L. Simon, V. Ventos.

Thème général de l'équipe

L'objectif de IASI/GEMO est l'étude des problèmes fondamentaux que posent les systèmes modernes de gestion et d'interrogation des données, des systèmes ouverts à des informations riches, orientés vers le réseau. Le thème central de l'équipe est l'intégration de toutes ces informations, distribuées et hétérogènes, qu'il s'agisse de données, de connaissances ou de services. Il s'agit de découvrir les informations et les services intéressants, comprendre leur contenu, les intégrer et gérer leur évolution dans le temps. Des solutions novatrices appropriées pour la conception d'environnements puissants et flexibles permettant d'accéder rapidement à des données significatives et distribuées sur le réseau sont nécessaires. L'équipe travaille sur la création d'entrepôts de données et de médiateurs, deux solutions offrant un large accès à de multiples sources hétérogènes.

Travaux en Ingénierie du Web et Intelligence Artificielle

1. Extraction d'information à partir de données semi-structurées

L'extraction d'arbres fréquents à partir de collections de données XML hétérogènes (travail réalisé conjointement avec Michèle Sebag de l'équipe I&A du LRI).

La découverte d'arbres fréquents dans d'énormes collections d'arbres de données est devenu un enjeu important pour l'extraction d'information avec

l'émergence des données XML. Deux nouveaux algorithmes ont été proposés pour découvrir des arbres fréquents. Ils sont originaux par la façon de gérer l'inclusion d'un arbre dans un autre. Les motifs fréquents arborescents considérés sont plus généraux que ceux distingués jusqu'ici dans la littérature, conduisant à la découverte d'arbres fréquents même en présence de variations dans l'imbrication des balises. Les deux algorithmes se distinguent par leur façon de gérer la complexité. La solution mise en oeuvre dans l'algorithme TreeFinder consiste à approximer la représentation de l'arbre par une représentation booléenne plate des arbres, en codant la présence ou l'absence de toutes les paires d'étiquettes possibles (nœuds liés par des relations d'ancestralité) dans chaque arbre. En contrepartie, TreeFinder n'est pas complet. La solution expérimentée dans l'algorithme Dryade est de calculer des arbres fréquents fermés, ce qui a l'avantage de fournir une représentation compacte des arbres fréquents sans perte d'information. Les performances de TreeFinder et de Dryade ont été testées sur des données artificielles de taille moyenne mais également sur des données réelles (projet Xyleme).

La construction semi-automatique d'une ontologie

La construction automatisée d'une ontologie à partir d'un ensemble de DTDs portant sur un même domaine d'application a été étudiée dans le cadre du projet PICSEL 2. Une ontologie composée de classes, de propriétés et de relations est nécessaire pour fournir le schéma global d'un système médiateur intégrant des sources de données hétérogènes. Les DTDs sont vues comme des arbres étiquetés mais, étant donné le manque de sémantique formelle, il n'existe aucune correspondance claire entre les tags étiquetant les différents nœuds de la DTD et les classes dans l'ontologie structurées par des relations de subsumption. La solution proposée consiste à extraire des classes, des propriétés et des relations de l'ontologie à partir des DTDs via des heuristiques. La phase d'extraction exploite la struc-

ture de DTDs. L'ensemble des éléments découverts est ensuite structuré. Les classes sont liées, les propriétés sont associées aux classes qu'elles caractérisent. Le processus d'extraction est entièrement automatique. Il est suivi d'une validation. La phase d'extraction est répétée jusqu'à ce que les résultats soient satisfaisants. Un prototype d'extraction de composants d'une ontologie, OntoMedia, a été implémenté. Une application a été conçue à partir de DTDs de l'OTA (Open Travel Alliance) dans le domaine du tourisme.

2. Approches et architectures d'intégration

Les approches de gestion de données pair-à-pair

Ces approches ont été récemment proposées pour généraliser l'approche centralisée d'intégration de systèmes d'information basée sur des portails uniques connectant des sources de données hétérogènes. Une telle vision centralisée est adaptée à l'intégration de plusieurs sources d'information mais le contexte du Web requiert des systèmes plus flexibles tels les systèmes pair-à-pair basés sur une médiation distribuée. L'accès à un système pair-à-pair n'est pas centralisé. Chaque pair a son propre schéma, ses propres données et peut se mettre en relation avec d'autres pairs pour répondre à des requêtes. Nous avons développé l'infrastructure pair à pair SomeWhere qui permet le partage sémantique de ressources du Web décrites par des ontologies. Les ontologies sont propres à chaque pair mais des correspondances sémantiques entre ontologies peuvent être déclarées sous forme de mappings. SomeWhere a été conçu conformément aux recommandations du W3C puisque le formalisme de représentation du modèle de données est le fragment propositionnel de OWL. Le calcul des réponses dans SomeWhere repose sur un algorithme distribué de raisonnement en logique propositionnelle. Selon les premières expérimentations, ce modèle de données, très simple, permet de gérer de l'ordre du millier de pairs. SomeWhere est la base du projet MediaD réalisé en collaboration avec France Telecom R&D sur la médiation distribuée.

L'intégration de données du Web

L'objectif est de peupler un entrepôt avec des données du Web trouvées dans des tableaux contenant des informations pertinentes et synthétiques se situant dans des documents HTML ou PDF. Les tâches relatives au format du document Web sont séparées des tâches spécifiques au domaine. Les tableaux sont automatiquement transformés en Xtab, une représentation générique XML. Leur intégration nécessite de les représenter différemment avec un vocabulaire bien particulier, c'est-à-dire d'adopter une représentation XML où la plupart des valeurs et des tags appartiennent à l'ontologie du domaine étudié. L'objectif du travail réalisé consiste à rendre cette transformation le plus automatique et flexible possible. Une DTD, appelée SML (Semantic Markup Language), qui peut être automatiquement générée à partir de l'ontologie, a été définie. Elle permet la gestion des données incomplètes, l'ajout de données supplémentaires dans une relation, la gestion des ambiguïtés ou des erreurs d'interprétation. Cette transformation a été partiellement implémentée et expérimentée sur des données réelles au sein du projet e.dot. Les données extraites sont exploitables par le moteur d'interrogation MIEL++ développé à l'INAPG.

3. Exploration ciblée du Web

La clé de la recherche ciblée de documents sur le web réside dans la capacité à caractériser clairement les documents d'intérêt pour pouvoir, par la suite, les distinguer sans ambiguïté des autres documents. Dans ce sens, nous avons proposé WeQueL, un langage de requêtes déclaratif qui permet de décrire les documents souhaités par combinaison de critères de recherche mots clés. La particularité de notre langage réside dans la possibilité de cibler la recherche des mots clés sur des parties spécifiques de la structure du document. Un score de pertinence est calculé pour chaque document à la suite de son évaluation sur une requête WeQueL. Ce score permet, par exemple, d'établir un classement des documents évalués pour présenter en priorité les documents totalisant les scores les plus élevés, i.e. les plus pertinents. En s'appuyant sur ce

langage, nous avons proposé une approche sélective d'exploration du Web guidée par une requête WeQueL (appelée requête d'exploration). Une exploration sélective guidée par une requête WeQueL repose sur une fonction qui permet d'estimer le score d'évaluation d'un document avant de rapatrier son contenu. Cette fonction d'estimation est mise en œuvre par un système d'apprentissage bayésien en ligne s'appuyant sur un ensemble de mesures statistiques recueillies au fur et à mesure de l'exploration.

Projets en lien avec les thèmes ci-dessus présentés

PICSEL2 (France Telecom R&D (Rennes) et IASI/LRI) : L'objectif de PICSEL2 est de faire passer à l'échelle l'approche médiateur développée dans PICSEL1 en permettant, en particulier, la construction automatique de l'ontologie qui intègre des sources XML. Des méthodes pour offrir des réponses coopératives à l'utilisateur ont également été proposées dans le cadre de ce projet.

e.dot (Verso/INRIA, IASI/LRI, BIA/INRA, Xyleme) : Le but est de proposer des solutions génériques pour enrichir un entrepôt thématique à l'aide des données du Web. Le domaine d'application est le domaine des risques alimentaires. Les aspects techniques traités sont : le crawling du Web, la classification et l'intégration de données hétérogènes, le stockage de masses de données.

Bibliographie

- [1] A. Termier, M.-C. Rousset, M. Sebag, DRYADE : a new approach for discovering closed frequent trees in heterogeneous tree databases. In International Conference on Data Mining (ICDM), 2004.
- [2] C. Reynaud, G. Giraldo, An Application of the Mediator Approach to Services over the Web, Special track "Data Integration in Engineering, Concurrent Engineering (CE), 26-30 July, Portugal, 2003.
- [3] P. Adjiman, P. Chatalic, F. Goasdoue, M.-C. Rousset, L. Simon, Scalability study of peer-to-peer consequence finding, IJCAI, Edin-

burgh, Scotland, 30 July – 5 August 2005.

- [4] H. Gagliardi, O. Haemmerlé, N. Pernelle, F. Sais, An automatic ontology-based approach to enrich tables semantically, In Proceedings of the first International Workshop on Context and Ontologies : Theory, Practice and Applications, AAAI, 2005.
- [5] A.-D. Mezaour, Focused search on the Web using WeQueL. In 10th International Workshop on Knowledge Representation meets Databases (KRDB), p. 63-74, 2003.

Laboratoire d'Informatique de Paris 6

Laboratoire d'Informatique de Paris 6
8, rue du Capitaine Scott
75015 Paris

Le LIP6 couvre un très large spectre des domaines de l'Intelligence Artificielle et de l'Ingénierie du Web. Ses activités de recherche autour de ces thématiques sont réparties en 5 équipes :

Acquisition de Connaissances et Apprentissage Symbolique Automatique (ACASA)**Membres**

Jean-Gabriel Ganascia, Vincent Corruble (permanents), Isabelle Audras, Frédéric Baudoin, Julien Boudaillet, Aydano Machado, Charles Madeira, Julien Velcin (Doctorants)

Mots-Clefs

Fouille de textes, Documents électroniques et multimédia, Analyse comportementale, Visualisation de données, Agents mobiles

L'équipe ACASA, dirigée par Jean-Gabriel Ganascia, travaille sur trois axes particuliers de l'intelligence artificielle : l'apprentissage, la découverte et la créativité. Initialement centrées sur l'acquisition de connaissances et l'apprentissage symbolique automatique, ses orientations scientifiques ont évolué vers la modélisation cognitive et la conception d'outils facilitant la lecture et l'accès aux connaissances à travers les médias électroniques.

Ainsi ont été récemment menées des recherches sur l'analyse littéraire (critique génétique, stylistique), sur la modélisation des représentations en sciences sociales (reconstruction de stéréotypes sociaux à partir d'articles de journaux), sur la découverte scientifique (modélisation de théories en sciences physiques et en médecine), sur la musicologie et la musique (détection de motifs, simulation d'improvisations), sur la didactique des langues et sur les jeux de stratégies.

Par ailleurs, des travaux sont en cours sur la visualisation d'information pour la lecture électronique et sur la philologie électronique.

Enfin, l'équipe ACASA s'est investie dans la conception du site « livre libre » qui vise à offrir à tous un accès facile à la littérature libre de droits. Ce projet recourt à des agents apprenant qui explorent automatiquement les sites littéraires afin d'en extraire les textes qu'ils contiennent et de les indexer automatiquement.

Apprentissage statistique et connexionnisme

Membres

Massih Amini, Thierry Artières, Patrick Gallinari (permanents)

Mots-clés

Recherche d'information, fouille de données, extraction d'information, hypermédias adaptatifs, modélisation d'utilisateurs, analyse de traces

L'équipe connexionniste du LIP6 travaille depuis une dizaine d'années sur des problématiques qui ont émergé avec le web, liées à la recherche d'information, l'extraction d'information, le résumé automatique etc.

De nombreuses publications, 5 thèses, des contrats témoignent de l'activité et de la compétence de l'équipe dans cette thématique. Le domaine de recherche lié au web est en perpétuelle et rapide évolution. Aujourd'hui, une partie des travaux de l'équipe concernent de nouvelles problématiques fondamentales liées aux applications du web, le traite-

ment de documents structurés, l'apprentissage en grande dimension et l'apprentissage semi-supervisé. Parallèlement, l'équipe connexionniste s'est investie depuis un peu moins d'une dizaine d'années dans le domaine de la modélisation d'utilisateurs.

Des travaux ont été menés, notamment dans le cadre de projets RNTL, autour de l'analyse de traces de navigation sur le web ou dans des hypermédias, sur la découverte de communautés d'internautes, l'inférence de profils d'utilisateurs etc. L'équipe a organisé un atelier autour de ce thème lors de la conférence EGC 2005 et organise deux ateliers sur l'apprentissage dans le cadre des conférences de référence dans le domaine, User Modeling 2005 et International Conference on Machine Learning 2005.

Base de données

Membres

Anne Doucet, Bernd Amann, Maha Abdallah (permanents), Nicolas Lumineau, Camelia Constantin (doctorants)

Mots-clés

Routage de requêtes, réseaux pair-à-pair, interrogation structurée, importance de services web

Une partie importante des activités de l'équipe Bases de Données du Laboratoire d'Informatique de Paris 6 (BD-LIP6) concerne les problèmes de passage à l'échelle dans le partage de données distribuées (clusters de données, web, réseaux pair-à-pair). Le premier axe de recherche lié au web sémantique s'intéresse à la modélisation sémantique de ressources dans les réseaux pair-à-pair. Les travaux de recherche effectués dans le contexte du projet ACI Grid PADOUE ont ainsi permis d'établir différentes techniques sémantiques pour améliorer le processus de localisation de ressources. D'une part, nous avons montré comment exploiter des descriptions (métadonnées) sémantiques concernant les ressources pour structurer le réseau [10].

D'autre part, nous avons proposé l'introduction de profils des communautés d'utilisateurs pour optimiser le routage

des requêtes. L'approche GridDB se situe dans la même problématique et propose l'utilisation de tables de hachage distribuées (DHT) pour l'indexation de données structurées. L'avantage de cette approche est la possibilité de définir des critères de localisation de ressources plus riches (à la SQL) qui prennent en compte la structure des données interrogées. Le deuxième axe de recherche lié au web sémantique s'intéresse au calcul importance de données et de services pour l'ordonnancement de résultats de requêtes. L'idée est d'étendre la notion d'importance de ressource web de l'approche PageRank (Google) aux documents ActiveXML qui intègrent des données XML et des résultats d'appels de services web.

Logique Floue et Traitement de l'Information (LOFTI)

Membres

Bernadette Bouchon-Meunier, Maria Rifqi, Nicolas Labroche (permanents)

Mots-clés

Web Mining (Usage, Content, Structure), mesure d'audience, profils utilisateurs, Web sémantique, sites adaptatifs, clustering, logique floue

L'équipe LoFTi s'intéresse à l'apprentissage de modèles utilisateurs pour aider à la découverte et la compréhension des comportements de recherche d'information sur Internet. Ainsi, des travaux ont été conduits dans le cadre de la mesure d'audience sur Internet (Web Usage Mining), de façon à extraire les profils de navigation types des internautes à partir des fichiers de traces présents sur les serveurs Web [8]. D'autres travaux [5] se sont attachés à caractériser automatiquement et dynamiquement les besoins d'information des internautes en s'appuyant sur le contenu des documents accédés (CEA, Clue Extraction Algorithm).

Notre recherche actuelle vise à donner plus de sens aux profils de navigation, en utilisant conjointement le contenu des documents accédés (texte, éléments multimédias), les éventuelles descriptions sémantiques des ressources accessibles via Internet et des ontologies

pour préciser le contexte des recherches d'information.

Dès lors que les profils sont établis, le champ des applications devient alors très large : prédiction fine et pertinente des documents pour les moteurs de recherche ou les systèmes de veille stratégique, réalisation de systèmes hypermédias adaptatifs, création d'index personnalisés (site portail), recommandations automatiques de produits pour les sites marchands, ...

L'équipe LoFTi travaille actuellement en collaboration avec le laboratoire des usages (LUTIN) pour aider à la caractérisation des parcours des internautes sur les sites Internet relevant de la culture scientifique et technique (projet Webcasti).

Systèmes Multi-Agents

Membres

Amal El Fallah-Seghrouchni, Nicolas Sabouret (permanents), Alexandru Suna, Tarek Meliti, Yasmine Charif (doctorants)

Mots-clés

Automatisation du raisonnement pour répondre à des requêtes, découvrir, composer des services, Agents mobiles

Notre équipe développe des outils et des méthodes pour l'interaction entre des entités autonomes (agents) situées sur l'Internet. Dans ce contexte, nous proposons deux langages et plate-formes d'agents :

- CLAIM + SYMPA [6] qui permet de définir des agents mobiles cognitifs. En effet, les plate-formes d'agents mobiles existantes n'intègrent généralement pas la capacité des agents à planifier leurs actions en fonction de buts. Le langage CLAIM permet de définir de tels agents. Cette plate-forme a été intégrée avec des fonctionnalités de web services : les agents CLAIM interagissent par envoi de messages SOAP dans des interfaces WSDL.
- VDL qui permet de définir des Services Web Sémantiques et Actifs (ASWS) [4].

Contrairement aux services classiques, les agents VDL sont proactifs, c'est-à-dire qu'ils peuvent interagir en dehors de toute interaction de l'utilisateur. En particulier, l'interopérabilité entre ASWS n'est pas définie a priori par le concepteur du système. Elle est définie au moyen d'un dialogue (c'est-à-dire d'un échange de requêtes formelles) à propos des capacités d'action et d'interaction entre les services. Du point de vue de l'utilisateur humain, ces services se comportent comme des pages web actives.

VirtualSysBio

Membres

Michel Soto, Bénédicte Le Grand (permanents)

Mots-clés

Recherche d'information dans des systèmes complexes, Visualisation du Web Sémantique, Analyse conceptuelle pour l'extraction de connaissances (concepts et relations), Construction et évolution des ontologies

L'objectif des recherches menées autour du Web Sémantique dans l'équipe VirtualSysBio visent à faciliter la compréhension et la prise en main de systèmes complexes. Nous avons exploré plusieurs pistes de recherche dans ce but :

- proposition de techniques de visualisation et de navigation permettant d'appréhender plus facilement des systèmes complexes,
- recherche de sémantique implicite par l'analyse des données avec des méthodes conceptuelles pour fournir différents niveaux de granularité et s'adapter aux besoins des utilisateurs
- méthodes d'extraction de concepts et de relations pour automatiser une partie du processus de construction d'ontologie et en faciliter la maintenance.
- Couplage de la sémantique implicite et de la sémantique explicite (lien entre les informations contenues dans un système et des connaissances plus générales, par exemple des ontologies).

Bibliographie

[1] M.-A. Aaufaure, B. Le Grand, M.

Soto, N. Bennacer, « Metadata- and Ontology-Based Semantic Web Mining »

- [2] M.-R. Amini, P. Gallinari « Semi-Supervised Learning with Explicit Misclassification Modeling », International Joint Conference on Artificial Intelligence (IJCAI), 2003.
- [3] Henri Binsztok, Thierry Artières, Patrick Gallinari, « A model-based approach to sequence clustering », ECAI 2004.
- [4] Yasmine Charif et Nicolas Sabouret (2004) « A Model of Interactions about Actions for Active and Semantic Web Services ». In Semantic Web Services workshop at 3rd International Semantic Web Conference.
- [5] Jean-Yves Delort, Bernadette Bouchon-Meunier, Maria Rifqi, « CEA: A Content-Based Algorithm to Detect Users' Shifts of Focus on the Web », In Proceedings of Hypertext, Hypermedia, Products, Tools and Methods (H2PTM), 2003, Paris, France
- [6] Amal El Fallah Seghrouchni et Alexandru Suna (2004) « CLAIM: A Computational Language for Autonomous, Intelligent and Mobile Agents ». In Dastani, Dix, El Fallah Seghrouchni (eds), Workshop on Programming Multi-agent Systems, Lecture Notes in Artificial Intelligence, Vol 3067, pp. 90--110, Springer Verlag.
- [7] Jean-Gabriel Ganascia, Irène Fenoglio, Jean-Louis Lebrave, « Manuscrits, genèse et documents numérisés EDITE : une étude informatisée du travail de l'écrivain », revue Document Numérique, numéro spécial sur le thème « Temps et Document » à paraître en 2004.
- [8] Nicolas Labroche, « Mesure d'audience sur Internet par populations de fourmis artificielles », a paraître aux éditions Cépaduès, 2005
- [9] Bénédicte Le Grand, Michel Soto, « Topic Maps, RDF Graphs, and Ontologies Visualization », chapitre du livre « Visualizing the Semantic Web, second edition », V. Geroimenko et C. Chen, ed. Springer, à paraître (1ère édition parue en 2002).
- [10] Nicolas Lumineau et Anne

Doucet, « Sharing Communities Experiences for Query Propagation in Peer-to-Peer Systems », In 8th International Database Engineering and Applications Symposium (IDEAS'04), Coimbra, Portugal, 7-9 July 2004.

- [11] Julien Velcin et Jean-Gabriel Ganascia, « Stereotype Extraction with Default Clustering », IJCAI 2005, à paraître.

LIPN – Institut Galilée, Université Paris 13

Laboratoire d'Informatique de Paris-Nord, UMR 7030 du CNRS
Institut Galilée, Université Paris 13
99, av. J.B. Clément 93430 Villetaneuse
<http://www-lipn.univ-paris13.fr>

La structuration et l'exploitation des données issues du web nécessitent le développement d'expertises multiples du fait de la diversité de la nature de ces données (comportement des internautes, données textuelles, services web, etc.). Ces expertises font appel à de nombreuses thématiques de recherches développées au laboratoire.

1. Apprentissage numérique

Membres

Y. Bennani (Professeur), K. Benaddeslem (Post-Doctorant CNRS), F. Zehraoui (Docteur²)

La modélisation du comportement des utilisateurs (internautes) sur un site web est un enjeu important. Elle permet par exemple dans le contexte d'un site web de commerce électronique, d'analyser et de prédire le cheminement d'un visiteur, permettant ainsi des améliorations ciblées (e.g. ventes) de la structure et du contenu du site. Cette thématique a été au centre d'un contrat de recherche entre le LIPN et la société Numsight-Consulting et a servi de domaine d'application pour la validation des travaux effectués dans le cadre de deux thèses de doctorat.

Un procédé de codage original des traces de navigation des internautes y a été proposé et utilisé pour construire une cartographie des usages du site web fournissant ainsi une représentation interprétable [1]. Cette technique a fait

l'objet d'un brevet international. Deux approches originales d'apprentissage connexionniste non-supervisé ont également été utilisées pour visualiser, classer et prédire le comportement des internautes. La première est fondée sur des techniques de mélange de modèles (Chaînes de Markov Cachées+Cartes topologiques) et la deuxième sur un modèle de Kohonen temporel original (M-SOM-ART) qui combine des propriétés de stabilité et de plasticité [2]. Cette dernière approche a permis d'implémenter un système hybride de prédiction du comportement des internautes combinant le raisonnement à partir de cas (RàPC) et l'apprentissage connexionniste (elle a été utilisée pour implémenter à la fois la phase de recherche et la phase de réutilisation du cycle RàPC).

2. Raisonnement à partir de cas

Membres

R. Kanawati (MC), H. Karoui (Doctorante), S. Salotti (MC), F. Zehraoui (Docteur¹)

La méthodologie RàPC est principalement employée au laboratoire dans la fouille de données d'usages du web. Différentes applications ont été réalisées. L'une d'entre elles concerne la prédiction du comportement des utilisateurs d'un site de commerce électronique (cf. section précédente).

Une collaboration avec l'EISTI a par ailleurs permis de définir une approche RàPC pour l'adaptation automatique des sites web dynamiques. Un schéma de codage original des fichiers de traces de sites web, basé sur des actions comportementales, y assure une meilleure prise en compte de la nature dynamique des sites. Le RàPC est également employé au sein de systèmes multi-agents de type égal à égal (peer to peer) pour permettre à un groupe d'utilisateurs de partager leurs expériences de recherches d'informations sur le web. Deux systèmes ont été développés autour de cette problématique : un système pour la gestion coopérative de signets [3] et une application visant à améliorer le tri des réponses rendues par les moteurs de recherche.

3. Traitement automatique des langues Membres

A. Nazarenko (MC³), T. Poibeau (CR1), T. Hamon (MC), T. Aït El Mekki (Docteur), S. Szulman (MC), B. Biébow (MC), C. Fouqueré (Professeur)

La masse et la diversité de nature des données textuelles récupérables sur le web rendent leur exploitation difficile pour les utilisateurs, les moteurs de recherche actuels s'avérant souvent insuffisants.

L'équipe Représentation des Connaissances et Langage Naturel du LIPN développe depuis plusieurs années des techniques d'extraction d'information, de recherche documentaire et de navigation dans les documents [4]. Le développement de tels outils, en particulier dans les domaines de spécialité, repose sur des techniques d'analyse robuste des documents mais aussi sur des méthodes d'acquisition de connaissances à partir de textes, celles-ci permettant de construire les lexiques, ontologies, et règles d'extraction qui sont exploités pour l'analyse des documents [5].

Ces recherches ont débouché sur le développement de différents outils d'analyse des textes du web. IndDoc, par exemple, est un outil d'aide à la construction d'index pour faciliter la navigation dans les documents. Les contrats Caderige, ExtraPloDocs et aujourd'hui le projet européen STREP ALVIS portent sur la problématique de l'analyse des bases de données textuelles de biologie (comme Medline ou Flybase), facilement accessibles sur le web, mais dont l'exploitation efficace reste un problème critique pour les chercheurs en biologie.

Le problème de la pertinence des pages récupérables sur le web [6] a été attaqué dans le projet franco-québécois UQAM - LLI - LIPN en évaluant l'apport possible des techniques de traitement du langage à une recherche d'information ne reposant pas sur une organisation de la base de données textuelles. L'étude a été réalisée en mesurant

2. Récemment nommée MC à Evry.

3. Récemment nommée Professeur à Paris 13.

l'effet d'une variation de la requête initiale sur la pertinence des pages récupérées. Elle constitue un premier pas vers la spécification de mécanismes de reformulation de requêtes intégrables aux moteurs de recherches.

Le web est également une source d'information pour l'acquisition de connaissances. Les travaux menés dans le domaine de l'ingénierie de connaissances à partir de textes ont conduit au développement de l'outil *Terminae* qui permet la construction d'ontologies à partir de textes. Cette approche a pu être appliquée sur des corpus extraits du web pour créer des ontologies de domaine possédant des biais applicatifs particuliers.

4. Programmation et services web

Membres

F. Lévy (Professeur), J-V. Loddo (MC), C. Fouqueré (Professeur), N. Amara-Hachmi (Doctorante, sous la direction de A. El Fallah-Segrouchni du LIP6)

La mise en oeuvre des services web et la programmation de sites web dynamiques posent de nombreuses difficultés, liées à la distribution et au caractère dynamique des données et des programmes. Divers travaux menés au laboratoire touchent à cette problématique. La planification d'actions complexes dans les services web est l'objet d'une collaboration avec le CNET. On s'attache ici à caractériser les plans d'actions redondants, selon plusieurs critères de redondance, dans le but de maintenir une bibliothèque de plans pré-établis (utilisés lors de la détection de requêtes similaires) et d'étudier le choix du meilleur plan.

Les difficultés de programmation des sites et services web sont également abordées sous un angle plus théorique. Le but de ce projet est la définition et l'implémentation d'un langage de programmation dévoué à la programmation de sites ou services web, les langages usuels s'avérant inadaptés à cette tâche. Ce développement s'inscrit dans le cadre d'un projet européen ambitieux intégrant un traitement de données multimédia (projet *HyperLearning*).

La technologie des agents mobiles apporte aussi des solutions attractives relativement à ces questions, mais elle est difficilement accessible aux programmeurs. Une plate-forme de développement et d'exécution d'agents mobiles a été proposée pour résoudre ce problème d'accessibilité. L'effort a ici porté sur la modularité et l'adaptabilité des agents mobiles, ainsi que sur leurs propriétés de reconfiguration dynamique.

Bibliographie

- [1] K. Benabdeslem K., Y. Bennani, E. Janvier, « Visualization and Analysis of Web Navigation Data », International Conference on Artificial Neural Networks (ICANN '02), pp. 486-491, Madrid, Espagne, Août 2002.
- [2] F. Zehraoui, Y. Bennani, « M-SOM-ART: Growing Self Organizing Map for sequence clustering and classification », European Conference on Artificial Intelligence (ECAI '04), Valencia, Spain, August 2004.
- [3] R. Kanawati, M. Malek, « CoWing: A Collaborative Bookmark Management System », International Workshop on Collaborative Information Agents CIA'01, Springer LANI 2182, pp. 38-44, Modena, septembre 2001.
- [4] T. Poibeau, *Extraction automatique d'information : du texte brut au web sémantique*, Hermès, Paris, 2003.
- [5] C. Nédellec, A. Nazarenko, « Ontology and Information Extraction: a necessary Symbiosis », In, P. Buitelaar, P. Cimiano and B. Magnini (eds.), *Ontology Learning from Text: Methods, Applications and Evaluation*, Advances in Artificial Intelligence, IOS Press, Amsterdam, à paraître 2005.
- [6] L. Emirkanian, C. Fouqueré, F. Issac, « Corpus issus du Web : analyse des pertinences thématique et informationnelle », JADT 2004, vol. 1, pp. 390-398.

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)

Membres

Rim Al-Hulou, Mathieu d'Aquin, Rokia Bendaoud, Adrien Coulet, Marie-Dominique Devignes, Sandrine Lafrogne, Jean Lieber, Nizar Messai, Malika Smail-Tabbone, Sylvain Ténier, Yannick Toussaint.

Les travaux de recherche de l'équipe Orpailleur s'articulent autour de trois axes majeurs : l'extraction de connaissances dans les bases de données, la gestion et la représentation de connaissances, et enfin le Web sémantique. Ces axes sont interdépendants et l'élément commun partagé par ces trois axes est la classification, qui intervient à tous les niveaux le long de ces trois axes. L'objectif in fine est de concevoir des systèmes intelligents pour résoudre des problèmes dans des domaines d'application comme l'agronomie, la biologie, la chimie, la médecine, la sidérurgie ...

Plusieurs projets sont concernés par l'ingénierie du Web et le Web sémantique, qui vont être décrits ci-après. Dans le projet *Kasimir*, les principes et les technologies du Web sémantique sont appliqués et étendus pour construire un portail sémantique pour la gestion des connaissances décisionnelles en cancérologie. Ce portail repose sur une représentation formelle des référentiels (protocoles de décisions en cancérologie) dans le standard de représentation d'ontologies du Web sémantique OWL. Les mécanismes de raisonnements associés à OWL permettent ainsi un accès intelligent à ces connaissances au travers d'outils d'aide à la décision sur le Web, à destination des médecins lorrains. Mais, si les principes du Web sémantique semblent répondre aux problématiques de formalisation et de diffusion de connaissances du projet *Kasimir*, les technologies telles que OWL montrent certaines limites vis-à-vis des besoins en représentation du projet *Kasimir*. Pour cette raison, ces technologies sont étendues pour la représentation multi-points de vue et la représentation floue ainsi que

pour l'adaptation des référentiels selon le principe du raisonnement à partir de cas.

Un second projet concerne l'extraction de connaissances à partir du Web avec deux objectifs, l'un est de construire puis d'enrichir des ontologies à partir du Web, et l'autre est de développer une méthodologie et des outils d'extraction des connaissances depuis des pages Web annotées sémantiquement. Pour cela, il faut combiner des mécanismes de raisonnement s'appuyant sur des connaissances du domaine, des techniques d'extraction d'information et d'extraction de connaissances. Ces travaux de recherche s'inscrivent dans la lignée des travaux de l'équipe sur la fouille de textes et la classification des règles extraites en fonction des connaissances du domaine. Ainsi, les méthodes actuelles de cartographie à partir de données du Web laissent une part d'interprétation très importante à l'utilisateur car la représentation du contenu est généralement trop pauvre. Nous allons pour notre part introduire dans l'extraction d'information à partir du Web la représentation formelle de connaissances en faisant appel notamment à OWL et RDF(S). Le cadre applicatif étudié consiste à mettre en oeuvre une analyse synthétique (cartographie) des compétences européennes en matière de recherche (travail actuel de thèse).

Un troisième projet est plus porté sur l'ingénierie du Web et la bioinformatique. L'un des grands défis de la bioinformatique aujourd'hui est de permettre aux biologistes d'accéder efficacement aux données gisant dans les centaines de sources de données réparties à travers le monde. Le grand nombre de sources, leur hétérogénéité et la complexité des objets biologiques auxquels elles font référence rendent souvent difficile la mise en relation d'une requête avec la source appropriée. Parfaitement compatible avec le Web sémantique, le projet BioRegistry fournit des éléments de solution à ce genre de problèmes. En effet, ce projet vise à associer aux sources biologiques des méta-données décrivant leur contenu et la qualité des données. Les méta-données sont organisées dans un modèle hiérarchique qui constitue une extension des éléments du DCMI (Dublin Core Metadata Initiative) dans lequel des onto-

logies ou vocabulaires standards du domaine sont utilisés pour renseigner les méta données. L'application des méthodes de l'analyse de concepts formels a permis d'organiser les sources de données biologiques sur la base du partage de propriétés. De plus, des mécanismes d'enrichissement sémantique de requêtes, fondés sur l'utilisation des ontologies, permettent d'améliorer encore la découverte des sources pertinentes.

Enfin, d'autres travaux sont également en cours dans l'équipe Orpailleur, équipe qui participe avec les projets Inria Exmo (Inria Rhône-Alpes) et Acacia (Inria Sophia) au réseau européen Knowledge Web (<http://knowledgeweb.semanticweb.org/>). Dans ce cadre sont étudiés des problèmes d'alignement d'ontologie, de recherche d'information guidée par des ontologies et de mise au point de mesures de similarité pour manipuler des documents par leur contenu.

Bibliographie

- [1] R. Al-Hulou, A. Napoli et E. Nauer, Une mesure de similarité sémantique pour raisonner sur des documents, Langages et modèles à objets, Lille (LMO'04), J. Euzenat et B. Carré éditeurs, (Hermès, L'objet 10(2-3)), pages 217-230, 2004.
- [2] M. d'Aquin, C. Bouthier, S. Brachais, J. Lieber and A. Napoli, Knowledge Edition and Maintenance Tools for a Semantic Portal in Oncology, International Journal on Human-Computer Studies, 62(5): 619-638, 2005.
- [3] M. d'Aquin, S. Brachais, J. Lieber and A. Napoli, Kasimir : portail sémantique pour la gestion des connaissances en cancérologie, Second séminaire francophone du Web Sémantique Médical - WSM 2004, Rouen, 2004.
- [4] D. Janetzko, H. Cherfi, R. Kennke, A. Napoli and Y. Toussaint, Knowledge-based Selection of Association Rules for Text Mining, 16h European Conference on Artificial Intelligence (ECAI'04, Valencia, Spain), R. Lopez de Màntaras and L. Saitta editors, pages 485-489, 2004.
- [5] N. Messai, M.-D. Devignes, A. Napoli et M. Smal"l-Tabbone M.,

Querying a bioinformatic data sources registry with concept lattices, In F. Dau, M.-L. Mugnier, G. Stumme, editor, 13th International Conference on Conceptual Structures (ICCS'05), Kassel, Germany, LNCS 3596, pages 323-336, 2005.

Laboratoire Paragraphe Université Paris VIII

Laboratoire Paragraphe,
Université Paris 8, Département
Hypermédiat
2 Rue de la liberté
93526 Saint-Denis Cedex

Contact : Imad Saleh, Professeur au
département Hypermédiat
Mél : imad.saleh@univ-paris8.fr

Paragraphe est un laboratoire habilité depuis 1983 (EA 349) attaché à l'Université Paris VIII, Saint-Denis. C'est un laboratoire de recherche interdisciplinaire rattaché à l'Ecole Doctorale Société Information Informatique Culture (SIIC). Il compte 22 enseignants-chercheurs permanents, 5 ATER et 54 doctorants.

Laboratoire de recherche interdisciplinaire actif et reconnu dans les domaines suivants : l'hypermédia distribué, la génération automatique de textes et d'images, le web sémantique, les moteurs de recherches textuels et images, bases de connaissances multimédia, indexation du contenu, visualisation de l'information, interface Homme/Machine etc., nombreuses publications internationales, plusieurs ouvrages publiés, organisation de sept colloques internationaux (conférence H2PTM), réalisation de plusieurs prototypes de recherche et plusieurs installations artistiques basées sur les générations de textes, d'images et de sons, nationales (Paris, Valenciennes, Montbéliard,...etc.) et internationales (Monterrey, Berlin, Florence,...etc.). Paragraphe est porteur du projet CITU avec le laboratoire CRECA de l'université Paris 1.

Le laboratoire Paragraphe participe à plusieurs contrats de recherches français

et européens (FEDER, RTP, ACI, CRAFT ...), industriels (Développement d'outils de génération automatique multilingues en partenariat avec Thomson-multimédia, 2 dépôts de brevets internationaux) et enfin des contrats directs avec des partenaires industriels (dont en particulier plusieurs thèses sous convention Cifre).

Dans ce cadre, les projets du laboratoire se développent dans six directions :

- 1 - Méthodes et outils de génération automatique de textes, d'images et de sons.
- 2 - Méthodes et outils d'hypertextualisations dynamiques.
- 3 - Bases de connaissances multimédia.
- 4 - Technologies de l'hypermédia distribué.
- 5 - Communication Homme/Machine.
- 6 - Conception et réalisation des outils multimédia.

La production de Paragraphe au cours du quadriennal est de 64 articles dans des revues avec comité de lecture, 40 contributions à des ouvrages et 150 communications à des conférences.

Quelques systèmes

HyperTectol

En réponse aux besoins des enseignants travaillant sur la conceptualisation et la publication des cours sur une plateforme virtuelle, HyperTectol est un outil qui a été développé dans le but de permettre aux utilisateurs de créer facilement un objet pédagogique en ligne et de le stocker dans un *Content Management System* (CMS) en forme de *Sharable Content Object* (SCO) partageable. La conception de l'interface visuelle considère une utilisation intuitive et simple qui ne demande pas des connaissances du langage HTML ni en informatique avancée. Par contre, une familiarisation avec les systèmes d'exploitation à base de fenêtres est conseillée.

Les documents hypermédia créés avec le système ont deux couches principales : la visualisation du contenu, qui est présentée en format HTML aux clients (enseignants et apprenants) depuis un navigateur conventionnel et la description de métadonnées, qui sont introduites par l'auteur dès le début d'une session.

Seules la manipulation et la modifica-

tion des objets pédagogiques sont sécurisées par un login et un mot de passe. L'accès aux documents est libre et la visualisation se fait à travers une interface qui pourra être adaptable selon les besoins spécifiques de l'action : niveau d'expertise de l'utilisateur (débutant, intermédiaire ou avancé), type de média à intégrer (texte, image, son, vidéo, animation ou hybride), type de travail à réaliser (création d'un nouveau document, modification d'un document existant ou téléchargement d'un document à partir de CMS). HyperTectol permet également l'affichage de tous les utilisateurs connectés dans un même réseau.

ICRS

ICRS est un système de recherche collaborative d'informations qui offre à la fois la possibilité à plusieurs utilisateurs de travailler ensemble dans le cadre de leurs recherches, tout en s'entraînant pour mettre en place un appareil critique pour chaque document trouvé à travers les commentaires, les annotations et l'évaluation. Ce travail collaboratif qu'offre le système permet aussi à l'utilisateur des ressources numériques de sortir de son état de solitude pour intégrer une communauté virtuelle de chercheurs qui lui sert entre autres comme un « bon » intermédiaire entre lui et le système. Dans ce système, nous avons proposé aussi un autre espace de travail collaboratif, celui du répertoire de modèles XML. Ce répertoire permet de tenir l'utilisateur au courant de tous les modèles existants dans son domaine, et notamment ceux de ses « partenaires ». Cet espace offre les moyens à l'utilisateur d'être à la fois « lecteur » (à travers la recherche et l'accès aux modèles déjà existants) et « auteur » (à travers les possibilités de créer son propre modèle et de le mettre à la disposition de tous les utilisateurs et/ou d'une communauté).

Par ailleurs, ICRS propose des documents structurés en format XML selon les modèles publiés dans le répertoire de schémas XML. La typologie de ces documents est très variée et sans aucune contrainte (cours, objets pédagogiques, mémoires de fin d'étude, des thèses, des rapports, etc.) et qui peuvent bien sûr contenir du texte, des images, de la vidéo, des graphiques, etc. Cette structuration

permettra certainement de faciliter la recherche d'informations, parce qu'avec le développement d'une ontologie dans le(s) domaine(s), nous pourrions évoluer vers un Web sémantique, où les informations seront compréhensibles aussi bien par les humains que les machines. Cette utilisation des standards XML (description des documents), XML Schema (modèles logiques des documents), RDF et RDFs (pour les métadonnées et les ontologies), SMIL (pour la création des contenus multimédias) favorise sans doute le travail collaboratif entre les différents utilisateurs et les différentes applications, tout en évitant un grand travail de conversion et de « parsing » des données provenant d'autres systèmes.

HyWebMap (Hypertext Web Mapping) : Un système hypertextuel sur le Web

HyWebMap est un système hypertexte pour la navigation et la structuration des documents sur le Web créé dans le cadre de l'élaboration et de la conception des systèmes hypermédia à vocation didactique au laboratoire Paragraphe à l'université Paris 8. Il est utilisé entre autres dans le domaine de l'éducation, la veille technologique et l'enseignement à distance et constitué, suite à une étude des besoins des utilisateurs, de plusieurs niveaux.

Nous reconnaissons 5 tâches essentielles à l'utilisateur : l'extraction d'information, le traitement de l'information, la « veille », la « suggestion » d'information, la personnalisation de l'information.

Ces fonctionnalités nous servent de modèle à base de couches logiques pour la conception d'HyWebMap. L'originalité d'HyWebMap repose d'une part sur ces cinq fonctionnalités qui s'appellent les unes les autres, et d'autre part sur la transversalité des usages ; l'éducation, la gestion de sites Web, la création de portail, la veille technologique, etc.

HNLS : créer, gérer et générer des espaces d'informations personnalisés en ligne

HNLS (Hypertext oN Line System) est un système qui a été créé dans le cadre de l'évolution du système HyWebMap (Hypertext Web Mapping) développé au sein du laboratoire Paragraphe.

HNLS peut être défini comme étant un

système de navigation et de structuration des documents sur le Web basé sur la notion de « document-réseau », mais surtout comme un système hypermédia interactif de gestion des connaissances en ligne. Un système destiné en premier lieu à favoriser l'appropriation des savoirs par l'utilisateur. Il est créé dans le cadre de l'élaboration et de la conception des systèmes hypermédiés à vocation didactique. Il est utilisé entre autres dans le domaine de l'éducation, la veille technologique et l'enseignement à distance et constitué, suite à une étude des besoins d'utilisateurs.

K-Web Organizer : un système de supervision de travail collaboratif

Le système K-Web Organizer rend possible la création d'une communauté d'utilisateurs fédérée autour de ressources documentaires thématiques. Elles sont produites directement par ses membres ou par des contributions réalisées à partir de compilations de documents pour lesquelles ils imaginent une organisation particulière. Ces productions originales ou composites mais toujours individuelles sont mises à la disposition de tous et participent à un seul et même fond documentaire dans lequel les interactions entre auteurs et documents sont signalées en temps réel à toute la communauté. La communauté dont nous traitons ici est une conséquence de l'association des réseaux virtuels engendrés par les utilisateurs-auteurs ; associations opérables en pratique par le système à partir d'objets documentaires (les réseaux virtuels personnels) dont les contenus peuvent être associés.

L'objectif que nous visons avec K-Web Organizer est un dispositif permettant d'agréger des connaissances homogènes au sein d'un environnement. Celles-ci se développent uniformément à partir d'un noyau stable et cohérent de connaissances expertisées sur lequel viendront se greffer de nouvelles contributions (l'ensemble noyau et contributions s'instanciant au travers de réseaux virtuels du type de ceux produits par *HyWebMap*). Les contributions satellitaires à ce noyau central pourront former de nouveaux points d'ancrages susceptibles d'être perçus comme autant de noyaux décentralisés. Ils inspireront de nouvelles contributions, formant alors

une sorte de maillage en volume dont l'expansion serait assurée à la fois par les contributions complémentaires se rattachant, au sein du même environnement collaboratif⁴, à des réseaux virtuels précis et par l'enrichissement permanent des réseaux virtuels composant l'environnement en question.

NeuroWeb : un moteur de recherche multilingue et cartographique.

Le laboratoire Paragraphe a développé des méthodes d'exploration de corpus à partir de cartographies textuelles, qu'il étend à des langues et écritures différentes par la technique des n-grammes. Un prototype de moteur de recherche couplant approche multilingue par les n-grammes et approche linguistique est présenté, qui permet à la fois des requêtes fines utilisant une analyse linguistique, et des vues d'ensemble à la demande au moyen de cartographies.

Roman interactif Trajectoires

« Trajectoires » roman interactif et génératif pour internet. Prix de la fondation Hachette 2000. Ce roman interactif intitulé « Trajectoires » va produire quotidiennement pendant un certain temps des fragments de textes. Ces fragments de textes seront générés à volonté par l'internaute et enregistrés pour lui, jusqu'à la résolution de l'intrigue ou jusqu'à la fin de la génération programmée. La génération est programmée mais la fin aussi.

Equipe Recherche Opérationnelle et Informatique (ROI) - LAMIH

Laboratoire LAMIH
(UMR CNRS 8530)
Le Mont Houy
593133 Valenciennes Cedex 9
<http://www.univ-valenciennes.fr/LAMIH/ROI>

Contact : Sylvain Lecomte
Mél : sylvain.lecomte@univ-valenciennes.fr
Tel : 03 27 51 19 49

4. C'est ici que se superpose la notion de communautés d'utilisateurs et de réseaux virtuels thématiques agrégés : un environnement collaboratif de réseaux virtuels représente les contributions des utilisateurs autour d'une thématique centrale. 2. Récemment nommée MC à Evry.

Membres

Nadia BENNANI (MCF), Thierry DELOT (MCF), Sylvain LECOMTE (MCF), Marie THILLIEZ (MCF)

Thème général de l'équipe

L'activité de recherche du thème Recherche Opérationnelle et Informatique - ROI s'est développée autour de la problématique générale visant à contribuer à la conception de logiciels performants permettant l'élaboration de scénarios pour éclairer la prise de décision. Le domaine d'applications principal est le transport. Les travaux se répartissent en deux sous thèmes principaux :

Recherche Opérationnelle et Aide à la Décision qui s'est développé autour de l'optimisation combinatoire. Situé à l'interface entre la programmation mathématique, l'informatique et la recherche opérationnelle, ce champ connaît une grande variété d'approches algorithmiques, que ce soit pour les méthodes exactes de type énumérative ou polyédrale, ou les méthodes appelées heuristiques qui ne garantissent pas à tout coup la qualité des solutions produites. Durant les dernières décennies, des progrès significatifs ont été réalisés dans l'élaboration de méthodes de plus en plus performantes. Ils autorisent une résolution satisfaisante (au sens du temps de calcul et de la qualité de solutions produites) de problèmes NP-difficiles de taille de plus en plus importante et intervenant dans la modélisation de problèmes réels de plus en plus complexes. L'activité du sous thème se focalise d'une part sur le développement de nouvelles approches algorithmiques pour la résolution de problèmes NP-difficiles académiques et d'autre part sur la modélisation de systèmes complexes via la programmation en nombres entiers et mixte et sur le développement d'algorithmes exacts et d'heuristiques efficaces pour résoudre des problèmes NP-complets.

Systèmes d'Information Distribués qui s'intéresse principalement à la distribution des données et des applications dans un contexte de plus en plus hétérogène. En effet, il est possible aujourd'hui d'interconnecter, à tout moment et

de manière ponctuelle, différents outils informatiques. Cependant, cette hétérogénéité amplifie des problèmes bien connus en informatique distribuée, comme la tolérance aux pannes, la localisation de services ou de données, ou la sécurité. Il est donc nécessaire à la fois créer de nouveaux services et d'adapter les services existants. Une autre direction de recherche explorée par l'équipe est l'adaptation de méthodes permettant la certification de composants logiciels. Les exigences de sécurité sont en effet des éléments essentiels du développement logiciel, en particulier pour des composants embarqués ou enfouis dans des systèmes de transports terrestres.

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

Les travaux en lien avec le thème du dossier sont effectués dans le thème Systèmes d'Information Distribués. Ils sont de 2 ordres : il y a tout d'abord les travaux qui consistent à étudier l'auto-adaptation des services techniques (tels que la gestion des transactions, la sécurité, la persistance des données, etc.) de manière à pouvoir utiliser ces services sur n'importe quel type de terminaux, en fonction des besoins des applications. Un des premiers buts de ces travaux a été d'étudier les besoins des nouvelles applications (et notamment les applications de commerce électronique de type B2B sur le web) pour fournir un moniteur transactionnel répondant à ces besoins (moniteur capable de gérer plusieurs modèles transactionnels, comme les transactions plates, les transactions emboîtées fermées ou les transactions emboîtées ouvertes). Ensuite, ces travaux consistent à généraliser l'approche d'adaptation qui a été menée pour le service transactionnel, pour l'appliquer à tous les services techniques, de manière automatique, en tenant compte à la fois des contraintes applicatives et des capacités des terminaux. Ce travail ne porte donc pas sur le service transactionnel, mais sur l'ensemble des services techniques mis à la disposition d'un programmeur d'applications (nommage, courtage, sécurité, transaction, persistance, etc.). Pour répondre au besoin d'adaptation, nous avons redéfini chaque étape du cycle de vie d'un service technique : conception, développement,

assemblage, déploiement et exécution. Concernant la conception et le développement, nous proposons de construire les services techniques eux-mêmes sous forme de composants Fractal et de donner une véritable représentation dans le système des différentes personnalités d'un service technique. Ces propositions ont été mises en oeuvre dans le service transactionnel proposé précédemment. Pour assurer l'automatisme de l'adaptation, nous devons définir un ensemble de composants de gestion de l'adaptation : contrats, coordinateur, moniteur et annuaire de services techniques. Ces composants permettent de localiser, choisir et utiliser les nouveaux services techniques.

Ensuite, sont également concernés par ce dossier, les travaux qui consistent à créer de nouveaux services pour l'informatique ubiquitaire. Dans la continuité des travaux portant sur l'évolution des terminaux, nous avons constaté que, dans certains cas, l'adaptation de services existants, comme présentée dans la partie précédente, ne suffirait pas. Le nomadisme grandissant des utilisateurs et les nouveaux réseaux mobiles et/ou sans fil favorisent le développement de nouveaux services et de nouvelles applications dédiées aux usagers mobiles. Parmi ces applications, les applications de proximité s'inscrivent dans un contexte fortement distribué. Un utilisateur va, suivant ses déplacements, appartenir à différentes sphères de communication et pouvoir participer à différentes applications de proximité avec d'autres utilisateurs physiquement proches de lui. Il doit être en mesure de découvrir et/ou de localiser les données disponibles. Dans cet environnement mobile et dynamique, les services de localisation actuels présentent d'importantes limites. Nous avons donc proposé un service complètement nouveau, adapté à la distribution de l'information et qui considère la mobilité des utilisateurs et donc, est capable d'évaluer des requêtes dépendantes de la localisation (LDQ) telles que « quel est l'arrêt de bus le plus proche de moi ? ».

De plus, afin de permettre l'évaluation des requêtes dépendantes de la localisation, ce service doit disposer de

la localisation géographique du client qui émet la requête. Aujourd'hui, les techniques de localisation ne sont pas toujours disponibles : par exemple, le GPS qui est la solution la plus répandue, ne fonctionne pas à l'intérieur d'un bâtiment, nous avons donc proposé une solution de localisation reposant sur les méta-données de l'environnement. Notre solution permet de localiser un utilisateur de façon approximative mais largement suffisante pour l'évaluation des requêtes dépendantes de la localisation. Cette solution a été optimisée de façon à minimiser la consommation des ressources sur les terminaux nomades et à réduire le nombre de communications entre les participants.

Ce service n'est pas une adaptation de services existants comme nous avons pu le faire précédemment avec le service transactionnel, mais bien un service nouveau, qui n'existait pas dans les applications distribuées traditionnelles.

En conclusion, les travaux du thème Systèmes d'Information Distribués de l'équipe Recherche Opérationnelle et Informatique dans le cadre du dossier « Ingénierie du Web et Intelligence Artificielle » sont de 2 ordres. Ils touchent tout d'abord « la Personnalisation et adaptation des applications Web », en fournissant des services techniques (comme la sécurité ou la gestion des transactions), capables de prendre en compte à la fois les contraintes des terminaux, mais également les besoins des applications pour fournir une qualité de service optimale à l'utilisateur. Ils touchent également au « Données et services distribués », en fournissant de nouveaux services, comme le service de requêtes dépendantes de la localisation, à des applications reposant sur une architecture pair à pair. Les domaines d'applications de ces travaux sont nombreux, mais on peut notamment citer la e-santé avec les travaux sur l'hospitalisation à domicile, le commerce électronique, avec les travaux sur le commerce électronique de proximité, ou encore les transports, avec les travaux sur la communication inter-véhicules.

Bibliographie

L'ensemble des articles peut être trou-

vés à l'adresse suivante : http://www.valenciennes.fr/LAMIH/site/commun/_gestion/publis/recherche/recherche_publis.php?

- [1] Thilliez M., Delot T., Lecomte S. An Original Positioning Solution to Evaluate Location-Dependent Queries in Wireless Environments. Special Issue on Distributed Data Management in Journal of Digital Information Management, 2005.
- [2] Anceaux F., Bennani N., Bricon-Souf N., Watbled L., Beuscart-Zéphir M., Lecomte S. Prise en charge de patients à domicile: un contexte nécessitant l'élaboration de systèmes d'informations coopératifs. Revue des Sciences et Technologies de l'Information, Série Ingénierie des Systèmes d'Information, 8, pp. 75-93. 2003.
- [3] Thilliez M., Delot T. A Localization Service for Mobile Users in Peer-To-Peer Environments. In Fabio Crestani, Mark Dunlop, Stefano Mizzaro (Ed.), Mobile and Ubiquitous Information Access, Springer-Verlag, LNCS 2954, pp. 271-282, ISBN 3-540-21003-2, 2004
- [4] BENNANI N., DONSEZ D., LECOMTE S., RAMOS F. Apport des agents mobiles dans le commerce électronique. In R. Mandiau, E. Grislin-le Strugeon, A. Péninou (Ed.), Organisation et applications des SMA, Hermès, Paris, pp. 321-331. 2002.
- [5] Hérault C., Lecomte S., Delot T. New technical services using the component model for applications in heterogeneous environment. In Innovative Internet Community Systems (I2CS), Guadalajara, Jalisco, Mexique, juin. 2004

Equipe RTIC (Réseaux et Technologies de l'Information et de la Communication)

Laboratoire d'Informatique de l'Université François-Rabelais de Tours (EA 2101)
Polytech'Tours, 64 avenue Jean Portalis,
37200 Tours

Tel : 02 47 36 14 33

Fax : 02 47 36 14 22

Web : <http://www.antsearch.univ-tours.fr>

Contact : Gilles Venturini
venturini@univ-tours.fr

Membres

G. Venturini (PR, responsable d'équipe), C. Guinot (HDR), F. Picarougne (Docteur), A. Oliver (Ingénieur), A. Alarabi (Doctorant), H. Azzag (Doctorante), D. Da Costa (Doctorant), F. Mokaddem (Doctorant).

Thématiques de l'équipe

Il s'agit d'étudier, d'un point de vue théorique et jusqu'à la réalisation pratique, des problématiques réelles ou industrielles situées à l'intersection des trois domaines suivants :

Le **Web** (domaine d'étude et d'application) : recherche d'information, veille stratégique, construction automatique de sites portails et d'hypertextes, analyse d'audience, sites Web adaptatifs, visualisation d'information au travers du Web.

La fouille visuelle et interactive de données en réalité virtuelle (méthodes utilisant une visualisation interactive d'information pour toutes les problématiques liées au Web et utilisant éventuellement du matériel de réalité virtuelle pour faciliter l'interaction avec l'utilisateur) : visualisation des résultats d'un moteur de recherche, de l'arborescence d'un site portail, de la navigation d'utilisateurs sur un site Web.

Les algorithmes biomimétiques (inspirés de principes biologiques et utilisés dans les problématiques précédentes comme outils d'apprentissage et d'optimisation) : algorithmes génétiques, fourmis artificielles, intelligence en essaim et nuages d'agents, automates cellulaires.

Projets de l'équipe

Les différents projets et thèses en cours d'étude ou étudiés récemment sont les suivants (plusieurs démonstrations en ligne sur notre site Web, beaucoup de projets ci-dessous liés à la fouille visuelle de données sont réalisés en collaboration avec le C.E.R.I.E.S., Chanel, et un projet dans le cadre d'une bourse CIFRE avec la société AGI-COM) :

Optimisation interactive de pages

Web : un algorithme génétique interactif propose différentes versions d'une même page (disposition/style) afin de laisser l'utilisateur choisir celles qui lui semblent les meilleures et personnaliser ainsi le « look » de sa page ou de son site.

Modélisation du système d'identification chimique des fourmis et application à la classification de données de navigation issues du Web

(thèse de N. Labroche, soutenue en 2003) : chaque navigation d'un internaute sur un site est interprétée comme une « odeur chimique ». A partir des mécanismes utilisés par les fourmis réelles pour construire collectivement une odeur coloniale, on regroupe ainsi les visiteurs d'un site Web afin de présenter des groupes de sessions.

Veille stratégique sur le Web et utilisation d'un algorithme génétique parallèle (thèse de F. Picarougne, soutenue en 2004) : un algorithme génétique parallèle manipule une population de pages afin de trouver en un temps limité les pages maximisant leur adéquation à la requête d'un utilisateur. Les opérateurs génétiques utilisés consistent à interroger les moteurs classiques ou à explorer des liens dans les pages.

Optimisation des liens dans un hypertexte

(thèse de A. Alarabi) : à partir d'un grand ensemble de documents, comment créer automatiquement des liens hypertextes entre ces documents de manière à faciliter la navigation d'un utilisateur et à publier ces documents sur le Web ? Un algorithme génétique est utilisé pour créer des liens maximisant la corrélation entre la distance thématique des documents et la distance hypertextuelle (nombre de liens faisant passer d'un document à un autre).

Modélisation du comportement d'auto-assemblage des fourmis pour la classification hiérarchique et la construction

automatique de sites portail pour le Web (thèse de H. Azzag) : des agents/fourmis représentent des données/textes et vont construire de manière distribuée un arbre de données. Cet arbre peut ensuite être interprété

comme une classification plane ou hiérarchique, ou comme un site portail complet avec système de navigation et d'interrogation.

Etude d'un nouvel algorithme de classification de données utilisant une population d'agents en déplacement : application à l'exploration interactive de données au travers du Web (thèse de D. Da Costa). Ce travail doit permettre de manipuler/explorer des données au travers d'une interface Web en utilisant des techniques graphiques comme JAVA ou SVG.

Visualisation des résultats d'un moteur de recherche par essaim particulière (thèse de F. Mokaddem) : plutôt que de présenter des résultats sous forme de listes qui sont difficiles à exploiter par l'utilisateur, cette thèse propose une visualisation graphique et interactive des résultats. Les documents retournés se positionnent en 2D en fonction de leurs ressemblances.

Système interactif de visualisation et d'interrogation d'une base documentaire issue du Web pour la veille stratégique : nous étudions des techniques de visualisation à base de points d'intérêt pour explorer un ensemble de documents (des textes, un site Web, plusieurs sites Web) pouvant se modifier au cours du temps

Etude et conception d'un système de visualisation de données multimédia (sons, images, vidéos, sites Web, etc) en réalité virtuelle : le système VRMiner permet d'explorer des données multimédia en utilisant du matériel de réalité virtuelle qui facilite l'interface avec l'utilisateur. Ce système a obtenu le prix de la Fondation VediorBis lors de la conférence EGC'2003.

Bibliographie

(voir www.antsearch.univ-tours.fr)

[1] H. Azzag, C. Guinot, A. Oliver, G. Venturini. A hierarchical ant based clustering algorithm and its use in three real-world applications. To appear in Special Issue on Applications of Metaheuristics, European Journal of Operational Research (EJOR). 2005. Wout Dullaert, Marc Sevaux, Kenneth

Sörensen and Johan Springael Editors. 2005.

- [2] F. Picarougne, H. Azzag, G. Venturini, C. Guinot. A new approach of data clustering using a flock of agents. To appear in Evolutionary Computation (2005). Marc Schoenauer Editor. MIT Press. 2005.
- [3] F. Mokaddem, F. Picarougne, H. Azzag, C. Guinot, G. Venturini, Techniques visuelles de recherche d'informations sur le Web, à paraître dans Revue des Nouvelles Technologies de l'Information, numéro spécial Visualisation en Extraction des Connaissances, Pascale Kuntz et François Poulet rédacteurs invités, Cépaduès. 2005.
- [4] H. Azzag, C. Guinot, G. Venturini, AntTree: web document clustering using artificial ants. Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 04), p. 480-484, IOS Press, Valencia, Spain. 2004.
- [5] A. Oliver, O. Regragui, N. Monmarché, G. Venturini. Optimisation génétique et interactive de sites web. Technique et Science Informatique, 21(7):965-984. 2002.

Equipe SMAC- IRIT

Université Paul Sabatier
118 route de Narbonne
31062 TOULOUSE CEDEX
<http://www.irit.fr/SMAC>

Contact : Pierre Glize
Mél : glize@irit.fr,
Tel : 05 61 55 82 95

Membres de l'équipe concernés par le thème

Camps Valérie, Gleizes Marie-Pierre, Glize Pierre, Machonin André, Ottens Kévin, Pécatte Jean-Marie, Singer Nicolas, Trouilhet Sylvie

Thème général de l'équipe

Les travaux de l'équipe SMAC s'insèrent dans le domaine des Systèmes Multi-Agents et s'intéressent plus particulièrement à la conception de systèmes complexes ouverts, incomplètement spécifiés, plongés dans un environnement dynamique et permettant de résoudre des problèmes n'admettant pas d'algorithme connu pour construire la solution. Ces

systèmes sont composés d'agents qui poursuivent un objectif individuel et essaient de maintenir en permanence des interactions locales coopératives avec leurs congénères en évitant les dysfonctionnements tels que l'ambiguïté, l'improductivité, l'inutilité, les conflits ... Grâce à la capacité des agents à s'auto-organiser, le système est capable de s'adapter par lui-même à toute perturbation et produit alors une fonction collective adéquate dans son environnement.

Travaux

Le monde du Web, environnement difficile à appréhender de par sa forte évolutivité, l'absence de toute vision synthétique globale et de toute règle de structuration des services, est un champ d'application particulièrement intéressant pour nos travaux car il met en évidence les caractéristiques essentielles auxquelles répond la théorie des AMAS (Adaptative Multi-Agent System).

Nos travaux s'intéressent donc à ce domaine et peuvent être déclinés en trois axes fortement complémentaires qui s'intègrent essentiellement sur trois des thèmes proposés dans ce dossier (Web mining, personnalisation et adaptation des applications Web, applications Web).

1) Recherche d'un ou plusieurs services répondant à la requête d'un utilisateur

L'objectif convoité, lors d'une recherche d'information, est de mettre en relation des utilisateurs et des services ayant des intérêts communs. Nous proposons d'atteindre cet objectif en utilisant une recherche coopérative et adaptative entre agents représentant un utilisateur/service réel.

Cette recherche s'appuie, pour un utilisateur ou un service réel donné, sur la création et le maintien du profil de cet utilisateur/service en fonction des transactions déjà réalisées et en cours de réalisation, ainsi que sur la création et le maintien du profil des utilisateurs/services déjà contactés dans de précédentes recherches. Le profil contient des informations sur l'utilisateur/service, notamment ses préférences, et il est utilisé soit :

- Lors du traitement d'une requête pour

rechercher de manière plus pertinente des informations pour le compte de l'utilisateur/service.

- Lors de la diffusion d'offres pour communiquer des informations susceptibles d'intéresser l'utilisateur/service.

Le maintien du profil ne peut qu'être effectué dynamiquement, en fonction des interactions qu'entretient l'utilisateur/service avec le système, en sachant que la grande variabilité des profils interdit beaucoup de techniques traditionnelles d'apprentissage fondées sur la connaissance de pré-supposés.

Nous proposons une architecture multi-agent composée de 3 types d'agents distincts [1] [2] se situant respectivement à un niveau différent de l'architecture du système envisagée et dont le fonctionnement respectif est régi par la théorie des AMAS :

- L'agent de niveau intermédiaire représente un utilisateur ou un service réel et extérieur au système. Il a en charge la résolution de la requête que l'utilisateur ou le service qu'il représente lui soumet. Ces agents doivent tenter de maintenir en permanence des relations coopératives avec leurs congénères.
- L'agent de plus haut niveau représente un site du système d'information répartie et se compose d'un certain nombre d'agents de niveau inférieur représentant des utilisateurs et/ou des services réels. Il a pour objectif d'aider la résolution du problème en participant activement à la mise en relation d'agents utilisateurs et d'agents services partageant des intérêts communs. Cette mise en relation repose sur des interactions coopératives entre agents concernés.
- Les agents de niveau inférieur permettent de représenter les profils des utilisateurs/services réels ainsi que les connaissances que possède un agent sur un autre agent. Ces connaissances sont appelées "croyances" et correspondent à des connaissances qui sont souvent partiales et partielles sur les compétences des congénères qu'il connaît.

2) Evaluation et la quantification des

navigations des utilisateurs du Web,

Partant du constat que les capacités coopératives du Web étaient très peu exploitées, nous avons développé un système multi-agent d'assistance à la navigation qui mutualise, le plus automatiquement possible, les expériences individuelles. Mawa [3] est un système ouvert, distribué sur le réseau Internet, capable de :

- récolter les parcours des utilisateurs (un parcours est une navigation Web relative au centre d'intérêt de l'utilisateur – quand il change de centre d'intérêt, l'utilisateur commence un nouveau parcours),
- retravailler ces parcours, par regroupement de certains et/ou renforcement de leur importance ; ceci afin de faire émerger des parcours-types,
- assister les utilisateurs lors de leurs navigations, en leur proposant des parcours-types ressemblants.

Les parcours des utilisateurs (appelés sessions de navigation) sont représentés par des n-uplets, chacun faisant figurer les ressources Web d'origine et de destination ainsi que leur force d'interconnexion.

Le système est composé d'agents assistants (associés aux utilisateurs de ressources) et d'agents connecteurs qui jouent le rôle de gestionnaire de sessions et mettent en correspondance les différents agents assistants. Un agent assistant modélise les navigations de son utilisateur (construction de sessions utilisateur), construit son profil et l'assiste ; un agent connecteur applique un algorithme de consolidation de sessions pour effectuer des appariements de sessions utilisateur et les fusionner en sessions collectives (sessions consolidées) [4].

Pour communiquer entre eux, les agents s'envoient des messages d'information (contenant une ou plusieurs sessions – pour la récolte et le partage de sessions) et des requêtes (contenant un ensemble de n-uplets caractérisant une session en cours de navigation – pour l'assistance aux usagers).

L'originalité du modèle réside sur sa

capacité à s'adapter à l'évolution du nombre d'agents : cette adaptation se fait grâce à la dynamique des spécialisations et des répartitions de connaissances. Ce sont les communications et les relaxations de connaissance entre agents qui créent des agrégations d'informations sur certains agents. La force d'attraction de ces derniers augmente avec la quantité de cette information. Ils deviennent de par ce fait, les agents faisant autorité sur ce type de connaissance. Ces caractéristiques sont nécessaires pour permettre un déploiement incrémental et « coloniser » le Web.

3) La conception d'ontologie

Les ontologies sont des représentations formelles de connaissances d'un domaine sous forme d'un réseau conceptuel. Elles sont considérées aujourd'hui comme un des moyens d'assurer la communication entre des agents logiciels ou d'accéder et annoter des sites Web. L'objet de nos recherches est de faire en sorte que des systèmes extraient des connaissances terminologiques et ontologiques à partir de textes et de leurs usages qui pourront ensuite se modifier et s'adapter à des environnements dynamiques.

L'objectif de ces recherches est d'automatiser le processus de conception d'ontologie à partir de corpus et le processus de maintenance de cette ontologie dans un contexte d'utilisation. Le mécanisme de création de l'ontologie est basé sur un système multi-agent adaptatif au sein duquel les mots sont les agents du système. Le réseau conceptuel produit par les agents est fourni à l'analyste qui propose d'éventuelles modifications. Ainsi le système multi-agent apprend aussi à partir des observations de ces modifications. Un outil de traduction permet à l'analyste de représenter en OWL le réseau d'agents obtenu. Pour la phase de maintenance de l'ontologie, le logiciel développé DynamO [5] travaille de manière autonome : il assure l'évolution de l'ontologie en fonction de son utilisation et des modifications du corpus textuel. C'est le système multi-agent qui a appris durant la phase de conception qui continue à apprendre et assister l'utilisateur en cours d'utilisation.

Bibliographie

- [1] Camps Valérie - « Vers une théorie de l'auto-organisation dans les systèmes multi-agents basée sur la coopération : application à la recherche d'information dans un système d'information répartie » - Thèse de Doctorat de l'Université Paul Sabatier, IRIT, N° d'ordre 2890, Toulouse, Janvier 1998.
- [2] Gleizes Marie-Pierre, Léger Alain, Athanassiou Eleutherios, Glize Pierre - « Abrose : Self Organization and Learning in Multi-Agent Based Brokerage Services » - 6th Inter-national Conference on Intelligence and Services in Networks, IS&N'99, Barcelona, Spain. Proceedings pp 41-54., Lecture Notes in Computer Science 1597, Springer, April 1999.
- [3] Pécatte Jean-Marie, Singer Nicolas, Trouilhet Sylvie - « MAWA : un outil multi-agent pour la recherche d'informations coopératif » - CNRIUT, Nice, Mai 2004.
- [4] Singer Nicolas, Trouilhet Sylvie - « Intégrer le comportement de l'utilisateur dans les documents Web » - 14ème congrès RFIA, Toulouse, Janvier 2004.
- [5] Ottens Kévin, Aussenac-Gilles Nathalie, Gleizes Marie-Pierre, Glize Pierre - « Systèmes Multi-Agents pour l'extraction d'ontologies à partir de textes : revue de questions » - Workshops of the ATALA - AGENTAL : Agents et langage.

Thème Systèmes de Bases de Données - PRISM Laboratoire PRISM

Université de Versailles et Saint-Quentin-en-Yvelines

Contact : Georges Gardarin

Membres

Chercheurs permanents : Tram Dang Ngoc, Benjamin Nguyen, Yann Loyer, Laurent Yeh, Karine Zeitouni

Thésards en cours : Florin Dragan, Clément Jamard, Nicolas Travers, Lionel Savary, Tao Wan,

Anciens thésards : Nadjim Chelghoum, these 2004

Thématique scientifique

XML émerge comme le langage universel pour représenter et échanger des données hétérogènes non seulement sur le web, mais dans les systèmes d'information en général. En effet, les données étant de plus en plus disparates dans les entreprises, XML s'impose comme un format commun d'échange de données. Il apparaît donc nécessaire de concevoir des outils permettant de fédérer et intégrer des sources de données hétérogènes en XML. En amont, la génération de documents XML par balisage automatique de textes ou de documents est une autre voie de recherche. Ce processus se base sur les techniques de fouille de données, essentiellement textuelles afin d'extraire le schéma des documents à partir d'instances en déterminant les catégories, les concepts et les mots-clés.

Associé à XML se développe une grande variété de standards, dont XQuery, nouveau langage de requêtes pour interroger des collections de documents XML. XQuery pourrait être dans un futur plus ou moins lointain le successeur de SQL. En effet, outre les fonctionnalités de SQL sur des documents XML plats, il permet l'interrogation de documents semi-structurés et les recherches plein texte par mots-clés et phrases, avec réponses rangées par degré de pertinence. Il est aussi plus qu'un langage de requêtes car son aspect fonctionnel permet de composer de véritables modules de programmes incorporant des fonctions internes ou externes.

L'extraction de connaissances depuis des données complexes est aujourd'hui un thème de recherche très actif en bases de données. Les données sont le plus souvent modélisées en relationnel étendu mais XML commence aussi à émerger. En effet, les données complexes telles que les données géographiques, textuelles ou spatio-temporelles ont des représentations et des méthodes d'accès très spécifiques. La découverte de connaissances depuis ces données nécessite d'interpréter leur contenu sémantique et d'optimiser leur manipulation. Or, la fouille de données (data mining) a été développée essentiellement dans le contexte de données tabulaires simples, donnant lieu à des logiciels et des applications opérant sur

un fichier ou une table relationnelle.

Ainsi, des mesures de similarités sophistiquées doivent être mises en oeuvre pour comparer des contenus textuels ; pour les données spatiales, leur analyse doit considérer le voisinage géographique et y accéder efficacement. Ces interprétations propres au domaine d'application ont amené à revoir les méthodes existantes de fouille de données pour les prolonger par des prédicats ou par des mesures de similarités spécifiques, définir de nouvelles méthodes et les optimiser. C'est là un des thèmes de recherche qui nous a intéressés ces dernières années.

Par ailleurs, le processus d'extraction de données dans une application décisionnelle ne se limite pas à l'étape de fouille de données, certes importante. Il nécessite au préalable la constitution d'un entrepôt de données et son alimentation par un processus d'intégration depuis des sources souvent hétérogènes, l'analyse en ligne par ce qui est communément appelé l'OLAP (On-Line Analytical Processing) et la transformation en vue de l'application d'une méthode de fouille de données. L'extension de ces concepts et techniques associées aux données spatio-temporelles soulève des problèmes pour l'intégration notamment de l'information géométrique, pour la modélisation spatiale et temporelle dans les entrepôts et pour le calcul efficace de cubes de données. Ceci constitue un autre sujet de notre recherche.

Perspectives

Optimisation de requêtes XQuery

L'optimisation des requêtes XQuery distribuées sur de larges collections de documents XML reste un problème ouvert. Les benchmarks nous ont montré les limites pour de larges collections distribuées. Nous souhaitons améliorer l'optimiseur actuel de notre médiateur XQuery en permettant l'élaboration de plans distribués plus efficaces et aussi en prenant en compte des structures d'accélérateurs plus variées.

Médiation XML pair à pair

Les architectures de médiation s'appuient aujourd'hui sur un médiateur cen-

tralisé connaissant les schémas des sources. Celui-ci apparaît comme un véritable goulot d'étranglement pour l'exécution de requêtes XQuery, tâche généralement assez lourde et nécessitant l'exécution d'opérateurs de type sélection, jointure, tri et agrégat sur des flux de documents. Quelques architectures ont déjà été présentées proposant de munir chaque nœud du réseau d'un médiateur et d'un SGBD XML minimum (ou d'un adaptateur). On aboutit alors à une architecture P2P sur une grille où chaque nœud joue à la fois le rôle de médiateur et de source de données XML.

Partant d'un système de médiation centralisé opérationnel, l'équipe SGBD du laboratoire PRiSM souhaite étudier sa distribution sur une grille de données XML avec des calculs XQuery réalisable dans chaque nœud. Diverses solutions aux problèmes mentionnées seront explorées, mise en œuvre et comparées du point de vue performance en utilisant des benchmarks XML adaptés. Dans ce cadre, nous devons participer au projet Grid Explorer qui semble vouloir mettre en œuvre un système concret de référence autorisant seul des expériences capables de valider les algorithmes et méthodes de distribution sur un véritable réseau à grande échelle.

Entrepôt XML orienté services

Le projet RNTEL e.dot a donné lieu à la réalisation d'un prototype pour concevoir à partir d'une interface graphique, le schéma de l'entrepôt, ainsi que de simplifier le mécanisme de connexion des services web entre eux. Nous travaillons actuellement sur un module de « pilotage » avancé du système, et des méthodes de « tableau de bord » permettant de visualiser les informations les plus importantes de l'entrepôt via un site web généré automatiquement. Le but de ces techniques est d'identifier et réaliser des fonctions génériques, utilisables dans le cas des entrepôts de données semi-structurées, comme primitives pour des techniques de recherche d'information.

Par ailleurs, l'application de ces méthodes dans le cadre de la Science Politique, en collaboration avec le Laboratoire d'Analyse des Systèmes

Politiques (Paris X) et l'Observatoire Politico-Economique des Structures du Capitalisme, et leur adaptation dans le cadre de bases de données sur les réseaux entre personnes est également une direction de recherche que nous comptons développer.

Entrepôt et fouille de données spatiales et/ou temporelles.

La fouille de données complexes est un thème nouveau qui a motivé la constitution du groupe FDC (GT Fouille de Données Complexes) et d'ateliers propres. Sa problématique a été soulevée dans les communautés d'analyse de données (symbolique) et de logique inductive (fouille de données relationnelles). L'analyse et la fouille de données spatio-temporelles illustre bien ces problèmes. Les premiers travaux ont concerné les données spatiales, mais il reste encore beaucoup à faire pour intégrer les données spatio-temporelles et mobiles.

Nous avons proposé et validé un modèle basé sur la discrétisation de l'espace et du temps qui présente de nombreux avantages pour les cubes de données. Seulement, d'un côté la discrétisation conduit à une perte d'information et de l'autre, un granule trop fin risque d'exploser la taille de stockage. Nous allons étudier d'autres solutions pour des modèles d'entrepôts d'objets spatiaux, spatio-temporels et mobiles. Aucune solution n'est connue aujourd'hui à ce problème. Nous devons ensuite proposer l'estimation de trajectoire basée sur ce modèle. Ceci fait appel à des procédures probabilistes avec des incertitudes. La prise en compte de ces incertitudes peut utiliser des logiques non binaires et des traitements du flou.

Bibliographie

- [1] Gardarin G. – XML : Des Bases de Données aux Services Web, Dunod Ed., novembre 2002, 521 p.
- [2] T.T. Dang-Ngoc, G. Gardarin. Federating Heterogeneous Data Sources. Dans les actes de publication de IASTED International Conference on Information and Knowledge Sharing (IKS 2003, Scottsdale, USA, Novembre 2003,

p. 193-198, ISBN 0-88986-396-2 (407).

- [3] Zeitouni K., Yeh L. - Le Data Mining Spatial et les bases de données spatiales. Revue internationale de géomatique, vol. 9, n° 4 (99), avril 2000, p. 389-394.
- [4] Halkidi M., Nguyen B., Varlamis I., Vazirgiannis M. Thesus, a Closer View on Web Content Management Enhanced with Link Semantics. Transactions on Knowledge and Data Engineering, vol. 16, n° 6, June 2004.
- [5] Loyer Y., Spyrtos N., Stamate D. Parametrized Semantics of Logic Programs - A Unifying Framework. Theoretical Computer Science (TCS), vol. 308, n° 1-3, novembre 2003, p. 429-447.

Bradley Efron

« Les statistiques sont la science de la collecte d'informations, en particulier lorsque ces informations arrivent par petits morceaux au lieu de gros. »

Traduction par Philippe Morignot (AXLOG) de l'interview en anglais disponible à <http://www.mhhe.com/business/opsci/bstat/efron.mhtml>

Bradley Efron est Professeur et Directeur du Département de Statistiques à l'Université de Stanford, et s'est vu décerner le Prix Ford, le Prix MacArthur, et la Médaille Wilks pour sa recherche sur les applications informatiques en statistique, en particulier sur les techniques connues sous le nom de bootstrap et de jackknife. Il a été orateur clé ou invité plus de 50 fois, et son dernier livre « An Introduction to the Bootstrap », a été publié par Chapman & Hall en 1993.

Professeur Efron, vous avez, probablement mieux que quiconque, apporté la puissance des ordinateurs modernes en statistique. Comment voyez-vous le futur de ce mariage ?

D'abord, je pense que [John] Tukey a beaucoup de mérite pour avoir fait entrer les ordinateurs dans le domaine. Au moins à mon sens. Et je crois que nous nous dirigeons vers un monde dans lequel les gens ne vont même plus remarquer que c'est de cette façon que cela a été fait. C'est-à-dire, toutes les statistiques vont être faites avec beaucoup de calculs sur ordinateur, tout simplement parce que vous n'aurez pas à faire beaucoup d'hypothèses ou de calculs théoriques. Les gens le font déjà, par la méthode Box-Jenkins par exemple. Je pense que l'idée va simplement se répandre, après un certain temps, que de plus en plus de théorie sera emballée de façon que le consommateur n'aura même plus à penser à chaque fois à travers la théorie.

Pensez-vous qu'ils (les ordinateurs modernes et les logiciels) puissent un jour remplacer notre profession (statisticien) ?

Non. Non, je crois que cela va étendre notre profession. Nous serons les gens qui emballons la théorie. C'est comme de dire : pensez-vous que les ordinateurs vont mettre les programmeurs au chômage ? Non, il y aura plus de personnes comme nous : les gens vont beaucoup plus utiliser les statistiques, parce qu'ils se sentiront mieux avec. Ils ne les ressentiront pas comme un sujet étranger, et il y aura beaucoup plus de demande.

Je veux vous poser une question à propos des méthodes à calculs intenses en statistiques ...

Je crois que j'ai – Persi [Diaconis] et moi – inventé le

terme de statistiques à calculs intenses. Je n'ai jamais trouvé de référence antérieure avant notre article dans le *Scientific American* ... donc je le revendique, ha ha ha ...

Bien ! Vous le méritez ! Avez-vous préconisé le bootstrap [une importante méthode à calculs intenses] parce que vous aviez une application en tête, ou étiez-vous à la poursuite d'une théorie ?

En fait ce qui s'est passé est que Rupert Miller, de notre département, était en train de travailler sur le jackknife. Il avait écrit un article intitulé le « Trustworthy Jackknife » dans lequel il essayait de comprendre quand le jackknife donnait des estimations de variance fiables. Le Jackknife était considéré comme très mystérieux. Ça marchait, mais personne n'était capable de comprendre pourquoi ça marchait. Et parfois cela ne marchait pas. Ce que j'ai pensé était que le Jackknife était un genre différentiel, local, d'approximation pour quelque chose d'autre. Et ainsi quand j'ai commencé à rechercher ce quelque chose d'autre, je suis tombé sur le bootstrap.

Quelle branche des statistiques voyez-vous devenir dans l'avenir comme le plus important pour la société ?

D'accord, là c'est une bonne question. Je ne suis pas sûr que ce soit des branches. Voulez-vous parler de branches à sujet, comme l'analyse de survie ? L'analyse de survie a été une merveilleuse histoire statistique à succès. Comme je parcours le pays, je suis frappé par la prospérité et le succès dans lequel les départements de biostatistiques se trouvent, par comparaison aux départements de statistiques pures. Et je pense que la biostatistique est vraiment en train d'arriver sur le devant de la conscience nationale, ou de la conscience internationale, et que cela fera autant partie du discours que les cures de santé le sont maintenant. C'est-à-dire, la biostatistique fera naturelle-

ment partie des histoires des journaux traitant de contrôles et de traitements, et traitant de la façon dont le niveau de signification avance, et des choses comme cela. Donc je pense que cela sera énormément important. Je travaille sur le genre de choses biostatistiques et biologiques, comme cela je ne suis pas amené à voir une grande partie des autres genres d'applications. L'histoire depuis 1900 a été que les statistiques prennent tout simplement le pouvoir domaine après domaine, en des termes d'être la méthodologie du choix, et je pense que ça continuera. J'ai récemment écrit des articles avec des gens en astronomie et en physique, et ils commencent à beaucoup plus utiliser les statistiques pour la simple raison que maintenant ils doivent être efficaces. C'est dur à dire, mais je ne vois aucun sous-domaine où il y a eu beaucoup de résistance.

Quels sont les problèmes et applications parmi les plus intéressants sur lesquels vous avez travaillé ?

Celui que j'ai récemment le plus apprécié était celui sur la loi de Hubble. Il y a un débat dans la communauté astronomique. La loi de Hubble est que plus une galaxie s'éloigne, plus elle ralentit de façon linéaire, que l'espace s'étend linéairement. De très bons astronomes en doutent. J'ai travaillé sur un ensemble de décalages vers le rouge de galaxies avec Vahi Petrosian à Stanford pour voir si c'était vrai ou pas. A ma grande surprise, après beaucoup de temps la loi de Hubble avait l'air de marcher plutôt bien. Je ne le croyais pas. Cela ne me semblait tout simplement pas bien de mon point de vue. Mais cela se passait vraiment bien.

J'obtiens beaucoup de bons problèmes en étant consultant en biostatistique. Je travaille pour des entreprises de médicament. Je suis en train de concevoir un test pour une entreprise qui fabrique des médicaments contre l'impuissance masculine, et c'est un test difficile à concevoir. Les faux effets peuvent très facilement se produire, et ainsi c'était très amusant de concevoir ce test et j'espère que ça sera aussi amusant de l'analyser. L'un des jeux de données les plus amusants sur lesquels j'ai jamais travaillé était ce jeu de données sur Shakespeare – le problème du « combien de mots Shakespeare connaît » avec Ron Thisted qui était un genre de jeu de données philologique. Aussi ils apparaissent partout.

Au passage, est-ce que la loi de Hubble est reliée à la relativité ? Que prédit la relativité ?

Quand Hubble a annoncé sa loi, Einstein avait prouvé auparavant que l'une des formes les plus stables de l'univers était une expansion constante. Et puis comme tout le monde lui a dit que c'était impossible, il s'est dédit. Puis Hubble a montré que c'était ce qu'il semblait se passer.

C'est le genre d'opinion majoritairement acceptée. Pourtant, il y a d'autres solutions sur lesquelles d'autres personnes s'accordent à dire que ce sont les bonnes. Je ne connais pas assez bien la physique pour en parler.

Et les statistiques ont prouvé la loi de Hubble ?

Au moins les statistiques étaient fortement cohérentes avec la loi de Hubble et pas avec celles des autres. Et j'ai été très surpris parce qu'il y a eu beaucoup de bonnes critiques à son sujet. Le jeu de données était constitué de 476 galaxies récupérées de bon nombre d'articles de synthèse différents --- il est difficile de faire des décalages vers le rouge. Cela ne donnait certainement pas l'impression que ça allait marcher, mais ça a marché.

Je veux vous poser une question à propos des fondements des statistiques. Pensez-vous qu'il y a des trous dans les fondements des statistiques, de la même manière qu'il y en a dans la théorie des ensembles ?

Je ne pense pas que les trous sont du même genre que ceux de la théorie des ensembles, en ce que la théorie des ensembles est sans doute perfectible, alors que les statistiques sont maintenant bien au delà de ce but. Je me retrouve à revenir encore et encore de la base des statistiques à des choses qui ne sont que des exemples, comme des problèmes normaux de traduction (NDT : normal translation problem). Si vous pouvez réduire un problème à un problème de traduction normal, vous connaissez quasiment la réponse. Donc si vous pouvez réduire un problème quelconque à cela, vous avez la réponse. Ou les intervalles de confiance. Si vous pouvez obtenir ce genre d'intervalle de confiance, cela semble la bonne réponse, donc si vous obtenez quelque chose d'autre qui fait à peu près cela, ça doit être plutôt bien. La seule théorie complète des statistiques est la théorie bayésienne, et même si elle est inattaquable, d'une certaine manière elle rate une partie de l'histoire, qui est que vous ne pouvez pas l'utiliser comme une véritable théorie directive sur des problèmes compliqués. Ensuite vous êtes toujours forcé de faire quelque chose de trop compliqué, et de vous faire une opinion sur des choses sur lesquelles vous n'avez pas d'opinion. Aussi d'une certaine manière la théorie bayésienne est merveilleuse, mais elle ne raconte pas toute l'histoire. La théorie fréquentiste (NDT : frequentist theory) a été lancée avec beaucoup de contradictions, mais elle semble marcher si bien. Essayer de réconcilier ces deux choses ... L'un de mes espoirs est que les statistiques à calculs intensifs vont faire se réconcilier plus facilement ces points philosophiques, parce qu'il n'y aura pas autant de problèmes techniques.

Vous êtes en train de parler des méthodes Bayésienne vs. Fréquentistes ?

Oui.

Donc vous pensez que les méthodes à calculs intensifs vont ...

Oui, ce n'est qu'un espoir. Je les utilise comme cela à certains moments. Les deux philosophies ne semblent pas si distinctes que cela ... Je suis actuellement encore en train de travailler sur ce genre de choses empiriques bayésiennes. Et je me suis servi du bootstrap sur l'empirique bayésien, ça commence à faire se rapprocher ces deux choses. Elles ne me semblent pas aussi distinctes qu'elles ne me le semblaient il y a longtemps. Peut-être que je suis en train de m'attendrir.

Vous avez commencé par un diplôme en mathématiques. Qu'est-ce qui vous a fait vous diriger ou était-ce un changement de direction vers les statistiques ?

Ma propre histoire était que mon père était un conducteur de camion et un mathématicien amateur, et aussi, à la maison, le statisticien sportif pour le club de bowling et le club de baseball (St. Paul). Et ainsi j'ai grandi entouré par les nombres et par un père assez sophistiqué qui en savait beaucoup sur la façon de calculer des choses. Je pensais que j'allais être mathématicien. Je suis allé à CalTech, et je pense que je serais resté mathématicien si les mathématiques étaient ce qu'elles étaient il y a cent ans quand vous calculiez des choses, mais je n'avais absolument aucun talent pour les mathématiques abstraites modernes. Et ainsi je voulais aller vers quelque chose qui était plus calculatoire. Après CalTech je suis allé à Stanford. Et les statistiques étaient assurément mieux.

Vous avez réussi votre passage ?

En fait, ils n'avaient pas de cours de statistiques à CalTech, mais certains membres permanents m'ont laissé lire le livre de Cramer. C'était un cours sous forme de lecture. Et j'ai vraiment adoré ce livre. Je l'ai lu de la page de couverture à la quatrième de couverture. Je ne vois personne qui fait des statistiques de la manière dont le fait Cramer. J'ai encore le livre complètement annoté.

Pouvez vous dire quelque chose à propos des statisticiens qui vous ont influencé dans vos jeunes années ?

Oui, facilement. A Stanford, il y avait Rupert Miller et Lincoln Moses, qui dirigeaient le cours de biostatistique.

J'y suis entré en tant que mathématicien. C'est dans

les cours de biostatistique que j'ai appris les statistiques en tant que ce genre de choses vivantes du quotidien qui aide les gens. Cela m'a beaucoup ouvert les yeux de voir comment ils faisaient des statistiques.

Vous avez reçu le prestigieux MacArthur Award, parmi beaucoup d'autres distinctions. Pouvez-vous nous dire quelque chose sur le projet que vous étiez en train de faire et qui a abouti au MacArthur Award ?

He bien, le MacArthur Award n'était pas une récompense pour faire des projets. Ils disent spécifiquement que ce n'est pas le cas. Ils disent spécifiquement qu'ils ne diront rien qui vous fassent faire quelque chose. Ils veulent que la récompense soit un prix, pas un empêchement pour votre travail futur. Et ce sont des personnes très sympathiques avec lesquelles être en relation, de ce point de vue. Donc, je n'ai jamais rien fait directement. Pourtant, cela a eu un gros effet sur moi d'une manière dont je ne me serais jamais douté. Les gens à la fois à l'intérieur de l'université et à l'extérieur de l'université faisaient beaucoup plus attention à ce que j'avais à dire.

Y a-t-il un prix Nobel en statistiques ? Ils les considèrent comme une branche des maths [pour lesquels il n'y a pas de prix Nobel] ?

Les statistiques n'ont pas de grand prix. Et c'est probablement une bonne chose. Avoir les médailles Field ou les prix Nobel n'a pas toujours eu un bon effet sur, disons, la physique ou les maths. Ils tendent à produire beaucoup de gros egos et de chamailleries. Les statistiques ont été, dans un sens, une sorte de terrain pour un homme travaillant intellectuellement, où vous travaillez pour l'amour du domaine et pas pour la réputation, parce qu'il n'y en a aucune.

Que sont les statistiques pour vous – est-ce des mathématiques, de l'informatique, de la philosophie, tout cela à la fois, ou quelque chose d'autre ?

Ma définition, qui se trouve dans notre livre, est que les statistiques sont la science de la collecte d'informations, spécialement lorsque les informations arrivent par petits morceaux plutôt que par un ou deux gros morceaux.

Comme un puzzle ?

Il est tout à fait étonnant qu'il puisse exister une théorie des statistiques. Vous pourriez penser que vous ne pouvez pas avoir de théories sur les choses ... Je veux dire, comme l'astronomie est pour les étoiles et la géologie l'est pour les rochers, he bien les statistiques sont à propos des

informations et de la collecte d'informations. Que ce soit une « chose » de la même façon que des rochers sont des rochers n'est pas clair.

Vous voulez dire comme un concept d'un nombre ou quelque chose comme cela ?

Oui, c'est assez philosophique. Les philosophes ont l'habitude de dire que les statistiques sont impossibles. Ils disent que vous ne pouvez pas apprendre à partir de l'expérience. Ils peuvent toujours penser à des contre-exemples. Mais nous vivons dans un monde où les exemples sont plus nombreux que les contre-exemples.

Où vous situez-vous sur l'estimation vs. le test d'hypothèses, et sur la critique des p-valeurs ? Comme quand un étudiant demande « pourquoi est-ce que alpha vaut 0.05 ? » ou « qu'est-ce qu'il se passe lorsque ma p-valeur est 0.06, pourquoi n'est-ce pas significatif alors que quand c'est 0.04, ça l'est ? ». Ou en général, que pouvez-vous dire sur le test d'hypothèses, les intervalles de confiance, le futur ... ?

Une estimation consiste à mettre les choses ensemble et tester est séparer les choses. Analyse vs. synthèse. J'ai toujours aimé les intervalles de confiance parce qu'ils vous laissent enjamber le terrain au milieu. Il y a un temps pour toutes ces choses. Tester des hypothèses est immensément efficace si vous êtes vraiment en train d'essayer de savoir si un effet est positif ou non. C'est la seule chose que vous pouvez utiliser quand vous avez $n=10$ beaucoup de fois. Mais ce n'est pas tout à fait aussi intéressant quand vous avez un plus grand n et que vous voulez vraiment en savoir plus. Il est habituellement assez clair que l'hypothèse est vraie ou pas. La question est quel est le degré de vérité. Et puis l'estimation et les intervalles de confiance arrivent de plus en plus ... J'utilise beaucoup plus les intervalles de confiance. Je pense que la plupart des bons statisticiens appliqués utilisent les intervalles de confiance plus qu'il n'est montré dans les livres. Si seulement ils étaient dans les livres. La raison pour laquelle ils ne sont pas autant dans les livres est qu'ils sont beaucoup plus difficiles à expliquer que l'estimation ou le test d'hypothèses. Mais c'est une de ces choses que j'espère que le futur nous apportera --- une acceptation bien plus facile de ces choses comme des intervalles parce que vous pourrez les obtenir à partir de la machine automatique et puis vous pouvez passer votre temps à comprendre ce qu'ils sont au lieu de cet horrible ... Les maths sont tout à fait impliquées.

Ils seront plus justes si vous utilisez le bootstrap,

exact ? Je suis vraiment fasciné par l'idée que le bootstrap vous donne n'importe quel niveau requis d'exactitude et de correction, spécialement le bootstrap répliqué, et en fonction de la taille de l'échantillon, et que vous êtes vraiment en train de faire mieux que la théorie normale, t-distribution.

Oui, vous êtes en train de faire un pas de mieux, et ce pas est un grand pas.

De 1 / racine (n) à 1 / n ...

Oui, je pense que c'est très important. Et la chose sur laquelle je suis en train de travailler --- et sur laquelle beaucoup de gens sont en train de travailler --- est d'essayer de rendre cela vraiment fiable, pour que vous l'obteniez vraiment à chaque fois --- et que ça marche. Pré-emballer cette théorie n'est pas si facile. C'est un travail difficile.

Encore une question. Vous mentionnez tout à l'heure un nouveau livre que vous avez écrit sur le bootstrap.

Oui. Maintenant que notre petit livre est sorti --- [Robert] Tibshirani et moi avons un petit livre intitulé "An Introduction to Bootstrap", publié par Chapman et Hall en 1993 --- cela va être plus simple de comprendre le bootstrap. Ce qui est en train de se passer est que le bootstrap a suivi son cours, son premier cours, dans le monde de la théorie, et maintenant je pense qu'il va aller vers les applications. Les gens travaillant sur les applications ont toujours semblé plus heureux avec cela que les gens travaillant sur la théorie.

Pouvez-vous nous dire quelque chose à propos de vos centres d'intérêt extérieurs ? Vous êtes un être humain, et un être humain très intéressant, pas seulement un statisticien.

J'adore les statistiques. Je vis exactement sur le campus de Stanford. J'y vais tout le temps. Je ne travaille pas dur tous les jours, pourtant. Je travaille tous les jours. Le soir j'aime aller au cinéma avec ma fiancée, et je vais voir quasiment tous les films. Je suis intéressé par l'astronomie et la science. Je revendique être un scientifique amateur tout autant qu'un scientifique professionnel. Et j'adore la science. Je pense que c'est la chose la plus grande à laquelle les gens ont jamais pensé. Aussi, je suppose que mon passe-temps est aussi la science. J'aime m'asseoir dans le coin et parler ... Les films, les stats, et la science.

PRÉSENTATION DE LABORATOIRES

Présentations de laboratoires dans le bulletin de l'AFIA

LIFIA, Grenoble	Bulletin n°1	IRIT - Toulouse (1)	Bulletin n°28
LRI, Orsay	Bulletin n°1	IRIT - Toulouse (2)	Bulletin n°29
Service Systèmes Experts, Renault	Bulletin n°1	LAAS - Toulouse (1)	Bulletin n°30
CEDIAG,	Bulletin n°2	Sony CSL	Bulletin n°31
CERT, ONERA, Toulouse	Bulletin n°2	LAAS - Toulouse (2)	Bulletin n°32
IRIT, Toulouse	Bulletin n°2	LIMSI - Département CHM	Bulletin n°33
LAAS, Toulouse	Bulletin n°2	LAMSADE	Bulletin n°34
HEUDIASYC, UTC	Bulletin n°3	Institut autrichien de recherches en I.A.	Bulletin n°36
IFP, Rueil Malmaison	Bulletin n°3	LIP6 – Université Pierre et Marie Curie	Bulletins n°37 & 38
DIAM, INSERM U194	Bulletin n°3	GREYC – Université de Caen	Bulletin n°40
Lab. Math. Info., Fac Médecine de Marseille ..	Bulletin n°4	LIFL – Université de Lille	Bulletin n°41
GMD, St. Augustin (RFA)	Bulletin n°4	LRI (équipes IA et IASI)	Bulletin n°43
ONERA, Chatillon	Bulletin n°4	IMAG - Grenoble	Bulletin n°44
KSL, Université de Stanford (USA)	Bulletin n°5	PSI (Perception, Système, Information - Rouen)	Bulletin n°45
Dépt Applications de l'IA au CNET, Lannion.	Bulletin n°5	INRIA – Sophia Antipolis	Bulletin n°46/47
LAFORIA, Univ. Pierre et Marie Curie	Bulletin n°6	LIH – Laboratoire d'Informatique du Havre	Bulletin n°46/47
L'institut FAW, ULM (RFA)	Bulletin n°6	Tech-CICO – Université de Technologie de Troyes	Bulletin n°51
Institut IIIA, Compiègne	Bulletin n°6	LIFO – Université d'Orléans – Équipe Contraintes et Apprentissage	Bulletin n°52
LAIR, OHIO State University (USA)	Bulletin n°7	LIIA – Ecole Nationale des Arts et Industries de Strasbourg	Bulletin n°52
ARAMIHS, Labo mixte MATRA-CNRS,	Bulletin n°7	LRL – Laboratoire de recherche sur le langage – Université Blaise Pascal Clermont 2	Bulletin n°53
CEA, Service SERMA, Saclay	Bulletin n°8	MIG – Mathématique, Informatique et Génome – INRA	Bulletin n°53
Société ILOG	Bulletin n°8	ESIEA Recherche, Laval et Paris	Bulletin n°54
LAIAC, Université de Caen	Bulletin n°9	Équipe Intelligence Artificielle et Applications (IAA) du Crip5, Paris 5	Bulletin n°55
Institut Français du Pétrole	Bulletin n°10	LAMIH – Univ. de Valenciennes et du Hainaut-Cambrésis	Bulletin n°55
DFKI (Centre allemand de recherches en IA) .	Bulletin n°11	Université de Lund	Bulletin 58
GRTC, Marseille	Bulletin n°11	CRIL – Centre de Recherche en Informatique de Lens – Université d'Artois	Bulletin 58
Inst. d'Analyse des Systèmes, Ac. Russe	Bulletin n°12	LIRIS – Laboratoire d'Informatique en Image et Systèmes d'Information	Bulletin 60
Georges Mason Univ., Center for AI (USA) ...	Bulletin n°13	KIN – Knowledge Innovation Center	Bulletin 60
IRISA, INRIA et Université de Rennes	Bulletin n°13		
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		
Institut autrichien de recherches en I.A.	Bulletin n°20		
ENST Bretagne	Bulletin n°21		
LIA - Université de Savoie	Bulletin n°22		
INRETS	Bulletin n°23		
IRIN Nantes	Bulletin n°24		
CRIN - INRIA Lorraine	Bulletin n°25		
DIRO - Université de Montréal	Bulletin n°26		

Computability and logic

G.S. Boolos, R.C. Jeffrey.

Cambridge University Press, 3ème édition 1989, réimprimée 1994.

Recension par Amedeo Napoli (LORIA)

Le chapitre 1 commence doucement avec quelques éléments sur les ensembles dénombrables et non-dénombrables, et le chapitre 2 n'est pas très violent non plus avec la diagonalisation, où l'on montre que l'on ne peut énumérer les suites d'entiers. Au chapitre 3 sont introduits des modèles de calcul (machines de Turing).

C'est au chapitre 4, avec les 3 premiers brefs chapitres, que l'on entre dans le vif du sujet. On y définit le "castor affairé" ; il s'agit de la fonction qui à un entier n associe la taille maximale d'un programme de taille n qui s'arrête proprement. La fonction du castor affairé ne peut être calculée par aucun programme. En effet, sinon on pourrait construire un programme qui regarde sa propre longueur l , applique le castor affairé à cette longueur obtenant ainsi $b(l)$ tel que tout programme de longueur $\leq l$ ait une sortie de taille $\leq b(l)$, multiplie le résultat par 2, et écrive une sortie de taille $2b(l)$; clairement, on arrive à la conclusion qu'un programme de taille l écrive une sortie de taille $>b(l)$, ce qui contredit la définition du castor affairé.

On en déduit au passage qu'il n'existe pas de programme qui sache déterminer si un programme donné s'arrête ou non ; sinon, en utilisant ce programme, on arriverait à construire un programme qui calcule la fonction du castor affairé. On a donc prouvé que "la halte n'est pas calculable".

Le chapitre 5 prouve de la non-calculabilité par diagonalisation. On peut voir simplement qu'il existe des fonctions de \mathbb{N} dans \mathbb{N} non-calculables par des programmes, puisque les fonctions de \mathbb{N} dans \mathbb{N} sont les suites d'entiers, et qu'on sait que ces suites sont non-énumérables, alors que les programmes sont énumérables ; les programmes calculent des applications de \mathbb{N} dans \mathbb{N} , mais ne sauraient les calculer toutes, puisqu'il y a "moins" de programmes que d'applications de \mathbb{N} dans \mathbb{N} ("moins", au sens où il y en a une infinité, mais une plus petite infinité).

Les chapitres 6, 7, 8 et 9 montrent l'équivalence

entre machines de Turing et divers formalismes.

Les chapitres 6 et 7 montrent que les machines de Turing sont Abacus-équivalentes, c'est-à-dire que l'on peut simuler des registres, ce qui est pratique. Le chapitre 8 montre que les fonctions récursives sont dans le même panier de fonctions. Le chapitre 9 montre le lien avec les formules logiques du premier ordre.

Le chapitre 10 traduit les résultats de non-calculabilité en résultats d'indécidabilité en logique : il existe des vérités qu'on ne saurait prouver (vérités dans un sens qu'il serait trop long de préciser dans ces lignes ; disons simplement qu'il s'agit de vérité intuitive (mais dans un sens bien formalisé) et non de vérité dans tous les modèles).

Le chapitre 11 formalise la notion de preuve en logique du premier ordre. Le chapitre 12 prouve le théorème dit de complétude : si un ensemble d'axiomes n'est pas satisfiable (n'a pas de modèle), alors il admet une réfutation (une preuve d'incohérence). On y trouve aussi le théorème de compacité : si un ensemble d'axiomes n'est pas satisfiable, alors un nombre fini de ses axiomes est insatisfiable. Enfin, un troisième résultat important, le théorème de Skolem-Löwenheim : si un ensemble d'axiomes a un modèle, alors il a un modèle au plus dénombrable. Cela fait que \mathbb{R} , s'il a un modèle, a un modèle dénombrable, si si - je sais, la première fois qu'on lit ça on a mal à la tête et on croit que c'est une blague.

Ces trois théorèmes, adéquatement groupés, sont prouvés avec efficacité et clarté. Ils font partie de la saine culture d'un honnête homme et justifient à eux seuls le prix du livre à mon humble avis. Pour donner un aperçu du principe des trois preuves d'un coup : faisons comme si un système d'axiomes avait un modèle. Construisons ce modèle par étapes, en ajoutant tout ce qu'on peut construire à partir de ce qu'on a déjà construit. Soit ça s'arrête un jour par incohérence, soit ça ne s'arrête jamais. Donc : soit le modèle ainsi construit ne "bloque jamais" et il est alors dénombrable, soit il a une contradiction, et nécessairement cette contradiction a été obtenue par un nombre d'étapes fini

(évidemment, la vraie démonstration est beaucoup plus compliquée).

Le chapitre 13 présente des variantes du théorème de Skolem-Löwenheim, vu au chapitre 12.

Le chapitre 14 présente l'arithmétique de Robinson, ici simplement appelée Q , et le fait qu'elle est "aussi forte qu'une machine de Turing". L'arithmétique Q est en fait proche de l'arithmétique usuelle, mais elle ne permet pas de prouver certains éléments simples, comme la commutativité de l'addition. Elle est néanmoins essentiellement indécidable, comme on le voit au chapitre 15, c'est-à-dire que toutes ses extensions (dont la bonne vieille arithmétique ou la théorie des ensembles et à peu près tout ce qu'on connaît d'un peu rigolo en maths) sont indécidables, i.e. comportent des choses dont on ne peut ni prouver qu'elles sont vraies ni prouver qu'elles sont fausses. L'arithmétique de Robinson est un peu le minimum en indécidable (enfin, il existe des tas de niveaux d'indécidabilité, de degré non-comparables, i.e. dont on ne peut dire que l'un est plus compliqué que l'autre, mais disons que Q est un cas particulièrement réduit et intéressant d'indécidable). Le chapitre 15 combine 14 et d'autres pour conclure que beaucoup de choses, enfin tout système un peu gros, contient de l'indécidable, en fait même de l'essentiellement indécidable que l'on ne saurait compléter à coups d'axiomes ou schémas récursifs d'axiomes pour atteindre du décidable.

Le chapitre 16 montre qu'on ne peut prouver la consistance de la plupart des systèmes intéressants. Dans la foulée, les auteurs montrent le théorème de Löb (préparez de l'aspirine) : si pour un certain A , on prouve que la prouvabilité de A entraîne A , alors en fait on a aussi une preuve de A (oui, je sais c'est pas clair quand on lit ça comme ça en débarquant), et le fait que si une théorie est consistante, alors elle n'a pas de prédicat désignant la vérité (théorème de Tarski).

Le chapitre 17 discute les modèles non-standards de l'arithmétique. L'introduction montre bien que l'on aimerait au moins que tous les modèles soient au moins "très ressemblants" mais que cela ne saurait être le cas. Évidemment, on pense immédiatement à renommer les choses pour rendre les modèles très ressemblants ; on se

dit aussi que, le théorème de Skolem-Löwenheim nous garantissant que des cardinalités différentes existent forcément (des modèles de plusieurs "tailles" d'infinies) (si du moins un modèle existe), on peut tenter de se restreindre aux modèles dénombrables. Peine perdue ! Même en se restreignant au dénombrable, rien à faire : l'arithmétique n'est pas « aleph-0 catégorique » (elle a des modèles dénombrables non-isomorphes). La page 195 raconte l'aspect des modèles non-standards : le bon vieux segment initial N comme on le connaît bien, suivi de copies de Z , en nombre dénombrable si le cardinal total est dénombrable, elles-mêmes ordonnées (roulement de tambour)... comme les rationnels !

Le chapitre 18 présente de manière percutante la logique du second ordre (en gros, la logique où l'on s'autorise à quantifier sur les fonctions, phrases, prédicats). Les 6 propriétés magiques suivantes sont en particulier prouvées et vont à l'encontre du cas de la logique du premier ordre :

- il n'y a pas de procédure effective pour décider la validité de formules du second ordre ;
- il y a un ensemble dénombrable de formules, non satisfiable, dont chaque ensemble fini est satisfiable (pas de théorème de compacité comme au premier ordre, cf chapitre 12) ;
- une phrase du second ordre permet de caractériser les modèles dénombrables (pas de théorème comme celui de Skolem-Löwenheim en logique du second ordre) ;
- il existe une phrase du second ordre dont les conséquences (A étant conséquence de B si tout modèle de A est modèle de B) sont exactement les phrases vraies dans le modèle standard ;
- il existe une phrase du second ordre dont les modèles sont exactement les modèles isomorphes au modèle standard (pas de modèles non-standards si l'on utilise cette phrase) ;
- il existe une formule du second ordre qui définit les phrases vraies dans le modèle standard (pas de théorème de Tarski).

Le point 5 surprend au début, mais il faut bien voir qu'il est tiédi par le point 1 ; certes, il y a une « formule magique » qui permet de se restreindre au modèle standard, malheureusement on ne peut déterminer les conséquences de cette formule magique (sinon

d'ailleurs on aurait une contradiction avec les théorèmes d'incomplétudes, car une procédure effective permettrait de déterminer la véracité dans le modèle standard).

Les chapitres suivants présentent des points plus spécifiques (mais très intéressants !) : la vérité en arithmétique (le théorème de Tarski montre qu'elle n'est pas définissable, ce chapitre montre qu'elle l'est en logique du second ordre), la décidabilité de l'arithmétique sans multiplication (arithmétique de Presburger, qui est donc un peu un exemple standard de 'gros' décidable comme l'arithmétique de Robinson est un exemple de 'petit' indécidable), logique dyadique, lemme d'interpolation de Craig (qui dit en gros que si on peut prouver $A-B$, alors on peut prouver $A-C$ et $C-B$ pour un certain C ne contenant que des symboles inclus dans A et B ; le sujet est introduit avec beaucoup de clarté), logique modale de la prouvabilité, paradoxe de Rosser pour la construction explicite d'énoncés non décidables dans les extensions de l'arithmétique de Robinson, la non-récurtivité de l'addition ou la multiplication des modèles non-standard (i.e., si on prend un modèle dénombrable non-standard de l'arithmétique, et qu'on récupère dans le modèle standard la fonction correspondant à l'addition, alors cette fonction n'est pas calculable, ni même définissable).

Ce livre est pas cher et difficile à poser ; attention, il peut nuire à votre travail : il peut vous faire rater votre arrêt de métro ou de Rer, et vous entraîner dans des chapitres où vous n'aviez pas prévu d'aller. Il présente l'inconvénient de notations et d'une terminologie un peu non-standard parfois ; quand vous y lisez « arithmétique Q », lisez « arithmétique de Robinson », quand vous lisez « arithmétique sans multiplication », lisez « arithmétique de Presburger », et méfiez-vous du symbole qui d'habitude signifie « depuis A on peut prouver B » et signifie ici « tout modèle de A est un modèle de B ». Mais on s'y fait vite, promis.

Quatrième Conférence Internationale sur l'Apprentissage et la Fouille de Données (MLDM'05)

9-11 juillet 2005, Leipzig, Allemagne
<http://www.mldm.de/frame.htm>

Compte-rendu rédigé par Marc Boule (France Télécom, R&D)

Remarques générales

Quatrième édition d'une petite conférence bi-annuelle, originellement un workshop, issu de la communauté du Pattern Recognition. La faible taille de la conférence permet des échanges intenses entre les participants.

Environ 50 participants (20 pays), 58 articles présentés (pour 103 soumissions).

Contenu scientifique

- Large diversité de domaines Machine Learning abordés: estimation de modèle paramétrique, réseaux de neurones, préparation des données, clustering, séries temporelles...
- Et de domaines d'application: médecine, vision par ordinateur, image, parole, texte...

Les actes

Les articles présentés sont regroupés dans un volume d'actes. Le programme est en ligne sur le site de la conférence.

Remarques particulières sur quelques communications

SSC: Statistical Subspace Clustering

I. Candillier, I. Tellier, F. Torre, O. Bousquet

La méthode de clustering présentée vise une interprétation aisée. Un clustering est ici constitué d'un ensemble d'hyper-rectangles, chaque hyper-rectangle utilisant un nombre réduit de dimensions. L'utilisation d'hyper-rectangles permet naturellement une interprétation aisée à l'aide de règles. Le nombre d'hyper-rectangles est un paramètre utilisateur (il pourrait être déterminé à l'aide d'un critère de type BIC). Les paramètres du modèle sont déterminés par un algorithme EM.

Prediction of Secondary Protein Structure Content from Primary Sequence Alone - a Feature Selection Based Approach

L. Kurgan and L. Homaeian

La présentation était intéressante essentiellement en tant que description d'un domaine d'application peu habituel. Une protéine est une molécule constituée d'une séquence d'acides aminés, et il y a 20 types d'acides aminés. Une protéine peut donc être modélisée comme une chaîne de symboles de longueur variable exprimée sur un alphabet de 20 lettres. Le problème est ici de prédire la structure secondaire de la molécule (brins, hélices...) à partir de sa structure primaire (la séquence d'acides aminés). Dans la méthode proposée, la modélisation de l'information contenue dans la chaîne de symbole se limite à l'utilisation de statistiques élémentaires sur les acides aminés (avec prise en compte de quelques connaissances du domaine). Remarque: on ne fait pas beaucoup mieux en text-mining quand on modélise un texte sous la forme d'un sac de mots.

Using Clustering to Learn Distance Functions for Supervised Similarity Assessment

Ch. F. Eick, A. Rouhana, A. Bagherjeiran, R. Vilalta

La méthode présentée utilise d'une part une méthode de clustering basée sur une distance (de type K-moyenne) et un algorithme d'ajustement des poids des attributs dans une fonction de distance. Après chaque étape de clustering, les poids des attributs sont modifiés de façon à rendre les clusters plus « purs ». Ces deux étapes sont répétées un grand nombre de fois. Les expérimentations montrent que la distance apprise permet d'améliorer la performance d'algorithmes de type plus-proche-voisin. Un autre type de classifieur « plus-proche-centroïde » qui associe à chaque instance la classe majoritaire du cluster de son plus proche centroïde a également été expérimenté. Les performances obtenues sont prometteuses.

Linear Manifold Clustering

R. Harpaz and R. Haralick

Dans la méthode présentée, un clustering est constitué d'un ensemble de sous-espaces de dimension quelconque orientés arbitrairement. Par exemple, un clustering en dimension 3 pourra être constitué d'un ensemble de plans, de droites et de points, les données étant distribués dans ces sous-espaces, avec un bruit gaussien. Ce type de représentation est intéressant par son expressivité. Dans la méthode, le nombre de cluster est un paramètre utilisateur. La méthode de recherche des clusters est particulièrement intéressante. Un sous-espace de dimension k étant déterminé par $k+1$ point, on peut par simple échantillonnage aléatoire contrôler la probabilité de trouver un tel sous-espace pour autant qu'il existe dans les données. Pour reconnaître si un échantillon de $k+1$ points correspond effectivement à un cluster dans sous-espace de dimension k , l'ensemble de tous les points du jeu de données est projeté sur un vecteur du sous-espace orthogonal afin de produire un histogramme. Si cet histogramme présente un pic de densité, le sous-espace correspondant est un bon candidat pour former un cluster.

Universal Clustering with Regularization in Probabilistic Space

V. Nikulin and A. J. Smola

La méthode présentée introduit une famille de fonctions de perte (de type Kullback-Leibler) permettant de travailler dans l'espace des probabilités. Le sujet paraît intéressant, mais l'article est assez hermétique. Lors de la présentation, l'auteur s'est contenté de projeter le texte de son article, ce qui n'a pas permis d'éclaircir notablement son approche...

MML-Based Approach for Finite Dirichlet Mixture Estimation and Selection

N. Bouguila and D. Ziou

La méthode présentée généralise aux fonctions de Dirichlet les techniques habituellement mises en oeuvre pour les modèles de mélanges de Gaussiennes. La fonction de Dirichlet étant plus expressive que la Gaussienne, on s'attend à une meilleure modélisation des données.

L'auteur s'attaque ici au problème du choix automatique du nombre de fonctions de Dirichlet à utiliser dans le modèle de mélange. Plusieurs approches sont évaluées (MML, MDL, Mixture MDL, AIC...). Le choix d'un prior « naturel » pour les paramètres des fonctions de Dirichlet est un problème difficile. Néanmoins, l'approche MML semble ici donner les meilleurs résultats.

Optimising the Choice of Colours of an Image Database for Dichromats

V. Kovalev and M. Petrou

L'intérêt de la présentation est ici plus culturel que scientifique. L'auteur s'intéresse ici au problème des daltoniens (8 % de la population « mâle »). Il propose de changer la palette des couleurs pour optimiser leur discrimination des couleurs.

Signature-based Approach for Intrusion Detection

Bon K. Sy

L'auteur s'intéresse ici au problème de la détection d'une tentative d'intrusion dans un système informatique en se basant sur un log des commandes du shell unix. Un des problèmes est qu'il y a beaucoup de commandes possibles. Un premier prétraitement a consisté à catégoriser (manuellement) ces commandes en 5 familles. Un second problème est que l'on dispose de nombreux logs « normaux » (les exemples), et de très peu de logs d'intrusion (les contre-exemples). Il paraît donc difficile d'adopter un approche supervisée classique. Même une approche-semi-supervisée paraît inadaptée. L'approche présentée collecte des statistiques sur des sous séquences (de longueur 4 ici). Ces statistiques permettent de proposer un modèle des comportements « normaux » (les signatures) et de déterminer le risque d'intrusion en temps réel par comparaison de la signature courante avec les signatures de référence.

Activity and Motion Detection Based on Measuring Texture Change

L. J. Latecki, R. Mieziako, D. Pokrajac

Les auteurs s'intéressent à la détection d'objets en mouvement dans une vidéo (application dans la télé-surveillance). Les techniques habituelles se focalisent sur les variations de couleurs de pixels entre les images de la vidéo. Elles sont efficaces mais largement bruitées. La méthode proposée travaille sur une représentation des images sous forme de textures. Les textures sont évaluées sous forme de blocs de pixels $8*8$ sur 3 images consécutives, l'information de texture étant extraite au moyen d'une analyse PCA. La méthode mesure alors la quantité de changement de texture en chaque point des images de la vidéo pour détecter les mouvements. Les expérimentations montrent une détection robuste des mouvements.



Reconnaissance des Formes et Intelligence Artificielle (RFIA'06)

Tours - 25-27 Janvier 2006

<http://www.antsearch.univ-tours.fr/rfia2006/>

Compte-rendu rédigé par Sébastien Laborie (INRIA Rhône-Alpes) et Jérôme Pierson (France-Télécom R&D, INRIA Rhône-Alpes), aidés par Jérôme Euzenat (INRIA Rhône Alpes).

C'est dans le centre international de congrès « Le Vinci » de Tours que les présidents du comité d'organisation, Hubert Cardot et Gilles Venturini, ont accueilli le 15^e congrès francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA 2006). Organisée par le laboratoire d'informatique de l'Université François-Rabelais de Tours, cette manifestation s'est déroulée du 25 au 27 janvier 2006. Pour cette année qui est l'occasion de célébrer le 50^{ème} anniversaire de l'intelligence artificielle, née en 1956 à la conférence de Dartmouth, les co-présidents du comité de programme, Patrick Bouthemy pour la communauté Reconnaissance des Formes et Yves Demazeau pour la communauté Intelligence Artificielle, ont sélectionné 111 communications sur plus de 230 soumissions dont 63 communications (pour 132 soumissions) en Reconnaissance des Formes et 48 communications (pour 100 soumissions) en Intelligence Artificielle. Le taux de sélection est donc de 48% pour chacune des deux disciplines. Les 237 participants ont pu profiter sur ces 3 jours de 4 sessions invités, et de 4 séries de sessions parallèles de présentations scientifiques.

Du côté Intelligence Artificielle, il nous a semblé que deux thèmes tenaient une place importante dans cette édition de RFIA : les systèmes multi-agents (sans doute l'impact de l'orientation du président du comité de programme) et les aspects décision.

Nous évoquons ci-dessous quelques présentations qui nous ont particulièrement intéressé. Elles ne représentent, bien entendu, qu'une partie des exposés auxquels nous avons assisté et ne reflètent que notre avis personnel. Vous pourrez vous en rendre compte par

vous-même car nos collègues de l'AFRIF ont mis en ligne les actes de la conférence à l'adresse suivante : <http://www.afrif.asso.fr/archive/rfia2006/> (déjà dans les archives, c'est dire que nous vivons une époque résolument moderne).

L'une des premières sessions de la première matinée traite de l'oral et de l'écrit et nous avons choisi de vous rapporter la communication de **Bahou Y., Belguith L., Aloulou C., Ben Hamadou A intitulée Adaptation et implémentation des grammaires HPSG pour l'analyse de textes arabes non voyellés**. La problématique présentée par l'article nous a été difficile à comprendre malgré l'application de l'auteur à nous présenter la problématique de l'arabe non-voyellé. Dans la langue arabe, une grande partie des corpus disponibles (principalement journaux et site web) sont non-voyellés. Pour en comprendre le sens, l'humain doit faire l'effort de les voyeller. L'article présente une méthode pour re-voyeller les textes sur des critères syntaxiques et grammaticaux. Ce travail automatisé est un prétraitement obligatoire sur tous ces corpus non voyellés pour toute analyse sémantique ou tout traitement automatique de la langue.

Une session en parallèle traite de fouille et extraction de connaissances. L'article de **Jacquet F., LARGERON C. sur la Prise en compte de la structure des documents pour la découverte d'informations inattendues** a suscité un grand intérêt du fait de son excellente présentation. Il est à noter que cet article a remporté le prix du meilleur papier en Intelligence Artificielle. L'objectif de ce travail est de trouver dans un corpus de documents textuels ceux qui comportent

des informations nouvelles. Par exemple, on peut imaginer un utilisateur sur le Web souhaitant rechercher des pages « nouvelles », c'est-à-dire des pages dont il ne connaît pas l'existence, par rapport à un site de référence. Les auteurs proposent d'améliorer leur système existant en prenant en compte la structure des documents comme les différentes sections d'un article scientifique (titre, mots-clefs, résumé...). À l'aide de cette structure, il est ensuite possible d'attribuer des poids d'importance diverse à certains mots présents dans le document. Par exemple, les mots contenus dans le titre sont plus importants que ceux situés dans le contenu du texte... Les expérimentations présentées montrent l'efficacité de leurs travaux.

Pour la deuxième série de sessions de la matinée, nous assistons à la session modélisation et situation d'agent où un membre de notre équipe présente un article : **Gestion dynamique de contexte pour l'informatique diffuse par Euzenat J., Pierson J., Ramparany F.** Nous y présentons une proposition d'architecture pour les environnements d'informatique diffuse et sensibles au contexte. Les technologies du web sémantique sont utilisées pour décrire les informations manipulées par les consommateurs (effecteurs) et les producteurs (capteurs) d'informations de contexte mais aussi pour trouver les correspondances entre elles.

Pour la première conférence invitée, **Jacques Pitrat** nous présente **notre collègue MALICE**. La démarche générale d'un système informatique consiste à fournir une succession de méthodes permettant de résoudre un problème. MALICE, s'inspirant du résolveur général de problème ALICE de Jean-Louis Laurière, aborde les problèmes comme le ferait un chercheur humain. Ce « chercheur automatique » travaille en premier lieu sur l'énoncé du problème posé. En modifiant ce problème, ou les méthodes de résolution, le système est capable de le résoudre efficacement.

La première journée s'achève autour de la table ronde de l'IA : « **Le futur de l'IA** », animée par Jérôme Euzenat qui la présente plus comme un débat avec la salle que comme une suite de monologues de spécialistes. Les différents intervenants sont Michèle Sebag (LRI), Henri Prade (IRIT), Yves Demazeau (IMAG) et Bertrand Braunschweig (IFP). Les différents orateurs ont souligné le manque de mutualisation entre les différentes disciplines de l'IA et le manque de représentation

de la recherche française au niveau européen au regard des compétences disponibles. De l'audace, de l'ambition et un esprit collectif doivent être le futur de la recherche en IA en France. Le débat s'est achevé sur une proposition concrète de création d'une école d'été en France sur le thème de l'IA.

Pour nous, le jeudi commence par le premier « petit déjeuner de l'IA » qui a pour but de présenter de façon informelle les différents appels à projet lancé par l'ANR (et en particulier l'Appel Blanc et le RNTL). Les informations sont accessibles à partir du site de l'ANR : <http://www.gip-anr.fr>.

Puis on assiste à la présentation invitée de **Samy Bengio : Challenges in Multimodal Sequence Processing**. L'orateur présente une méthode générique pour l'analyse de différents signaux correspondant à un même événement ou une même scène. Après une présentation des différentes méthodes d'analyse de signal probabiliste et d'apprentissage (HMM, filtre de Kalman, réseau Bayésien dynamique, etc.), l'auteur présente la principale difficulté de l'analyse multi-sources, c'est-à-dire la désynchronisation entre les différents signaux pour un même événement. Il propose un modèle de HMM appelé asynchronous HMM permettant de gérer les dépendances temporelles. Enfin, il expose les cadres pratiques d'application de ses travaux comme l'analyse d'une réunion de travail.

Pour terminer cette deuxième matinée, nous assistons à la session Ontologies, modèles et usages d'où nous résumons la communication **Interrogation et vérification de diagrammes de classes UML : une approche fondée sur le modèle des graphes conceptuels par Raimbault T., Genest D., Loiseau S.** Les auteurs proposent une méthode de vérification et d'interrogation de diagramme de classes UML. Pour cela, ils proposent de traduire automatiquement un diagramme de classes UML en un graphe conceptuel et de raisonner sur ce graphe. On peut alors interroger le graphe conceptuel avec des requêtes du type : Quelle est la relation entre la classe X et la classe Y ?, Quelles sont les sous-classes de la classe X ?, Quelle classe a pour propriété Z ?, etc.

En début d'après-midi, **Jean Charlet**, conférencier invité a présenté : **Les connaissances médicales à l'épreuve de l'informatisation : entre documents non-**

structurés et ontologies. Il a introduit les problèmes liés à l'informatisation du dossier médical du point de vue de l'usage des médecins. La problématique globale consiste à fournir aux médecins un outil garantissant la même pratique de la médecine ainsi que la même souplesse d'utilisation (par exemple pour l'annotation de document) que les inscriptions sous forme papier. Plus précisément, il s'agit de mettre à leur disposition des services de traçabilité et d'interopérabilité entre les différents types de documents. Pour cela, les dossiers médicaux sont stockés dans des fichiers XML. L'utilisation d'ontologies permet de formaliser ces données pour que le système puisse par la suite les interpréter.

La session suivante à laquelle nous assistons ensuite s'intitule Composition et raisonnement. Nous retenons la présentation de **Kaci S.**, volant littéralement de session en session, intitulée **Première comparaison des réseaux de conditionnelles « CP-nets » et de la logique possibiliste en modélisation des préférences par Dubois D., Kaci S., Prade H.** Les auteurs comparent deux approches pour la représentation de préférences, les CP-nets qui peuvent modéliser des ordres partiels et la logique possibiliste qui permet d'exprimer des pré-ordres totaux. Les auteurs montrent qu'il n'y a pas d'équivalence entre les deux représentations, mais qu'il est possible d'approximer un CP-net par la logique possibiliste. De plus, ils mettent en évidence les limitations des CP-nets.

En fin de journée, nous assistons à la session Plates-formes et environnements où un membre de notre équipe y présente un papier: **Adaptation spatiale efficace de documents SMIL par Laborie S., Euzenat J., Layaida N.** Pour exécuter un document multimédia sur différentes plates-formes (PDA, téléphones portables...) il est nécessaire de l'adapter. Les auteurs présentent une méthode permettant d'adapter des documents multimédia SMIL dans leur dimension spatiale. Pour cela, ils proposent une nouvelle représentation spatiale assurant un compromis entre expressivité et rapidité de calcul.

C'est déjà le jeudi soir avec le repas de gala et la remise des prix aux meilleurs articles. Pour ce repas de Gala, nous dînons à l'Hôtel de ville de Tours construit à la fin du XIXe siècle dans un style que l'on peut admirer sur la photo. On peut féliciter le comité d'organisation pour cette soirée dans ce magnifique lieu: ce n'est

pas si fréquent de pouvoir danser sur le parquet des salles de conseils municipaux.

La journée du vendredi débute par un « petit déjeuner de l'IA ». Aujourd'hui, le petit déjeuner est l'occasion pour ceux qui le désirent de présenter des idées de projets, ou d'informer la communauté d'idées à creuser. Ainsi, Gérard Sabah présente un projet en cours de montage qui a pour but la mise en place de plates-formes françaises pour le partage de corpus. Puis Olivier Boissier et Yves Demazeau présentent la problématique des aspects légaux et éthiques impliqués par les systèmes multi-agents. Enfin Marie-Odile Cordier évoque un débat éthique autour de l'implication de l'IA dans les technologies émergentes de type « Big Brother » en proposant d'organiser une table ronde.

Ce petit déjeuner est suivi d'une conférence invitée en reconnaissance de forme : **Peut-on être sûrs d'avoir reconnu une forme ?** par **Jean-Michel Morel.**

La session Décisions multi-agents sera ensuite marquée par la présentation de l'article: **Un modèle décisionnel orienté comportement utilisant le vote, application à la navigation autonome en environnement simulé par Hannon D., Grislin-LeStrugeon E., Mandiau R.** Les auteurs proposent un modèle décisionnel pour un piéton dans un environnement virtuel basé sur les comportements et non sur la planification. Un agent est un piéton. Le choix d'une action est le résultat d'un vote entre les différents comportements qui s'exécutent en parallèle et qui régissent son état mental. Chacun d'entre eux vote pour l'action qu'il juge la plus appropriée, mais il peut aussi émettre un veto contre certaines actions ou proposer des options différentes de l'action sélectionnée.

Parmi les dernières sessions de la conférence, nous assistons à celle traitant d'apprentissage et agents où l'article suivant était présenté: **Des agents cognitifs situés apprennent à désigner des actions par Poudade J., Landwerlin L., Paroubek P.** Les auteurs exposent un mécanisme de jeux de langages par lequel des agents cognitifs situés apprennent à désigner des objets. Un lexique commun émerge dans cette population d'agents. Ils montrent qu'en ajoutant une mémoire de travail aux agents, on réduit les ambiguïtés. Enfin, l'expérience est enrichie en ajoutant l'action et sa description aux jeux de langage.

CONFÉRENCES

Un avantage certain de la conférence RFIA réside dans le fait que deux communautés y sont représentées. Dans cet esprit d'ouverture, nous avons participé à certaines présentations côté Reconnaissance des Formes. Ainsi, de la session Documents graphiques et traitements en ligne où l'article **Apprentissage progressif pour la reconnaissance de symboles dans les documents graphiques par Barrat S., Tabbone S.** nous a semblé intéressant. Le but de ce travail est de reconnaître des symboles dans des documents graphiques comme, par exemple, des schémas de circuits électroniques. Disposant d'une base de symboles à reconnaître, les auteurs proposent plusieurs méthodes permettant de les identifier tout en prenant en considération leur orientation ainsi que les

différents bruits du symbole à reconnaître comme l'épaisseur du symbole...

En conclusion, on ne peut que se réjouir d'avoir participé et assisté à RFIA 2006, d'autant que l'organisation était remarquable et nous ne parlons pas seulement des délicieux repas. Pour les nostalgiques et les curieux, nous leur conseillons d'aller faire un tour sur le site de la conférence (<http://www.antsearch.univ-tours.fr/rfia2006/>) afin d'y admirer un magnifique montage vidéo de certains moments de la conférence et quelques photos dites non censurées.

La prochaine édition de la conférence RFIA aura lieu à Amiens en 2008, souhaitons lui d'être aussi réussie.



24^e Journées d'Arithmétiques Faibles Journées complexité, modèles finis et bases de données

IUT de Fontainebleau, 26-28 mai 2005

<http://www.univ-paris12.fr/lacl/durand/cm2005.html>

Compte-rendu rédigé par Olivier Teytaud (Paris-Sud).

Présentation

Les 24^{èmes} journées d'arithmétique faible et les journées du groupe de travail « Complexité, modèles finis et bases de données » du GDR ALP se sont tenues en même temps à l'IUT de Fontainebleau (université Paris 12), les jeudi 26, vendredi 27 et samedi 28 mai 2005.

Les rencontres du groupe de travail « Complexité, modèles finis et bases de données » ont pour but de permettre à tout chercheur qui le souhaite de présenter ses travaux en complexité, en théorie des modèles finis ou dans un de leur nombreux champs d'application comme : les bases de données, la vérification de programmes, la satisfaction de contraintes, les jeux formels, etc... Les éditions précédentes ont eu lieu en 1995 à Paris XI (organisateur : M. Santha), en 1996 à l'IUT de Fontainebleau (P. Cégielski), en 1997 à Paris II (M. de Rougemont), en 1998 à l'ENS de Lyon (P. Koïran), en 1999 à Paris VI (I. Guessarian), en 2000 à Paris XII (A. Durand), en 2001 à Paris II (M. Hermann), en 2002 à Arcachon (D. Janin), en 2003 à Paris-Dauphine (C. Bazgan et A. Durand) et en 2004 à Lausanne (J. Duparc).

Les arithmétiques faibles jouent un rôle fondamental dans plusieurs domaines des mathématiques, de l'informatique et de la philosophie en étudiant les propriétés des nombres entiers d'un point de vue logique. Ces journées sont l'occasion pour les chercheurs du domaine qui étudient ou appliquent les arithmétiques faibles de se rencontrer et d'échanger des idées. Les JAF ont eu lieu précédemment à : Lyon (1990), Paris (1990, 1991, 1996), Clermont (1991, 1992, 1993, 1994, 1995, 1999, 2000), Fontainebleau (1994, 2001), Metz (1996), St-Petersburg (1997, 2002), Mons (1997), Warsaw (1998), New-York (1999), Naples (2003), et Yerevan (2004).

Une grande partie des sessions des deux événements sont susceptibles d'intéresser les deux communautés et ont donc été communes. Les organisateurs de

cette année sont Patrick Cégielski, Arnaud Durand, Michel de Rougemont.

Liste des participants

Danièle Beauquier Université Paris 12 (France)
Alexis Bès Université Paris 12 (France)
Olivier Bournez LORIA Nancy (France)
Thomas Brihaye Université Mons-Hainaut (Belgique)
Patrick Cégielski Université Paris 12 (France)
Andres Cordon Sevilla University (Spain)
Bruno Courcelle Université Bordeaux 1 (France)
Paulin de Naurois Mines de Nancy (France)
Arnaud Durand Université Paris 12 (France)
Alex Esbelin Université Clermont 1 (France)
Olivier Finkel Université Paris VII (France)
Hervé Fournier Université de Versailles (France)
Daniel Graça LORIA Nancy (France)
Etienne Grandjean Université de Caen (France)
Serge Grigorieff Université Paris VII (France)
Miki Hermann Ecole Polytechnique (France)
Emil Jerabek Académie des sciences (Rép. Tchèque)
Janos Makowsky Technion - Israel Institute of Technology
Yuri Matiyasevich Moscou (Russie)
Malika More Université Clermont 1 (France)
Anca Muscholl Université Paris VII (France)
Selma Naboulsi Université Paris 12 (France)
Joachim Niehren INRIA, Lille (France)
Frédéric Olive Université Aix-Marseille 1 (France)
Romain Péchoux Mines de Nancy (France)
Mélanie Ponchie Université de Versailles (France)
Alexandre Rambaud Université Paris VII (France)
Michel de Rougemont Université Paris II (France)
Luc Segoufin INRIA, Rocquencourt (France)
Victor Selivanov Novosibirsk University (Russie)
Anatol Slissenko Université Paris 12 (France)
Olivier Teytaud Université Paris XI (France)
Jerzy Tomasik Université Clermont 1 (France)
Pierre Valarcher Université de Rouen (France)

Abstracts

Thomas Brihaye, Université de Mons-Hainaut

Model-Checking, bisimulation and word combinatorics.

Abstract. In this talk we explain briefly what model-checking is. Then we recall the notion of bisimulation and explain why the study of bisimulations is relevant in the context of model-checking. In particular we focus on bisimulations of dynamical systems and we explain how the dynamics of these systems can be recovered through some words encoding. This technique leads to nice finiteness results for o-minimal dynamical systems.

Andres Cerdón-Franco, University of Seville (Spain)

On axiom schemes for $D(n+1)(T)$ -formulas (joint work with A. Fernández-Margarit and F.F. Lara-Martín)

Abstract. Motivated by J. Paris' problem on the equivalence between the schemes of induction and minimization for $\Delta(n+1)$ -formulas, in [3] and [6] we introduced theories $\mathbf{I}\Delta(n+1)(\mathbf{T})$, $\mathbf{L}\Delta(n+1)(\mathbf{T})$ and $\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ (in what follows referred to as *relativized fragments*), where \mathbf{T} is an arbitrary theory in the language of first-order Arithmetic extending $\mathbf{I}\Delta 0$. These fragments are obtained by restricting the schemes of induction, minimization and (a version of) collection to $\Delta(n+1)(\mathbf{T})$ -formulas (that is, $\Sigma(n+1)$ -formulas which are equivalent in \mathbf{T} to some $\Pi(n+1)$ -formula).

The work developed in [3]–[6] provides evidence that relativized fragments are interesting subsystems of Peano Arithmetic (= \mathbf{PA}). Firstly, relativized fragments are closely related to fragments of Arithmetic described in terms of inference rules (see [1], [2]). In fact, $\mathbf{T}+\mathbf{I}\Delta(n+1)(\mathbf{T})$ and $\mathbf{T}+\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ are deductively equivalent to the closure of \mathbf{T} by unnested applications of $\Delta(n+1)$ -induction rule and $\Sigma(n+1)$ -collection rule, respectively. Thus, the study of relativized fragments in [3], [6] can be considered to be a model-theoretic analysis of inference rules. Secondly, fragments for $\Delta 1(\mathbf{T})$ -formulas are relevant to some fundamental problems on Complexity Theory. As an example, in [4] it is shown that the provably total recursive functions of the parameter free $\Delta 1(\mathbf{T})$ -induction scheme are, precisely,

the primitive recursive functions which are provably total in \mathbf{T} ; and sufficient conditions for separating the Polynomial Time Hierarchy and \mathbf{PSPACE} are established. Finally, relativized fragments provide a general framework for the analysis of extensions of bounded quantifier complexity of *classical* fragments of Arithmetic. The results in [5] show that fragments for parameter free $\Sigma(n+1)$ and $\Pi(n+1)$ formulas have extensions of *small* quantifier complexity (that is, of quantifier complexity less than that of their axiomatizations) while their parameter counterparts do not. This divergent behaviour can be explained from the study of fragments for $\Delta(n+1)(N)$ -formulas (that is, $\Sigma(n+1)$ -formulas equivalent in the standard model of Arithmetic, N , to some $\Pi(n+1)$ -formula).

In this work we present a new method for the study of relativized fragments. The crucial property is that, under certain assumptions on the theory \mathbf{T} , these fragments give axiomatizations of the classes of structures which are either models of the $\Pi(n+2)$ -consequences of \mathbf{T} or models of a certain classical fragment of Arithmetic. More concretely, it holds that (let us denote by $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ the $\Pi(n+2)$ -consequences of \mathbf{T}):

Theorem 1. (Transfer Theorem)

1. If M is a model of $\mathbf{I}\Delta(n+1)(T)$, either M is a model of $\mathbf{Th}(\Pi(n+2))(T)$ or M is a model of $\mathbf{I}\Sigma(n+1)$.
2. If M is a model of $\mathbf{B}^*\Delta(n+1)(T)$, either M is a model of $\mathbf{Th}(\Pi(n+2))(T)$ or M is a model of $\mathbf{B}\Sigma(n+1)$.

In addition, we establish versions of the theorem above for the parameter free counterparts of the relativized fragments: $\mathbf{I}\Delta(n+1)(\mathbf{T})$ -, $\mathbf{L}\Delta(n+1)(\mathbf{T})$ - and $\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ -.

By applying these Transfer Theorems, we develop a systematical analysis of the relative strength, finite axiomatizability and quantifier complexity properties of relativized fragments (for both their parametric and parameter free versions). In fact, Transfer Theorems allow one to reduce the study of these properties to the analysis of the existence of extension of bounded quantifier complexity of classical fragments of Arithmetic.

(A) On relative strength.

From Theorem 1 and the fact that $\mathbf{B}\Sigma(n+1)+exp$ does not have consistent $\Sigma(n+3)$ -extensions (Proposition 4.3 in [5]), we can deduce the following result:

Theorem 2.

1. Assume that $\mathbf{T1}, \mathbf{T2}$ are closed under $\Delta(n+1)$ -induction rule. Then

$$\mathbf{I}\Delta(n+1)(\mathbf{T1}) \uparrow \mathbf{I}\Delta(n+1)(\mathbf{T2}) \text{ if, and only if,} \\ \mathbf{Th}(\Pi(n+2))(\mathbf{T1}) \uparrow \mathbf{Th}(\Pi(n+2))(\mathbf{T2}).$$

2. Assume that $\mathbf{T1}, \mathbf{T2}$ prove exp and are closed under $\Sigma(n+1)$ -collection rule. Then

$$\mathbf{B}^*\Delta(n+1)(\mathbf{T1}) \equiv \mathbf{B}^*\Delta(n+1)(\mathbf{T2}) \text{ if, and only if,} \\ \mathbf{Th}(\Pi(n+2))(\mathbf{T1}) \equiv \mathbf{Th}(\Pi(n+2))(\mathbf{T2}).$$

As a consequence, we deduce Hierarchy Theorems for $\Delta(n+1)(\mathbf{T})$ -induction and $\Delta(n+1)(\mathbf{T})$ -collection.

Theorem 3. (Hierarchy Theorem)

1. $\mathbf{I}\Delta(n+1)(\mathbf{I}\Sigma(n)) \subset \mathbf{I}\Delta(n+1)(\mathbf{I}\Sigma(n+1)) \subset \mathbf{I}\Delta(n+1)(\mathbf{I}\Sigma(n+2)) \subset \dots \subset \mathbf{I}\Delta(n+1)(\mathbf{PA}) \subset \mathbf{I}\Delta(n+1)(N).$
2. $\mathbf{B}^*\Delta(n+1)(\mathbf{I}\Sigma(n)) \subset \mathbf{B}^*\Delta(n+1)(\mathbf{I}\Sigma(n+1)) \subset \mathbf{B}^*\Delta(n+1)(\mathbf{I}\Sigma(n+2)) \subset \dots \subset \mathbf{B}^*\Delta(n+1)(\mathbf{PA}) \subset \mathbf{B}^*\Delta(n+1)(N).$

Part 1 in the theorem above was previously established in [6] with a different proof. Nevertheless, Part 2 answers an open problem posed in [6].

Similar results for parameter free relativized fragments are also obtained by using Transfer Theorems and results in [5] on properties of extensions of parameter free (classical) fragments.

(B) On finite axiomatizability.

Theorem 4. Assume that \mathbf{T} is closed under $S(n+1)$ -collection rule.

1. $\mathbf{I}\Delta(n+1)(\mathbf{T})$ is finitely axiomatizable if, and only if, $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ is finitely axiomatizable.
2. If $\mathbf{I}\Delta(n+1)(\mathbf{T})$ - is finitely axiomatizable then so is $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$.
3. ($n \geq 1$) $\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ is finitely axiomatizable if, and only if, $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ is finitely axiomatizable.

From Theorem 4, we deduce that if \mathbf{T} is a consistent extension of $\mathbf{I}\Sigma(n+1)$, then none of the following theories is finitely axiomatizable: $\mathbf{I}\Delta(n+1)(\mathbf{T})$, $\mathbf{B}^*\Delta(n+1)(\mathbf{T})$, $\mathbf{I}\Delta(n+1)(\mathbf{T})$ -.

Our results for parameter free minimization and collection schemes are not as general as the previous ones, but they allow us to prove that neither

$\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ - nor $\mathbf{L}\Delta(n+1)(\mathbf{T})$ - is finitely axiomatizable for $\mathbf{T} = \mathbf{Th}(N), \mathbf{PA}, \mathbf{I}\Sigma(n+k)$ ($k \geq 1$).

(C) On quantifier complexity properties.

Theorem 5. Let \mathbf{T} be a consistent theory.

1. If $\mathbf{I}\Sigma(n+1)$ does not imply $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ then $\mathbf{I}\Delta(n+1)(\mathbf{T})$ is $\Pi(n+3)$ -axiomatizable but not $\Sigma(n+3)$. If $\mathbf{I}\Sigma(n+1)$ implies $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ then $\mathbf{I}\Delta(n+1)(\mathbf{T})$ is $\Pi(n+2)$ -axiomatizable and, for $n \geq 1$, it is not $\Sigma(n+2)$ -axiomatizable.
2. If $\mathbf{I}\Sigma(n)+exp$ does not imply $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ then $\mathbf{B}^*\Delta(n+1)(\mathbf{T})$ is $\Pi(n+3)$ -axiomatizable but not $\Sigma(n+3)$. Moreover, if $\mathbf{I}\Sigma(n)$ implies $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$ then $\mathbf{B}^*\Delta(n+1)(\mathbf{T}) \uparrow \mathbf{I}\Sigma n$.

Observe that our methods provide a complete description of the quantifier complexity of $\Delta(n+1)(\mathbf{T})$ -induction and $\Delta(n+1)(\mathbf{T})$ -collection schemes (except for the case $n=0$).

We also obtain similar (but weaker) results on the quantifier complexity of parameter free relativized fragments.

Finally, we discuss some questions left unanswered in our work. While Transfer Theorems for induction and collection schemes are best possible, we have obtained only a partial result for $\mathbf{L}\Delta(n+1)(\mathbf{T})$ -. Namely, if M is a model of $\mathbf{L}\Delta(n+1)(\mathbf{T})$ -, either M is a model of $\mathbf{Th}(\Sigma(n+1) \cup \Pi(n+1))(\mathbf{T})$ or M is a model of $\mathbf{I}\Pi(n+1)$ -.

Problem 6. Does Transfer Theorem for $\mathbf{L}\Delta(n+1)(\mathbf{T})$ - hold if we replace $\mathbf{Th}(S(n+1) \gg P(n+1))(\mathbf{T})$ by $\mathbf{Th}(P(n+2))(\mathbf{T})$?

Problem 6. Does Transfer Theorem for $\mathbf{L}\Delta(n+1)(\mathbf{T})$ - hold if we replace $\mathbf{Th}(\Sigma(n+1) \cup \Pi(n+1))(\mathbf{T})$ by $\mathbf{Th}(\Pi(n+2))(\mathbf{T})$?

In addition, we obtain interesting consequences by considering relativized fragments for $\mathbf{T} = \mathbf{I}\Delta 0 + exp$, or $\mathbf{I}\Pi(n+1)$ -. Taking $\mathbf{T} = \mathbf{I}\Delta 0 + exp$, we get a reformulation of an open problem raised by Wilkie and Paris [7] asking if every model of $\mathbf{I}\Delta 0$ which is not closed under exponentiation is a model of the $S1$ -collection scheme. On the other hand, in [2] Beklemishev proved that, for each $n \geq 1$, $\mathbf{I}\Pi(n+1)$ - is conservative over $\mathbf{I}\Sigma n$ - with respect to boolean combinations of $\Sigma(n+1)$ -sentences. For $\mathbf{T} = \mathbf{I}\Pi(n+1)$ -, the study of theories $\mathbf{L}\Delta(n+1)(\mathbf{I}\Pi(n+1))$ -

is related to the problem of determining whether that conservation result is best possible.

References

- [1] Beklemishev, L.: Induction rules, reflection principles and provably recursive functions. *Ann. Pure Appl. Logic* 85 (3), 193-242 (1997)
- [2] Beklemishev, L.: Parameter-free induction and provably total computable functions. *Theor. Computer Sci.* 224, 13-33 (1999)
- [3] Cordon-Franco, A., Fernández-Margarit, A., Lara-Martín, F.F.: On the quantifier complexity of $D(n+1)(T)$ -induction. *Archive for Math. Logic* 43, 371-398 (2004)
- [4] Cordon-Franco, A., Fernández-Margarit, A., Lara-Martín, F.F.: Provably Total Primitive Recursive Functions: Theories with Induction. *Computer Science Logic 2004. Lecture Notes in Computer Sci.* 3210, 355-369 (2004)
- [5] Cordon-Franco, A., Fernández-Margarit, A., Lara-Martín, F.F.: Fragments of Arithmetic and true sentences. *Math. Logic Quarterly* 51 (3), 313-328 (2005)
- [6] Fernández-Margarit, A., Lara-Martín, F.F.: Induction, minimization and collection for $D(n+1)(T)$ -formulas. *Archive for Math. Logic* 43, 505-541 (2004)
- [7] Wilkie, A., Paris, J.: On the existence of end extensions of models of bounded induction. *Methodology and Philosophy of Science VIII*, 143-161. North-Holland, Amsterdam, 1989

Bruno Courcelle, LaBRI, Université Bordeaux 1

On Seese's Conjecture : Recent Advances (joint work with S. Oum)

Abstract. D. Seese conjectured that if a set of graphs has a decidable monadic theory, then it is the image of a set of trees under a Monadic Second-order transduction, i.e. a transformation of relational structures expressed in Monadic Second-order logic. This is equivalent to saying that it has bounded clique-width. Clique-width is a graph complexity measure relevant to the construction of polynomial graph algorithms.

In my lecture in Lausanne in 2004, I explained that relativized versions of the conjecture, in particular to partial orders, to comparability graphs and to bipartite

graphs, are equivalent to its full version. I presented also some provable relativizations, to line graphs, to partial orders of dimension 2 and to interval graphs.

In Fontainebleau, I will present a common work with S. Oum (Princeton) establishing the weak form of the conjecture where the hypothesis is that the satisfiability problem for monadic second-order formulas with even cardinality set predicates is decidable. The proof technique is completely different from the one used for the previously known results. It uses the notion of « vertex-minor » which parallels for clique-width the notion of « minor » intimately related to tree-width.

Olivier Finkel, Équipe de Logique Mathématiques, Université Paris 7

Infinite computations and highly undecidable problems

Abstract. We consider infinite computations of simple finite machines, like pushdown automata or one counter automata. The class of context free (respectively, 1-counter) ω -languages is the class of languages of infinite words which are accepted by pushdown (respectively, 1-counter) automata with a Büchi or a Muller acceptance condition. It is well known that any Turing machine can be simulated by a 2-counter automaton. We show the following surprising result. When considering acceptance of infinite words, from different points of view, a 1-counter automaton is sufficient to get the whole complexity of a Turing machine.

More precisely, we show first that the topological complexity of context free or of 1-counter w -languages is the same as that one of ω -languages accepted by Turing machines with a Büchi or a Muller acceptance condition. In particular, for every recursive ordinal $\alpha < \omega_1^{CK}$, where ω_1^{CK} is the first non recursive ordinal, there are some Σ^0_α -complete and some Π^0_α -complete ω -languages accepted by Büchi 1-counter automata, [Fin04].

We show also that many classical decision problems about context free ω -languages, or even about 1-counter ω -languages, are located in the analytical hierarchy and are Π^1_2 -complete. For instance, the universality problem, the inclusion problem and the equivalence problem are Π^1_2 -complete. Some other decision problems are shown to be Π^1_2 -hard, as the

problem to determine whether a given context free w -language is in a given Borel class Σ^0_α .

References

- [ABB96]
J-M. Autebert, J. Berstel and L. Boasson, Context Free Languages and Pushdown Automata, in Handbook of Formal Languages, Vol 1, Springer Verlag 1996.
- [Ber79]
J. Berstel, Transductions and Context Free Languages, Teubner Studienbücher Informatik, 1979.
- [CC89]
J. Castro and F. Cucker, Nondeterministic $_$ -Computations and the Analytical Hierarchy, Zeitschr. J. Math. Logik und Grundlagen d. Math, 35, 1989, p.333-342.
- [Dup03]
J. Duparc, A Hierarchy of Deterministic Context Free c -Languages, Theoretical Computer Science, Volume 290 (3), p. 1253-1300.
- [DY92]
P. Darondeau and S. Yoccoz, Proof Systems for Infinite Behaviours, Information and Computation, Volume 99 (2), 1992, p. 178-191.
- [EH93]
J. Engelfriet and H. J. Hoogeboom, X-Automata on c -Words, Theoretical Computer Science, Volume 110 (1), 1993, p. 1-51.
- [Fin01]
O. Finkel, Topological Properties of Omega Context Free Languages, Theoretical Computer Science, Vol. 262 (1-2), 2001, p. 669-697.
- [Fin03a]
O. Finkel, Borel Hierarchy and Omega Context Free Languages, Theoretical Computer Science, Vol 290 (3), 2003, p. 1385-1405.
- [Fin04]
O. Finkel, Borel Ranks and Wadge Degrees of Context Free c -Languages, in the Proceedings of the International Conference CIE 2005, Computability in Europe: New Computational Paradigms, Amsterdam, June 8-12, 2005, Lecture Notes in Computer Science, to appear.
- [FS03]
O. Finkel and P. Simonnet, Topology and Ambiguity in Omega Context Free Languages, Bulletin of the Belgian Mathematical Society, Volume 10 (5), 2003, p. 707-722.
- [Kec95]
A.S. Kechris, Classical Descriptive Set Theory, Springer-Verlag, 1995.
- [Lan69]
L. H. Landweber, Decision Problems for c -Automata, Math. Syst. Theory 3 (1969) 4, p. 376-384.
- [LT94]
H. Lescow and W. Thomas, Logical Specifications of Infinite Computations, In:"A Decade of Concurrency" (J. W. de Bakker et al., eds), Springer LNCS 803 (1994), p. 583-621.
- [Mos80]
Y. N. Moschovakis, Descriptive Set Theory, North-Holland, Amsterdam, 1980.
- [PP04]
D. Perrin and J.-E. Pin, Infinite Words, Automata, Semigroups, Logic and Games, Volume 141 of Pure and Applied Mathematics, Elsevier, 2004.
- [Sis89]
A. Prasad Sistla, On Verifying That A Concurrent Program Satisfies A Nondeterministic Specification, Information Processing Letters, Volume 32 (1), 1989, p.17-23.
- [Sta86]
L. Staiger, Hierarchies of Recursive c -Languages, Jour. Inform. Process. Cybernetics EIK 22 (1986) 5/6, p. 219-241.
- [Sta97a]
L. Staiger, c -Languages, Chapter of the Handbook of Formal languages, Vol 3, edited by G. Rozenberg and A. Salomaa, Springer-Verlag, Berlin.
- [Tho90]
W. Thomas, Automata on Infinite Objects, in: J. Van Leeuwen, ed., Handbook of Theoretical Computer Science, Vol. B (Elsevier, Amsterdam, 1990), p. 133-191.

Daniel Graça, LORIA, Nancy

Computing with continuous-time analog circuits

Abstract. We present an overview of the General Purpose Analog Computer (GPAC). This model, based on analog circuits, was introduced by C. Shannon in 1941 as a mathematical model of an analog device, the Differential Analyzer. In this talk, we will review the existing theory and discuss recent developments on this model, including results about its computational power.

Serge Grigorieff, LIAFA, Université Paris 7

Infinite computations and Kolmogorov complexities

Abstract. We introduce the classes $\text{MaxPR}[X \rightarrow D]$ and $\text{MaxRec}[X \rightarrow D]$ of functions $X \rightarrow D$ which are pointwise maximum of partial or total recursive sequences of functions where $(D, <)$ is some computable partially ordered set and X is the set of binary words. The enumeration theorem and the invariance theorem always hold for $\text{MaxPR}[X \rightarrow D]$, leading to a variant $K_{\text{max}D}$ of Kolmogorov complexity. We characterize the orders D such that the enumeration theorem (resp. the invariance theorem) also holds for $\text{MaxRec}[X \rightarrow D]$. It turns out that $\text{MaxRec}[X \rightarrow D]$ may satisfy the invariance theorem but not the enumeration theorem. Also, when $\text{MaxRec}[X \rightarrow D]$ satisfies the invariance theorem then the Kolmogorov complexities associated to $\text{MaxRec}[X \rightarrow D]$ and $\text{MaxPR}[X \rightarrow D]$ are equal (up to a constant). Letting $K_{\text{min}D} = K_{\text{max}D'}$, where D' is the reverse order, we prove that either $K_{\text{min}D} = K_{\text{max}D} = K$ (up to a constant) or $K_{\text{min}D}, K_{\text{max}D}$ are incomparable and $< K$ and $> K'$. We characterize the orders leading to each case. We also show that $K_{\text{min}D}, K_{\text{max}D}$ cannot be both much smaller than K at any point. These results are proved in a more general setting with two orders on D , one extending the other.

Miki Hermann, LIX, École Polytechnique

Complexity of Clausal Constraints Over Chains (joint work with Nadia Creignou (Marseille), Andrei Krokhin (Durham) and Gernot Salzer (Vienna))

Abstract. We investigate the complexity of the satisfiability problem of constraints over finite totally ordered domains. In our context, a clausal constraint is a disjunction of inequalities of the form $x \leq d$ and $x \geq d$. We classify the complexity of constraints based on clausal patterns. A pattern abstracts away from variables and contains only information about the domain elements and the type of inequalities occurring in a constraint. Every finite set of patterns gives rise to a (clausal) constraint satisfaction problem in which all constraints in instances must have an allowed pattern. We prove that every such problem is either polynomially decidable or NP-complete, and give a polynomial-time algorithm for recognizing the tractable cases. Some of these tractable cases are new and have not been previously identified in the literature.

Emil Jerabek, Academy of sciences (Rép. Tchèque) and Utrecht University

Hardness amplification in bounded arithmetic

Abstract. One of the important achievements in derandomization theory is the following theorem: given a truth table of a Boolean function f on n variables, which cannot be approximated by circuits of size $2^{\epsilon n}$ with nonnegligible advantage, one can construct in polynomial time a function f' on $O(n)$ variables with worst-case circuit complexity $2^{W(n)}$. We show how to formalize this statement in S^1_2 . To achieve this goal, we develop in S^1_2 basics of the theory of finite fields and list-decoding of Reed-Muller codes, and we present a modification of Soltys' theory LAP for linear algebra.

Janos Makowsky, Technion Haifa, Israel

Application of Logic in Combinatorics (Joint work with Eldar Fisher)

Abstract. We discuss linear recurrence relations for various combinatorial counting functions. Among these we have the density function for classes of labeled graphs, and various graph polynomials.

In each of the cases discussed, a condition ensuring definability in CMSOL (Monadic Second Order Logic with modular counting) suffices to prove the existence of such recurrence relations.

Anca Muscholl, LIAFA, Université Paris 7

Logic with data

Abstract. In a data word each position carries a label from a finite alphabet and a data value from some infinite domain. We show that two-variable logic (with successor and order relation) on such strings is decidable and that the complexity is as hard as Petri net reachability. In the case where only the successor relation is used, the complexity is shown to become tractable.

Paulin de Naurois, LORIA, Nancy

The Complexity of Semilinear Problems in Succinct Representation

Abstract. We prove completeness results for twenty three problems in semilinear geometry. These results involve semilinear sets given by additive circuits as input data. If arbitrary real constants are allowed in the circuit, the completeness results are for the Blum-Shub-Smale additive model of computation. If, in

contrast, the circuit is constant-free, then the completeness results are for the Turing model of computation. One such result, the $\Delta(2, | \cdot |)$ -completeness of deciding Zariski irreducibility, exhibits for the first time a problem with a geometric nature complete in this class.

Joachim Niehren, INRIA-LIFL, Lille

Querying XML-Documents by Tree Automata

Abstract. Information extraction from semi-structured documents requires to find queries in XML trees that define appropriate sets of n -tuples of nodes. In this talk, I will discuss how to represent such queries by tree automata. I will first recall notions of tree automata for unranked trees as in XML, present representation formalisms for n -ary queries by tree automata, and discuss efficiency, expressiveness, and learnability. Finally, I will illustrate an application to information extraction.

Romain Péchoux, LORIA, Nancy

Resource analysis by sup-interpretation

Abstract. Sup-interpretation is a new tool to control memory resources by Static analysis which bounds from above the size of function output. This method applies to first order functional programming with pattern matching. This work is related to quasi-interpretations but we are now able to determine resources of more algorithms and it is easier to perform an analysis with this new tools.

Olivier Teytaud, LRI, Université Paris 11

Inductive-deductive systems : a statistical learning point of view

Abstract. The theorems about incompleteness of arithmetic have often been cited as an argument against automatic theorem proving and expert systems. However, these theorems rely on a worst-case analysis, which might happen to be overly pessimistic with respect to real-world domain applications.

For this reason, a new framework for a probabilistic analysis of logical complexity is presented in this paper. Specifically, the rate of non-decidable clauses and the convergence of a finitely describable set of axioms toward the target one when the latter exists in the language are studied, by combining results from mathematical logic and from statistical learning.

The theoretical analysis gives practical hints into the pros and cons of optimizing the set of axioms along

learning, with respect to the convergence rates above.

Jerzy Tomasiak, LLAIC, Université Clermont 1

Synthesis theories of finite structures of arithmetic (joint work with Micha[1] Krynicki and Konrad Zdanowski (Warsaw University))

Abstract. Experiences show that computer arithmetic is not the theory of the standard (infinite) model of arithmetic since computer may accede to a finite part of natural numbers only. Whence it is important to study the family of theories of finite fragments of natural numbers that looks like the most natural approximation of the theory the whole infinite model. More general question arises how to reconstruct the information (specification) on the whole structure from the network of its finite approximations. We examine here some examples of reconstruction operators like lower limit, upper limit (sl) of theories as well as reduced product and ultraproduct operations on finite models.

Consider the family FIN of all theories of initial segments of a standard model of arithmetic. In this approach one replaces an actual infinity of a standard model by the (potentially) infinite family of finite models. M. Mostowski propose to study the family FIN under the operation sl which is a kind of limit operation on theories. The operation sl can be compared with some other limit operations, well known in the model theory, like reduced product and ultraproduct constructions (or limit of directed system). Our approach may find applications in theoretical foundations of computing, in information systems, in information fusion theory and in software engineering.

Let $A=(N, R_1, \dots, R_s, f_1, \dots, f_t, a_1, \dots, a_r)$ be a model having as a universe the set N of natural numbers, and s.t. R_1, \dots, R_s are relations on N , f_1, \dots, f_t are operations on N and $a_1, \dots, a_r \in N$. We consider the family $FM(A)$ of all finite initial segments $A^n = (\{0, \dots, n\}, R^n_1, \dots, R^n_s, f^n_1, \dots, f^n_t, a^n_1, \dots, a^n_r, n)$ of the model A , where R^n_i is the restriction of R_i to the set $\{0, \dots, n\}$, $a^n_i = a_i$ if $a_i \leq n$, otherwise $a^n_i = n$ and f^n_i is defined by $f^n_i(b_1, \dots, b_n) = f_i(b_1, \dots, b_n)$ if $f_i(b_1, \dots, b_n) \leq n$, otherwise $f^n_i(b_1, \dots, b_n) = n$. The signature of A_n contains one new constant Max s.t. $Max^n = n$. We say that ϕ is satisfied by b_1, \dots, b_p in all sufficiently large finite models of $FM(A)$, what is denoted by $FM(A) \text{ models } \text{sl} \phi[b_1, \dots, b_p]$, (or simply $\text{models } \text{sl} \phi[b_1, \dots, b_p]$) if there is $k \in \mathbb{N}$ such that for all $n \geq k$ $A_n \text{ models } \phi[b_1, \dots, b_p]$. We call sl -Theory of A the following set of sentences $sl-Th(A) = \{ \phi : \exists k \forall n \geq k A_n \text{ models } \phi \}$.

The theory $sl-Th(A)$ is a consistent theory without finite models and it can be described in terms of limits of the sequence of sets $(Th(A_n):n \in \mathbb{N})$.

Theorem 1 The following holds :

(i) $sl-Th(A) = \limsup(n \rightarrow \infty) Th(A_n)$.

(ii) Sentence ϕ is consistent with the theory $sl-Th(A)$ iff $\phi \in \liminf(n \rightarrow \infty) Th(A_n)$.

(iii) $sl-Th(A)$ is a complete theory iff the sequence $(Th(A_n):n \in \mathbb{N})$ is convergent.

Recall that a sequence $(A_n):n \in \mathbb{N}$ of sets is convergent if it has a limit:

$$\lim_{n \rightarrow \infty} (A_n) = \limsup(n \rightarrow \infty) (A_n) = \liminf(n \rightarrow \infty) (A_n).$$

A semantical characterization of the theory $sl-Th(A)$ can be done in terms of ultraproducts. This theory is the intersection of all theories of ultraproducts of the family $(A_n):n \in \mathbb{N}$ under nonprincipal ultrafilters over \mathbb{N} .

Theorem 2 We have the following identity :

$$sl-Th(A) = U\{Th(PA_n/U):U \text{ nonprincipal ultrafilter over } \mathbb{N}\}.$$

Another properties of the operation sl will be illustrated by examples of models A over \mathbb{N} .

Victor Selivanov, Novosibirsk University (Russia)

Some Hierarchies and Reducibilities on Regular Languages

Abstract. We discuss some known and introduce some new hierarchies and reducibilities on regular sets. We establish some facts on the corresponding degree structures and relate the reducibilities to hierarchies. As an application, we characterize regular languages whose leaf-language classes (in the balanced model) are contained in the polynomial hierarchy. For any reducibility we try to give some motivation and interesting open questions, in a hope to convince the reader that study of these reducibilities is important for automata theory and computational complexity.

Un portail sémantique pour la gestion des connaissances en cancérologie

Mathieu d'Aquin

Thèse de Doctorat

Soutenue le 14 décembre 2005 à l'Université Henri Poincaré – Nancy 1.

Jury : Rose Dieng-Kuntz (Directrice de recherche INRIA – Sophia Antipolis), Jean-Paul Haton (Professeur à l'Université Henri Poincaré – Nancy 1), Marie-Christine Jaulent (Directrice de recherche INSERM, Paris), Jean Lieber (maître de conférences à l'Université Henri Poincaré – Nancy 1), Amedeo Napoli (Directeur de recherche CNRS, Nancy), Michaël Rusinowitch (Directeur de recherche CNRS, Nancy)

Résumé : Cette thèse présente une application et des extensions des principes et des technologies relatives au Web sémantique pour la construction d'un portail dédié à la gestion des connaissances en cancérologie. OWL, langage de représentation d'ontologies, est notamment employé pour la représentation des connaissances contenues dans les référentiels, sortes de protocoles de décision médicaux. Les inférences associées à OWL sont utilisées pour l'aide à la décision dans le cadre de l'application des référentiels. Mais les connaissances contenues dans les référentiels ne permettent pas de prendre en compte toutes les situations et nécessitent d'être adaptées pour certains cas particuliers. Pour cette raison, nous proposons une extension des technologies du Web sémantique afin de permettre la représentation de connaissances d'adaptation et l'implémentation d'un mécanisme de raisonnement à partir de cas (RàPC) intégré à l'infrastructure du Web sémantique. Par ailleurs, dans des domaines complexes tels que la cancérologie, coexistent différents points de vue, c'est-à-dire différentes façons de considérer les connaissances du domaine et différentes façons de les utiliser. La notion d'ontologie ne permet pas de prendre en compte cette multiplicité de points de vue. Pour cette raison, nous avons étudié les formalismes de représentation des connaissances intégrant de façon explicite la notion de point de vue. Nous nous appuyons sur le langage C-OWL afin d'établir une représentation multi-points de vue des connaissances contenues dans les référentiels en accord avec les technologies du Web sémantique. Sur la base de cette représentation multi-points de vue en C-

OWL, un mécanisme de RàPC décentralisé est défini. Celui-ci exploite les multiples points de vue pour distribuer le RàPC en différents processus. Les relations existant entre les points de vue sont quant à elles utilisées pour l'échange de connaissances et la collaboration entre ces différents processus. Au travers de l'application en cancérologie, la représentation multi-points de vue montre ainsi des apports concernant les difficultés inhérentes à l'ingénierie des connaissances (acquisition, évolution), la mise en oeuvre du mécanisme de RàPC (utilisation de plusieurs cas, combinaison de points de vue) et le « passage à l'échelle » des mécanismes de raisonnements pour le Web sémantique, RàPC compris.

Mots clés : Représentation et gestion des connaissances, Web sémantique, représentation multi points de vue, raisonnement à partir de cas, cancérologie, OWL, logiques de descriptions, C-OWL.

Ontologies et aide à l'utilisateur pour l'interrogation de sources multiples et hétérogènes

Hassen Kefi

Thèse de Doctorat

Soutenue le 29 mars 2005 à l'Université de Paris-Sud, Orsay.

Jury : Bernd Amann (Professeur à l'Université Pierre et Marie Curie), Mohand-Said Hacid (Professeur à l'Université Claude Bernard Lyon I), Christine Froidevaux (Professeur à l'Université de Paris-Sud XI), Alain Léger (Directeur de Programme Scientifique France Télécom R&D Rennes), Chantal Reynaud (Professeur à l'Université de Paris-Sud XI), Brigitte Safar (Maître de conférences à l'Université de Paris-Sud XI).

Résumé : L'explosion du nombre de sources d'information accessibles via le Web multiplie les besoins de techniques d'intégration des sources de données multiples et hétérogènes. Ces techniques reposent sur la construction d'une vue uniforme des données distribuées permettant de donner l'impression à l'utilisateur qu'il interroge un système homogène et centralisé. Les travaux menés dans cette thèse portent sur les ontologies en tant qu'outils d'aide à l'interrogation d'un serveur d'information. Les deux aspects des ontologies que nous

avons traités sont les ontologies en tant qu'outil d'aide à l'affinement de requêtes, d'une part, et en tant qu'aide à l'interrogation unifiée, d'autre part.

Concernant le premier aspect, nous proposons de construire progressivement, avec l'utilisateur, des requêtes plus spécifiques, plus contraintes que la requête initiale, jusqu'à l'obtention de réponses moins nombreuses et plus pertinentes. Notre approche est basée sur l'utilisation conjuguée d'une ontologie du domaine et de la technique des treillis de Galois.

Concernant le second aspect, nous proposons une approche générique d'alignement d'ontologies mise en œuvre au travers d'un processus semi-automatique. L'approche que nous proposons s'applique en présence d'une dissymétrie dans la structure des taxonomies comparées. Nous proposons d'assembler, dans un ordre bien déterminé, des techniques terminologiques, structurelles et sémantiques.

Ces deux aspects ont fait l'objet de travaux distincts effectués au sein de deux projets. Le premier est le projet Picse 2, mené en collaboration avec France Télécom R&D dont le domaine d'expérimentation est le tourisme. Le second est le projet RNTL eDot, dont le domaine d'application porte sur l'analyse du risque bactériologique de contamination d'aliments.

Planification de mission pour un véhicule aérien autonome

Elodie Chantery

Thèse de Doctorat

Soutenue le 30 septembre 2005 à l'École Nationale Supérieure d'Aéronautique, Toulouse.

Jury : Dominique Luzeaux (Directeur du Centre Technique des Systèmes d'Information, DGA, Rapporteur), Erik Sandewall (Professeur, Université de Linköping, Suède, Rapporteur), Philippe Morignot (Responsable scientifique, Axlog Ingénierie, Examineur), Magali Barbier (Ingénieur de recherche, ONERA/DCSD, Co-Directrice de thèse), Raja Chatila (Directeur de recherche, LAAS/CNRS, Directeur de thèse), Jean-Loup Farges (Ingénieur de recherche, ONERA/DCSD, Membre invité).

Résumé : De nos jours, les engins autonomes suivent un plan de mission donné, éventuellement réactualisé par un opérateur. Les missions de longue durée, avec des communications limitées entre l'opérateur et l'engin, requièrent de l'engin une autonomie décisionnelle : le véhicule doit être capable de prendre des décisions et de réaliser les actions associées. Ce travail porte sur la replanification embarquée, nécessaire dès qu'un événement survenant en cours de mission invalide le plan courant. L'illustration choisie est une mission militaire d'observation effectuée par un drone Moyenne Altitude Longue Endurance en zone ennemie. Cette thèse a pour but d'élaborer un module de planification de mission intégré dans une architecture embarquée.

La première partie propose un formalisme adapté au problème de planification de mission. La planification doit sélectionner un ensemble d'objectifs auxquels sont associées des récompenses variables selon la date et les ressources. Il doit aussi choisir la meilleure manière de réaliser chaque objectif dans le temps et l'espace en optimisant un critère prenant en compte les récompenses et les coûts de réalisation associés aux objectifs.

La deuxième partie expose le cadre algorithmique, inspiré de l'algorithme A*, et différentes méthodes d'évaluation de coût, d'élagage et de rangement. La résolution consiste d'une part en un choix discret de points de passage et d'autre part en une optimisation continue d'un critère non linéaire sous contraintes mixtes.

L'intégration du module de planification embarquée dans une architecture fait l'objet de la troisième partie de cette thèse. Une architecture hybride hiérarchisée en quatre niveaux est développée de manière à gérer des niveaux d'autonomie allant de la gestion de la mission jusqu'au guidage.

La dernière partie présente le résultat des tests de la partie algorithmique : 36 scénarios sont simulés sur 16 combinaisons de méthodes, en fonction de l'instant de replanification et du type d'événement déclencheur (carburant, danger, objectifs). L'analyse des résultats permet de dégager les méthodes présentant les meilleurs compromis qualité/temps de calcul, notamment un algorithme basé sur une heuristique évaluant les récompenses non encore obtenues et sur un rangement au meilleur d'abord.

Mots clés : engin autonome, planification, recherche heuristique, architecture embarquée.

Apprentissage et circulation d'information

Antoine Cornuéjols

Habilitation à Diriger des Recherches

Soutenue le 8 décembre 2005 à l'Université de Paris-Sud à Orsay.

Jury : François Denis (Rapporteur), Stan Matwin (Rapporteur), Lorenza Saitta (Rapporteur), Christine Froidevaux (Examineur), Jean-Gabriel Ganascia (Examineur), Mirta Gordon (Examineur), Andrée Tiberghien (Examineur).

Résumé : L'apprentissage est évidemment lié à la notion d'information. D'une part, parce qu'il s'agit d'un mécanisme de filtrage de régularités dans les données, et de ré-expression de ces régularités. D'autre part, parce que l'apprentissage, au lieu d'être le phénomène en une passe habituellement étudié, est surtout un processus dynamique se déroulant dans le temps avec transfert d'information d'un état au suivant. Finalement, parce que l'apprentissage implique aussi l'articulation de domaines de connaissances entre eux. Paradoxalement, le mot information n'apparaît quasiment pas dans la littérature sur l'apprentissage.

Mon mémoire de HDR présente une vision de l'étude de l'apprentissage recentrée sur la notion d'information et sa circulation au cours de l'apprentissage. Il est organisé autour de trois parties.

La première examine le processus d'induction d'hypothèses à partir d'exemples et un phénomène de transition de phase associé.

Si l'analyse théorique de l'induction a surtout mis l'accent sur le nécessaire contrôle de la capacité ou richesse de l'espace des hypothèses afin d'obtenir des garanties, en probabilité, sur la performance des hypothèses sélectionnées, des expériences récentes ont montré que d'autres propriétés jouent également un rôle déterminant sur la possibilité de l'apprentissage.

Un phénomène de transition de phase a été en effet mis en évidence, d'abord en induction de programmes logiques, puis, par nous-même, en inférence grammaticale.

le. Ce phénomène permet d'expliquer les difficultés rencontrées par tous les algorithmes connus sur certaines familles de problèmes. Ce phénomène semble aussi fondamental que les propriétés liées à la capacité des espaces d'hypothèses. En particulier les conséquences en terme de codage des entrées et de représentation des hypothèses sont examinées.

La deuxième partie du mémoire porte sur des éléments à mettre en place afin de pouvoir définir une géométrie des évolutions des systèmes d'apprentissage. La mesure de corrélation, qui doit pouvoir prendre des valeurs négatives, entre programmes est discutée et illustrée par une méthode de comparaison de techniques de sélection d'attributs. Le problème des changements de repères et de transport parallèle le long d'une trajectoire sont ensuite illustrés par des travaux sur le raisonnement par analogie.

La troisième partie du mémoire est consacrée à l'étude de l'apprentissage en tant que système dynamique évoluant sous l'effet d'une séquence d'entrées (apprentissage en-ligne ou incrémental). Le cas limite des apprentissages insensibles à l'ordre des entrées est discuté, ainsi que la place de l'apprentissage actif dans l'étude des apprentissages en-ligne. Le cas complexe de l'apprentissage comme construction et articulation entre plusieurs domaines conceptuels est finalement abordé à travers un mécanisme de raisonnement que nous avons proposé pour expliquer l'apprentissage de concepts de physique par des élèves de lycée.

Comprendre pour apprendre ... et apprendre pour comprendre

Brigitte Grau

Habilitation à Diriger des Recherches

Soutenue le 13 décembre 2005 au LIMSI, Orsay.

Jury : Rapporteurs : Patrice Enjalbert, Professeur d'Université, GREYC, Patrick Gallinari, Professeur d'Université, LIP6, Guy Lapalme, Professeur, RALI. Examineurs : Mohand Boughanem, Professeur d'Université, IRIT, Christian Jacquemin, Professeur d'Université, LIMSI, Gérard Sabah, Directeur de Recherche, LIMSI, Pierre Zweigenbaum, Chargé de recherches, AP-HP.

Résumé : La compréhension profonde de texte vue comme un processus d'inférence visant à expliciter tout l'implicite d'un texte a buté sur le problème de la modélisation des connaissances nécessaires. Une approche alternative consiste à redéfinir ce processus et à le décomposer en sous-tâches potentiellement plus faciles à résoudre. Considérant que le but premier n'est pas de comprendre un texte dans l'absolu, mais de comprendre un texte afin d'atteindre un but précis dépendant de la tâche, la compréhension de texte peut être redéfinie selon différents points de vue sur le texte qui vont permettre de répondre à ces besoins spécifiques (par exemple évaluer la pertinence d'un texte par rapport à un thème, rechercher une réponse précise à une interrogation). Suivant le besoin et le type de connaissance cherché dans un document, le lecteur n'extraira du texte que l'information qui l'intéresse et s'appuiera pour cela sur les indices qui lui permettent de réaliser sa tâche de lecture, et donc de compréhension, sans avoir à tout assimiler. On peut alors parler de « compréhension à niveaux variables » ou « d'analyse de texte », analyse qui englobe tous les types de processus appliqués au texte.

Si l'on veut construire un continuum entre ces méthodes et disposer des connaissances requises par chacune, il faut aussi créer un continuum entre ces connaissances. Une partie de celles-ci peut être automatiquement acquises des textes, textes qui ne constituent pas seulement des objets à analyser mais aussi une réserve de connaissances qu'il faut pouvoir extraire. Aussi, la voie de recherche que nous explorons consiste à utiliser des méthodes robustes pour effectuer une analyse des textes et en extraire un premier type de connaissances qui va servir de base à un apprentissage automatique. Les connaissances ainsi produites pourront ensuite être réutilisées dans des processus d'analyse plus fins, produisant par là-même des connaissances plus précises et de plus en plus structurées. On peut alors parler de compréhension pour l'apprentissage et d'apprentissage pour la compréhension.

Nous présenterons les travaux que nous avons effectués en analyse de texte et qui portent sur l'analyse thématique et la réponse à des questions et nous montrerons comment les résultats en analyse ont donné lieu à un apprentissage automatique de signatures thématiques et à leur structuration en vue d'apprendre des descriptions de situations. L'exposé de ces travaux sera

complété par la présentation de projets permettant de poursuivre dans cette voie.

Algorithmes pour la planification

Pierre Régnier

Habilitation à Diriger des Recherches

Soutenue le 8 septembre 2005 à l'IRIT.

<http://www.irit.fr/~Pierre.Regnier>

Jury : Claudette Cayrol (Directeur de Recherche, Professeur UPS), François Charpillet (rapporteur, Directeur de Recherche INRIA, Nancy), Marie-Odile Cordier (rapporteur, Professeur, Rennes), Henri Farreny (Professeur, ENSEEIHT, Toulouse), Frédéric Garcia (Directeur de Recherche, INRA Toulouse), Monique Thonnat (rapporteur, Directeur de Recherche INRIA, Sophia).

Mots clé : planification, cadre classique, recherche heuristique, GRAPHPLAN, graphe de planification, codages SAT de planification, ressources numériques...

Résumé : Dans le cadre classique de la planification, et depuis 1992 environ, stimulé par les idées issues des planificateurs SATPLAN (1992) et GRAPHPLAN (1995), l'algorithmique de la planification a réalisé des progrès spectaculaires :

- Les idées qui sont à la base de SATPLAN ont ouvert la voie aux méthodes de planification par compilation qui consistent à construire une structure disjonctive (une base de clauses pour SATPLAN, un graphe de planification pour GRAPHPLAN) susceptible de contenir tous les plans-solutions d'une taille donnée puis à en extraire un plan-solution en utilisant les techniques classiques (solveur SAT, techniques CSP). Les planificateurs SATPLAN et GRAPHPLAN ainsi que les planificateurs qui en dérivent (IPP, STAN, BLACKBOX, GP-CSP, DPP, LCGP, LCDPP) sont ainsi optimaux en nombre de niveaux d'actions indépendantes ou autorisées mais ces propriétés, bien qu'intéressantes, ne garantissent pas l'optimalité des plans-solutions en nombre d'actions. De nombreuses techniques (SAT, CSP, ILP, recherche locale) ont été utilisées, souvent avec succès pour optimiser la procédure d'extraction de GRAPHPLAN. Leur utilisation est intéressante car tout progrès dans ces tech-

niques permet une amélioration quasi-automatique des performances des algorithmes de planification qui les utilisent (BLACKBOX a ainsi de remporter la compétition IPC04 dans la catégorie des planificateurs optimaux).

- Les planificateurs actuels par recherche dans les espaces d'états utilisent souvent le graphe de planification de GRAPHPLAN pour calculer une heuristique qui leur permet de résoudre rapidement des problèmes de très grande taille (FF, YAHSP). Malheureusement, les heuristiques les mieux informées ne sont pas admissibles et ils n'offrent donc aucune garantie sur l'optimalité des plans-solutions en nombre d'actions.

Je mettrai en perspective les travaux auxquels j'ai contribué dans ce domaine depuis la soutenance de ma thèse, en 1992, et dont voici les grandes lignes :

Pendant la thèse de Christine Bastié, nous avons conçu un système de planification et d'exécution de haut niveau pour le contrôle d'exécution d'actions parallèles pour des agents autonomes en milieu dynamique. Notre système, SPEEDY, comprend ainsi trois modules distincts, fonctionnant en parallèle et de manière incrémentale ; il permet des réactions plus diversifiées que des systèmes comparables. Les tests que nous avons mené à cette époque sur les algorithmes de planification disponibles nous ont montré qu'ils étaient (à ce moment là) trop peu performants pour envisager une implémentation efficace. A la suite de ce travail sur SPEEDY, nous avons donc recentré nos recherches sur l'algorithme de la planification.

Nous avons d'abord décidé d'implémenter un planificateur pour le langage de description d'actions ADL basé sur un algorithme classique de recherche heuristique dans les espaces d'états. Les résultats de l'étude expérimentale détaillée que nous avons menée entre les planificateurs UCPOP, IPP, et notre planificateur ont contribué à remettre en cause l'opinion (courante à l'époque) selon laquelle cette technique était peu performante. Nous avons clairement montré qu'elle permettait d'obtenir, dans la très grande majorité des cas, de bien meilleures performances qu'un algorithme de recherche dans les espaces de plans partiels grâce à l'utilisation d'un test de contrôle de bouclage des états et d'une heuristique indépendante du domaine. Nous avons également exhibé quelques domaines de test dans

lesquels notre algorithme s'avérait également plus performant que des planificateurs comme IPP.

Cette étude nous ayant conduit à travailler sur les algorithmes basés sur GRAPHPLAN (IPP en est une optimisation), nous avons décidé d'étudier ce dernier de manière plus détaillée et ce travail a été le sujet essentiel de la thèse de Vincent Vidal, sous la direction de Michel Cayrol. Nous avons ainsi contribué de manière significative à l'amélioration de GRAPHPLAN sur plusieurs points :

- L'introduction de la relation d'autorisation qui permet le développement d'un espace de recherche plus compact et une amélioration importante des performances, mesurée grâce à notre planificateur LCGP ;
- La mise au point de deux heuristiques level-based et noops-first-level-based qui ont amené une amélioration significative des performances de la procédure d'extraction ;
- L'amélioration de la procédure d'extraction de plans DPPLAN, par sa simplification, l'ajout de plusieurs propagations et l'utilisation de la relation d'autorisation. Notre planificateur, LCDPP, qui implémente ces perfectionnements, peut ainsi résoudre des problèmes difficiles sur lesquels des planificateurs comme FF échouent ;
- L'amélioration des codages utilisés dans BLACKBOX par la prise en compte des seules actions et mutex et l'utilisation de la relation d'autorisation. Les tests que nous avons menés grâce à notre système TSP ont mis en évidence la supériorité de nos codages.

Nous avons enfin mené une large étude sur le domaine des techniques SAT appliquées à la planification. Nous avons apporté plusieurs simplifications aux codages existants dans les espaces d'états, dans les espaces de plans, et introduit le parallélisme dans les codages pour les espaces de plans, ce qui nous a conduit à des codages encore plus compacts. Nous avons également remplacé, dans les codages qui le permettaient, la relation d'indépendance par celle, moins contraignante, d'autorisation. Le système TSP que nous avons implémenté nous a permis de montrer une amélioration souvent très importante des performances avec nos codages.

ANNONCES

Journée scientifique

organisée en l'honneur de Jean-Louis Laurière

le 22 mars 2006
au 12, rue de l'École de Médecine, Paris 5^{ème}.

Les présentations des conférences sont disponibles à <http://afia.lri.fr/node.php?node=1125>

9h. Introduction. Jean-Charles Pomerol (Président de l'Université Pierre et Marie Curie), Patrick Gallinari (Directeur du LIP6).

9h30 Première session. Président Jacques Pitrat (LIP6-Paris 6), Jean-Yves Lucas (EDF) Quand le programme est un problème. Dominique Pastre (Crip5-Paris 5) Démonstration automatique de théorèmes, méthodes naturelles ou principe de résolution. Basile Starynkevitch (CEA), Méta-connaissance pour générer des programmes pour le web. Tristan Cazenave (Paris 8) Transformation des représentations de connaissances et recherche heuristique dans les jeux. Monique Baron, Hélène Giroire, Odile Paliès et Alain David (Paris6) Enseigner les travaux de Jean-Louis Laurière.

12h. à 14h. Pause

14h. Deuxième session Président Daniel Kayser (LIPN-Paris 13), Jean-Luc Dormoy (CEA) ALICE,

SNARK et RABBIT : l'efficacité par la généralité. Patrick Albert (ILOG) De SNARK et ALICE aux règles métier. Mehmet Dincbas (COSYTEC) Des contraintes de ALICE aux contraintes de CHIP. Jacques Pitrat (LIP6-Paris 6), ALICE, un outil fondamental pour un chercheur artificiel en Intelligence Artificielle.

16h. Pause

16h30 Troisième session. Président Jean-François Perrot (LIP6-Paris 6), Michel Gondran (EDF). Modélisation de l'Éthique de Spinoza en SNARK. Michèle Sebag (LRI-Paris 11), L'apprentissage supervisé comme un problème de satisfaction de contraintes. Jean-Gabriel Ganascia (LIP6-Paris 6), Les changements de représentation et la reformulation en résolution de problèmes. François Pachet (Sony) Des idées venues d'ALICE pour trouver des "bonnes" fonctions du signal.

18h30 Réception

INSECTES ARTIFICIELS 1^{er} OCTOBRE 2005 – 7 MAI 2006



Puces électroniques, fourmis virtuelles et robots insectes au muséum de Tours !

Depuis plusieurs années déjà, le monde fascinant des insectes inspire informaticiens et roboticiens dans leurs travaux de recherche. L'exposition « Insectes artificiels » vous propose de pénétrer dans l'univers étonnant des fourmis virtuelles et des robots autonomes dont les étonnantes capacités permettent de résoudre de façon inédite des problèmes de robotique, de classification automatique de données, de design...

A l'aide de simulations informatiques et de dispositifs interactifs conçus par l'Ecole Polytechnique de Tours, vous pourrez manipuler et voir évoluer ces singulières et passionnantes créatures !

Le muséum d'histoire naturelle et l'Ecole Polytechnique de l'Université de Tours s'associent pour montrer aux visiteurs des développements en informatique et en robotique se basant sur les capacités collectives et individuelles des insectes et notamment des fourmis. L'objectif est de vulgariser les recherches effectuées dans ce domaine, et principalement au Laboratoire d'Informatique de l'Ecole. L'exposition est présentée sous la forme d'un parcours composé de plusieurs ateliers. Chacun d'eux met en lumière un concept particulier avec deux approches. La première est orientée sur la robotique autonome et présente l'animation d'un ou plusieurs robots avec des possibilités d'interactions sous forme ludique. La deuxième est orientée sur la simulation et présente par l'intermédiaire d'un ordinateur différentes possibilités de simulation. Enfin, pour chaque atelier, on pourra trouver des exemples d'applications « industrielles » présentées sous forme de posters.

Parcours de l'exposition :

Module n°1 : suivre des traces de phéromones

L'objectif est de montrer qu'avec du matériel accessible à tous, le robot Lego Mindstorm, on peut réaliser un comportement qui s'apparente au déplacement des insectes suivant des traces de phéromones. L'utilisateur doit pouvoir ajouter manuellement des traces de phéromones et ainsi jouer lui aussi le rôle d'un robot-fourmi.



Robot fourmi réalisé à partir d'éléments Lego.

Dans cette surface de jeu, le robot se déplace en suivant des lignes blanches. Le terrain est constitué de plaques de bois fixes, sur lesquelles sont peints des chemins. On peut modifier le trajet en introduisant des plaques. Chacune est de forme carrée avec des coins arrondis pour faciliter leur coulissement, l'aire de déplacement du robot est protégée par des plaques transparentes. Un autre atelier permet d'expérimenter, par l'intermédiaire de l'ordinateur, l'impact des paramètres des phéromones (évaporation) et des robots-fourmis (nombre, vitesse) sur ce mode de navigation.

Module n°2 : retour au nid !

Cet atelier a pour but de démontrer que l'on peut s'inspirer des capacités de navigation des fourmis pour concevoir des robots capables de se déplacer dans un environnement variable et incertain. Il présente à la fois la navigation autonome d'un robot équipé d'une caméra et la simulation du comportement d'un robot-fourmi.



Le robot Boe-bot peut grâce à ses pinces transporter des objets comme les fourmis.

ANNONCES

La simulation apporte la possibilité d'interagir, en modifiant l'environnement du robot par l'ajout d'obstacles par exemple. Le visiteur peut également avoir le point de vue du robot, ce qu'il perçoit avec sa caméra.

Module n°3 : travail d'équipe

L'objectif de cet atelier est de montrer l'importance des comportements collectifs des insectes pour la résolution des problèmes pouvant intervenir en robotique. Il présente donc une colonie de robots. Mais ces simulations doivent également montrer que le travail simultané de plusieurs robots peut être plus efficace que celui d'un seul robot plus coûteux.



Une colonie de petits robots.

Module n°4 : Résolution de problèmes combinatoires

Cet atelier illustre le plus simplement possible la résolution collective de problèmes (optimisation, classification...) en s'inspirant des comportements naturels des fourmis.

Module n°5 : Design artistique

Cet atelier montrera des développements originaux qui utilisent les mouvements collectifs des insectes et

leurs capacités à se servir des phéromones pour construire des structures collectives.

Deux réalisations pourront être exposées. Des peintures de fourmis artificielles, elles se déplacent sur une toile et déposent des phéromones de couleur. Les différentes caractéristiques des fourmis produisent des peintures abstraites en perpétuelle évolution. La musique de fourmis artificielles : les fourmis se déplacent entre les notes et construisent une mélodie en déposant des phéromones entre les notes qu'elles traversent.



Une œuvre réalisée par des fourmis virtuelles.

Informations / Renseignements :

MUSÉUM D'HISTOIRE NATURELLE
3 rue du Président-Merville
37000 Tours
Téléphone : 02-47-64-13-31
Télécopie : 02-47-66-80-14
Courriel : museum@ville-tours.fr
Site internet : www.ville-tours.fr

Contacts :

d.lastu@ville-tours.fr
a.jeandrot@ville-tours.fr
nicolas.monmarche@univ-tours.fr

Une brève introduction aux méthodes d'optimisation de forme par simulation numérique des écoulements

Renan HILBERT

renan.hilbert@lbdr-ingenierie.com

http://www.lbdr-ingenierie.com

On propose ici une introduction à l'optimisation de forme pour le contrôle des écoulements, utilisant la simulation numérique en mécanique des fluides. Cette vaste problématique aux retombées industrielles concrètes présente un intérêt au niveau de l'élaboration de méthodes d'optimisation issues de l'I.A. L'exemple plus parlant pour le grand public est probablement l'optimisation de la forme d'une carrosserie automobile (par exemple une voiture de course) pour réduire la traînée de l'air et ainsi, comme le dit le langage courant, « augmenter l'aérodynamique » du véhicule. Ceci peut être réalisé par des essais en soufflerie mais aussi par des calculs numériques couplés à des algorithmes d'optimisation. On peut citer aussi par exemple la réduction du poids et l'amélioration des propriétés des moteurs, des avions ou des véhicules terrestres. Grâce à la fois à l'augmentation significative des capacités de calculs et à la mise en œuvre de méthodes numériques adaptées, l'optimisation de forme automatisée comme outil d'aide à la conception connaît un succès croissant dans l'industrie. Elle offre un outil d'amélioration du dimensionnement et de la conception de produits qui aide à réduire nombre d'essais et donc à limiter les coûts et les durées du cycle de développement. Ces méthodes restent néanmoins des outils d'aide à la décision et doivent être vues comme un complément, et non un concurrent, au savoir-faire des ingénieurs et techniciens du domaine concerné. Leur application et l'obtention de résultats significatifs et réellement utilisables nécessitent le suivi d'une méthodologie adaptée dont les principales étapes sont décrites dans cet article.

Plusieurs points importants permettent de caractériser ce type de problèmes en terme de problème d'optimisation. Tout d'abord, les variables du problème sont pour la plupart des variables paramétrisant les formes à optimiser pour contrôler l'écoulement considéré. Leur choix est primordial pour assurer une bonne faisabilité des calculs. Ensuite, les impératifs industriels en terme de taille, poids, encombrement des pièces considérées se traduisent souvent par de fortes contraintes mathé-

matiques. De plus, le fait que l'évaluation de la fonction objectif nécessite une simulation de mécanique des fluides implique un besoin important en temps de calcul, souvent d'au moins deux ordres de grandeur supérieur au temps CPU requis par l'algorithme d'optimisation lui-même. Enfin, les problèmes considérés mettent bien souvent en œuvre au minimum deux caractéristiques à minimiser/maximiser, souvent contradictoires, ils sont à classer dans la catégorie des problèmes multiobjectifs.

Une présentation des concepts, des difficultés, de la démarche et des outils en simulation numériques des écoulements dans des applications industrielles est proposée ici. Le principe du couplage entre un logiciel de simulation et un algorithme d'optimisation est ensuite décrit, ainsi que les possibilités d'utilisation du calcul parallèle pour réduire les temps. L'accent est mis sur les problèmes liés au paramétrage des calculs, au choix de l'algorithme d'optimisation et sur l'utilisation des résultats. Un exemple d'application est enfin proposé.

La simulation numérique des écoulements fluides

On parle souvent de CFD (acronyme anglais pour Computational Fluid Dynamic) ou, en français, de Mécanique des Fluides Numérique pour ce qui concerne la simulation numérique des écoulements fluides. La procédure mise en œuvre présente des spécificités qu'il est important de garder à l'esprit lors du couplage avec un algorithme pour l'optimisation automatique de forme et dont les principales étapes sont décrites ici.

La première, et non la moindre, consiste à poser le problème et, souvent, de le réduire afin de le rendre accessible à la simulation en un temps raisonnable. Il est nécessaire de définir la géométrie de base (souvent à partir d'une CAO) et l'ensemble des paramètres variables (vitesse d'admission du fluide, forme du coude d'un tuyau, forme d'un profil défini par des points et des splines...), dont ceux retenus comme

variable du problème d'optimisation. Une précision des résultats obtenus par le calcul (coefficient caractéristique de la pièce, température maximale, perte de charge...) sera nécessaire pour définir le ou les objectifs du problème d'optimisation, parfois sous forme d'une combinaison entre eux. Dans l'étape suivante, le volume occupé par le fluide, domaine de résolution du problème, est discrétisé en cellules, dont l'ensemble constitue le maillage. Cette opération étant un préalable à la discrétisation mathématique des équations du problème, elle peut avoir une répercussion importante sur la qualité des résultats, engendrant par exemple des problèmes de stabilité, de convergence ou de sous-résolution. Des logiciels adaptés sont utilisés pour les géométries complexes. Une optimisation du problème passera donc par une automatisation du maillage, qui devra en effet être re-effectué pour chaque nouvelle forme déterminée par les nouveaux paramètres.

Vient ensuite la simulation de l'écoulement : les équations régissant la mécanique des fluides, connues sous le nom d'équation de Navier Stokes. Leur résolution pose certaines difficultés numériques spécifiques et des méthodes adaptées ont du être développés. Il faut aussi définir les modèles physiques associés au problème : la turbulence dans la plupart des cas, l'estimation des propriétés thermo-physiques des fluides, les transferts thermiques, les phénomènes de changements de phase (vaporisation, solidification), des réactions chimiques, etc.... ce qui peut augmenter considérablement les temps de calcul par ajout d'EDP ou sous-modèles couplés.

Les aspects numériques sont pris en compte en spécifiant au code de calcul les conditions aux limites (conditions aux frontières du domaine pour les différentes variables), les conditions initiales et les méthodes de résolution (ordre de discrétisation, algorithmes de résolution

implicites ou explicites, paramètres de sous-relaxation). Les équations sont alors résolues itérativement par un solveur mettant en œuvre des méthodes numériques adaptées (volumes finis, éléments finis,...). Dans les cas réalistes, les calculs peuvent être extrêmement longs. Les simulations sont souvent lancées sur des calculateurs parallèles ou, de plus en plus, des grilles de calculs distribuées, dont chaque nœud calcule une partie du domaine (décomposition de domaine).

Enfin, le post-traitement des résultats consiste à visualiser les écoulements simulés, à les analyser et à extraire des résultats bruts les paramètres désirés. Les opérations de visualisation 3D, de coupes, de moyennages et les calculs des quantités physiques annexes sont nécessaires et devront aussi être automatisés pour coupler le problème à un algorithme d'optimisation automatique.

Paramétrisation et automatisation

La Figure 2 propose une représentation schématique du couplage entre une simulation de mécanique des fluides et un algorithme d'optimisation. Une fois le problème proprement paramétré, à chaque itération de l'algorithme d'optimisation, un certain nombre de jeu de paramètres devra être évalué, nombre qui peut varier de un dans les cas d'algorithmes séquentiels à un nombre important si toute une population est considérée (algorithmes génétiques par exemple). Pour chacun des cas, le jeu de paramètres est lu par un script qui doit lancer automatiquement le logiciel de maillage. La forme maillée résultante est utilisée par le solveur CFD qui calcule l'écoulement. Ces résultats sont post-traités pour extraire la valeur du ou des objectifs associés aux paramètres. Si l'optimum n'est pas atteint, l'algorithme d'optimisation peut itérer jusqu'à convergence.

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho \cdot u)}{\partial x} + \frac{\partial(\rho \cdot v)}{\partial y} + \frac{\partial(\rho \cdot w)}{\partial z} = 0$$

$$\rho \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} \right) = k_x - \frac{\partial p}{\partial x} + \frac{\partial}{\partial x} \left[\mu \cdot \left(2 \cdot \frac{\partial u}{\partial x} - \frac{2}{3} \cdot (\nabla \cdot \vec{v}) \right) \right] + \frac{\partial}{\partial y} \left[\mu \cdot \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \right] + \frac{\partial}{\partial z} \left[\mu \cdot \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \right]$$

$$\rho \left(\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} \right) = k_y - \frac{\partial p}{\partial y} + \frac{\partial}{\partial y} \left[\mu \cdot \left(2 \cdot \frac{\partial v}{\partial y} - \frac{2}{3} \cdot (\nabla \cdot \vec{v}) \right) \right] + \frac{\partial}{\partial z} \left[\mu \cdot \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) \right] + \frac{\partial}{\partial x} \left[\mu \cdot \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \right]$$

$$\rho \left(\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} \right) = k_z - \frac{\partial p}{\partial z} + \frac{\partial}{\partial z} \left[\mu \cdot \left(2 \cdot \frac{\partial w}{\partial z} - \frac{2}{3} \cdot (\nabla \cdot \vec{v}) \right) \right] + \frac{\partial}{\partial x} \left[\mu \cdot \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) \right] + \frac{\partial}{\partial y} \left[\mu \cdot \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) \right]$$

Figure 1 : Les équations de Navier-Stokes sont des EDP qui décrivent l'écoulement d'un fluide newtonien. Les inconnues sont les trois composantes (u,v,w) du vecteur vitesse, la pression p et la densité .

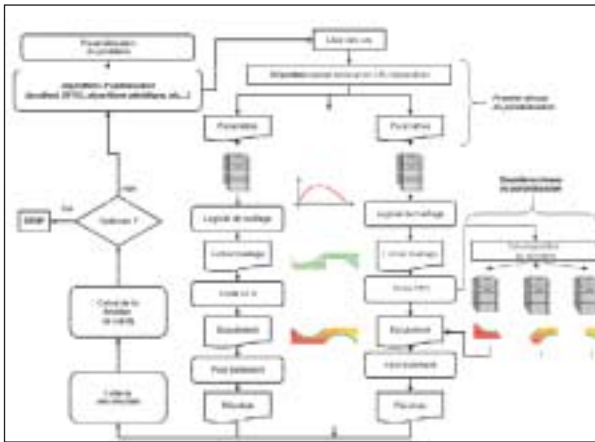


Figure 2 : Couplage entre un algorithme d'optimisation et un code de calcul CFD

Dans le cas de l'évaluation simultanée de l'ensemble d'une population, il est intéressant de répartir les calculs qui sont indépendants entre eux sur l'ensemble des ressources d'une grille de calcul ou des nœuds d'un cluster de PC. Le ratio de parallélisation est alors extrêmement avantageux car la quantité d'information à communiquer entre le processus principal et ses subordonnés étant minime (valeurs des paramètres dans un sens, valeurs des objectifs dans l'autre sens), les temps de communications sont quasi-négligeables devant les temps de calculs. On parle ainsi souvent de parallélisation à deux niveaux : le premier concerne la répartition des calculs à effectuer simultanément, le second le calcul d'un domaine décomposé sur plusieurs ressources.

Choix de l'algorithme d'optimisation

Comme expliqué précédemment, le problème à résoudre lors de l'optimisation de forme par simulation numérique pour le contrôle d'un écoulement présente des caractéristiques spécifiques qui vont guider le choix des algorithmes d'optimisation à utiliser : évaluation des objectifs très longs en terme de temps de calcul, paramétrisation de la forme pouvant mener à des erreurs de maillage et/ou de convergence, problèmes le plus souvent multi-objectifs avec contraintes sur les paramètres. Un panorama non exhaustif des algorithmes disponibles est proposé sur la Figure 3. Les méthodes les plus utilisées en optimisation de forme sont, en l'état actuel les méthodes classiques basées sur des estimations des dérivées de la fonction objectif, et les algorithmes génétiques. Les avantages/inconvénients de ces deux grandes familles sont discutés ici.



Figure 3 : Panorama (non exhaustif) des grandes familles de méthodes d'optimisation

Les méthodes basées sur des méthodes par gradients, ont été appliquées avec un succès limité sur les problèmes couplés à une simulation numérique d'écoulement. En tant que méthode séquentielle, elles offrent l'inconvénient de voir le processus itératif stoppé si une configuration ne renvoie pas de résultat, ce qui, malheureusement, peut s'avérer courant soit à cause d'un problème numérique (maillage automatique non adapté, problème de convergence), soit à cause d'un problème physique comme l'arrêt d'un serveur de calcul. De plus, ce type de méthodes, plutôt adaptées à des problèmes avec un nombre réduit de paramètres, nécessitent souvent, pour fonctionner, de partir d'un point de départ relativement bon, c'est-à-dire assez proche de l'optimum. Enfin, elles présentent le défaut de souvent tomber dans les minima locaux et d'offrir une exploration réduite de l'espace de recherche. Ces méthodes gardent néanmoins leur place dans la boîte à outils de l'ingénieur, d'une part pour leur rapidité de convergence, et d'autre part car, basées sur de solides fondements mathématiques, elles offrent souvent un choix adapté et au problème et au temps disponible, et permettent de bien contrôler le processus de convergence.

Parmi les méthodes qui ne sont basées sur des évaluations du gradient de la fonction objectif, les méthodes dites évolutionnaires simulent l'évolution d'une population d'individus porteurs des paramètres du problème. Les plus connus et les plus couramment utilisés dans l'industrie sont les algorithmes génétiques (AG). Chaque paramètre du problème (x_n) est codé dans un gène, souvent pour les cas qui nous concernent sous forme d'un réel. L'ensemble des paramètres (x_1, \dots, x_n) forme ainsi un chromosome décrivant un individu, qui correspond à un point de l'espace de recherche, c'est à dire à une forme géométrique particulière et donc à un écoulement. Un

ensemble d'individus forme une population. A partir d'une population initiale, les algorithmes de recherche simulent les lois de l'évolution des espèces en réalisant des reproductions entre individus, qui produisent des enfants, par une combinaison de moyenne et de cross-over (échange) sur les chromosomes, c'est-à-dire d'opérations mathématiques simples sur les valeurs des paramètres. Un paramètre de mutation (légère modification aléatoire des valeurs des gènes des enfants) permet d'introduire une notion de diversité et de gérer ainsi le degré d'exploration de l'espace de recherche. Les enfants sont introduits dans la population, les meilleurs individus sélectionnés pour former la nouvelle population. Par meilleur, il faut entendre les individus les mieux adaptés, c'est-à-dire ceux dont les critères d'objectifs sont le plus proches des critères recherchés. Dans le cas du couplage avec une simulation numérique d'écoulement, une simulation correspond *in fine* à l'évaluation des objectifs associés à un individu. C'est de loin, l'étape la plus coûteuse du processus en terme de temps de calcul. Au fur et à mesure des générations, les individus avec les objectifs les moins adaptés (par exemple concrètement, les formes de carrosseries offrant la moins bonne résistance à l'air) sont remplacés par des individus de plus en plus optimaux.

Les AG nécessitent beaucoup plus d'itérations qu'une méthode séquentielle à base de gradient d'une part car ils opèrent sur une population d'individu de taille typique 20-30 dans les cas concernés, et d'autre part pour converger. Mais leurs avantages dans les cas qui nous concernent ici sont multiples. Tout d'abord, du fait qu'ils opèrent sur une population répartie sur l'ensemble de l'espace de recherche, les AG permettent d'éviter la plupart du temps les optima locaux. Ils sont aussi connus pour leur robustesse : si l'évaluation de l'objectif associé à un individu produit une erreur (maillage impossible, non convergence), le processus n'est pas bloqué et l'optimisation peut suivre son cours. Ensuite, du fait que l'évaluation des objectifs des différents individus d'une même génération peut être réalisée en simultanément, les algorithmes génétiques sont facilement parallélisables comme expliqué ci-dessus.

Un autre intérêt dans les cas d'ingénierie est qu'il existe un nombre important d'algorithmes génétiques adaptés aux problèmes multi-objectifs. Leur but est de faire converger la population vers une représentation discrète du front optimal de Pareto, représentant la famille des solutions optimales en vertu des différents critères. Par

exemple, dans le cas de l'optimisation de la forme d'une carrosserie automobile, le front de Pareto pourra être l'ensemble des formes offrant le meilleur compromis poids/résistance à l'air. Les algorithmes génétiques multi-objectifs sont basés sur le concept de domination permettant de classer les individus en vertu des différents objectifs à minimiser/maximiser.

Exemple

L'exemple proposé ici est un cas réalisé lors d'une collaboration entre la société Renuda et le laboratoire de mécanique des Fluides de l'université de Magdebourg en Allemagne (www.uni-magdeburg.de/isut/LSS). On considère ici un échangeur de chaleur composé de tubes au sein desquels circule de l'eau portant la température pariétale des ailettes à une certaine température fixe. De l'air avec une certaine vitesse circule au sein du faisceau de tubes et est refroidi par convection au contact des ailettes. Le but du problème est d'optimiser la forme des tubes afin de garantir un échange maximal de chaleur tout en minimisant la perte de pression entre l'entrée et la sortie du faisceau de tubes.

Pour ce faire, la forme des tubes a été paramétrée par quatre variables géométriques définissant des points qui sont ensuite reliés via une fonction B-spline. Les variables du problème sont quatre nombres et le problème possède deux objectifs concurrents, la variation de température entre l'entrée et la sortie $_T$ et la variation de pression $_P$. En effet, une augmentation de la surface d'échange des ailettes a tendance à améliorer les échanges convectifs mais, en même temps, réduit la section de passage de l'air, augmentant ainsi la perte de charge

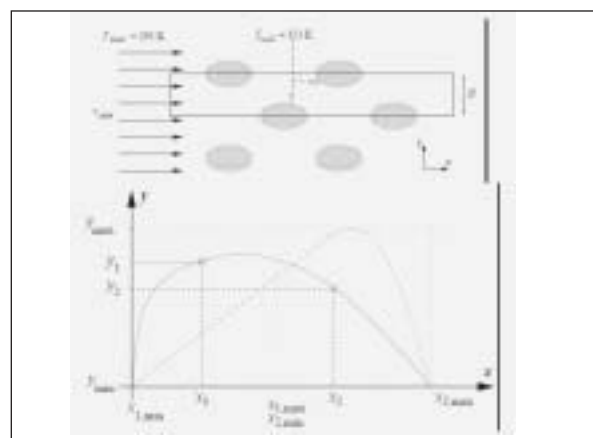


Figure 4 : Configuration d'échangeur de chaleur étudiée. La forme des ailettes de refroidissement est para-

métrée en fonction des coordonnées de deux points (figure de droite) reliés entre eux par une fonction d'interpolation de type B-spline.

Les codes commerciaux Gambit et Fluent ont été utilisés pour paramétrer la génération du maillage et pour résoudre l'écoulement. Un algorithme génétique multi-objectif basé sur le concept de domination de Pareto a été utilisé. Une parallélisation de premier niveau a été utilisée, c'est-à-dire qu'à chaque génération, l'évaluation des objectifs des individus de la population a été répartie sur un ensemble de 16 PC Linux en cluster, le nœud maître se chargeant de l'évolution de la population au cours des itérations. Les résultats se présentent sous la forme d'une approximation du front optimal de Pareto, présenté sur la Figure 5. Ils offrent accès à la gamme des formes optimales qui pourront ensuite être testés expérimentalement

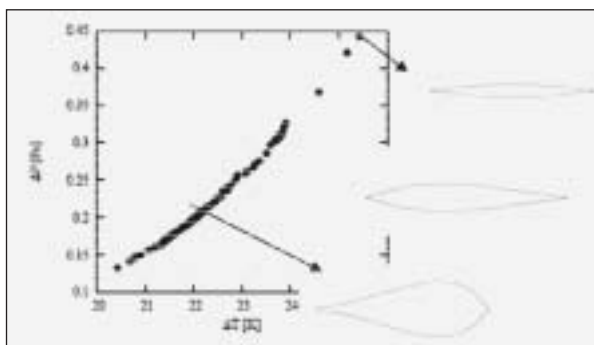


Figure 5: Approximation du front de Pareto et exemples de formes « optimales ».

Conclusions

De nombreux problèmes existent dans le domaine de l'optimisation de forme par simulations numérique d'écoulements, qui n'ont pas été évoqués dans cette présentation introductive. Par exemple, le cas dans lequel l'écoulement est instationnaire, la forme à optimiser peut aussi avoir à varier au cours du temps. On parle alors de contrôle actif d'écoulement et il convient d'envisager dans certains cas des méthodes temps réel avec les problèmes spécifiques associés en terme de modélisation du problème et de rapidité d'exécution. Les pistes envisagées à l'heure actuelle pour améliorer les méthodes évoquées dans cet article sont de deux ordres. D'une part, les efforts peuvent porter sur l'algorithme d'optimisation lui-même. Deux voies potentiellement intéressantes sont envisageables. La première serait de tester les possibilités offertes par les approches

à base d'essaim de particules. La deuxième, plutôt active en ce moment, consiste à envisager des méthodes mixtes algorithmes génétiques/ gradients pour utiliser les avantages des deux approches. D'autre part, les techniques de surfaces de réponse et/ou des approches à base de réseaux de neurones artificiels pourraient se révéler extrêmement intéressantes pour réduire le temps de calcul lié à l'évaluation des objectifs.

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/>

ARTIFICIAL INTELLIGENCE

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

AI V166 N°1 AUGUST 2005

Temporal prepositions and their logic, I. PRATT-HARTMANN

Incremental qualitative temporal reasoning: Algorithms for the Point Algebra and the ORD Horn class, A. GEREVINI

Knowledge and communication: A first-order theory, E. DAVIS

Explorations in engagement for humans and robots, C. L. SIDNER, C. LEE, C. D. KIDD, N. LESH, C. RICH

Unifying tree decompositions for reasoning in graphical models, K. KASK, R. DECHTER, J. LARROSA, A. DECHTER

Processes and continuous change in a SAT-based planner, J.-A. SHIN, E. DAVIS

CHEVRIER, F. CHARPILLET

AI V167 N°1-2 SEPTEMBER 2005

Special issue

Connecting language to the world, D. ROY, E. REITER

Word sense disambiguation with pictures, K. BARNARD, M. JOHNSON

Learning to talk about events from narrated video in a construction grammar framework, P. F. DOMINEY, J.-D. BOUCHER

Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context, J. KELLEHER, F. COSTELLO, J. VAN GENABITH

Protocols from perceptual observations, C. J. NEEDHAM, P. E. SANTOS, D. R. MAGEE, V. DEVIN, D. C. HOGG, A. G. COHN

Choosing words in computer-generated weather forecasts, E. REITER, S. SRIPADA, J. HUNTER, J. YU, I. DAVY

Semiotic schemas: A framework for grounding language in action and perception, D. ROY

The emergence of compositional structures in perceptually grounded language games, P. VOGT

morphologique et phonologique, *M. VOGA-REDLINGER*

AI V168 N°1-2 OCTOBER 2005

Weak AGM postulates and strong Ramsey Test: A logical for-

malization, L. GIORDANO, V. GLIOZZI, N. OLIVETTI
In defense of PDDL axioms, S. THIÉBAUX, J. HOFFMANN, B. NEBEL

Reasoning on UML class diagrams, D. BERARDI, D. CALVANESE, G. DE GIACOMO

Weak nonmonotonic probabilistic logics, T. LUKASIEWICZ
SCC-recursiveness: a general schema for argumentation semantics, P. BARONI, M. GIACOMIN, G. GUIDA

AI V169 N°1 NOVEMBER 2005

An incremental algorithm for generating all minimal models, R. BEN-ELIYAHU - ZOHARY

A theoretical framework on proactive information exchange in agent teamwork, X. FAN, J. YEN, R. A. VOLZ

AI MAGAZINE

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

AI MAGAZINE V26 N°1 SPRING 2005

Semantic Integration, N. F. NOY, A. DOAN, A. Y. HALEVY
Semantic Integration through Invariants, M. GRÜNINGER, J. B. KOPENA

Automatic Ontology Matching Using Application Semantics, A. GAL, G. MODICA, H. JAMIL, A. EYAL

Automatically Utilizing Secondary Sources to Align Information Across Sources, M. MICHALOWSKI, S. THAKKAR, C. A. KNOBLOCK

Semantic Integration in Text: From Ambiguous Names to Identifiable Entities, X. LI, P. MORIE, D. ROTH

Data Integration: A Logic-Based Perspective, D. CALVANESE, G. DE GIACOMO

Ontology Translation for Interoperability Among Semantic Web Services, M. H. BURSTEIN, D. V. MCDERMOTT

Semantic Integration Research in the Database Community: A Brief Survey, A. DOAN, A. Y. HALEVY

AI MAGAZINE V26 N°2 SUMMER 2005

AAAI-05: Twentieth National AI Conference Is a Panoply of Content, S. REESE HEDBERG

Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment, M. E. POLLACK

The 2004 Mobile Robot Competition and Exhibition, W. D. SMART, S. TEJADA, B. MAXWELL, A. STROUPE, J. CASPER, A. JACOFF, H. YANCO, M. BUGAJSKA

RoboCup 2004 Competitions and Symposium: A Small Kick for Robots, a Giant Score for Science, P. LIMA, L. CUSTÓDIO, L. AKIN, A. JACOFF, G. KRAETZSCHMAR, N. B. KIAT, O. OBST, T. RÖFER, Y. TAKAHASHI, C. ZHOU

General Game Playing: Overview of the AAAI Competition, M. GENESERETH, N. LOVE, B. PELL

Description Logics and Planning, Y. GIL

Embodied Communication in Humans and Machines, I. WACHSMUTH, G. KNOBLICH

COGNITIVE SCIENCE

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

APPLIED ARTIFICIAL INTELLIGENCE

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

APPLIED AI V19 N°6 JULY 2005

PA Phrase Grammar-Based Conceptual Indexing Paradigm, P. C. REGHU RAJ, S. RAMAN

Finding Simplified Fuzzy If-Then Rules for Function Approximation Problems Using a Fuzzy Data Mining Approach, Y.-C. HU

Hypertext-Based Representation of Knowledge and Reasoning for User Guidance in Diagnosis of Complex Systems, P. MORIZET-MAHOUDEAUX, C.-C. PAUPE

Software and Performance Measures for Evaluating Multi-Agent Frameworks, D. CAMACHO, R. ALER

APPLIED AI V19 N°7 AUGUST 2005

EEvolving Text Classification Rules with Genetic Programming, L. HIRSCH, M. SAEEDI, R. HIRSCH

ARMGA: Identifying Interesting Association Rules with Genetic Algorithms, X. YAN, C. ZHANG, S. ZHANG

Multilingual, Ontology-Driven, Content-Based Search and Navigation of Information Items, G. A. VOUIROS, E. EUMERIDOU, P. TSELIOS, K. KOTIS

APPLIED AI V19 N°8 SEPTEMBER 2005

Lotteries as a Means of Financing for Preservation of the Global Commons and Agent-Based Simulation Analysis, I. NISHIZAKI, T. SASAKI, Y. UEDA

The Language Component of the Fast Text Prediction System, H. TROST, J. MATIASEK, M. BARONI

Machine Learning in Hybrid Hierarchical and Partial-Order Planners for Manufacturing Domains, S. FERNÁNDEZ, R. ALER, D. BORRAJO

APPLIED AI V19 N°9 NOV-DEC 2005

Toward Learning Grid Infrastructures: An Overview of Research on Grid Learning Services, R. NKAMBOU, G. GOUARDÈRES, B. P. WOOLF

Customizing the Instructional Grid, B. WOOLF, C. ELIOT

Competence Ontology for Domain Knowledge Dissemination and Retrieval, B. LEFEBVRE, G. GAUTHIER, S. TADIÉ, T. HUU DUC, H. ACHABA

Services, Semantics, and Standards: Elements of a Learning Grid Infrastructure, C. ALLISON, S. A. CERRI, P. RITROVATO, A. GAETA, M. GAETA

Collaboration in the Semantic Grid: a Basis for E-Learning, K. R. PAGE, D. T. MICHAELIDES, D. C. DE ROURE, N.

R. SHADBOLT, Y.-H. CHEN-BURGER, J. DALTON, S. POTTER, A. TATE, S. J. BUCKINGHAM SHUM, M. EISENSTADT, M. BACHLER, J. KOMZAK

Ontology-Based User Modeling and E-Portfolio Grid Learning Services, L. RAZMERITA, G. GOUARDÈRES, E. CONTÉ

Interactive Knowledge Construction in the Collaborative Building of an Encyclopedia, P. LEMOISSON, S. A. CERRI

The Strobe Model: Dynamic Service Generation on the Grid, C. JONQUET, S. A. CERRI

Analysis and Synthesis of Learning Agent's Communicative Behavior, N. N. BINTI ABDULLAH, S. A. CERRI

The Grid-E-Card: Architecture to Share Collective Intelligence on the Grid, G. GOUARDÈRES, M. SABER, R. NKAMBOU, R. YATCHOU

APPLIED INTELLIGENCE

<http://www.kluweronline.com/issn/0924-669X/>

APPLIED INTELLIGENCE V23 N°1 2005

Developments in Applied Artificial Intelligence: Guest Editor's Introduction, T. HENDTLASS

Genetic Algorithm Coding Methods for Leather Nesting, A. CRISPIN, P. CLAY, G. TAYLOR, T. BAYES, D. REEDMAN

Genetic Algorithm Optimisation of Mathematical Models Using Distributed Computing, S. DUNN, S. PEUCKER, J. PERRY

Dynamic Ant Colony Optimisation, D. ANGUS, T. HENDTLASS

A Case Based System for Oil and Gas Well Design with Risk Assessment, S. KRAVIS, R. IRRGANG

Evolutionary Multiobjective Design in Automotive Development, M. LAUMANN, N. LAUMANN

APPLIED INTELLIGENCE V23 N°2 2005

Temporal Bayesian Network of Events for Diagnosis and Prediction in Dynamic Domains, G. ARROYO-FIGUEROA, L. E. SUCAR

Temporal Relevance in Dynamic Decision Networks with Sparse Evidence, A. Y. TAWFIK, S. KHAN

Extension of Petri Nets for Representing and Reasoning with Tasks with Imprecise Durations, S. KURKOVSKY, R. LOGANANTHARAJ

A Formalism for Representing and Reasoning with Temporal Information, Event and Change, N. OBEID

Towards a Logic of Perishable Propositions, F. SOARES CORREA DA SILVA

COMPUTATIONAL INTELLIGENCE

<http://www.blackwellpublishing.com/journal.asp?ref=0824-7935&site=1>

SOMMAIRES DES REVUES

COMPUTATIONAL INTELLIGENCE V21 N°3 AUGUST 2005

- Using Reinforcement Learning to Coordinate Better, C. B. EXCELENTE-TOLEDO, N. R. JENNINGS
- Automatically Generating Tree Adjoining Grammars from Abstract Specifications, F. XIA, M. PALMER, K. VIJAYSHANKER
- Preprocessing Rules for Triangulation of Probabilistic Networks*, H. L. BODLAENDER, A. M.C.A. KOSTER, F. VAN DEN EIJKHOF
- Hybrid Ace: Combining Search Directions for Heuristic Planning, D. VRAKAS, I. VLAHAVAS

COMPUTATIONAL INTELLIGENCE V21 N°4 NOVEMBER 2005

- Learning to Improve Reasoning, M. AFZAL UPAL, S. ROGERS
- Learning to Support Constraint Programmers, S. L. EPSTEIN, E. C. FREUDER, R. J. WALLACE
- Applying Machine Learning to Low-Knowledge Control of Optimization Algorithms, T. CARCHRAE, J. C. BECK
- Learning Preconditions for Planning from Plan Traces and HTN Structure, O. ILGHAMI, D. S. NAU, H. MUNOZ-AVILA, D. W. AHA
- Incremental Learning of Procedural Planning Knowledge in Challenging Environments, D. J. PEARSON, J. E. LAIRD
- Learning to Improve Plan Quality, M. AFZAL UPAL
- The Disciple-RKF Learning and Reasoning Agent, G. TECUCI, M. BOICU, C. BOICU, D. MARCU, B. STANESCU, M. BARBULESCU
- A Constrained Architecture for Learning and Problem Solving, R. M. JONES, P. LANGLEY

MINDS AND MACHINES

<http://www.springeronline.com/sgw/cda/frontpage/0,0,5-0-70-35534114-0,0.html>

MINDS AND MACHINES V15 N°2 2005

- An Active Symbols Theory of Chess Intuition, A. LINHARES
- Is There Life After the Death of the Computational Theory of Mind?, J. HERSHFIELD
- Of Robots and Believing, C.T.A. SCHMIDT
- Computation and Intentionality: A Recipe for Epistemic Impasse, I. SHANI
- Canonical Naming Systems, L. HORSTEN
- Book Reviews
- T. CRANE, *The Mechanical Mind*, 2nd edition, London and New York: Routledge, 2003, by J. KNOWLES
- R. CORDESCHI, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Dordrecht, The Netherlands: Kluwer Academic

Publishers, 2002, by S. BEGEER

- D. K. SIMONTON, *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*, Cambridge: Cambridge University Press, 2004, by M. L. SCHAGRIN

ARTIFICIAL INTELLIGENCE REVIEW

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

AI REVIEW V24 N°2 2005

- Introduction to the Special Issue on Explanation in Case-Based Reasoning, D. LEAKE, D. MCSHERRY
- Explanation in Case-Based Reasoning-Perspectives and Goals, F. SØRMO, J. CASSENS, A. AAMODT
- The Explanatory Power of Symbolic Similarity in Case-Based Reasoning, E. PLAZA, E. ARMENGOL, S. ONTANON
- A Case-Based Explanation System for Black-Box Systems, C. NUGENT, P. CUNNINGHAM
- Explanation in Recommender Systems, D. MCSHERRY
- Explaining Compound Critiques, J. REILLY, K. MCCARTHY, L. MCGINTY, B. SMYTH

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

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

IJHCS V62 N°4 APRIL 2005

- Time-related behaviour in multitasking situations, Y. ZHANG, R. S. GOONETILLEKE, T. PLOCHER, S.-F. M. LIANG
- Supporting the cognitive process of user interface design with reusable design cases, H. KIM, W. C. YOON
- An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices, L. BARNARD, J. S. YI, J. A. JACKO, A. SEARS
- On the passage of time: Temporal differences in video-mediated and face-to-face interaction, R. VAN DER KLEIJ, R. PAA-SHUIS, J. M. SCHRAAGEN

IJHCS V62 N°5 MAY 2005

Special issue

- Protege: Community is Everything, M.A. MUSEN
- Cognitive support for ontology modelling, N.A. ERNST, M.-A. STOREY, P. ALLEN
- EZPAL: Environment for composing constraint axioms by instantiating Templates, C.S.J. HOU, M.A. MUSEN, N.F. NOY
- Experiences in reusing knowledge sources using Protege and PROMPT, M. TABOADA, D. MARTINEZ, J. MIRA
- Knowledge editing and maintenance tools for a semantic portal in oncology, M. D'AQUIN, C. BOUTHIER, S. BRACHAIS, J. LIEBER, A. NAPOLI
- Protege as a vehicle for developing medical terminological systems, A. ABU-HANNA, R. CORNET, N. DE KEIZER, M. CRUBEZY, S. TU

SOMMAIRES DES REVUES

GAS Ontology: An ontology for collaboration among ubiquitous computing devices, E. CHRISTOPOULOU, A. KAMEAS

IJHCS V62 N°6 JUNE 2005

GUI objects with impenetrable borders: Instruction (not practice) makes perfect, K.S. JONES, B.R. JOHNSON, J.S. FARRIS

Browsing a document collection represented in two- and three-dimensional virtual information space, S.J. WESTERMAN, J. COLLINS, T. CRIBBIN

An empirical investigation of sources of application-specific computer-self-efficacy and mediators of the efficacy-performance relationship, R.D. JOHNSON

Focusing on what might happen and how it could feel: can the anticipation of regret change students' computing-related choices?, C. WRIGHT, P. AYTON

An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study, I.-L. WU, J.-L. CHEN

IJHCS V63 N°1-2 JULY 2005

Special issue: HCI research in privacy and security

Why HCI research in privacy and security is critical now, C. KARAT, J. KARAT, C. BRODIE

In the eye of the beholder: A visualization-based approach to information system security, R. DE PAULA, X. DING, P. DOURISH, K. NIES, B. PILLET, D.F. REDMILES, J. REN, J.A. RODE, R. SILVA FILHO

Improving user-interface dependability through mitigation of human error, R.A. MAXION, R.W. REEDER

Security and usability engineering with particular attention to electronic mail, V. ROTH, T. STRAUB, K. RICHTER

Mechanisms for increasing the usability of grid security, B. BECKLES, V. WELCH, J. BASNEY

PassPoints: Design and longitudinal evaluation of a graphical password system, S. WIEDENBECK, J. WATERS, J.-C. BIRGET, A. BRODSKIY, N. MEMON

Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems, A. DE ANGELI, L. COVENTRY, G. JOHNSON, K. RENAUD

Privacy in information technology: Designing to enable privacy policy management in organizations, J. KARAT, C.-M. KARAT, C. BRODIE, J. FENG

Bridging the gap between organizational and user perspectives of security in the clinical domain, A. ADAMS, A. BLANDFORD

Privacy practices of Internet users: Self-reports versus observed behaviour, C. JENSEN, C. POTTS, C. JENSEN

Keeping ubiquitous computing to yourself: A practical model for user control of privacy, B.A. PRICE, K. ADAM, B. NUSEIBEH

Public space systems: Designing for privacy?, L. LITTLE, P. BRIGGS, L. COVENTRY

IJHCS V63 N°4-5 OCTOBER 2005

Special issue

Computer support for creativity, E.A. EDMONDS, L. CANDY

How can computers be partners in the creative process: Classification and commentary on the Special Issue, T. LUBART

Creativity or creativities?, R.J. STERNBERG

Informing the design of computer-based environments to support creativity, T.T. HEWETT

Fostering motivation and creativity for computer users, T. SELKER

Towards supporting evocation processes in creative design: A cognitive approach, N. BONNARDEL, E. MARMECHE

Developing creativity, motivation, and self-actualization with learning systems, W. BURLESON

The studio as laboratory: Combining creative practice and digital technology research, E.A. EDMONDS, A. WEAKLEY, L. CANDY, M. FELL, R. KNOTT, S. PAULETTO

Beyond binary choices: Integrating individual and social creativity, G. FISCHER, E. GIACCARDI, H. EDEN, M. SUGIMOTO, Y. YE

Interaction design of tools for fostering creativity in the early stages of information design, Y. YAMAMOTO, K. NAKAKOJI

IJHCS V63 N°6 DECEMBER 2005

Age differences and the acquisition of spatial knowledge in a three-dimensional environment: Evaluating the use of an overview map as a navigation aid, M. SJOLINDER, K. HOOK, L.-G. NILSSON, G. ANDERSSON

Ecological interface design and computer network management: The effects of network size and fault frequency, P. DUEZ, K.J. VICENTE

Reducing cognitive workload of a computer-based procedure system, Y.-L. LEE, S.-L. HWANG, E. MIN-YANG WANG

Evaluation of integrated software development environments: Challenges and results from three empirical studies, R.B. KLINE, A. SEFFAH

COMPUTATIONAL LINGUISTICS

<http://mitpress.mit.edu/catalog/item/default.asp?sid=8563C099-9701-4DD2-85C88F3502E9C8AE&tttype=4&tid=10>

COMPUTATIONAL LINGUISTICS V31 N°2 JUNE 2005

A General Technique to Train Language Models on Language Models, M.-J. NEDERHOF

Tree-Local Multicomponent Tree Adjoining Grammars with Shared Nodes, L. KALLMEYER

A Mathematical Model of Historical Semantics and the

SOMMAIRES DES REVUES

Grouping of Word-Meanings into Concepts, M. C. COOPER

Representing Discourse Coherence: A Corpus-Based Study, F. WOLF, E. GIBSON

NATURAL LANGUAGE ENGINEERING

<http://journals.cambridge.org/action/displayJournal?jid=NLE>

NLE V11 N°2 JUNE 2005

The head-modifier principle and multilingual term extraction, A. HIPPISEY, D. CHENG, K. AHMAD

Finite-state multimodal integration and understanding, M. JOHNSTON, S. BANGALORE

Machine learning-based named entity recognition via effective integration of various evidences, G. ZHOU, J. SU

The Penn Chinese TreeBank: Phrase structure annotation of a large corpus, N. XUE, F. XIA, F.-D. CHIOU, M. PALMER

NLE V11 N°3 SEPTEMBRE 2005

Special issue

Parallel texts, R. MIHALCEA, M. SIMARD

Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus, L. BENTIVOGLI, E. PIANTA

Constrained EM for parallel text alignment, D. TALBOT

Optimization of word alignment clues, J. TIEDEMANN

Comparing example-based and statistical machine translation, A. WAY, N. GOUGH

Bootstrapping parsers via syntactic projection across parallel texts, R. HWA, P. RESNIK, A. WEINBERG, C. CABEZAS, O. KOLAK

Automatic bilingual lexicon acquisition using random indexing of parallel corpora, M. SAHLGREN, J. KARL-GREN

NLE V11 N°4 DECEMBER 2005

Implementing clarification dialogues in open domain question answering, M. DE BONI, S. MANANDHAR

Segmenting documents by stylistic character, N. GRAHAM, G. HIRST, B. MARTHI

A decision-theoretic framework for the evaluation of language models used in speech recognizers, J. R. DELLER, K. H. DESAI, Y. P. YANG

Paraphrasing spoken Chinese using a paraphrase corpus, Y. ZHANG, K. YAMAMOTO

Book review

R. MITKOV, Anaphora Resolution. Longman. 2002, by J. PREISS

M. RAYNER, D. CARTER, P. BOUILLON, V. DIGALAKIS, M. WIREN, editors. The Spoken Language Translator. Cambridge University Press. 2000, by M. OSBORNE

USER MODELING AND USER-ADAPTED INTERACTION

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

UMUAI V15 N°1 MARCH 2005

Special issue

Preface on Special Issue on Language-Based Interaction, S. CARBERRY, I. ZUKERMAN

A Probabilistic Approach for Argument Interpretation, I. ZUKERMAN, S. GEORGE

A Model to Disambiguate Natural Language Parses on the Basis of User Language Proficiency: Design and Evaluation, L. N. MICHAUD, K. F. MCCOY, R. Z. DAVIS

Using Dialogue Features to Predict Trouble During Collaborative Learning, B. A. GOODMAN, F. N. LINTON, R. D. GAIMARI, J. M. HITZEMAN, H. J. ROSS, G. ZARRELLA

Tailoring Automatically Generated Hypertext, K. BONT-CHEVA, Y. WILKS

User Modeling in Spoken Dialogue Systems to Generate Flexible Guidance, K. KOMATANI, S. UENO, T. KAWAHARA, H. G. OKUNO

UMUAI V15 N°3-4 AUGUST 2005

Special issue

Preface to the Special Issue on User Modeling in Ubiquitous Computing, A. JAMESON, A. KRÜGER

Consistent Modelling of Users, Devices and Sensors in a Ubiquitous Computing Environment, D. J. CARMICHAEL, J. KAY, B. KUMMERFELD

Exploring Issues of User Model Transparency and Proactive Behaviour in an Office Environment Control System, K. CHEVERST, H. E. BYUN, D. FITTON, C. SAS, C. KRAY, N. VILLAR

Personalization and Context Management, A. ZIMMERMANN, M. SPECHT, A. LORENZ

User-Centred Design of Flexible Hypermedia for a Mobile Guide: Reflections on the HyperAudio Experience, D. PETRELLI, E. NOT

Ontology-Based User Modeling in an Augmented Audio Reality System for Museums, M. HATALA, R. WAKKARY

COMPUTER SPEECH AND LANGUAGE

http://www.sciencedirect.com/science?_ob=JournalURL&_issn=08852308&_auth=y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=be0614a6a8826664cf3538182118628

COMPUTER SPEECH & LANGUAGE V19 N°4 OCTOBER 2005

Special issue on Multiword Expression

SOMMAIRES DES REVUES

Introduction to the special issue on multiword expressions: Having a crack at a hard nut, A. VILLAVICENCIO, F. BOND, A. KORHONEN, D. MCCARTHY

Comparing and combining a semantic tagger and a statistical tool for MWE extraction, S. S. PIAO, P. RAYSON, D. ARCHER, T. MCENERY

Deep lexical acquisition of verb-particle constructions, T. BALDWIN

The availability of verb-particle constructions in lexical resources: How much is enough?, A. VILLAVICENCIO

Multiword expressions in spoken language: An exploratory study on pronunciation variation, D. BINNENPOORTE, C. CUCCHIARINI, L. BOVES, H. STRIK

Using small random samples for the manual evaluation of statistical association measures, S. EVERT, B. KRENN

Learning about the meaning of verb-particle constructions from corpora, C. BANNARD

On the semantics of noun compounds, R. GIRJU, D. MOLDOVAN, M. TATU, D. ANTOHE

Disambiguating Japanese compound verbs, K. UCHIYAMA, T. BALDWIN, S. ISHIZAKI

Creative discovery in the lexical "validation gap", J. HAYES, N. SECO, T. VEALE

A symbolic approach to automatic multiword term structuring, E. SANJUAN, J. DOWDALL, F. IBEKWE-SANJUAN, F. RINALDI

COMPUTER SPEECH & LANGUAGE V20 N°1 JANUARY 2005

Multiple resolution analysis for robust automatic speech recognition, R. GEMELLO, F. MANA, D. ALBESANO, R. DE MORI

Product of Gaussians for speech recognition, M.J.F. GALES, S.S. AIREY

MAP adaptation of stochastic grammars, M. BACCHIANI, M. RILEY, B. ROARK, R. SPROAT

A prosodic phrasing model for a Korean text-to-speech synthesis system, K. YOON

The architecture and the implementation of a finite state pronunciation lexicon for Turkish, K. OFLAZER, S. INKELAS

Improved automatic speech recognition through speaker normalization, D. GIULIANI, M. GEROSA, F. BRUGNARA

MACHINE LEARNING

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

MACHINE LEARNING V59 N°1-2 2005

Evolving Soccer Keepaway Players Through Task Decomposition, S. WHITESON, N. KOHL, R. MIIKKULAINEN, P. STONE

A Reinforcement Learning Scheme for a Partially-Observable Multi-Agent Game, S. ISHII, H. FUJITA, M. MITSUTAKE,

T. YAMAZAKI, J. MATSUDA, Y. MATSUNO

PAC-Bayesian Compression Bounds on the Prediction Error of Learning Algorithms for Classification, T. GRAEPEL, R. HERBRICH, J. SHAWE-TAYLOR

Multicategory Proximal Support Vector Machine Classifiers, G. M. FUNG, O. L. MANGASARIAN

Maximizing Agreements with One-Sided Error with Applications to Heuristic Learning, N. H. BSHOUTY, L. BURROUGHS

Internal Regret in On-Line Portfolio Selection, G. STOLTZ, G. LUGOSI

Logistic Model Trees, N. LANDWEHR, M. HALL, E. FRANK

MACHINE LEARNING V59 N°3 2005

Learning Bayesian Network Classifiers: Searching in a Space of Partially Directed Acyclic Graphs, S. ACID, L. M. DE CAMPOS, J. G. CASTELLANO

Latent Classification Models, H. LANGSETH, T. D. NIELSEN

On Discriminative Bayesian Network Classifiers and Logistic Regression, T. ROOS, H. WETTIG, P. GRUENWALD, P. MYLLYMAEKI, H. TIRRI

Structural Extension to Logistic Regression: Discriminative Parameter Learning of Belief Net Classifiers, R. GREINER, X. SU, B. SHEN, W. ZHOU

TAN Classifiers Based on Decomposable Distributions, J. CERQUIDES, R. LOPEZ DE MANTARAS

MACHINE LEARNING V60 N°1-3 2005

Introduction: Machine Learning in Speech and Language Technologies, P. FUNG, D. ROTH

Support Vector Learning for Semantic Argument Classification, S. PRADHAN, K. HACIOGLU, V. KRUGLER, W. WARD, J. H. MARTIN, D. JURAFSKY

Filtering-Ranking Perceptron Learning for Partial Parsing, A. XAVIER CARRERAS, B. LLUIS MARQUEZ, C. JORGE CASTRO

Ranking and Reranking with Perceptron, L. SHEN, A. K. JOSHI

Online Multiclass Learning with k-Way Limited Feedback and an Application to Utterance Classification, H. ALSHAWI

Moment Kernels for Regular Distributions, C. CORTES, M. MOHRI

Maximum Entropy Modeling: A Suitable Framework to Learn Context-Dependent Lexicon Models for Statistical Machine Translation: Basic Instructions, I. GARCIA-VAREA, F. CASACUBERTA

Maximum Entropy Models with Inequality Constraints: A Case Study on Text Categorization, J. KAZAMA, J. TSUJII

A Neural Syntactic Language Model, A. EMAMI, F. JELINEK

Combining Statistical Language Models via the Latent

Maximum Entropy Principle, S. WANG, D. SCHUURMANS, F. PENG, Y. ZHAO

Corpus-based Learning of Analogies and Semantic Relations, P. D. TURNEY, M. L. LITTMAN

NEURAL NETWORKS

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

NEURAL NETWORKS V18 N°4 MAY 2005

Special issue: Emotion and Brain

Emotion and brain: Understanding emotions and modelling their recognition, J.G. TAYLOR, K. SCHERER, R. COWIE

A systems approach to appraisal mechanisms in emotion, D. SANDER, D. GRANDJEAN, K.R. SCHERER

The interaction of attention and emotion, J.G. TAYLOR, N.F. FRAGOPANAGOS

Beyond emotion archetypes: Databases for emotion modelling using neural networks, R. COWIE, E. DOUGLAS-COWIE, C. COX

Emotion recognition in human-computer interaction, N. FRAGOPANAGOS, J.G. TAYLOR

Challenges in real-life emotion annotation and machine learning based detection, L. DEVILLERS, L. VIDRASCU, L. LAMEL

Emotion recognition through facial expression analysis based on a neurofuzzy network, S.V. IOANNOU, A.T. RAOUZAIYOU, V.A. TZOUVARAS, T.P. MAILIS, K.C. KARPOUZIS, S.D. KOLLIAS

ASR for emotional speech: Clarifying the issues and enhancing performance, T. ATHANASELIS, S. BAKAMIDIS, I. DOLOGLOU, R. COWIE, E. DOUGLAS-COWIE, C. COX
Emotion understanding from the perspective of autonomous robots research, L. CANAMERO

NEURAL NETWORKS V18 N°5-6 JUNE 2005

Special issue

IJCNN 2005, Edited by: D. PROKHOROV, D. LEVINE, F. HAM, W. HOWELL

A model of STDP based on spatially and temporally local information: Derivation and combination with gated decay, A. GORCHETCHNIKOV, M. VERSACE, M.E. HASSELMO

Stochastic resonance in noisy spiking retinal and sensory neuron models, A. PATEL, B. KOSKO

Quantifying information and performance for flash detection in the blowfly photoreceptor, P. XU, P. ABSHIRE

Optimizing conductance parameters of cortical neural models via electrotonic partitions, K. BUSH, J. KNIGHT, C. ANDERSON

Metastability, instability, and state transition in neocortex, W.J. FREEMAN, M.D. HOLMES

A mixed analog/digital chaotic neuro-computer system for quadratic assignment problems, Y. HORIO, T. IKEGUCHI, K. AIHARA

FPGA implementation of self organizing map with digital phase locked loops, H. HIKAWA

Rich dynamics of pulse-coupled spiking neurons with a triangular base signal, Y. KON'NO, T. SAITO, H. TORIKAI

Non-homogenous neural networks with chaotic recursive nodes: Connectivity and multi-assemblies structures in recursive processing elements architectures, E. DEL MORAL HERNANDEZ

A hierarchical classifier using new support vector machines for automatic target recognition, D. CASASENT, Y.C. WANG
Neural network model for extracting optic flow, K. TOHYAMA, K. FUKUSHIMA

On-chip visual perception of motion: A bio-inspired connectionist model on FPGA, C. TORRES-HUITZIL, B. GIRAU, C. CASTELLANOS-SANCHEZ

Interactive image data labeling using self-organizing maps in an augmented reality scenario, H. BEKEL, G. HEIDEMANN, H. RITTER

Incremental learning of feature space and classifier for face recognition, S. OZAWA, S.L. TOH, S. ABE, S. PANG, N. KASABOV

Generalized 2D principal component analysis for face image representation and recognition, H. KONG, L. WANG, E.K. TEOH, X. LI, J.G. WANG, R. VENKATESWARLU

Training neural networks with heterogeneous data, J.A. DRAKOPOULOS, A. ABDULKADER

Framewise phoneme classification with bidirectional LSTM and other neural network architectures, A. GRAVES, J. SCHMIDHUBER

Subfield variations in hippocampal processing-components of a spatial navigation system, M. HARTLEY, N. TAYLOR, J. TAYLOR

Modelling divided visual attention with a winner-take-all network, D.I. STANDAGE, T.P. TRAPPENBERG, R.M. KLEIN

Dynamical systems and cognitive linguistics: toward an active morphodynamical semantics, R. DOURSAT, J. PETITOT

Neural dynamics of psychotherapy: what modeling might tell us about us, A.M.C. ALEKSANDROWICZ, D.S. LEVINE

Model of multi-modal cortical processing: Coherent learning in self-organizing modules, O. MENARD, H. FREZZA-BUET

The contribution of active body movement to visual development in evolutionary robots, M. SUZUKI, D. FLOREANO, E.A. DI PAOLO

Application of neural network to humanoid robots-development of co-associative memory model, K. ITOH, H. MIWA, H. TAKANOBU, A. TAKANISHI

Constructing Bayesian formulations of sparse kernel learning methods, G.C. CAWLEY, N.L.C. TALBOT

Handling missing values in support vector machine classifiers, K. PELCKMANS, J. DE BRABANTER, J.A.K. SUYKENS, B. DE MOOR

Bayesian approach to feature selection and parameter tuning

for support vector machine classifiers, C. GOLD, A. HOLUB, P. SOLLICH

Improving dimensionality reduction with spectral gradient descent, R. MEMISEVIC, G. HINTON

Unifying cost and information in information-theoretic competitive learning, R. KAMIMURA

A new classifier based on information theoretic learning with unlabeled data, K.H. JEONG, J.W. XU, D. ERDOGMUS, J.C. PRINCIPE

Cross-entropy embedding of high-dimensional data using the neural gas model, P.A. ESTEVEZ, C.J. FIGUEROA, K. SAITO

A model of evolution and learning, V.G. RED'KO, O.P. MOSALOV, D.V. PROKHOROV

A regenerating spiking neural network, D. FEDERICI

Recurrent neural networks with backtrack-points and negative reinforcement applied to cost-based abduction, A.M. ABDELBAR, M.A. EL-HEMALY, E.A.M. ANDREWS, D.C. WUNSCH

Broadcast scheduling in wireless multihop networks using a neural-network-based hybrid algorithm, H. SHI, L. WANG

The control of a parallel hybrid-electric propulsion system for a small unmanned aerial vehicle using a CMAC neural network, F.G. HARMON, A.A. FRANK, S.S. JOSHI

A comparative study of autoregressive neural network hybrids, T. TASKAYA-TEMIZEL, M.C. CASEY

Efficient streaming text clustering, S. ZHONG

Efficient information theoretic strategies for classifier combination, feature extraction and performance evaluation in improving false positives and false negatives for spam e-mail filtering, V. ZORKADIS, D.A. KARRAS, M. PANAYOTOU
Modelling ordinal relations with SVMs: An application to objective aesthetic evaluation of breast cancer conservative treatment, J.S. CARDOSO, J.F. PINTO DA COSTA, M.J. CARDOSO

Ovarian cancer diagnosis by hippocampus and neocortex-inspired learning memory structures, T.Z. TAN, C. QUEK, G.S. NG

Individualization of pharmacological anemia management using reinforcement learning, A.E. GAWEDA, M.K. MUEZZINOGLU, G.R. ARONOFF, A.A. JACOBS, J.M. ZURADA, M.E. BRIER

Identification of motifs with insertions and deletions in protein sequences using self-organizing neural networks, D. LIU, X. XIONG, Z.G. HOU, B. DASGUPTA

Bayesian model selection for mining mass spectrometry data, A. SAKSENA, D. LUCARELLI, I.J. WANG

Refining competition in the self-organising tree map for supervised biofilm image segmentation, M. KYAN, L. GUAN, S. LISS

ARTIFICIAL INTELLIGENCE IN MEDICINE

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

AI IN MEDICINE V34 N°3 JULY 2005

Part-whole representation and reasoning in formal biomedical ontologies, S. SCHULZ, U. HAHN

Description logic-based methods for auditing frame-based medical terminological systems, R. CORNET, A. ABU-HANNA

An expert study evaluating the UMLS lexical metaschema, L. ZHANG, G. HRIPCSAK, Y. PERL, M. HALPER, J. GELLER

Logical analysis of diffuse large B-cell lymphomas, G. ALEXE, S. ALEXE, D.E. AXELROD, P.L. HAMMER, D. WEISSMANN

Factors affecting automated syndromic surveillance, L. WANG, M.F. RAMONI, K.D. MANDL, P. SEBASTIANI

An optimized experimental protocol based on neuro-evolutionary algorithms, M. BUSCEMA, E. GROSSI, M. INTRALIGI, N. GARBAGNA, A. ANDRIULLI, M. BREDA

AI IN MEDICINE V35 N°1-2 SEPTEMBER 2005

Special issue

Computational Intelligence Techniques in Bioinformatics, K. J. CIOS, H. MAMITSUKA, T. NAGASHIMA, R. TADEUSIEWICZ

Computational intelligence in solving bioinformatics problems, K.J. CIOS, H. MAMITSUKA, T. NAGASHIMA, R. TADEUSIEWICZ

Finding the biologically optimal alignment of multiple sequences, H. MAMITSUKA

Highly accurate and consistent method for prediction of helix and strand content from primary protein sequences, J. RUAN, K. WANG, J. YANG, L.A. KURGAN, K. CIOS

Discovering reliable protein interactions from high-throughput experimental data using network topology, J. CHEN, W. HSU, M.L. LEE, S.K. NG

SiMCAL 1 algorithm for analysis of gene expression data related to the phosphatidylserine receptor, D. DVORKIN, V. FADOK, K. CIOS

Efficient RNAi-based gene family knockdown via set cover optimization, W. ZHAO, M.L. FANNING, T. LANE

A model of motor control of the nematode *C. elegans* with neuronal circuits, M. SUZUKI, T. TSUJI, H. OHTAKE

Fuzzy attributes of a DNA complex: Development of a fuzzy inference engine for codon-"junk" codon delineation, T.V. ARREDONDO, P.S. NEELAKANTA, D. DE GROFF

Computational modeling of oligonucleotide positional densities for human promoter prediction, V. NARANG, W.K. SUNG, A. MITTAL

Granular support vector machines with association rules mining for protein homology prediction, Y. TANG, B. JIN, Y.Q. ZHANG

Application of tabu search strategy for finding low energy structure of protein, J. BLAZEWICZ, P. LUKASIAK, M. MILOSTAN

Prediction of MHC class II binders using the ant colony search strategy, O. KARPENKO, J. SHI, Y. DAI

A landmark extraction method for protein 2DE gel images based on multi-dimensional clustering, J.E. SHIM, W.S. LEE

Knowledge-assisted recognition of cluster boundaries in gene expression data, Y. OKADA, T. SAHARA, H. MITSUBAYASHI, S. OHGIYA, T. NAGASHIMA

Gene extraction for cancer diagnosis by support vector machines-An improvement, T.M. HUANG, V. KECCMAN

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

<http://www.elsevier.com/inca/publications/store/5/0/5/7/8/7/>

IJAR V40 N°3 NOVEMBER 2005

A generalization of local divergence measures, C. BERTOLUZZA, P. MIRANDA, P. GIL

An uncertainty interchange format with imprecise probabilities, P. BARONI, P. VICIG

Belief functions on real numbers, P. SMETS

H[∞]-estimation for fuzzy membership function optimization, D. SIMON

Dynamic projection network for supervised pattern classification, C.J. LI, C. JANSUWAN

On two types of discrete implications, M. MAS, M. MONSERRAT, J. TORRENS

Soft computing optimization methods applied to logistic processes, C.A. SILVA, J.M.C. SOUSA, T. RUNKLER, R. PALM

A new approach to fuzzy wavelet system modelling, E. KARATEPE, M. ALCI

LES DOSSIERS DU *BULLETIN DE L'AFIA*

Raisonnement à partir de cas.....	Bulletin n°59
Multimédia et I.A.....	Bulletin n°58
Planification et heuristiques.....	Bulletin n°56/57
Systèmes d'Information.....	Bulletin n°55
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 connexionnisme.....	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 connexionnisme.....	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

Dossier « Intelligence Artificielle et Entreprises »

Coordonnateur :

Gérald Petitjean (AXLOG, gerald.petitjean@axlog.fr)

Objectifs

Nous préparons actuellement un dossier sur les entreprises et l'Intelligence Artificielle (sociétés de services, de conseil, éditeurs de logiciels ou de composants, services R&D ou « utilisateurs » de grands groupes faisant appel à des méthodes et technologies issues de l'Intelligence Artificielle et de l'Aide à la Décision).

Toutes les entreprises (sociétés de services, de conseil, éditeurs de logiciels ou de composants, services ou départements de grands groupes) utilisant des techniques issues de l'Intelligence Artificielle et de l'Aide à la Décision, développant des outils reposant sur ces techniques, ou menant des activités de R&D sur ces thématiques, peuvent participer à ce dossier en nous faisant parvenir un descriptif. Ce descriptif devra être fourni sous la forme d'un document Word de trois pages au maximum et pourra contenir par exemple :

- des informations générales :

- nom de la société (et du service éventuel)
- Contacts :
Adresse postale
Numéros de téléphone / fax
Site Internet

E-mails et noms des personnes à contacter

- une présentation succincte de la société ou du service (historique, secteur(s) d'activité, clients, effectif, ...)
- une description des activités liées à l'Intelligence Artificielle et à l'Aide à la Décision :
 - Produits vendus et/ou services offerts
 - Applications développées
 - Technologies d'Intelligence Artificielle et/ou d'Aide à la Décision utilisées
 - Exemples de problématiques, de projets, d'applications et/ou de clients liés à l'Intelligence Artificielle et à l'Aide à la Décision
- la liste des partenariats avec des laboratoires universitaires, des écoles, ...

Voici une liste (non exhaustive) de techniques et de thèmes concernés par ce dossier :

Optimisation, Aide à la décision, Planification, Classification, Systèmes-experts, Auto-adaptation, Apprentissage, Analyse de données, Fouille de données, Extraction de connaissances, Ingénierie des connaissances, Traitement du langage naturel, Traitement de la parole, Traitement de l'image, Reconnaissance de formes, Interface et Communication Homme-Machine, Web sémantique, Ontologies, Robotique, Systèmes et agents autonomes, Systèmes Multi-Agents, Simulation, Programmation Par Contraintes, Programmation Linéaire, Recherche Opérationnelle, Méta-heuristiques, Algorithmes évolutionnaires (algorithmes génétiques, colonies de fourmis, essais de particules), Emergence, Apprentissage par renforcement, Réseaux de neurones, Réseaux bayésiens, Arbres de décision, Logique Floue, Systèmes à base de règles, Raisonnement à partir de cas, Programmation Logique, Logique formelle, ...

Dans le cas où vous souhaiteriez figurer dans ce dossier, merci de me faire parvenir une réponse avant le 15 mai 2006. Vous pouvez envoyer votre réponse à mon adresse e-mail professionnelle (gerald.petitjean@axlog.fr).

Je suis à votre disposition pour toute information complémentaire.

Du coté des éditeurs ...

Voici quelques livres que des éditeurs proposent aux membres de l'AFIA de reviewer.

Si vous souhaitez effectuer une recension d'un de ces ouvrages, envoyez un email avec le titre, les auteurs et l'éditeur du livre que vous souhaitez reviewer, ainsi que vos coordonnées postales à philippe.morignot@axlog.fr : nous vous enverrons ce livre (que vous garderez pour prix de votre recension). Généralement, la recension sera envoyée aux auteurs du livre, afin qu'il puisse répondre à la recension, s'ils le souhaitent (la recension et la réponse s'il y en a une seront publiées dans le Bulletin).

1. Jacques Richalet, Guy Lavielle et Joëlle Mallet. La commande prédictive. Eyrolles, Paris, décembre 2004, 256 pages.
2. Patrick Naïm, Pierre-Henri Willemin, Philippe Leray, Olivier Pourret et Anna Becker. Réseaux bayésiens. Eyrolles, Paris, mai 2004, 224 pages.
3. Gérard Dreyfus, Manuel Samuelides, Jean-Marc Martinez, Mirta B. Gordon, Fouad Badran, Sylvie Thiria et Laurent Hérault. Réseaux de neurones. Eyrolles, Paris, Avril 2004, avril 2004, 408 pages.
4. Philippe Lacomme, Christian Prins et Marc Sevaux. Algorithmes de graphes. Eyrolles, Paris, octobre 2003, 422 pages.
5. Johann Dréo, Alain Pérowski, Patrick Siarry et Eric Taillard. Métaheuristiques pour l'optimisation difficile. Eyrolles, Paris, septembre 2003, 356 pages.
6. Yann Collette et Patrick Siarry. Optimisation multiobjectif. Eyrolles, Paris, octobre 2002, 322 pages.
7. Antoine Cornuéjols, Laurent Miclet et Yves Kodratoff. Apprentissage artificiel. Eyrolles, Paris, août 2002, 638 pages.
8. Christelle Gueret, Christian Prins et Marc Sevaux. Programmation linéaire. Eyrolles, Paris, octobre 2000, 384 pages.

Vos recensions dans le Bulletin

Vous avez récemment lu un livre sur l'I.A. et vous l'avez aimé ? Faites-nous part de ce coup de coeur ! Nous la publierons (avec review minimale, bien évidemment) pour en faire bénéficier le plus grand nombre.

Vos cours sur le PortAI

Les enseignants et chercheurs sont invités à mettre leur cours à disposition des utilisateurs du PortAI :

[http : //www.afia-france.org](http://www.afia-france.org)

Si vous êtes membre de l'AFIA, cela consiste à mettre un fichier (par exemple, ppt) dans votre espace « valise » et à mettre un lien hypertexte dessus à partir de votre page personnelle. Signalez nous par email votre contribution !

L'AFIA recrute !

Nous recherchons une personne qui pourrait tenir la rubrique « Présentation de laboratoires » pour le Bulletin de l'AFIA. Cette personne fera partie du Comité de Rédaction du Bulletin, et sera responsable de joindre des chercheurs/enseignants/industriels dans les mondes académiques ou industriels, afin de récolter une présentation de laboratoire tous les 3 mois. Toute personne intéressée est invitée à envoyer un email aux Rédacteurs en Chef du Bulletin (voir page 2). Et nous remercions Catherine BARRY-GREBOVAL d'avoir tenue cette rubrique pendant toutes ces années !

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 (*razer 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	150 Euros	gratuit pour les 5 premières adhésions
<input type="checkbox"/> Personnes morales non universitaires	450 Euros	gratuit
<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'ECCAI pour constituer un fichier européen.

Veillez trouver un règlement (à l'ordre de l'AFIA) de Euros

Trésorier AFIA : Jérémie MARY, LRI, Bâtiment 490, Université Paris-Sud, 91405 ORSAY.

Mode d'adhésion :

De préférence, en ligne via le site Internet de l'AFIA : <http://www.afia-france.org>
 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 clé RIB 3.

TVA non applicable, article 293B du CGI

SOMMAIRE DU BULLETIN N° 61

Editorial.	3
Prix AFIA	4
l'I.A. au Québec	7
Ingénierie du web et intelligence artificielle	11
Interviews	35
Présentation de Laboratoires	39
Revue bibliographique	40
Conférences	43
Resumé de thèses et d'HDR	57
Annonces	62
Article	65
Sommaires des revues	70
Les dossiers du bulletin de l'Afia	79
Appels à dossiers	80

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