

BULLETIN

DE

L'AFIA



JANVIER 2003

N° 53

Présentation du bulletin

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

Pour contacter l'AFIA

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

Contributions au bulletin
Jean-Paul SANSONNET
voir ci-contre

Serveur WEB
<http://www.afia-france.org>

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

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

Membres d'honneur

Jacques Pitrat, Jean-Paul Haton, Marie-Odile Cordier
Jean-Marc David, Daniel Kayser, Claude Vogel, Alain Colmerauer

Personnes morales adhérentes à l'AFIA

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

Bureau de l'AFIA

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

Comité :

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

Comité de rédaction

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

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

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

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

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

Antoine CORNUEJOLS
Rubrique « livres »
LRI, bâtiment 490
Université Paris-Sud
91405 - ORSAY Cedex
cornuejols@ie.cnam.fr

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

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

Jérôme EUZENAT
Rubrique : « Groupes de travail »
INRIA Rhône-Alpes, 655 avenue de
l'Europe,
Montbonnot Saint Martin,
38334 Saint-Ismier, France
Jerome.Euzenat@inrialpes.fr

Rubrique « Débats »
Nous recherchons un éditeur.



Réa. Numérocopie – 01 46 31 64 53

ISSN 1273-1323

Dépôt légal octobre 2002

L'IA et le RNTL : trois questions à Bertrand Braunschweig

Ce numéro de début d'année du bulletin de l'AFIA a été décalé afin de pouvoir tenir compte des contraintes du calendrier du RNTL pour la présentation d'un dossier sur des projets qui se réclament de l'IA et qui font partie des activités soutenues au sein du RNTL. A cette occasion, nous avons posé trois questions à Bertrand Braunschweig qui s'est chargé de l'édition du dossier « IA et RNTL »

JPS (Jean-Paul Sansonnet) : *Comment juges-tu la place des techniques d'intelligence artificielle dans les appels et dans les projets du RNTL ?*

BLB (Bertrand Braunschweig) : Comme son nom l'indique, Le RNTL s'intéresse avant tout aux logiciels. L'intersection de la thématique "logiciel" avec la thématique "intelligence artificielle" est importante, mais force est de reconnaître qu'il existe des sujets de recherche et de développement autour du logiciel n'ayant rien à voir avec l'IA (c'en est même la grande majorité), et que, réciproquement, faire de l'IA ne signifie pas nécessairement développer des logiciels. Dans ce contexte, l'IA représente une part significative des thèmes du RNTL, comme nous le montrons dans l'introduction au dossier thématique de ce bulletin. Les thèmes du RNTL ne sont pas organisés par discipline de l'informatique mais par domaine d'applications où plusieurs disciplines peuvent intervenir. Si dans la liste des mots clefs il n'y a pas le mot IA il y a par contre des termes qui renvoient à des techniques d'IA. Il faut dire - et tu connais mon discours sur ce point - que l'IA est de plus en plus présente dans les recherches et développements informatiques, grâce à l'explosion du web, et grâce aux rapprochements entre les domaines de l'IA et d'autres domaines de l'informatique. Il ne nous reste plus qu'à proposer des bons projets, pour être encore plus présents !

JPS : *Que devrait faire la communauté IA pour renforcer sa présence internationale sur le plan industriel et commercial ?*

BLB : J'ai l'impression que nous avons pris un bon tournant. Il y a quelques années, les recherches académiques françaises en IA ne débouchaient que très rarement sur des produits exploitables, susceptibles de créer de la valeur et de l'emploi sur notre territoire. Je vois autour de moi plusieurs exemples de jeunes pousses françaises qui fondent leur développement sur des résultats de recherche de la communauté IA. Dans le contexte économique actuel, bien difficile pour les sociétés informatiques, on ne peut jurer de rien, et les exemples récents m'incitent à la prudence. Mais nos bases scientifiques et techniques sont bonnes, solides, et peuvent très bien être exploitées si l'on sait s'organiser pour les promouvoir et les valoriser suffisamment. Les projets RNTL sont un moyen possible pour transformer ces recherches en produits commercialisables. Ensuite, il faudra se bouger...

JPS : *Quel conseil donnerais-tu aux membres de l'AFIA, dans l'optique du prochain appel se terminant en juin 2003 ?*

BLB : Je leur dirai de préparer leurs propositions avec beaucoup de soin. Le processus d'évaluation du RNTL est sérieux, les experts sont vos pairs, ils connaissent la chanson ; une proposition bâclée, contenant des ambiguïtés, non fondée scientifiquement, sans potentiel d'exploitation, ou avec un consortium déséquilibré, ne passera pas, car la concurrence est rude. Il faut veiller à satisfaire tous les critères. Le RNTL publie ses documents d'évaluation depuis deux ans, comme l'avait fait auparavant la Commission Européenne. Alors, faites l'exercice de vous auto-évaluer, ou, mieux, demandez à un collègue de le faire pour vous. Et n'oubliez pas de consulter l'existant, les projets déjà soutenus, les plates-formes, car le RNTL est soucieux de la réutilisation de ses résultats. Si vous pouvez appuyer votre proposition sur un projet existant, à condition évidemment de bien montrer ce que vous apportez en plus, elle sera regardée avec bienveillance. Il y a de la place pour les projets IA dans tous les thèmes du RNTL : à nous de la prendre.

Dans les prochains numéros du bulletin de l'AFIA nous insisterons sur cette approche et nous envisageons des contributions pour des dossiers qui sont à la fois au cœur de l'IA et qui intéressent plus largement la communauté scientifique. Ce sera le cas avec l'annonce des prochains dossiers comme par exemple : « Web Sémantique » ou encore « Systèmes d'Information ».

Jean-Paul Sansonnet

Rédacteur en chef du Bulletin de l'AFIA

IMALAIA Intégration de Modèles ALIant Automatique et IA

Animateurs : Louise Travé-Massuyès , Marie-Odile Cordier, François Levy

Equipes Académiques

ENSIEG-LAG (Grenoble) : Sylviane Gentil, Stéphane Ploix

LAAS-CNRS (Toulouse) : Louise Travé-Massuyès

LAIL (Lille): Marcel Staroswiecki

IRISA (Rennes): Marie-Odile Cordier

LIPN-UMR 7030 (Villetaneuse): François Lévy, Philippe Dague

Industriels

CEA Marcoule : Jacky Montmain

IFP : Sylvie Cauvin

Thales : Patrick Taillibert

Le groupe IMALAIA concerne des chercheurs de l'Automatique et de l'Intelligence Artificielle. Il bénéficie de l'acquis accumulé des groupes Alarme (PRC IA), Diagnostic (GDR Automatique) et MQ&D (PRC IA / GDR Automatique). Il est reconnu par les GDR I3 et Automatique et constitue un groupe S3 ainsi qu'un groupe AFIA.

L'objectif général de ce groupe d'une douzaine de personnes est d'explorer les points communs ou les liens et les complémentarités des modèles et méthodes des deux communautés pour le diagnostic. Sur un travail de fond consistant à faire l'inventaire des modèles et méthodes proposés de part et d'autre et leurs motivations, le groupe mène des travaux plus techniques. Citons dernièrement une analyse comparative de l'approche dite des résidus structurés et de la théorie logique du diagnostic qui a donné lieu à une publication à l'ECAI 00 et à Safeprocess 00. Des réflexions ont été engagées concernant les aspects temporels du diagnostic, la prise en compte des incertitudes, les systèmes hybrides.

IMALAIA a des interactions fortes avec MONET (I et II) Réseau d'Excellence « Model based and qualitative systems » du programme européen ESPRIT (<http://monet.aber.ac.uk/>) et des liens étroits avec le « Task Group » BRIDGE de MONET II, qui a les mêmes objectifs qu'IMALAIA au niveau européen.

Fonctionnement du groupe

Le groupe fonctionne à composition constante, et

accueille de nouveaux participants lorsqu'il renouvelle ses centres d'intérêt. Il tient environ 4 réunions d'un à deux jours par an. Il a publié en tant que tel à l'Ecai 2000 et SafeProcess 2000, et a un article de revue en préparation.

Description des Activités

Pour une description rapide des activités d'Imalaia : sur le site AFIA, en suivant groupes de travail, soit <http://www.afia.polytechnique.fr/node.php?lang=fr&node=7> puis IMALAIA

Pour une description plus approfondie, sur le site du GDR Automatique, voir S3, projets en cours, projet 1, soit <http://www.univ-lille1.fr/s3/fr/fenetre.htm>

Les deux thèmes prioritaires en ce moment sont décrits ci-dessous.

Aspects temporels du diagnostic

La prise en compte du temps au niveau de l'isolation de fautes est un vrai problème, notamment dans le cadre du suivi de systèmes dynamiques intégrant une boucle surveillance, diagnostic, actions de reconfiguration. La dynamique des fautes est à prendre en compte et se superpose à la dynamique propre du système, sans parler des fautes évolutives ou transitoires. Ce problème est loin d'être bien formalisé de manière générale dans l'une tout autant que dans l'autre des communautés. Cependant, un ensemble de travaux existent de part et d'autre.

En IA, une classification des approches de diagnos-

tic « temporel » a été fournie par (Console & Theseider-Dupré, 1998). Les critères pertinents concernant le comportement du système physique à diagnostiquer ainsi que la dynamique des fautes sont dégagés ainsi que les différentes ontologies du temps. Par ailleurs, (Struss, 1997) (Struss et al., 1997) montrent que, sous certaines conditions, il est possible de produire un diagnostic uniquement à partir de mesures prises au même instant (state-based diagnosis). De leur côté (Panati & Theseider-Dupré, 2000) montrent que ces conditions sont très restrictives et illustrent les déficiences de l'approche « state-based » par rapport à l'approche « simulation-based » sur un exemple. En Automatique, on peut noter les travaux de (Combastel, 2000) qui considèrent des résidus temporels pour élaborer les diagnostics. Cette approche est à rapprocher de l'approche DX par reconnaissance de chroniques (Dousson et al., 1993). Cependant l'approche la plus communément utilisée procède en fait à une « élimination du temps » puisque l'on cherche à donner aux résidus des propriétés invariantes dans l'espace de parité. Cette approche contourne en quelque sorte le problème mais son intérêt est à évaluer par rapport aux approches DX de représentation explicite du temps. Les systèmes hybrides. Les algorithmes de surveillance et de diagnostic doivent reconnaître, indiquer et expliquer, le plus rapidement possible et avec le maximum de certitude, les anomalies de comportement (ou défaillances) à partir des informations disponibles sur le système. Les mesures et les historiques constituent la partie de ces informations générée au cours de l'exploitation du système. Les modèles et les données de conception en constituent la partie connue a priori. Or ces connaissances sont de types très variés, et sont exprimées dans des formalismes très différents.

1) Différents types de connaissances sont à représenter : points de vue téléologique (quels sont les objectifs pour lesquels il a été construit ?), fonctionnel (quelles sont les fonctions mises en œuvre pour les atteindre ?), comportemental (quels sont les mécanismes de génération des trajectoires temporelles des variables décrivant l'évolution du système ?), topologique, structurel, etc.

2) Ces connaissances sont à exprimer à différents niveaux de détail (modèles hiérarchiquement structurés),

3) Les variables utilisées peuvent intervenir à différents niveaux de granularité (numérique, qualitatif), et donc nécessiter l'utilisation de formalismes différents pour décrire leur comportement (équations différentielles, automates, règles). Dans ce sens, l'objectif visé peut se résumer de la manière suivante : rechercher des modélisations pertinentes pour la surveillance et le diagnostic, c'est-à-dire de complexité réduite, ayant un sens physique, induisant un bon pouvoir discriminant, permettant la prise en compte et la manipulation conjointe de différents niveaux de connaissance (fonctionnel, structurel, comportemental), de différents types de description de cette connaissance (numérique, symbolique, qualitatif) et de différents types de systèmes (à états continus, à événements discrets, hybrides), générant des informations adaptées à leurs différents utilisateurs (algorithmes, opérateurs), etc. Questions envisagées : Comment faire co-habiter des représentations différentes (mais pas indépendantes) du même système : il est nécessaire d'en gérer la cohérence (mécanismes de détection des contradictions, mécanismes de traduction, de passage d'un point de vue à un autre, d'un niveau à un autre) et la complémentarité (mécanismes d'interfaçage). Comment utiliser ces différentes représentations pour la conception d'algorithmes de surveillance et de diagnostic faisant appel à toutes les connaissances disponibles sur le système ? (par exemple en utilisant un formalisme unique - réseaux de Petri - pour la représentation de systèmes hybrides, ou encore en associant raisonnement temporel et fonctionnel pour le diagnostic). Quelles stratégies de contrôle mettre en place, le cas échéant, pour le choix du modèle adéquat dans un contexte temps réel ? Pertinence scientifique. Le projet proposé tend à lier la problématique du diagnostic à celle de la surveillance. Cette problématique est en effet assez mal prise en compte par les approches de surveillance développées par les automaticiens, qui se limitent à la détection et à la localisation des défaillances, car elles n'utilisent que des représentations du comportement du système (numériques ou qualitatives), et n'ont pas accès à d'autres types d'informations (graphes causaux, interprétations fonctionnelles par exemple). Il conduit également à développer des algorithmes de surveillance de niveaux différents, en considérant par exemple des modèles de comportement à différents niveaux d'abstraction (signaux et équations numériques au niveau bas, séquences d'événements ou chroniques à un niveau plus élevé).

Nouvelles du Québec

Nicole Tourigny

Bonjour!

Un laboratoire des systèmes intelligents (LSI) vient tout juste d'être formé au département d'informatique et de génie logiciel (<http://www.ift.ulaval.ca>) de l'Université Laval (<http://www.ulaval.ca>), regroupant une cinquantaine de personnes effectuant des travaux de recherche en intelligence artificielle. Je vous en ferai une description lors du prochain numéro du bulletin de l'AFIA.

Dans le présent numéro, vous retrouverez une description du laboratoire du professeur Sylvain Giroux, de l'Université de Sherbrooke (<http://www.usherb.ca>). Vous pourrez ensuite lire le projet de thèse de Luc Lamontagne, étudiant au niveau Ph.D au laboratoire RALI (voir bulletin de l'AFIA de janvier 2002), au département d'informatique et de recherche opérationnelle (<http://www.iro.umontreal.ca>) de l'Université de Montréal (<http://www.umontreal.ca>).

Laboratoire DOMUS (Laboratoire de DOmotique et d'informatique Mobile de l'Université de Sherbrooke)

Responsable : Professeur Sylvain Giroux, Ph.D.
Contact : sgiroux@dmi.usherb.ca
<http://www.usherb.ca/~sgiroux/domus>

Le laboratoire DOMUS est un laboratoire multidisciplinaire dédié à la recherche en DOmotique et en l'informatique Mobile à l'Université de Sherbrooke. Ses membres sont Sylvain Giroux (informatique), Hélène Pigot (informatique et ergothérapie), André Mayers (informatique et psychologie), Philippe Mabilieu (génie informatique) et Philippe Gagnon (éthique).

« The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it. » Mark Weiser, *The Computer for the 21st Century*, Scientific American, 1991.

Problématique

L'informatique telle que nous la concevons aujourd'hui est en profonde mutation. D'aucuns auront noté la prolifération des micro-processeurs dans notre environnement quotidien. L'interconnexion et l'intégration de ces micro-processeurs à travers les réseaux sans fil autorisent le développement d'applications inconcevables, il y a peu. Ces développements technologiques

ont signifié l'apparition et l'expansion de domaines de recherche et d'application tels l'informatique mobile, l'informatique vestimentaire et l'informatique diffuse. Avec l'informatique mobile, l'ordinateur part en ballade. Les petits appareils portables (assistants personnels (PDAs), téléphones cellulaires...) demeurent malgré tout traditionnels, dans le sens où reste palpable la parenté avec leur ancêtre, l'ordinateur personnel (PC) muni de son écran, de son clavier et de sa souris. Alors que l'informatique mobile libère l'ordinateur de ses attaches fixes, l'informatique vestimentaire le délivre de sa forme. Disparus écran, clavier et souris ! L'« ordinateur » est porté, tout comme le sont les lunettes ou les vêtements. Avec la domotique, le micro-processeur investit les objets de la vie quotidienne. Toutefois à l'inverse de l'informatique mobile et de l'informatique vestimentaire, les processeurs intégrés dans les objets d'utilité courante ne se déplacent guère. L'informatique diffuse met un point final à l'identification de l'informatique à un appareil physique. L'ordinateur ne sera plus personnel, mais un ensemble d'applications seront réparties sur plusieurs appareils hétéroclites et diverses interfaces permettront d'y accéder. D'un côté, il y aura le monde réel, celui de l'individu, et de l'autre le monde

virtuel, celui des systèmes d'information. Les deux mondes se superposeront et les appareils fixes et mobiles serviront de points de contact, d'interfaces entre ces deux mondes.

Dans un environnement informatique diffus ou mobile, la personne, intégrée dans un environnement par essence complexe, est assistée par une infrastructure informatique afin de réaliser ses activités quotidiennes en toute sécurité. D'une part, l'informatique diffuse installe des capteurs et des appareils intelligents dans l'environnement. D'autre part, l'informatique mobile et vestimentaire fait porter par la personne un code mobile inséré dans un minuscule processeur. Tous deux, intégrés dans l'environnement, n'indisposent pas l'utilisateur. Saurons-nous concevoir un environnement où l'on puisse se servir de l'ordinateur sans même y penser ? La « réalité virtuelle » mettrait l'interface au centre en laissant le monde derrière, consacrant son énergie à simuler un environnement, plutôt que de rendre plus agréable l'usage du monde tel qu'il nous est donné sans interférence. L'informatique diffuse se modèle « docilement » sur l'espace habité plutôt que de l'adapter à un plan préconçu.

L'informatique mobile et l'informatique diffuse sont complémentaires. Même plus, dans beaucoup d'applications, elles ne peuvent exister l'une sans l'autre.

UN SYSTÈME D'INFORMATION DIFFUS OU MOBILE N'EST PAS UN SYSTÈME TRADITIONNEL

Bien entendu, l'infrastructure requise pour l'informatique diffuse et mobile diffère grandement de celle des systèmes d'information « traditionnels ». Tout d'abord, elle doit intégrer des appareils fortement hétérogènes : appareils électroménagers, senseurs, assistants personnels (PDAs), écrans tactiles, TV digitale, téléphones cellulaires, serveurs, etc. Bien que cette hétérogénéité semble difficile à circonscrire et à maîtriser, elle peut s'avérer un levier : les faiblesses de cer-

taines composantes sont compensées par les forces des autres. Des coalitions entre appareils doivent spontanément se former et agir de manière concertée et efficace. Il faut concevoir les systèmes d'information et les implanter en faisant abstraction d'appareils spécifiques, puis lier automatiquement ces implantations « abstraites » à des environnements réels différents qui évoluent dynamiquement. Il faut décentraliser et répartir les algorithmes de raisonnement et de prise d'actions. Finalement la tolérance aux fautes reste un aspect crucial.

LE LABORATOIRE DOMUS

Afin d'étudier les problèmes de recherche en informatique diffuse et en informatique mobile, une infrastructure de recherche doit être mise en place et un programme de recherche doit structurer les travaux. C'est là le rôle du laboratoire de recherche DOMUS. Ses objectifs sont :

- 1 - mettre en place les fondements théoriques, logiciels et matériels pour la conception et la réalisation de services mobiles et diffus ;
- 2 - concevoir et implémenter une architecture logicielle pour la réalisation de services mobiles personnalisés, géo-référencés et répartis capables de se déployer dans des environnements diffus ;
- 3 - développer et expérimenter des prototypes novateurs en grandeur nature d'environnements diffus et d'applications mobiles ;
- 4 - favoriser l'étude multidisciplinaire des environnements diffus par la réalisation d'artefacts pour mieux comprendre les impacts de l'informatique diffuse et de l'informatique mobile sur la vie quotidienne et la société.

Les principaux domaines d'applications sont la santé, l'éducation, la protection civile et le tourisme. Parmi les projets en cours, mentionnons l'assistance cognitive aux personnes souffrant de démences de type Alzheimer, de schizophrénie ou d'autres déficits cognitifs. Cette assistance est dispensée à travers la réalisation d'habitats intelligents et d'applications mobiles.

Titre de la thèse : **Approche de raisonnement à base de cas textuels pour la réponse au courrier électronique**
Auteur : Luc Lamontagne
Directeur de recherche : Dr Guy Lapalme, Département d'informatique et de recherche opérationnelle, Université de Montréal
Laboratoire : Laboratoire de Recherche Appliquée en Linguistique Informatique
<http://www-rali.iro.umontreal.ca/>
Statut de la thèse : en cours de réalisation. Date de fin prévue : été 2003.

RÉSUMÉ.

Dans ces travaux, nous proposons une approche de réponse automatique au courrier électronique à l'aide de techniques de raisonnement à base de cas textuels (CBR textuel). Un système CBR est un système qui tente de résoudre de nouveaux problèmes à partir d'expériences passées. Dans un système CBR textuel, les expériences (i.e. les problèmes et les solutions) sont décrites en langue naturelle. La faible structuration des descriptions textuelles requiert l'extension des techniques traditionnelles du CBR (i.e. CBR structurel). Notre système CBR de gestion de réponse s'appuie sur un corpus de messages antécédents, constitué de paires <requête, réponse>. Présentée selon une perspective "client courriel", l'approche proposée réutilise les messages de la boîte courrier SENT pour proposer des réponses aux nouveaux messages du INBOX. Notre cadre d'application est le service aux investisseurs, un domaine qui présente une certaine répétitivité dans le contenu des messages échangés entre les investisseurs et les analystes financiers.

La première étape de l'approche consiste à repérer avec précision des messages antécédents pour répondre à une nouvelle requête. Traditionnellement, la sélection de cas CBR est basée sur la similarité entre les descrip-

tions de problèmes (requêtes). Pour guider le processus de recherche vers des solutions potentiellement intéressantes, nous utilisons en plus des associations entre les mots des requêtes et les mots des réponses. Nous modélisons les associations à l'aide de modèles statistiques de co-occurrences et de traduction. Le modèle de traduction se révèle le plus intéressant car il permet de mieux éliminer les associations indirectes entre les mots.

Par la suite, pour identifier les portions des réponses antécédentes qui sont réutilisables, nous découpons chaque phrase des réponses retenues et nous évaluons leur degré de pertinence par rapport à la nouvelle requête. Par la suite, nous éliminons les redondances et nous appliquons un algorithme de recherche afin de sélectionner le sous-ensemble de ces phrases qui recouvre le mieux les thèmes de la nouvelle requête. Finalement, nous déterminons les informations pouvant être modifiées. Pour notre application de service aux investisseurs, les principales différences entre les messages portent sur des informations factuelles telles que les dates, numéros de téléphone, noms de personnes. Ces informations, des entités nommées, sont repérées par un processus d'extraction et de nouvelles valeurs sont attribuées à certaines de ces entités nommées par des règles du domaine.

Il me fait toujours plaisir de répondre à vos questions, et vos commentaires sont grandement appréciés.

Nicole Tourigny, Ph.D., professeur
Directrice du laboratoire LSI
Département d'informatique et de génie logiciel
Pavillon Adrien-Pouliot
Université Laval
Québec, CANADA
G1K 7P4
tourigny@ift.ulaval.ca
<http://www.ift.ulaval.ca/~ericae/>

PRÉSENTATION DE LABORATOIRES

Présentations de laboratoires dans le bulletin de l'AFIA

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

Présentation d'orientations de recherche du Laboratoire de Recherche sur le Langage – LRL

Michel Chambreuil, Gabriel G. Bes

LRL
Maison de la Recherche
4 rue Ledru
63057 Clermont-Ferrand
<http://lrlweb.univ-bpclermont.fr>

Michel Chambreuil
Tél : 04 73 34 68 35
Mél :
chambreuil@lrl.univ-bpclermont.fr
Gabriel G. Bes
Tél : 04 73 91 54 44
Mél : Gabriel.Bes@univ-bpclermont.fr

Le Laboratoire de Recherche sur le Langage (LRL) de l'Université Blaise Pascal Clermont 2 a été mis en place en 1991. Il a le statut d'Equipe d'Accueil (EA 999) et regroupe actuellement une cinquantaine de chercheurs, enseignants-chercheurs et doctorants. Il est laboratoire d'accueil du DEA « Linguistique, Logique et Informatique », et intégré à l'Ecole Doctorale « Lettres, Sciences Humaines et Sociales ».

Le LRL ne s'est pas constitué à partir d'une discipline en tant que telle mais à partir de thématiques déjà présentes dans des recherches conduites dans le cadre de l'Université Blaise Pascal.

Le LRL a retenu trois grandes thématiques qui ne sont pas considérées comme indépendantes, mais au contraire en étroite interdépendance sous plusieurs aspects. Ces trois thématiques concernent respectivement :

- la construction de la signification,
- le traitement informatique de la langue,
- les environnements informatiques d'aide à l'apprentissage et à l'enseignement de la langue.

En liaison avec les trois thématiques retenues, deux autres caractéristiques du

LRL peuvent être soulignées : sa pluridisciplinarité et la complémentarité de recherches fondamentales et appliquées.

Le LRL est effectivement un laboratoire pluridisciplinaire susceptible d'intégrer, au quotidien, sur un même projet, des linguistes, des informaticiens, des mathématiciens, des spécialistes de l'apprentissage. Cette pluridisciplinarité, présente dans le groupe des enseignants-chercheurs membres du LRL est encore plus marquée dans le groupe des doctorants travaillant sur les projets à l'intérieur de chacune des thématiques. Elle est également présente dans le DEA « Linguistique, Logique et Informatique » associé au LRL. La pluridisciplinarité du LRL sur chacune de ces trois thématiques est simplement la concrétisation locale de la composition de la communauté internationale travaillant aujourd'hui sur ces trois domaines.

La complémentarité des recherches fondamentales et appliquées se traduit, elle, dans le développement de logiciels concernant, en particulier :

- des aspects du traitement informatique de la langue
- des environnements multimédias pour une aide à l'apprentissage du Français Langue Etrangère
- des environnements multimédias pour une aide à l'apprentissage et à l'enseignement de la lecture

1. Construction de la signification (contact : M. Chambreuil)

Sur cette thématique, dans le cadre de cette présentation, ce sont essentiellement des orientations de recherche à l'intérieur du projet ElaDyS (Elaboration Dynamique de la Signification) qui seront abordées. L'objectif fondamental de ce projet est celui d'une modélisation de principes organisateurs de réseaux information-

nels associés aux expressions linguistiques et de processus traitant dynamiquement ces réseaux en fonction de différents éléments contextuels.

Cet objectif peut être illustré, naïvement, dans une activité de compréhension. Par exemple, dans le traitement en réception de l'expression « Jacques a acheté une voiture noire », on peut imaginer que l'auditeur activera à partir de 'noire' le concept de 'couleur' et à partir de 'voiture' le concept de 'carrosserie' (et non celui de 'moteur', de 'roue', de 'siège', ...). Ceci signifie que 'voiture' donne accès à 'carrosserie' et que 'noir' donne accès à 'couleur'. Ceci signifie également que le processus combinatoire, en fonction de différents éléments contextuels (dont les deux expressions traitées) active sélectivement des informations particulières dans les réseaux associés respectivement à 'voiture' et à 'noir'. C'est cette activation dynamique et sélective qui sera susceptible de déclencher des inférences traduites, par exemple, par des questions posées au locuteur.

Par exemple encore, dans le traitement, dans un contexte donné, de l'expression « Jacques a écrit un livre noir », 'noir' n'activera pas le concept de 'couleur' mais un concept de 'pessimisme' ou de 'policier' associé là encore à une sélection parmi les éléments informationnels associés à 'livre', sélection qui retiendra des aspects du contenu du livre et non pas, par exemple, sa couverture.

De même, on peut penser que le fait de demander à une personne « d'ouvrir une porte » ou « d'ouvrir une lettre » ou encore « d'ouvrir une boîte de conserve » ne déclenchera pas chez cette personne les mêmes planifications d'actions, ni, bien sûr, les mêmes potentiels inférentiels.

Par rapport aux trois aspects de l'acti-

tivité langagière, l'acquisition, la réception et la production, l'objectif du projet ElaDyS, s'appuie sur une hypothèse générale du type suivant.

– Les réseaux informationnels associés à des expressions lexicales et les processus de traitement de ces réseaux, en acquisition, en réception ou en production, sont organisés selon des régularités pouvant être l'objet d'une modélisation.

Par rapport à une activité particulière, par exemple l'activité de réception d'une expression complexe, l'objectif du projet s'appuie aussi sur des hypothèses générales telles que :

- Le résultat du traitement en réception sera lui-même une structure informationnelle construite à partir de différents éléments, dont les structures informationnelles associées aux expressions lexicales de l'expression traitée.
- Ce traitement active de façon dynamique et sélective des informations dans les réseaux informationnels associés aux expressions lexicales.
- Cette activation dynamique et sélective est sous-jacente au potentiel inférentiel associé au traitement de l'expression complexe.
- Ce traitement dynamique et sélectif est dépendant de différents éléments contextuels.

A partir des hypothèses générales retenues, les questions posées sont multiples, mais renvoient aussi à des enjeux de plus en plus importants concernant par exemple la compréhension de différents aspects de l'activité langagière : acquisition, réception, production, voire dysfonctionnement. Par rapport à l'IA ces enjeux renvoient aux problématiques de la représentation des connaissances, à la constitution évolutive de bases de connaissances, à différents processus dynamiques et contextuels traitant ces connaissances.

Dans les limites de cette présentation du LRL, nous évoquerons juste

quelques questions ouvertes concernant les réseaux informationnels associés aux expressions lexicales prises comme entités élémentaires, ces questions renvoyant en IA à des questions sur la représentation des connaissances.

Une première question, par exemple, concerne les liens entre les principes organisateurs du réseau informationnel associé à une expression lexicale et la catégorie de cette expression. Les représentations conceptuelles associées, par exemple, à un nom, à un verbe, à un adjectif, ou encore à un déterminant ou à une entité de coordination, sont-elles de même nature, mettent-elles en jeu les mêmes types d'entités et les mêmes principes organisateurs ? Ce type de questions doit aussi être introduit à l'intérieur même d'une catégorie. Par exemple le réseau informationnel associé à un nom d'artéfact tel que « voiture » est-il de même nature que le réseau associé à un nom abstrait tel que « liberté » ? On rejoint là, en IA, des questions abordées aujourd'hui dans les recherches de plus en plus nombreuses conduites sous le terme d'ontologie.

Une autre question concerne, bien sûr, la nature même des constituants du réseau informationnel associé à une catégorie d'entité donnée. On peut penser, par exemple, aux recherches sur des liens mérologiques entre entités, sur des liens fonctionnels ou encore sur des liens de classes à sous-classes. On peut penser aussi à des questions sur l'opposition entre connaissances déclaratives et connaissances procédurales ou encore sur l'association de « théories » (naïves) à différentes entités, liens ou nœuds d'un réseau informationnel. On rejoint encore des questions sur les liens entre différents types d'informations telles que, par exemple, informations langagières, perceptives, de planification et de réalisation d'actions.

En relation directe avec les questions posées sur les régularités organisatrices des réseaux informationnels se posent des questions sur les processus sous-jacents à l'acquisition de ces réseaux et

sur les processus combinatoires sous-jacents à la compréhension et à la production. Concernant, par exemple, la compréhension, en dehors des questions liées aux régularités d'activation dynamique et sélective d'informations dans les réseaux informationnels associés à des expressions linguistiques élémentaires ou complexes se posent, par exemple, des questions sur les éléments porteurs de ces processus combinatoires. On peut penser là à des recherches conduites à partir de la notion de mots pris en tant que mots experts, mais on peut aussi penser aux liens entre syntaxe et sémantique ou encore aux fonctionnalités des rôles thématiques pouvant être envisagés comme rôles experts déclencheurs de combinatoires spécifiques et de potentiels inférentiels.

C'est donc à partir d'hypothèses générales et de questions du type de celles évoquées ci-dessus que se développent les recherches de modélisation dans le cadre du projet ElaDyS. A titre d'illustration on pourrait citer ici des travaux sur la sémantique des adjectifs et la combinatoire adjectif-nom, sur la sémantique de verbes de changement d'état et sur leur combinatoire avec des entités nominales, sur la sémantique d'expressions nominales d'artéfacts, ou encore des travaux sur la sémantique d'entités de coordination, de prépositions et d'articles. Sur un plan méthodologique, non indépendant de la pluridisciplinarité du projet ElaDyS, cette orientation de recherche s'appuie sur différents domaines de références non considérés en opposition. On peut mentionner, par exemple, le domaine des approches logico-formelles de la sémantique, ou encore celui des approches cognitives de plus en plus nombreuses.

2. Traitement informatique de la langue

Cette thématique recouvre deux

grandes orientations. L'une se situe dans le cadre du projet ElaDyS et concerne la simulation informatique d'aspects de l'activité langagière du locuteur-auditeur. La seconde porte sur une modélisation des systèmes linguistiques qui est associée à un traitement informatique, mais qui reste autonome par rapport à ce traitement. Cette orientation continue les travaux du GRIL (Groupe de Recherches sur les Industries de la Langue), intégré au LRL depuis 1996.

2.1 Simulation d'aspects de l'activité langagière (contact : M. Chambreuil)

En fonction des recherches conduites dans le projet ElaDyS, l'orientation de simulation concerne actuellement des aspects de la construction de la signification en situation de réception. Les recherches conduites selon cette orientation se situent dans une perspective symbolique, sans que cette perspective soit considérée en totale opposition à une perspective connexionniste. Cette orientation de simulation renvoie, bien sûr, à la quasi totalité des questions de modélisation évoquées ci-dessus dans la présentation du projet ElaDyS. Mais elle renvoie aussi à de nombreuses questions concernant l'aspect simulation en tant que tel. On peut penser là, par exemple, à des questions sur les structures de représentation des connaissances, sur la mémorisation et l'évolution de bases de connaissances. Ces questions concernent également, toujours par exemple, la spécification des processus de combinatoire de ces connaissances et d'activation dynamique et sélective d'informations, ou encore la spécification des processus d'activation potentielle de processus inférentiels. Les enjeux, dans le contexte actuel, sont aussi importants par les applications susceptibles de prolonger une simulation validant les hypothèses de modélisation. Ces applications, en dehors d'outils de représentation des connaissances, concernent différents problèmes du traitement informatique de la langue intégrant la construction de la signification. On peut penser, par

exemple, à des problèmes posés en traduction automatique ou en aide à la traduction par la polysémie des expressions linguistiques, ce terme étant pris en un sens très général susceptible de correspondre à différents niveaux d'activation sélective d'informations dans les réseaux informationnels associés aux expressions linguistiques.

2.2 Modélisation et traitement informatiques des systèmes linguistiques (contact : Gabriel G. Bes)

Les choix méthodologiques de cette orientation peuvent se résumer ainsi :

- (1) La linguistique, comme, parmi d'autres, la physique, la chimie, la biologie ou l'économie, devrait pouvoir être pratiquée comme une science de l'empirie, c'est-à-dire une science dont l'objet est un réel observable.
- (2) Dans une science de l'empirie, le test de corroboration est l'élément clé qui met en relation (qui permet de falsifier, dans une terminologie poppérienne) les observations avec les déductions obtenues formellement à partir des hypothèses.
- (3) Les observations ne sont pas une donnée première. Elles se construisent de manière explicite et elles doivent être accessibles et intersubjectives. Selon le choix effectué dans l'orientation, elles sont associées à des suites de codes (par exemple, Ascii ou Unicode). Ces suites de codes sont considérées comme l'objet qui doit être traité par l'outil informatique où les hypothèses sont implantées. Dans l'option, on ne considère donc pas le langage oral non transcrit. Les observations doivent s'exprimer dans un langage adéquat afin de les confronter, dans le test de corroboration, avec les déductions. Déductions et hypothèses, dans la situation idéale vers laquelle on tend, sont exprimées dans un langage formel, et donc explicite et calculable.
- (4) Suivant l'exigence newtonnienne de

ne pas feindre la formulation d'hypothèses (*hypothesis non fingo*), aucun appel n'est fait à une supposée relevance cognitive des hypothèses ou des descriptions proposées, et aucun langage mental n'est invoqué.

Cette orientation, de par la méthodologie qui la structure, prétend se démarquer de l'utilisation d'exemples ciblés, en général choisis pour illustrer le bien fondé de l'utilisation de tel ou tel formalisme, que ce soit en syntaxe ou en sémantique. Elle prétend se démarquer aussi de l'obtention, par les approches texto-algorithmiques, de résultats chiffrés de l'analyse de tel ou tel type de construction et/ou d'étiquetage, approches qui laissent dans une quasi totale opacité les descriptions linguistiques sous-jacentes qui ont permis l'obtention de ces résultats, et que l'on doit supposer non strictement distinguées de la machinerie algorithmique, elle aussi opaque, opérant sur des textes effectifs, choisis selon des critères, toujours opaques eux aussi. Ni dans les approches dites "théoriques" utilisant des exemples ciblés ni dans les approches texto-algorithmiques, aucune application sincère du test de corroboration n'est possible et, par conséquent, la linguistique n'est pas pratiquée comme science de l'empirie.

Dans ce cadre méthodologique général est développé le Paradigme 5P avec : *P1* ou *P* de Protocoles, les Protocoles étant des enregistrements systématiques des observations ; *P2* ou *P* de *Propriétés*, dont on reconnaît trois grands types (Existence, Linéarité et Fléchage), qui, de manière factorisée et déclarative, décrivent formellement des structures linguistiques ; *P3* ou *P* de *Projections*, pour exprimer des régularités sur les *P2* d'une langue ; *P4* ou *P* de *Principes*, pour exprimer des contraintes interlinguistiques sur l'organisation des *P2* ou des *P3*, et, finalement, *P5* ou *P* des *Processus*.

Les Processus (P5) permettent le traitement informatique effectif des suites de codes : textes effectifs ou réels, c'est-à-dire attestés en tant que tels et non en tant qu'échantillons du langage-objet que l'on prétend décrire, mais aussi des suites-test, relevant des textes que l'on peut dire *virtuels* (c'est-à-dire non effectivement utilisés en tant que tels) que l'on souhaite activer. Les P5 comprennent une machinerie algorithmique effectivement implantée en machine, associée à une source déclarative, mais les sources déclaratives des P5 ne sont pas les Propriétés (P2) : elles se déduisent des P2 avec la granularité requise par le traitement que l'on veut effectuer.

À proprement parler, il n'y a pas, en 5P, de grammaire, au sens où on a des grammaires LFG, GPSG, HPSG ou autres. Celles-ci sont des objets hybrides, avec une double fonction, pas toujours bien explicitée : elles décrivent le langage-objet et, en même temps, elles sont la source déclarative (les données au sens informatique) de la machinerie algorithmique qui va traiter des textes. En 5P on décrit en P2, on traite informatiquement en P5 en obtenant la ou les sources déclaratives adéquates à partir des P2, on exprime les observations en P1 et on essaye d'expliquer en P3 ou en P4. Les Propriétés P2 sont ainsi une *Base déclarative de données linguistiques plutôt qu'une grammaire*.

Le traitement informatique, déjà implanté et opérationnel, s'organise aujourd'hui autour de trois composantes : (i) segmentation et analyse morphologique de la suite d'entrée, (ii) traitement syntaxique local, (iii) traitement syntaxique de la phrase. En développement, on étudie l'exploitation des Paires de fléchage des Propriétés P2. Ces Paires de fléchage permettent la définition d'un graphe, partiellement connecté dans son état actuel, qui est l'entrée de la fonction sémantique permettant d'associer la suite de codes analysée à une représentation sémantique (partielle) sous forme de DRS. Les travaux en

cours portent sur la reconnaissance du syntagme verbal noyau de la phrase racine et les coordinations des syntagmes verbaux noyau de la phrase racine et des phrases enchâssées et subordonnées. En développement aussi, pour l'analyse automatique des textes et l'extraction de l'information, on étudie l'exploitation possible des balises des documents structurés avec un langage comme XML

5P est le résultat d'une quinzaine d'années d'expérience dans les Industries de la Langue dans une dizaine de contrats européens ou avec l'industrie, où des modèles des grammaires d'unification ou des grammaires catégorielles ont été pratiqués ; cela a permis de mesurer le fossé qui existe entre eux et une ingénierie efficace, obligée à traiter du texte « tout venant » et qui ne peut donc pas se satisfaire d'un échantillonnage d'exemples utilisés pour illustrer tel ou tel formalisme. La même magnitude de fossé a été observée entre des sémantiques véri-conditionnelles et les traitements textuels nécessaires pour rendre compte des phénomènes aussi largement attestés que, par exemple, les articles définis singuliers ou pluriels, par rapport auxquels les traitements à la Russel échouent complètement. Il est fondé sur la conviction que l'ignorance réciproque, réelle sinon affichée, entre approches qui se veulent théoriques et approches rangés comme relevant d'une ingénierie à court terme, est pernicieuse et pour la théorie au sens noble de ce terme et pour l'efficacité et le développement des outils d'ingénierie. Le pari en est que c'est l'application sincère du test de corroboration sur des exemples non ciblés de manière ad hoc, qui va permettre d'avancer et dans la linguistique conçue en tant que science de l'empirie, et dans le domaine de l'ingénierie, sans que, dans l'état actuel du problème, aucune « primauté » ne soit donnée ni à l'une ni à l'autre voie. Cela exige une explicitation exhaustive des descriptions linguistiques et des outils informatiques utilisés, et une analyse

sincère des résultats obtenus, afin de pouvoir détecter les limites des résultats obtenus et de pouvoir les interpréter correctement.

3. Environnements informatiques d'aide à l'apprentissage et à l'enseignement de la langue (contact : M. Chambreuil)

Cette thématique regroupe trois projets.

Un premier projet, « Discours oral pour l'apprentissage d'une langue étrangère », concerne la mise en place d'une nouvelle génération de laboratoires de langues, centrés sur la langue orale, s'appuyant sur les nouvelles technologies d'information et de communication ainsi que sur les recherches sur les environnements multimédias d'aide à l'apprentissage.

Un second projet, « AMAL » (Aide Multimédia à l'Apprentissage des Langues), porte sur des recherches théoriques sous-jacentes à la réalisation de logiciels favorisant la compréhension orale et l'acquisition de vocabulaire pour des étrangers apprenant le Français. Il porte aussi sur la réalisation effective de ces logiciels multimédias.

Le troisième projet, « AMICAL » (Architecture Multi-agents Interactive Compagnon pour l'Apprentissage de la Lecture), a pour objectif de mettre, au service de l'apprentissage et de l'enseignement de la lecture, les recherches qui peuvent être conduites aujourd'hui dans les perspectives des sciences cognitives et des environnements informatiques, à base de connaissances, d'aide à l'apprentissage.

Dans le cadre de cette présentation du LRL, ce sont essentiellement des orientations de recherche du projet AMICAL qui seront évoquées, sachant que différentes questions abordées dans ce projet se retrouvent dans les deux autres.

Face à la diversité des situations d'apprentissage de la lecture, pour de multiples raisons, scientifiques et sociales,

depuis 1991 les recherches du projet AMICAL portent essentiellement sur une situation correspondant à un apprentissage initial de la lecture, en milieu médiateur, pour des enfants au tout début de l'apprentissage.

Les recherches fondamentales conduites dans le projet AMICAL correspondent à un double choix. Le premier, évoqué ci-dessus, est celui d'essayer d'apporter une aide à l'apprentissage de la lecture. Le second, lui, a pour objectif une théorisation de différents éléments nécessairement présents dans un environnement informatique pour un apprentissage humain individualisé. Ce choix est d'aborder ces éléments et leur complexité à partir de l'étude approfondie du domaine d'apprentissage retenu dans le premier choix.

Ces recherches fondamentales conduisent également à des recherches appliquées correspondant, en particulier, à la réalisation de prototypes susceptibles de permettre une évaluation d'hypothèses faites, mais aussi des expérimentations en milieu classe.

Un environnement AMICAL regroupe trois grands types de module considérés comme complémentaires dans une situation d'apprentissage.

- Un module de type ressource donne un accès libre, accompagné d'un guide, à des informations.
- Un module de type exploration donne accès à un micro-monde d'entités et de processus combinatoires de ces entités.
- Un module de type tutoriel conduit un enseignement individualisé pour chaque apprenant.

A ce jour, les recherches ont porté principalement sur les modules ressource et tutoriel. Par rapport à l'IA et dans les limites de cette présentation, ce sont des éléments des problématiques liées à un module tutoriel qui seront évoquées à partir de la présentation d'un cycle fonctionnel d'enseignement.

En simplifiant, un cycle fonctionnel d'enseignement peut être décrit de la façon suivante. L'objectif individualisé d'une session de travail avec un apprenant est déterminé à partir de la représentation de cet apprenant et de différentes connaissances liées au domaine d'apprentissage, à son apprentissage et à des connaissances pédagogiques plus générales. Une séquence d'activités didactiques susceptible de permettre d'atteindre l'objectif retenu est ensuite déterminée, là encore de manière spécifique à l'apprenant concerné. On est là, en fait dans la modélisation de la préparation d'une session de travail avec un apprenant particulier. La séquence retenue sera ensuite réalisée avec l'apprenant et chaque activité didactique donnera lieu à un compte-rendu correspondant à la mémorisation des interactions entre le module et l'apprenant. C'est ensuite une interprétation qualitative de ces compte-rendus qui conduira à une nouvelle représentation de l'apprenant concerné et donc bouclera un cycle fonctionnel.

Un cycle fonctionnel d'enseignement est donc défini à partir de trois grands types d'entités correspondants:

- à la représentation que le module a de l'apprenant,
- à l'objectif d'une session de travail avec l'apprenant,
- à une activité didactique.

En fonction de ces trois entités, des différentes prises de décision sous-jacentes au cycle fonctionnel et des connaissances expertes mises en jeu dans ces prises de décision, un module tutoriel est conçu comme une architecture multi-agents. A un premier niveau sont définis sept agents rationnels :

- Agent Gestionnaire de la Planification Didactique (AGPD) déterminant et construisant l'objectif d'une session de travail et la séquence d'activités didactiques associée à l'objectif,
- Agent Gestionnaire des Situations Didactiques (AGSD) exécutant la

séquence d'activités didactiques et mémorisant les comptes-rendus,

- Agent Gestionnaire des Représentations de l'Elève (AGRE) interprétant qualitativement les comptes-rendus dans une construction dynamique de la représentation de l'élève,
- Agent Expert Linguiste Lecture (AELL) donnant accès aux connaissances expertes en linguistique liées à l'apprentissage de la lecture,
- Agent Expert en Apprentissage de la Lecture (AEAL) donnant accès aux connaissances sur l'apprentissage de la lecture,
- Agent Expert Pédagogue (AEP) donnant accès à des connaissances pédagogiques indépendantes du domaine d'apprentissage,
- Agent de Contrôle des Sessions Didactiques (ACSD) jouant aussi le rôle de médiateur gérant la communication entre les agents de premier niveau.

Chacun de ces agents peut lui-même être envisagé comme une architecture multi-agents. C'est par exemple le cas de AGPD et de AGRE. Ce peut aussi être le cas des interfaces apprenant-système.

Par rapport à la modélisation et à la réalisation de ces agents, c'est, en fait, chacune des entités du cycle fonctionnel et chacune des prises de décisions pour la réalisation du cycle qui définissent les problématiques de recherches dans AMICAL.

Par exemple, pour la représentation de l'apprenant, des recherches déjà développées ou en cours portent sur :

- les types de représentation (sur le domaine d'apprentissage, comportementale, ...),
- les connaissances présentes dans ces représentations,
- la structure de ces représentations,
- les processus de construction de ces représentations, de gestion de leur évolution, de gestion des contradictions,
- les processus d'analyse d'erreurs,

d'analyse de stratégies.

Par exemple encore, pour l'objectif d'une session de travail, on peut noter des travaux sur les points suivants :

- la structure d'un objectif comme ensemble de couples ou de triplets mettant en jeu des connaissances du domaine d'apprentissage, des actions sur ces connaissances, des statuts d'évaluation de ces connaissances,
- l'ensemble des connaissances expertes mises en jeu dans les différentes phases de détermination de l'objectif (détermination des possibles, choix parmi l'ensemble des possibles),
- les processus de modifications de l'objectif à différentes étapes du cycle fonctionnel qui en fait n'a pas la linéarité de la présentation simplifiée évoquée ci-dessus, ...

On peut aussi évoquer différentes questions et des travaux liés à l'entité fondamentale constituée par une activité didactique. Une activité didactique doit être envisagée sous différents points de vue tels que, par exemple :

- une structure d'actions exercées sur l'apprenant,
- un problème à résoudre par l'apprenant,
- un espace d'interaction avec l'apprenant,
- un espace d'observation de l'apprenant.

Ces points de vue renvoient à leur tour à de multiples questions, par exemple, sur les paramètres susceptibles d'instanciations individualisées dans une activité, sur le dialogue pédagogique (consigne, aide, commentaires, ...), sur l'ergonomie des interfaces, sur des outils de conception et de développement des interfaces, ...

Par rapport aux travaux sur les activités didactiques, une bibliothèque d'activités susceptibles d'être mises en œuvre pour le tout début d'apprentissage a déjà fait l'objet de développements. Cette bibliothèque a donné lieu à différentes expérimentations en milieu classe permettant, en particulier d'analyser :

- le comportement des apprenants par rapport au domaine d'apprentissage,
- le comportement des apprenants par rapport aux interfaces d'interaction,

- sur plusieurs cycles fonctionnels, la pertinence de séquences d'activités didactiques individualisées par rapport à différents profils d'apprenants.

Les quelques points évoqués ci-dessus laissent imaginer la complexité d'un module tutoriel gérant un enseignement individualisé. Les recherches conduites dans le cadre du projet AMICAL renvoient donc à des recherches fondamentales multiples et complémentaires correspondant en fait à la réalité du problème abordé. Ces recherches portent nécessairement sur l'apprentissage de la lecture, mais au-delà elles rejoignent des problèmes largement indépendants du domaine d'apprentissage retenu et présents dans un cycle fonctionnel d'un enseignement individualisé. Elles peuvent aussi conduire à des hypothèses sur des outils d'ingénierie des connaissances orientés vers les connaissances expertes qui doivent être présentes dans un module tutoriel d'enseignement individualisé. Des recherches, dans le cadre du projet, portent également sur une théorisation d'agents rationnels intentionnels susceptibles de correspondre à différents agents de l'architecture du

MIG : Unité Mathématique, Informatique et Génome, Institut National de la Recherche Agronomique Activité en IA - le projet Caderige

Introduction

Ce chapitre décrit les activités de l'unité MIG dans le domaine de l'IA. Plus de détails sur les membres de MIG, les collaborations et publications pourront être trouvés sur le site www.mig.jouy.inra.fr. Ces activités sont indissociables du projet Caderige (<http://caderige.imag.fr>), projet bioinformatique inter-EPST lancé en 1999 et auquel collaborent les laboratoires informatiques suivants, LEIBNIZ CNRS de l'IMAG, Grenoble, LIPN

CNRS, Université de Villeneuve Paris 13, LRI CNRS, Université d'Orsay Paris 11 pour le traitement automatique de la langue et l'apprentissage automatique, le Projet Symbiose de l'IRISA, INRIA, Rennes et l'Unité MIG, INRA, Jouy-en-Josas, pour la bioinformatique, transcriptome, protéome et *biologie* et enfin, l'unité INRA-ENSAR de Génétique Animale, Rennes. L'objectif du projet Caderige est la conception d'outils logiciels pour l'exploitation de documentation scientifique par la sélection de documents

pertinents et l'extraction d'information.

De nouvelles technologies de la recherche documentaire et d'extraction d'information

Aujourd'hui les moyens automatiques d'accès à l'information scientifique sont essentiellement limités à la recherche documentaire à l'aide de requêtes, de mots-clefs et d'index. L'imprécision des résultats rend la tâche d'exploitation des documents sélectionnés coûteuse voire infaisable. Pourtant l'exploitation de la documentation scientifique par les

scientifiques eux-mêmes et par les techno-veilleurs est un facteur crucial de progrès et d'innovation. L'accès automatisé à l'information réclame de nouvelles technologies dans les domaines de la recherche documentaire et de l'extraction d'information.

Dans un cas comme dans l'autre, deux problèmes clefs sont à résoudre.

→ Pour répondre à l'exigence de pertinence de la recherche documentaire comme de l'extraction d'information, il faut mettre en œuvre des moyens d'analyse du contenu reposant sur des ressources lexicales, syntaxiques et sémantiques spécifiques au domaine étudié. Des représentations normalisées et enrichies des documents peuvent alors être automatiquement construites sur lesquelles peuvent s'appliquer des moyens de sélection de documents et d'extraction d'information. Outre une meilleure précision, ces moyens d'accès à l'information permettent alors de répondre à d'autres besoins que d'exploration thématique, par exemple, des besoins de fiabilité, d'originalité, et de synthèse.

→ Les ressources sont spécifiques au domaine et les moyens de sélection et d'extraction sont spécifiques à la tâche. Les uns et les autres sont longs et coûteux à acquérir manuellement et l'acquisition manuelle pêche généralement par manque d'exhaustivité. L'application de l'apprentissage automatique à partir de corpus textuels dans les domaines spécifiques est une alternative reconnue mais encore débutante.

Les travaux récents en apprentissage de ressources ontologiques et terminologiques et de règles d'extraction vont dans ce sens, les partenaires de Caderige y ont contribué. Leur dispersion et leurs limitations empêchent encore une utilisation automatisée et efficace. Des étapes clefs sont manquantes. Principalement, il s'agit,

→ de la prise en compte des ressources existantes et de leur nature, pour l'apprentissage de ressources spécifiques

à la tâche, que ces ressources soient d'ordre lexical, conceptuel ou bibliographique.

→ et de l'utilisation des ressources, acquises ou apprises, pour l'apprentissage de règles d'extraction, et de critères de sélection de documents.

Caderige a pour objectif de concevoir et d'adapter des outils logiciels d'apprentissage automatique de ressources lexicales, syntaxiques et sémantiques à partir de textes, capables de prendre en compte des ressources préexistantes, et des outils logiciels d'apprentissage de critères de sélection de documents et d'extraction d'information basés sur ces ressources [Bessières et al., 2001]. MIG est particulièrement actif dans les activités de recherche en apprentissage automatique appliqué au texte et en génomique.

L'exploitation de la documentation scientifique dans le domaine exemplaire de la génomique

Les approches développées dans Caderige sont guidées et validées par les besoins identifiés en génomique fonctionnelle, tout en conservant leur applicabilité dans les domaines scientifiques présentant les mêmes spécificités, domaine scientifique expérimental, communication scientifique organisée en bases de connaissances distribuées et hétérogènes, tentatives de normalisation terminologique et ontologique. Citons par exemple la médecine et l'astrophysique. Du point de vue documentaire, la génomique fonctionnelle forme un domaine exemplaire à plusieurs titres.

L'essor exponentiel des méthodes expérimentales en biologie moléculaire a profondément modifié les pratiques des chercheurs au cours de la décennie précédente. Traditionnellement, un projet de recherche en biologie moléculaire s'intéresse à un petit nombre d'objets, gènes ou protéines, responsables d'une fonction biochimique ou physiologique précise. Auparavant, quelques centaines de publications scientifiques en moyenne se rapportaient au système étudié, et pouvaient être traitées manuellement.

Cette manière de procéder n'est plus praticable avec la mise en œuvre d'outils expérimentaux qui permettent des analyses globales à l'échelle des organismes, comme les microbes, ou d'organes complexes, comme le foie ou le cerveau. À ce niveau, le volume de la bibliographie concernée par le sujet de l'expérience atteint des milliers, voire quelques dizaines de milliers de publications. La cause en revient au démarrage de projets ambitieux comme le séquençage du génome humain, c'est-à-dire la lecture complète de son information génétique, et qui ont exercé une très forte demande sur les techniques. Aujourd'hui, il est désormais possible, pour un individu ou une petite équipe, d'analyser simultanément l'activité de plusieurs milliers de gènes dans une cellule ou un organisme biologique, comme c'est le cas avec les « puces à ADN ». De plus, beaucoup de connaissances fines, comme les régulations de l'expression des gènes et de l'activité des protéines, ne sont décrites que dans ces publications, sous la forme de texte non structuré, or, ces connaissances sont absolument nécessaires pour interpréter et analyser les interactions dans les systèmes biologiques de façon pertinente, à partir des millions de mesures élémentaires que peut produire annuellement un laboratoire muni de ces nouveaux dispositifs expérimentaux.

C'est donc l'adaptation des techniques du laboratoire aux analyses à grande échelle, et corrélativement l'explosion de leurs performances, qui a suscité l'intérêt des biologistes pour l'extraction de connaissances textuelles.

C'est pourquoi Caderige vise à fournir des outils qui contribuent à faire le lien entre les données factuelles issues des expériences à grande échelle et structurées en bases de données, les connaissances synthétiques disponibles et les connaissances de la littérature scientifique. Notons que les méthodes mises au point pour repérer, extraire et représenter l'information pertinente doivent posséder une grande précision et une grande sensibilité pour être compatibles

avec le degré d'automatisation élevé requis par le volume des données textuelles à traiter.

Le problème ne se pose pas particulièrement pour la génomique fonctionnelle : il est général dans le champ de la biologie. Bien d'autres domaines des sciences et des techniques sont porteurs d'une telle demande, et la biologie n'est qu'un des secteurs dans lequel elle s'y est d'abord manifestée. Plusieurs facteurs expliquent cette précocité, et ils justifient en même temps que la génomique soit l'objet d'une attention particulière de la part des milieux de la recherche en informatique. Tout d'abord l'information textuelle en biologie moléculaire est traditionnellement très bien structurée dans des journaux scientifiques, et ces derniers sont bien repérés par la communauté des biologistes. L'essor d'Internet et des services WWW a également favorisé l'apparition de collection de données factuelles et de connaissances en ligne. Elles sont aussi bien identifiées par les biologistes, et des accords entre leurs auteurs et les éditeurs des revues scientifiques en assurent l'alimentation par les laboratoires (une publication reposant sur l'analyse d'une nouvelle séquence ne sera pas publiée si la séquence n'a pas auparavant été déposée par les biologistes dans la collection ad hoc).

Pratique collaborative et intégrative

Ces données factuelles de séquences, d'expression de gènes, etc., incorporent également beaucoup d'annotations textuelles, résultats du travail d'analyse par les chercheurs, ainsi que des liens vers la bibliographie. Enfin, les collections en ligne renfermant des informations complémentaires se réfèrent mutuellement de manière systématique, ce qui explique l'existence aujourd'hui d'une fédération navigable de plusieurs centaines de collections WWW pour la biologie moléculaire. A la lueur des efforts de cette communauté pour créer et maintenir des accès publics aux résultats de la recherche scientifique, c'est donc tout naturellement qu'ont progressive-

ment émergé des projets collaboratifs de développement de logiciels, stimulés par l'apparition de nouvelles normes de médiation informatique contribuant à l'échange, l'interopérabilité et l'intégration de l'information, comme CORBA et XML. Le projet SpiD développé par MIG en est un exemple [Hoebecke et al., 2000]

Au-delà de l'exploitation endogène de la documentation scientifique proposée par l'état de l'art, le problème posé ici de façon exemplaire par la biologie moléculaire est la prise en compte de connaissances externes abondantes, faciles d'accès, partiellement redondantes et très hétérogènes quant à leur représentation, leur stabilité, leur généralité et leur fiabilité.

La contribution des outils d'exploitation documentaire à l'interopérabilité et l'intégration de l'information dans les domaines scientifiques ne pourra être efficace qu'à la condition d'exploiter les ressources existantes dès le niveau amont par l'apprentissage de moyens de sélection et d'extraction bibliographique et pas seulement, au niveau aval, par la seule confrontation des résultats de sélection et d'extraction à ces ressources.

Caderige se concentre donc sur la conception d'outils d'apprentissage de moyens de sélection de documents et d'extraction d'information dans ces documents (règles, distances, index, etc.) à partir de corpus documentaire d'apprentissage et de ressources externes.

Objectifs de recherche technologique du projet

État de l'art et intérêt du projet

Dans certains domaines, la documentation scientifique est abondante et joue un rôle clef dans le travail des acteurs. L'exploitation de cette documentation est nécessaire à l'activité de veille scientifique, permet de confronter des résultats expérimentaux entre eux et contribue à l'innovation scientifique.

Un effort important est d'ailleurs fait pour rendre accessible cette documentation scientifique au plus grand nombre. La médecine et le génomique ont sans doute été des domaines précurseurs à cet égard : constitution de bases bibliographiques mettant en ligne les résumés de millions d'articles scientifiques (Medline), indexation de ces bases documentaires à l'aide de thesaurus spécialisés (MeSH, Snomed...). D'autres domaines comme l'astronomie suivent ce chemin, en prenant appui sur l'expérience de Medline.

Recherche documentaire

Actuellement, les outils d'exploration de ces bases documentaires reposent essentiellement sur des méthodes de recherche d'information, sur des mesures statistiques et, dans le meilleur des cas, sur une indexation à base de mots-clefs. Ils répondent mal aux besoins des utilisateurs. Ces méthodes à bases de mots-clefs s'avèrent inadaptées à la fois en termes de couverture et de fiabilité. Elles ramènent souvent beaucoup trop de documents par rapport à ce que l'utilisateur est en mesure de parcourir et des documents pertinents peuvent être omis.

Le projet Caderige propose une approche plus ambitieuse de l'exploration de la documentation scientifique. Il ne s'agit pas seulement de retrouver les documents pertinents, mais de localiser l'information pertinente, et de la formaliser. L'approche de Caderige repose sur des techniques d'extraction d'information. Nous défendons l'idée que cette approche est viable dès lors qu'elle est guidée par des connaissances sur la langue et le domaine considéré.

Extraction de connaissances

L'extraction d'information (*information extraction*) désigne l'activité qui consiste à remplir automatiquement des formulaires ou une banque de données à

partir de textes écrits en langue naturelle. Elle s'oppose classiquement à la recherche documentaire (*information retrieval*), qui vise à retrouver dans une base de documents un ensemble de documents pertinents au regard d'une question. L'extraction met en œuvre une analyse du texte pour produire du factuel (remplissage d'un formulaire prédéfini, en anglais *template*) : il ne s'agit pas de donner du texte brut à l'utilisateur mais d'apporter des réponses précises aux questions qu'il pose. De ce point de vue, l'extraction se situe dans le prolongement des travaux en compréhension automatique de textes.

Si l'extraction d'information a, de fait, vu le jour dès les années 60, ce sont les « Message Understanding Conferences » (MUC) [MUC Proceedings], une série de conférences organisées depuis la fin des années 1980 à l'initiative de différentes institutions américaines pour l'évaluation des systèmes de compréhension de messages, qui ont considérablement réactivé ce courant de recherche. Par exemple, pour MUC-4 (1992), la quatrième conférence, l'objectif était d'extraire des informations concernant les scénarios terroristes à partir de textes relatant des récits d'attentats. Il s'agissait de retrouver des informations sur l'attentat (date, lieu, type), sur ses auteurs (nom et type d'organisation) et sur ses victimes (nom, qualités, degré de la blessure), de manière à instancier un schéma informationnel fixé au départ.

Le processus d'analyse est guidé par la connaissance *a priori* du type des informations recherchées. Il s'appuie sur des structures et des indices de surface pour sélectionner dans les textes, les passages pertinents qui sont ensuite analysés en détail. La stratégie généralement adoptée pour répondre aux campagnes d'évaluation MUC repose sur une analyse strictement locale centrée sur des mots déclencheurs connus, à partir d'automates locaux en cascade [Hobbs et al. 1997]. Ces automates encodent des patrons ou schémas de phrases (*extraction patterns*) qui servent

à extraire des textes les informations recherchées. Il s'agit souvent de patrons lexico-syntaxiques, i.e. de structures syntaxiques comportant certains éléments lexicaux.

Les approches en la matière achoppent sur l'élaboration très souvent manuelle de ces bases de patrons d'extraction. Écrire ces patrons est un travail long et fastidieux et pour lequel il est difficile de garantir une bonne couverture. Le coût de ces approches en rend donc l'opérationnalisation difficile sur des applications importantes.

Dans des domaines scientifiques et techniques, la nature de la tâche d'extraction change en raison du type d'utilisateur, de la nature des informations recherchées et de la nature de la base documentaire rendant à la fois nécessaire et réaliste l'utilisation de méthodes plus sophistiquées. Par rapport à l'état de l'art, l'objectif du projet Caderige est plus ambitieux en termes de la précision et l'expressivité de l'extraction. Les résultats devant être exploités sur des domaines limités et des langues spécialisées, Caderige peut mettre en œuvre des techniques plus sophistiquées d'acquisition des ressources linguistiques et du domaine nécessaires à l'extraction.

Acquisition et apprentissage des ressources

Les systèmes d'exploration textuelle basés sur une analyse superficielle ont donc révélé leurs limites en termes de portabilité et d'évolution par rapport aux besoins des utilisateurs. Parallèlement les nouveaux courants de recherche centrés sur la notion de corpus soulignent que chaque type de texte possède ses particularités propres. La conséquence principale de ce double constat est la nécessité d'élaborer des systèmes dont les ressources et les traitements peuvent s'adapter dynamiquement au corpus. Les connaissances nécessaires aux systèmes d'extraction d'information pour reconnaître l'information pertinente et l'interpréter au regard d'un objectif,

dépendent à la fois du domaine et du besoin.

L'apprentissage automatique, discipline qui a pris son essor dans les années 80, permet d'acquérir de telles connaissances à partir d'exemples. Les systèmes d'apprentissage symboliques et relationnels ont montré leur capacité à acquérir des connaissances structurées tout en prenant en compte les connaissances du domaine et des besoins quand elles sont disponibles. Ces approches présentent également l'avantage d'offrir une bonne compréhensibilité des méthodes et des résultats, compréhensibilité qui facilite d'autant la mise au point et la validation. Parallèlement, les méthodes d'apprentissage numériques basées, peu ou prou, sur l'analyse de données sont capables de chercher des régularités dans de grosses masses d'exemples augmentant ainsi la validité statistique des résultats. Ces aspects, expressivité, compréhensibilité et robustesse, sont nécessaires à l'extraction d'information dans la documentation scientifique. Ainsi, les premiers travaux sur la définition de systèmes hybrides pour la fouille de données textuelles ou non (*Data / Text Mining*) qui intègrent les avantages des deux classes sont un succès. L'application de ces méthodes pour l'extraction d'information est une des conditions nécessaires à la réutilisabilité et à l'adaptation des systèmes d'extraction.

Par ailleurs, la spécificité des langues de spécialités rend les dictionnaires et ressources lexicales existantes de la langue générale peu ou pas exploitables. Comme l'a mis en évidence Harris [Harris et al. 89] en immunologie, la variabilité des sous-langages utilisés dans les domaines de recherche spécifiques est limitée à la fois du point de vue du vocabulaire, de la polysémie, des formes syntaxiques et du nombre des concepts représentés. Il est donc à la fois nécessaire et réaliste de vouloir acquérir automatiquement les ressources lexicales, sémantiques et conceptuelles nécessaires à une analyse profonde, ceci à partir des régularités observées dans un corpus. Ensuite, grâce à ces res-

sources, par l'intermédiaire de transformations, on peut ramener toute phrase d'un corpus de spécialité à une forme canonique qui représente son contenu informationnel en termes d'opérateurs (les prédicats) et d'arguments. Cette approche n'a jamais été mise en œuvre de manière automatique.

Différents types de ressources doivent être intégrées dans les outils d'exploration de documentation : les noms propres ou noms des entités mentionnées dans les documents (la nomenclature), les expressions clefs du domaine (la terminologie), les concepts du domaine et les relations qu'ils entretiennent (l'ontologie), une grammaire décrivant comment s'expriment les concepts du domaine sur le plan linguistique (les schémas prédicatifs) et enfin les règles d'extraction (patrons d'extraction).

Nomenclature

La mise au point de nomenclatures ou de dictionnaires d'entités nommées est l'une des tâches qui a été le plus étudiée dans le cadre des travaux en extraction d'information [Bikel et al., 97]. Pour un corpus de documentation technique, dans des domaines où il existe des nomenclatures (noms de médicaments, noms de protéines, ...), l'enjeu consiste à étendre la nomenclature aux éventuels nouveaux noms et à reconnaître les variantes de ces noms (variation orthographique, synonyme...).

En biologie moléculaire par exemple, une base de donnée telle que SwissProt référence de nombreux noms de gènes et les variantes les plus courantes. Reste que le biologiste utilise pour rédiger des variantes personnelles d'ordre généralement typographique (un tiret au lieu d'une parenthèse) et que toutes les entités ne sont pas référencées. MIG est impliqué dans le projet inter-EPST bioinformatique BioMire dont l'objectif est de développer des moyens automatiques d'identifier les entités non référencées

dans les bases de données et leurs variantes.

Terminologie

Les années 1990 ont connu un effort important en matière d'acquisition de terminologie spécialisée (cf. les travaux du groupe TIA¹). Les outils d'extraction terminologique (Lexter [Bourigault, 1986], Faster [Jacquemin, 1996], Acabit [Daille, 1995], Xtract [Smadja, 1995], etc.) ont fait leurs preuves. Ces outils produisent des listes d'expressions qui sont des candidats termes plausibles.

L'enjeu aujourd'hui consiste à valider cette première liste pour ne retenir que les termes effectivement pertinents du domaine et non fortuits (par exemple, *enzymatic pathway*, mais pas *SpoIIID expression*). La lourdeur de ce travail de validation est souvent, en pratique, un frein à l'essor de la terminologie computationnelle.

Dans le cadre du projet Caderige, une part importante de ce travail de validation peut être réalisée par des techniques de filtrage ou d'élimination des candidats termes. Plusieurs techniques de filtrage ont été proposées : sur la base de critères statistiques (Acabit, Xtract) ou syntaxiques (Xtract).

Une piste peu explorée mais prometteuse pour notre projet consiste à mettre au point les critères de pertinences statistiques à partir de liste de certifiées de termes, ce qui est possible dans des domaines scientifiques où des thesaurus existent.

Depuis le milieu des années 1990, les recherches ont également abordé la question de la structuration de terminologie par le repérage de variantes de termes, de relations de synonymie [Hamon & Nazarenko 2001] ou d'hyponymie [Morin & Jacquemin 1999]. L'intérêt et la pertinence de ces méthodes de structuration sont évalué dans le cadre de notre application en génomique fonctionnelle.

Ontologie

De très nombreux travaux portant sur les ontologies ont vu le jour depuis 15 ans. Parallèlement à des projets très

ambitieux de construction d'ontologies générales mais généralement peu exploitables pour des applications particulières (CYC, par exemple), un courant de recherche a défendu l'idée de l'acquisition d'ontologie à partir de corpus (travaux du groupe TIA, notamment). On a ainsi montré dans la lignée des travaux de Harris qu'une analyse distributionnelle d'un corpus spécialisé, en rapprochant les mots qui ont des emplois similaires, permet de faire émerger des classes sémantiques [Pereira et al., 1993] [Bensch & Savitch, 1995] [Habert & Nazarenko, 1996] [Faure & Nédellec, 1999] [Bisson et al., 2000]. Les méthodes de classification conceptuelle hiérarchiques telle que celle de ASIUM [Faure & Nédellec, 1999] permettent d'obtenir des graphes de classes acycliques orientés pour représenter la polysémie.

Aujourd'hui, ces travaux présentent une double limitation. Les outils existants construisent des classes de mots et non des classes de termes. Les unités de sens pertinentes sont pourtant souvent des termes complexes, expressions composées de plusieurs mots. Il faut donc pouvoir construire des classes de termes pour obtenir une description sémantique adéquate. De plus, il faut pouvoir combiner, voire amorcer ce processus d'acquisition en corpus avec des catégories prédéfinies de thesaurus ou d'ontologie quand elles existent. Nos recherches visent à dépasser ces limitations en intégrant la terminologie et les connaissances préalables à la méthode de classification conceptuelle de ASIUM. Le traitement de ces deux problèmes nécessite des adaptations des méthodes de classification conceptuelle utilisées de manière à prendre en compte à la fois la question de l'incrémentalité posé par l'utilisation de connaissances préexistantes et la question de la localité de la terminologie en fonction du contexte.

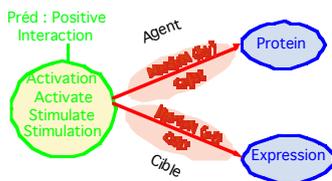
Dans le cadre de Caderige, ce type d'approche doit permettre de construire des classes sémantiques utilisées pour faire un étiquetage sémantico-conceptuel du corpus.

1. Terminologie et Intelligence Artificielle (<http://www.biomath.jussieu.fr/TIA/>)

Schémas prédicatifs

Associer des concepts à certains mots du corpus est une première étape de normalisation sémantique réalisée grâce à l'ontologie. On peut cependant aller plus loin grâce aux schémas prédicatifs, pour s'affranchir de la diversité des formulations des différents rédacteurs et construire des représentations enrichies et normalisées des documents. Un schéma prédicatif décrit la manière dont un prédicat ou une classe de prédicats s'expriment linguistiquement (nombre et type sémantique de ses arguments, nature et mode de construction syntaxique).

Par exemple,



La maîtrise de ces schémas prédicatifs permet ainsi de donner une unique représentation à un ensemble de phrases comme :

The expression of spoIIID stimulates the expression of sigK
 SpoIIID expression activates sigK expression
 SigmaK production is stimulated by the spoIIID gene product.

Si de nombreuses recherches en sémantique lexicale se sont consacrées à la formalisation et à la représentation de ces schémas prédicatifs (cf. [Pustejovsky, 1995]), peu de travaux se sont intéressés au problème de l'automatisation de leur acquisition, sinon au prix prohibitif d'un étiquetage sémantique de corpus d'apprentissage [Thomson, 95].

Les méthodes d'apprentissage étudiées par MIG sont combinaison d'inférence grammaticale pour apprendre les contraintes de structure des schémas, et de classification conceptuelle d'ordre un pour généraliser ces structures.

Apprentissage de règles d'extraction

L'originalité de notre projet tient au travail de normalisation et d'enrichisse-

ment du texte préalable à l'apprentissage des règles d'extraction. Ce prétraitement permet de souligner les régularités des textes au-delà de la diversité des formes de surface et ainsi faciliter l'apprentissage.

Les approches récentes en matière d'apprentissage de règles d'extraction [Nédellec, 2000] convergent vers l'utilisation de méthodes d'apprentissage descendantes relationnelles de type FOIL [Quinlan, 90], appliquées à des exemples de phrases *annotés*, [Califf & Mooney, 98], [Freitag, 98], [Soderland, 99], [Ciravegna, 2000], c'est-à-dire dont les valeurs des champs à remplir (information à extraire) sont marquées, par exemple, les protéines, gènes, fonctions et natures des interactions. Conçues pour s'appliquer dans des domaines généraux ou pour extraire des informations superficielles (date, noms, etc.), ces méthodes n'utilisent que peu ou pas de ressources lexicales. Caderige ambitionne de traiter des domaines où la subtilité de l'informa-

occurrence statistique [Craven, 99], [Stapley & Benoit, 2000], [Pillet, 2000].

Technologies

Approche générale

Le projet Caderige aborde trois méthodes différentes d'exploitation d'une documentation scientifique : il s'agit de sélectionner les textes pertinents, de sélectionner les fragments les plus intéressants et enfin d'extraire de la connaissance véhiculée par les textes pour en donner une représentation formelle.

Ces trois méthodes sont complémentaires et reposent sur des techniques apparentées. Caderige les intègre dans un processus en trois étapes (fig. 1). Des traitements de complexité croissante sont effectués à chaque étape, mais les traitements intermédiaires apportent déjà une aide appréciable en matière d'exploration documentaire :

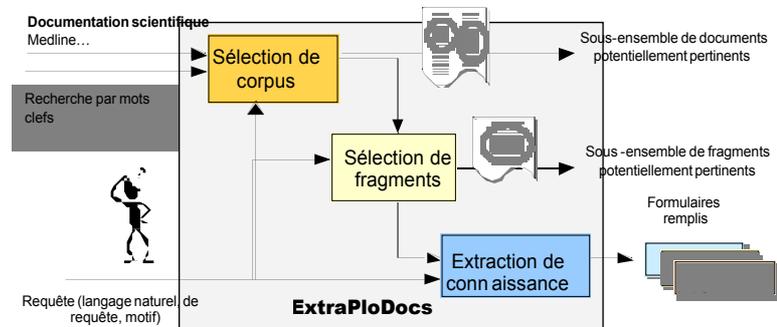


Figure 1 : Exploitation de documentation dans Caderige

tion nécessite l'exploitation de ces ressources et d'adapter les méthodes courantes à cette nouvelle problématique.

En biologie moléculaire, l'état de l'art est bien en-deçà [Nédellec, 2002]. Les deux approches majoritaires consistent à construire des patrons manuellement [Blaschke *et al.*, 99], [Thomas *et al.*, 2000], [Ono *et al.*, 2001], et le taux de rappel (de couverture), se situe par exemple entre 0 % et 30 % pour le problème d'identification des interactions géniques, ou à construire les patrons de faible précision sur des bases de co-

- La première étape « Sélection de corpus » consiste à constituer un corpus en sélectionnant au sein de la base documentaire les documents effectivement pertinents au regard de la requête formulée par l'utilisateur. Ces corpus sont directement exploitables par les utilisateurs comme source d'information et ils constituent le corpus approprié sur lequel devra porter l'extraction automatique d'information. Cette sélection repose sur un classifieur de documents qui dépend lui-même des types de documents consi-

dérés et des critères de pertinence de l'utilisateur.

- La deuxième étape « Sélection de fragments » permet d'identifier dans les documents du corpus sélectionné à l'étape précédente, les portions de texte qui concernent effectivement le problème applicatif posé. Ils sont généralement très dispersés, par exemple les phrases sur les interactions géniques forment 3 % des résumés sur le sujet. Cette étape de filtrage est donc essentielle, que ce soit pour présenter directement les fragments sélectionnés à l'utilisateur sous la forme de texte surligné et comme une aide à la lecture des documents, ou pour aborder le traitement lourd d'extraction de connaissance proprement dit.

Prenons comme exemple le domaine de la génomique fonctionnelle qui est l'application test de Caderige. Pour identifier les interactions fonctionnelles entre les objets biologiques, problème clef de la génomique fonctionnelle, il est important de savoir repérer dans les résumés d'articles scientifiques de Medline des fragments tels que celui de la figure 2 (colonne de gauche).

De la même manière que pour la sélection de corpus, cette sélection de fragments repose sur un classifieur de phrases [Nédellec et al, 2001a, 2001b].

- La troisième étape « Extraction de connaissance » permet d'extraire et de donner une représentation formelle de l'information pertinente. La figure 2 montre quel type de connaissance le module d'extraction vise à extraire. Cette connaissance est représentée sous la forme de formulaires (colonne de droite sur la figure 2). Du fragment 2, on retient ainsi l'information que HIPK2 appartient à la famille des kinases sérine/threonine et que HIPK2 interagit positivement (activation) sur le gène p53.

De manière classique, cette étape d'extraction repose sur un moteur d'ex-

Fragment 1

Furthermore, HIPK2 and p53 cooperate in the activation of p53-dependant transcription and apoptotic pathways.

Fonctionnal role

Source : protein HIPK2
Target : apoptosis

Fonctionnal role

Source : protein p53
Target : apoptosis

Fragment 2

Here we demonstrate that homeodomain-interacting protein kinase-2, a member of a novel family of nuclear serine/threonine kinases, activates p53 by directly phosphorylating it at Ser 46

Membership

Element : protein kinase-2
Family : serine /threonine kinases

Figure 2. Des fragments de texte, aux formulaires remplis.

traction qui confronte au texte un ensemble de règles, de patrons ou schémas de phrases (*extraction patterns*) qui modélisent les informations recherchées (exemple, figure 4).

Extraction guidée par des ressources

Les trois logiciels qui effectuent ces traitements (classifieur de document, classifieur de phrase, et moteur d'extraction) intègrent des connaissances. Nous parlerons ici de *ressources* pour distinguer clairement les connaissances intégrées dans les outils d'exploitation de documents et les connaissances que ces outils servent à extraire.

La couverture et la pertinence des résultats obtenus par l'exploitation dépendent de manière cruciale de la richesse et de la couverture de ces ressources. Ces ressources à leur tour dépendent à la fois du domaine considéré, du type des documents traités et de la nature des informations recherchées. Généralement, ces ressources se limitent à des mots-clefs ou sont construites manuellement. Dans le premier cas, les ressources sont pauvres et rapides à construire. Dans le deuxième cas, les ressources sont beaucoup plus riches, mais très coûteuses à construire et il est difficile d'en garantir la couverture et la fiabilité.

L'enjeu du projet Caderige est d'éla-

borer ces ressources par des techniques d'apprentissage automatique combinées avec des méthodes de traitement automatique de la langue pour permettre de construire à un coût raisonnable des ressources riches et avec une bonne couverture.

De manière plus technique, l'architecture du projet Caderige s'articule autour de deux grands axes, un axe exploitation et un axe acquisition. L'*extraction de connaissance* extrait les informations des textes et fragments de textes pertinents et les formalise. En d'autres termes, ce module réalise les trois étapes de sélection et d'extraction que nous venons de présenter. Le module d'*acquisition des ressources*, d'autre part, permet d'apprendre ou de modéliser de manière semi-automatique l'ensemble des ressources nécessaires à cette extraction. En pratique, l'utilisateur exploite le module d'extraction lequel fait appel aux ressources acquises. Le module d'acquisition est utilisé en amont chaque fois qu'il est nécessaire de modéliser un domaine ou de remettre à jour ce modèle.

Dans le projet Caderige, nous mettons l'accent sur le module d'acquisition des ressources. Le module d'extraction de connaissances sera mis au point à partir des techniques de l'état de l'art pour valider notre approche et évaluer les résultats obtenus dans le module d'acquisition.

Apprentissage de ressources

L'apprentissage des ressources est donc l'objet principal du projet Caderige. Deux types de ressources sont apprises.

Les ressources liées à l'extraction proprement dite sont les suivantes :

- Pour le classifieur de documents et le classifieur de phrases, il s'agit essentiellement des critères de pertinence (ou critères de sélection).
- Pour le moteur d'extraction, il s'agit de patrons d'extraction. L'ambition de Caderige est de concevoir un outil d'apprentissage de patrons qui permette de construire à un coût raisonnable des patrons expressifs (au-delà des mots-clefs ou de la cooccurrence de mots-clefs).

Dans Caderige, ces ressources sont apprises à partir de corpus qui sont pré-traités. Le pré-traitement de ces corpus fait lui-même appel à d'autres ressources, c'est-à-dire :

- les nomenclatures et terminologies spécialisées,
- l'ontologie décrivant les concepts du domaine,
- la description des schémas prédicatifs de la langue considérée.

Normalisation et enrichissement de corpus

L'une des originalités de l'approche de Caderige tient aux différentes étapes de normalisation et d'enrichissement de corpus. On peut ainsi donner différentes images d'un même texte, chacune présentant un certain niveau de représentation et d'abstraction par rapport à la for-

mulation originale. La figure 3 ci-dessous montre quelle représentation peut être construite pour le fragment de texte « csp Ap directs the expression of the cspA gene ».

La normalisation soulignant les régularités dans les textes, elle facilite les étapes d'apprentissage. Elle permet également d'exprimer les règles de sélection ou d'extraction à un niveau de généralité approprié et les rend ainsi plus précises et plus compréhensibles.

L'apprentissage de règles d'extraction de bonne qualité est facilité par l'explicitation de contraintes lexicales dans les exemples et cela évite de décliner le patron pour prendre en compte la diversité des formes lexicales.

Techniquement, cette normalisation repose sur un étiquetage terminologique, une analyse syntaxique, un étiquetage sémantico-conceptuel et une procédure de transformation syntaxico-sémantique pour réduire la complexité et la diversité syntaxique. Nous exploitons un analyseur syntaxique existant, Link Parser (www.link.cs.cmu.edu/link) mais les étiqueteurs du module de transformation sont développés dans le cadre du projet.

Acquisition des ressources nécessaires à la normalisation et à l'enrichissement des corpus

Il s'agit ici de l'acquisition de la terminologie, de l'ontologie et des schémas prédicatifs. Chacune des étapes est guidée par des ressources qui sont acquises automatiquement à partir de corpus et combinées aux connaissances du domaine.

L'acquisition terminologique de Caderige repose sur des outils existants et une méthodologie classique. L'essentiel de l'effort portera sur la mise au point de méthodes de filtrage de termes en combinant des critères statistiques, des contraintes structurelles et un vocabulaire contrôlé issu des thesaurus du domaine

Les innovations principales de Caderige concernent essentiellement l'acquisition d'ontologie et l'apprentissage de schémas prédicatifs.

Pour l'acquisition de l'ontologie du domaine, nous exploitons le logiciel ASIUM [Faure & Nédellec, 1999] et Mo'K [Bisson & Nédellec, 2001] en intégrant la terminologie à l'approche distributionnelle et en adaptant méthodes de classification conceptuelle pour prendre en compte certaines connaissances initiales existantes et ainsi intégrer une démarche incrémentale.

Notre méthode d'acquisition des schémas prédicatifs procède en deux étapes (cf. figure 4).

- Il s'agit d'abord d'apprendre des schémas de sous-catégorisation (contraintes sur le nombre, les types et les rôles des arguments d'un verbe ou d'un nom prédicatif) à partir des contraintes de sélection élémentaires qui sont apprises en parallèles des classes sémantiques par ASIUM. Des méthodes d'inférence grammaticale seront mises en œuvre pour apprendre ces contraintes de structures autour des verbes et des noms.

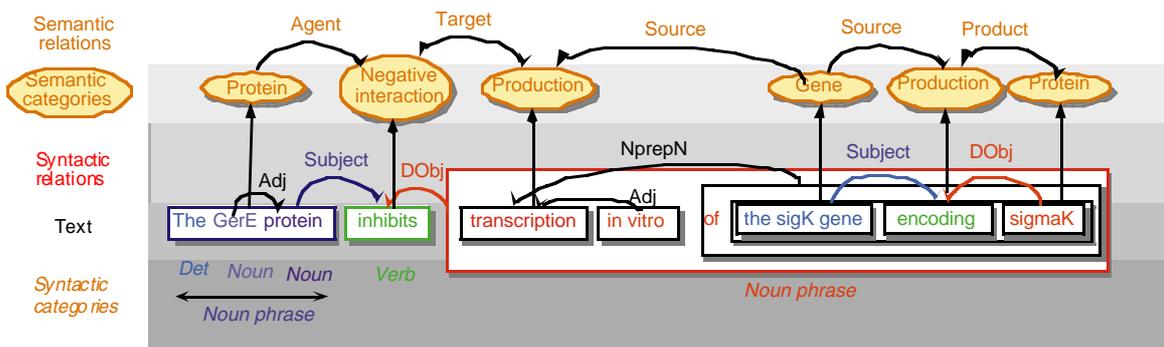


Figure 3 : Enrichissement d'un fragment textuel par l'ajout d'information de nature linguistique et conceptuelle

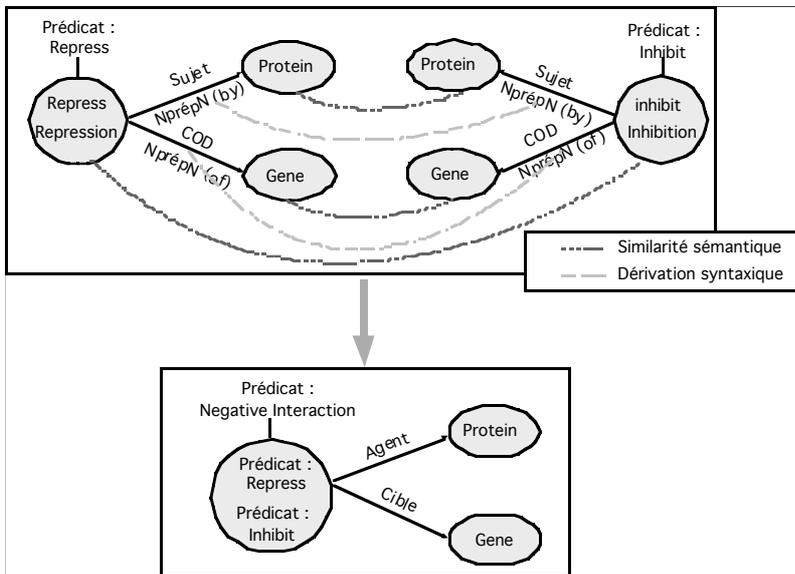


Figure 4. Exemple d'apprentissage de schémas prédicatifs sur la base de similarités sémantiques entre prédicats et de dérivations syntaxiques

• Il faut ensuite apprendre des schémas prédicatifs à partir de ces schémas de sous-catégorisation. Il s'agit ici d'un problème de classification d'ordre un sur des objets relationnels. Cet apprentissage nécessite d'une part, la conception de nouvelles méthodes de classification conceptuelle relationnelle qui pourront s'inspirer des rares exemples de méthodes existants [Bisson, 92], et d'autre part, la prise en compte de différents types de similarités (similarité morphologique et sémantique entre différents prédicats, similarité de structures syntaxiques entre différents schémas...).

Apprentissage des ressources nécessaires à l'extraction

Dans Caderige, la normalisation et l'enrichissement des textes permettent à la fois d'exprimer les critères de sélection, et les patrons d'extraction, à un niveau de généralité et d'expressivité supérieur à ce qui est généralement fait.

La figure 5 montre un exemple d'automate qui implémente un patron d'extraction permettant de repérer et de baliser une interaction entre une protéine et l'expression d'un gène. Il combine l'appel à d'autres automates (<Gene expression>) et des contraintes sur les classes sémantiques (<Positive interaction>), sur les catégories grammaticales (verbe) et sur les relations syntaxiques (Sujet).

En fonction de la représentation des documents obtenue, des méthodes d'apprentissage sont adaptées et développées de manière à apprendre les classificateurs de sélection de corpus et les règles d'extraction. Par exemple, la présence de relations syntaxiques dans la représentation nécessite la mise en œuvre de méthodes d'apprentissage relationnelles telles que FOIL ou plus généralement issues de la Programmation Logique Inductive. L'exploitation d'une ontologie sous la forme de « théorie du domaine »

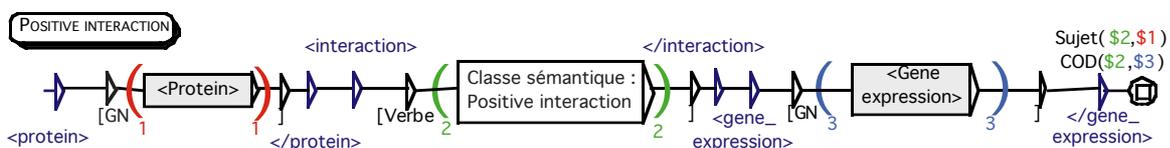


Figure 5 : Exemple d'automate reconnaissant et annotant une interaction

réclame l'utilisation de méthodes dites *knowledge intensive*. La qualité variable des annotations et de l'analyse syntaxique jointe au volume important de donnée nécessite l'utilisation de méthodes robustes. Le projet explore la combinaison de méthodes hybrides présentant ces différentes spécificités.

Conclusion

L'automatisation de l'extraction d'information dans les domaines scientifiques et techniques requiert l'intégration et l'application de nombreuses méthodes d'apprentissage automatique et de traitement automatique de la langue. L'activité de MIG en IA et le projet Caderige sont des illustrations de la recherche pluridisciplinaire qui peut être menée sur le sujet.

Références

Bensch P.A., Savitch W. J., « An occurrence-based model of word categorization ». *Annals of Mathematics and Artificial Intelligence* 14(1): (1995).

Bessières P., Nazarenko N. et Nédellec C., « Apport de l'apprentissage à l'extraction d'information : le problème de l'identification d'interactions géniques. », in *Actes du Colloque International sur le Document Electronique, Méthodes, Démarches et Techniques Cognitives, CIDE'2001*, Toulouse Octobre 2001.

Bikel D. M., Miller S., Schwartz R. and Weischedel R., « Nymble: a High-Performance Learning Name-finder », *Conference on Applied Natural Language Processing*, 1997.

Bisson G., « Learning in FOL with a similarity measure ». In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI)*, San Jose, CA., p. 82-87, AAAI Press, 1992.

Bisson G. et Nédellec C., « Aide à la conception de méthodes de classification pour la construction d'ontologies : l'atelier Mo'K » in

PRÉSENTATION DE LABORATOIRES

Actes des *Journées Francophones d'Extraction et de Gestion des Connaissances (EGC'2001)*, Briand H. (Ed.), Hermès (Pub.), Nantes, Janvier 2001.

Bisson G., Nédellec C. and Canamero D. « Designing clustering methods for ontology building - The Mo'K workbench » in Proceedings of the workshop on *Ontology Learning*, workshop of the European Conference on Artificial Intelligence (ECAI-2000), S. Staab, et al (Eds.). pp. 13-19, Berlin, August 2000.

Blaschke C., Andrade M. A., Ouzounis C. and Valencia A., « Automatic Extraction of biological information from scientific text: protein-protein interactions », in Proceedings of *International Symposium on Molecular Biology, (ISMB'99)*, 1999.

Bourigault D. « A Natural Language Processing Tool for terminology Extraction ». In Proceedings of *EURALEX*. 1986.

Califf M. E. and Mooney R. J., « Relational Learning of Pattern-Match Rules for Information Extraction. » In Proceedings of *AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, CA, March, 1998.

Ciravegna F., « Learning to Tag for Information Extraction. » In Proceedings of the *ECAI-2000 Workshop on Machine Learning for Information Extraction*, F. Ciravegna et al. (Eds.), Berlin, August 2000.

Craven M. and Kumlien J., « Constructing Biological Knowledge Bases by Extracting Information from Text Sources. » In Proceedings of the *7th International Conference on Intelligent Systems for Molecular Biology (ISMB-99)*, 1999.

Daille, B., « Study and Implementation of Combined Techniques for Automatic Extraction of Terminology ». In *The balancing act combining symbolic and statistical approaches to language*, J. Klavans & P. Resnik, eds., MIT Press, 1995.

Faure D. and Nédellec C., « Knowledge Acquisition of Predicate-Argument Structures from technical Texts using Machine Learning » in Proceedings of *Current Developments in Knowledge Acquisition: EKAW-99*, pp. 329-334, Fensel D. and Struder R. (Eds.), Karlsruhe, Germany, April 1999.

Freitag D., « Multistrategy learning for information Extraction. » In the Proceedings of the *Fifteenth Machine Learning Conference (ML-98)*, J. Shavlik (Ed.), Madison, USA, 1998.

Habert, B et Nazarenko, A., « La syntaxe comme marche-pied de l'acquisition des connaissances : bilan critique d'une expérience ». Actes de *Journées sur l'acquisition*

des connaissances, pp. 137-142, Sète, mai 1996.

Hamon T. and Nazarenko A., « Detection of synonymy links between terms: experiment and results », *Recent Advances in Computational Terminology*. John Benjamins, 2001.

Harris Z., Gottfried M., Ryckman T., Mattick Jr P., Daladier A., Harris T. and Harris S., « The form of Information in Science, Analysis of Immunology Sublanguages », vol. 104 of *Boston Studies in the Philosophy of Science*. Dordrecht, the Netherlands, Kluwer Academic (Pub.), 1989.

Hobbs, J, Appelt, D., Bear, J. Israel, D., Kameyana, Stickel, M., M. Tyson, M., « FASTUS : a cascaded finite-state transducer for extracting information in natural language text ». in *Finite state Language Processing*, E. Roche a E. Schaves (eds), MIT Press, Cambridge, 1997.

Mark Hoebeke, Hélène Chiapello, Philippe Noiro, Philippe Bessières: SPiD: a subtilis protein interaction database. 1209-1212, *Bioinformatics*, Volume 17, Number 12, December 2001.

Jacquemin C., « A symbolic and surgical acquisition of terms through variation ». In *Connectionist, Statistical and Symbolic Approaches for Natural Language Processing*, pp. 425-438, Heidelberg, Springer, 1996.

MUC, Proceedings of the *Message Understanding Conference (MUC-4-7)*, Morgan Kaufman, San Mateo, USA, 1992-98.

Morin E. and Jacquemin C., « Projecting Corpus-Based Semantic Links on a Thesaurus », in Proceedings of the *37th Annual Meeting of the Association for Computational Linguistics (ACL'99)*, pp. 389-396, Maryland, USA, 1999.

Nédellec C., « Bibliographical Information Extraction in Genomics » in *IEEE Intelligent Systems: Trends & Controversies - Mining Information for Functional Genomics*, N. Shadbolt (éd.), p. 76-78, mai-juin, 2002.

Nédellec C., Ould Abdel Vetah M., et Bessières P., « Sentence Filtering for Information Extraction in Genomics, a Classification Problem ». In Proceedings of the Conference on Practical Knowledge Discovery in Databases, PKDD'2001, p. 326-338, Freiburg, septembre 2001.

Nédellec C., « Knowledge Extraction from Text, a Machine Learning Approach », communication invitée. In Proceedings of the Third International Conference on Human-System Learning, CAPS'3, Learning WWW, Euroipa Production (Pub.), Paris, France, décembre 2000.

Nédellec C., Ould Abdel Vetah M., Bessières

P., Brun C. et Jacq B., « Reconnaître des fragments de phrase pertinents pour l'extraction d'information dans les textes de génomiques, un problème de classification ». In Actes de la Conférence Française d'Apprentissage, Bisson G. (Ed.), Grenoble, Juin 2001.

Ono T., Hishigaki H., Tanigami A. and Takagi T., « Automated extraction of information on protein-protein interactions from the biological literature », in *Bioinformatics* vol 17, n° 2, pp. 155-161, Oxford Press, 2001.

Pereira F., Tishby N. and Lee L., « Distributional clustering of English words » in *Proceedings of ACL'93*, pp. 183-190, 1993.

Pillet V., *Méthodologie d'extraction automatique d'information à partir de la littérature scientifique en vue d'alimenter un nouveau système d'information*, thèse de l'Université de droit, d'économie et des sciences d'Aix-Marseille, 2000.

Poibeau T. et Nazarenko A., « L'extraction d'information, une nouvelle conception de la compréhension de texte ? » *Traitement Automatique des Langues*, 40(2), 1999.

Poibeau, T. *Extraction d'information à base de connaissances hybrides*. Thèse de doctorat en Informatique, Université Paris-Nord. Mars 2002.

Pustejovsky, J. *The Generative Lexicon*, Cambridge, MIT Press, 1995.

Quinlan J., « Learning logical definitions from relations. » In *Machine Learning Journal*, vol 5(3) p. 239-266, 1990.

Smadja F., « Retrieving Collocations From Text : Xtract. » *Computational Linguistics*, 19(1), pp. 67-74. 1993.

Soderland S., « Learning Information Extraction Rules for Semi-Structured and Free Text » in *Machine Learning Journal*, vol 34, 1999.

Stapley B. J. and Benoit G., « Bibliometrics: Information Retrieval and Visualization from co-occurrence of gene names in MedLine abstracts ». In Proceedings of the *Pacific Symposium on biocomputing (PSB'2000)*, 2000.

Thomas, J., Milward, D., Ouzounis C., Pulman S. and Carroll M., « Automatic Extraction of Protein Interactions from Scientific Abstracts ». In Proceedings of the *Pacific Symposium on biocomputing (PSB'2000)*, vol.5, p. 502-513, Honolulu, 2000.

Thompson C. A., « Acquisition of a Lexicon from Semantic Representations of Sentences », in Proceedings of the *33rd Annual Meeting of the Association of Computational Linguistics, (ACL'95)*, p. 335-337, Boston, M A, Juillet, 1995.

Les projets IA dans le RNTL et les autres réseaux nationaux

Bertrand Braunschweig, Jean-Luc Dormoy, Patrick Gallinari

Objectif de ce dossier

Le Réseau National des Technologies Logicielles existe depuis trois ans. L'appel à propositions 2003 est le quatrième du nom. La première phase de l'appel est passée depuis le 13 Janvier 2003, elle sera complétée par une deuxième phase se terminant en juin prochain.

Plutôt que de les reformuler, nous reprenons ci-après des extraits du texte fondateur de présentation du RNTL, qui explique parfaitement les ambitions du réseau.

« Le RNTL a été mis en place par les ministères de la Recherche et de l'Industrie, et a reçu comme mission de favoriser les coopérations entre les équipes de la R&D industrielle et les équipes de la recherche publique dans le domaine des technologies logicielles...Il se positionne en complément et en appui, d'une part des programmes européens dans le cadre du PCRD (principalement IST) et des actions EUREKA (en particulier le programme ITEA), et d'autre part des actions nationales déjà lancées où le développement logiciel prend une part importante...

Les ambitions à moyen et long terme du réseau sont :

- d'une part, d'aider à la création et au développement en France d'entreprises dans les technologies de l'information et de la communication à partir des points forts actuels de la recherche publique et de l'industrie en matière de développement logiciel,
- d'autre part, d'améliorer la compétitivité des équipes françaises en matière de développement logiciel par la mise au point de nouvelles méthodes et de nouveaux outils pour faire face aux enjeux de la société de l'information.

La réussite du réseau se mesurera à l'importance des retombées économiques que ces projets auront :

- en termes de création de produits et services nouveaux, que les logiciels développés soient génériques ou spécifiques à un secteur applicatif,
- en matière de création et de développement d'entreprises,
- en gain de productivité en développement logiciel dans les sociétés de service informatique et dans les départements de R&D des entreprises.

Viser cette ambition nécessite que le réseau s'attache à augmenter les performances de la R&D en France par les actions de fond suivantes :

- augmenter la diffusion et l'influence des logiciels développés en France, que ces logiciels soient commerciaux ou à usage libre, pour conforter la position française dans l'industrie des composants logiciels et de leur intégration dans les systèmes d'information,
- augmenter l'ambition technologique des projets de recherche coopératifs entre laboratoires académiques et entreprises, pour augmenter la maturité technologique des entreprises en matière de développement logiciel, pour mieux insérer les laboratoires dans les projets à moyen et long terme des acteurs économiques, et pour bâtir des partenariats durables.
- contribuer à une meilleure connaissance de la gestion de projets de développement en entreprise de la part des chercheurs et des doctorants et faciliter ainsi leur participation à la création ou à la croissance d'entreprises innovantes,
- améliorer pour les entreprises la visibilité et la lisibilité de la communauté de recherche académique en informatique, mais aussi dans l'ensemble des sciences et technologies de l'information et de la communication, et

dans les disciplines connexes, en mathématiques appliquées ou en sciences sociales en particulier,

- améliorer la position française dans la définition des normes et des standards au niveau européen et au niveau mondial et contribuer à organiser la protection des intérêts des acteurs français en matière de propriété industrielle et intellectuelle,
- contribuer à la structuration de la communauté de recherche académique dans le domaine des STIC et à sa coordination avec la recherche industrielle et avec les besoins des PME. »

L'IA dans le RNTL

Les technologies IA sont très présentes dans la thématique du RNTL, et de nombreux projets comportent un volet IA consistant. Les thèmes 2 et 3, traditionnellement, sont un peu les thèmes « IA » du RNTL. Mais des projets des deux autres thèmes mettent également en œuvre certains outils de l'IA, de manière explicite – dans les objectifs – ou bien cachée – dans les moyens.

Le but de ce dossier est de faire le point sur la présence des méthodes et outils de l'IA dans les projets RNTL existants. Par extension, des projets de deux autres réseaux, le RNRT (réseau national de recherche en télécommunications), et RIAM (Réseau pour la recherche et l'innovation dans l'audiovisuel et le multimédia) ont également invités à présenter leurs volets IA, avec un succès limité, nous l'avouons.

Avec ce dossier, d'une part, nous permettons aux équipes IA présentes dans les projets RNTL en cours, de faire connaître leurs contributions et de valoriser leur recherche ; d'autre part, nous visons également les porteurs de proposition, pour leur donner des exemples de

sujets considérés comme importants par le RNTL, et leur permettre de se référer à ces projets existants si nécessaire, par exemple en exploitant les résultats.

Si vous voulez en savoir plus sur le RNTL, le site www.industrie.gouv.fr/rntl/ vous donnera toute l'information sur son histoire, ses objectifs, ses modalités, les projets labellisés et financés, les appels à propositions, les manifestations à venir, etc.

Le RNTL dans l'IA

En fait, ce titre rappelle simplement la visite de Gérard Roucairol, président du comité d'orientation, instance de décision du RNTL, lors de la journée plénière de la plate-forme AFIA à Grenoble en juin 2001. La présentation de G. Roucairol avait pour objectif d'inciter la communauté IA à initier ou à participer à des projets RNTL. Placée après une série de présentations à caractère prospectif ou scientifique, cette communication a eu le mérite de démontrer à tous les présents que le RNTL était ouvert aux propositions à contenu IA, même si, naturellement, ces propositions seraient mises en compétition avec d'autres et qu'il était hors de question d'en garantir le succès! Comme pour tous, les propositions « IA » doivent satisfaire aux exigences d'excellence scientifique, d'exploitation, et d'organisation de projet qu'impose le RNTL. Nous revenons sur ce point à la fin de cette introduction.

Mots-clés du Thème 2

- 2.1 Systèmes d'Information d'Entreprise
- 2.2 **Bases de Données**
- 2.3 **Bases de connaissances**
- 2.4 Systèmes à base d'agents, à base de règles
- 2.5 **Fouille de Données (Data Mining)**
- 2.6 « Semantic Web »
- 2.7 Architecture de systèmes répartis, infrastructure d'intégration, langages et outils de description d'architecture
- 2.8 « Web services »

- 2.9 **Applications coopératives / collectives**
- 2.10 Architecture de sécurité
- 2.11 Architecture à haute disponibilité,
- 2.12 **Systèmes auto adaptatifs, auto reconfigurables**
- 2.13 Workflow
- 2.14 Commerce électronique, relations avec les clients et les fournisseurs
- 2.15 Systèmes d'information personnels, systèmes pair-à-pair, communautés coopérantes
- 2.16 Connexion d'équipements à l'Internet, architectures locales et de passerelle, applications
- 2.17 Intégration, administration des applications
- 2.18 Caractérisation et description de composants, de services (spécifications formelles, semi-formelles, UML,...)
- 2.19 Assistance à la construction, au déploiement et à l'assemblage de composants, de services (interopérabilité)
- 2.20 **Validation de composants logiciels (interprétation abstraite, vérification par model checking, preuve automatique et assistée, simulation, certification, sûreté de fonctionnement, aspects relatifs à la sécurité)**
- 2.21 Réutilisation de composants
- 2.22 **Autres**

La place de l'IA dans le RNTL

Nous examinons brièvement la place de l'IA dans les appels à proposition du RNTL. Les thèmes ont évolué depuis la création du réseau (par exemple, on est passé de cinq à quatre thèmes depuis 2002, en répartissant le thème des composants logiciels). Ces thèmes ont été définis par un processus ouvert de consultation, au cours duquel des représentants de la communauté IA ont pu s'exprimer et indiquer leurs priorités. Nous regardons les thèmes 2003, ceux qui sont applicables au prochain appel à propositions.

On retrouve dans les thèmes 2 plusieurs mots-clés intéressant directement l'IA. Nous les avons surlignés dans le tableau ci-dessous. Mais la place de l'IA dans le thème dépasse cette simple liste de mots surlignés, car le tableau com-

porte aussi bien des technologies (« systèmes à base d'agents », « fouille de données ») que des applications (« commerce électronique », « systèmes d'information d'entreprise »). Il est donc possible de proposer des solutions IA aux problèmes applicatifs présentés dans cette liste. Enfin, le mot-clé 22, « autres », laisse la place à l'imagination... à replacer dans le contexte de la définition du thème. Mais vous avez tout à fait le droit, si cela porte sens, de proposer une solution à base d'évolution artificielle pour les systèmes d'information d'entreprise, ou une solution à base de raisonnement par cas pour l'administration des applications...

La thématique IA se retrouve également dans les trois autres thèmes: le deuxième tableau donne les mots-clés, extraits des thèmes 1, 3, et 4, qui relèvent directement de l'appellation IA. Nous ne reproduisons pas le reste, pour lequel la remarque précédente sur les solutions IA s'applique de manière identique. Le thème 3 fait aussi très fortement références aux techniques d'IA.

Il ne nous reste plus qu'à faire accepter au RNTL d'inclure l'expression « intelligence artificielle » dans les documents de l'appel, ce que nous avons poussé dans le passé, sans trop de succès jusqu'à présent. Sachant que les diverses instances d'évaluation et de décision du RNTL comportent plusieurs membres de la communauté IA, la partie n'est pas perdue d'avance !

Combien de projets « IA » ont été ou sont financés par le RNTL ?

C'est une question à laquelle il est difficile de répondre, car le classement d'un projet ou d'une partie d'un projet en IA est forcément flou ! Un dossier sur les projets en cours avait été préparé par le RNTL pour ses journées nationales d'octobre 2000 à Toulouse. Nous avons identifié vingt neuf projets, sur une centaine, pouvant relever de l'appellation IA. Notre appel à contribution a été envoyé directement à ces projets, en complément de sa diffusion dans le bulletin de l'AFIA et sur le portAI. Quelques-uns ont répondu que le contenu IA était modeste, d'autres qu'il était trop tôt pour en parler. La dossier que

vous avez entre les mains (ou sur votre écran), présente finalement les quelques projets qui ont jugé important de marquer leur lien avec la communauté de l'anneau AFIA. Pour information, nous donnons cette liste de projets sélectionnés, forcément subjective, en annexe, avec un mot clé en indiquant - très grossièrement - le contenu IA. Vous trouverez des informations sur ces projets sur le site web du réseau.

Et dans RNRT et RIAM

Les réseaux RNRT (télécoms) et RIAM (audiovisuel) financent également quelques projets à contenu IA, en moins grand nombre. Un des projets de ce dossier vient du RNRT. Notre appel vers ces autres réseaux a remporté moins de succès, nous vous encourageons donc à consulter les sites web correspondants.

1 - Concevoir des logiciels enfouis, critiques ou temps réel pour les objets et systèmes

1.8 Modélisation, vérification de comportement, validation

1.9 Transformations de spécifications / programmes

1.10 Cohérence, conformité

3 - Inventer de nouvelles interfaces personnes - systèmes - environnement

3.18 Reconnaissance et synthèse des sons

3.19 Reconnaissance et synthèse de la parole

3.20 Reconnaissance et synthèse de l'écriture et des formes

3.21 Interaction en langue naturelle, multilinguisme

3.22 Interaction émotionnelle

3.23 Agents intelligents, avatars

3.27 Entrepôts de données, fouille de données, médiateurs

Thème 4 - Elaborer une nouvelle conception pour de nouveaux objets

4.9 Mise en partage des savoir-faire individuels (un métier) pour une exploitation aisée et généralisée comme un savoir-faire de groupe (plusieurs métiers)

Ce dossier

Vous savez lire : inutile donc de résumer ici les présentations qui suivent, qui sont elles-mêmes des résumés de documents bien plus importants. Nous ne faisons qu'un seul commentaire: celui de la diversité. Rarement un dossier du bulletin aura présenté des sujets aussi divers, qui ont en commun d'avoir été « bénis » par les représentants des ministères de l'industrie, de la recherche et des télécommunications. Dans ces dix projets, on passe des systèmes à agents à l'extraction de connaissances dans les documents, de la programmation par contrainte à la planification de mouvements de robots...avec des utilisations également très diverses, regardez par exemple DRA MACHINA, un domaine d'application peu commun !

Soumettre une proposition

Les propositions pour le RNTL, qu'elles soient « IA » ou non, sont soumises à un ensemble de critères touchant essentiellement à leur contenu scientifique et technique, au potentiel d'exploitation, à l'organisation et à la conduite de projet. Il est indispensable de satisfaire tous ces critères : une proposition à contenu scientifique excellent, mais sans perspective de création de valeur, ne sera pas labellisée, pas plus qu'une proposition à l'organisation impeccable mais ne comportant pas la moindre innovation. Ces remarques sont à décliner selon la nature des projets (exploratoires, pré-compétitifs, plates-formes). Les critères d'innovation sont particulièrement importants pour les projets exploratoires, tandis que les critères de marché sont prépondérants pour les projets pré-compétitifs. Les plates-formes, considérées comme cruciales par le RNTL, se doivent d'être des facteurs de fédération et d'entraînement pour toute une communauté. Enfin, les logiciels libres sont plus que bienvenus, et sont jugés avec des critères spécifiques (existence d'une communauté de développeurs, processus de co-développement du logiciel, organisation de la maintenance, position sur le marché etc.).

Les critères d'évaluation des propositions sont largement communiqués dans les appels, et les documents utilisés par les experts sont consultables sur le site du réseau. Nous en soulignons quelques points. N'hésitez pas à réagir si vous êtes en désaccord avec notre analyse !

Objectifs de recherche technologique du projet : c'est un critère essentiel ; vous devez clairement montrer que vous connaissez l'état de l'art, et que vous innovez. Le domaine de l'IA est très vaste, mais vous devez connaître votre domaine aussi bien (mieux) que les experts qui évalueront la proposition. Pensez que l'expert académique sera sans doute un chercheur du même domaine que vous. Montrez que vous connaissez les projets et plates-formes RNTL proches de votre sujet, ainsi que les projets européens du même tonneau. Mieux, basez vous sur les résultats disponibles de ces projets.

Objectifs industriels du projet : un critère essentiel également; vous devez présenter des perspectives réelles d'exploitation interne (par les partenaires du projet) ou externe (par une société chargée de distribuer le logiciel) ; soyez précis sur les marchés, la concurrence. L'expert industriel sera certainement attentif à ces aspects. Dans le cas du logiciel libre, les critères sont un peu différents, il faut démontrer qu'il existe une communauté prête à s'approprier le logiciel, et un processus de co-développement maîtrisé.

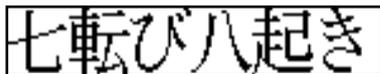
Réalisation technique du projet : si vous avez un bon objectif de recherche technologique, vous devez également démontrer que vous savez comment l'atteindre. Dans cette perspective, justifiez bien vos choix technologiques (encore une fois, les experts sont vraiment experts!), et mentionnez votre rapport avec les normes et standards, c'est important. Un projet de SMA devra certainement citer le standard FIPA, qu'il l'utilise ou non. Un projet lié au Web Sémantique devra (c'est évident) se positionner par rapport aux standards émergents du W3C (pas tous, il y en a tellement...).

Composition du consortium, res-

sources, coûts, délais : ces critères n'appellent pas de commentaires particuliers pour les projets « IA », mais ils sont évidemment pris très sérieusement en considération dans l'évaluation. Faites attention à la maîtrise des risques, peut-être plus importante dans des projets nécessitant de représenter des connaissances, activité qui peut être plus consommatrice de ressources que prévu initialement...

Conclusion

Le RNTL peut permettre de gagner un ou deux ordres de grandeur dans vos projets, en mettant à votre portée des partenariats et des ressources humaines et matérielles significativement supérieures à ce que vous pouvez mobiliser habituellement. Plusieurs équipes d'IA ont déjà bénéficié de ce soutien, et nous vous encourageons à suivre leur exemple. Bien évidemment, pour cela, il faut consacrer un effort important à la préparation d'un dossier de proposition impeccable, et vous ne réussirez peut-être pas du premier coup. Mais, comme disent les japonais¹ « nana korobi, ya oki », tombe sept fois, relève-toi huit fois (un peu de persévérance, le jeu en vaut la chandelle).



La suite de ce dossier, qui en est la partie essentielle, présente plusieurs exemples de projets. Bonne lecture!

Les sites web

RNTL: ww.industrie.gouv.fr/rntl
RNRT: www.telecom.gouv.fr/rnrt
RIAM: www.cnc.fr/riam

Liste de projets RNTL à contenu IA ou assimilé, avec un mot-clé sur la technique employée. Les projets marqués d'une * sont présentés dans ce dossier.

* ADELFE (Agents & composants)
AMIBE (Inférence bayésienne)
AudioSurf (Indexation audio)

1. Merci à Fabien Gandon pour la référence

- * CLACIC (Calage de codes par Ingénierie des Connaissances)
- * CO2 (Conception par contraintes)
- DICO (Base de Connaissances attaques & vulnérabilité)
- DocMining (Interprétation de documents)
- Domus Videum (Apprentissage & audiovisuel)
- * DraMachina (Base de connaissance en dramaturgie)
- * E.dot (intégration de sources XML)
- * EOLE (environnement d'optimisation pour télécoms) (RNRT)
- EPIA (Portail adaptatif)
- * ExtraPloDocs (Extraction de connaissances dans documentation)
- FADO (contraintes)
- * InKa (Génération de cas de test)
- IntelliSurf (Web & sémantique)
- IPPOP (Formalisation de connaissance technologique)
- * LUTIN (Outillage sémantique des patrons de conception)
- MACAO (Réorganisation de hiérarchies de classes)
- MAGIE (Management intelligent d'information de veille)
- MASCV (Mutliagents)
- Monastir (Réseaux neuronaux pour analyse stochastique)
- * NeMove3D (planification de mouvements de robots)
- OADymPPaC (Analyse dynamique de programmes avec contraintes)
- Outilex (Traitement du langage naturel)
- Perceval (Traitement de la parole)
- TRAMS (Transformations sur métamodèles)
- Usiquick (connaissances de conception)
- WACIF (Apprentissage de robot par réseaux neuronaux)
- XMiner (Balise XML de textes)

Tél : 05 61 55 82 94
Mél : gleizes@irit.fr
Contact : Marie-Pierre Gleizes
Contributeurs : Carole Bernon, Marie-Pierre Gleizes, Pierre Glize, Gauthier Picard, Jean-Pierre Mano
L3I
Adresse : Université de La Rochelle,
Av. Michel Crépeau,
17042 La Rochelle Cedex1
Tél : 05 46 45 87 32
Mél : vcamps@univ-lr.fr
Contact : Valérie Camps
Contributeur : Valérie Camps

2. Autres partenaires

ARTAL Technologies

Adresse : 10 avenue de l'Europe,
31520 Ramonville St Agne
Tél : 05 61 28 56 71
Mél : pierre.duverneuil@artal.fr
Contact : Pierre Duverneuil
Contributeurs : Pierre Duverneuil,
Raphaël Giraud, Sylvain Peyruqueou

IRIT – Toulouse – Equipe génie logiciel

Adresse : Université Paul Sabatier,
118 route de Narbonne,
31062 Cedex Toulouse
Tél :
Mél : millan@irit.fr
Contact : Thierry Millan
Contributeurs : Jean-Paul Bodeveix,
Thierry Millan, Christian Percebois

TNI Valiosys – Brest

Adresse : 120, rue René Descartes,
Technopôle Brest-Iroise, BP
70801 29608 BREST Cedex
Tél : 02 98 05 27 44
Mél : philippe.hordonneau@tni-valiosys.com
Contact : Philippe Hordonneau
Contributeurs : Philippe Hordonneau,
Jean-Marie Lions, Didier Simoneau

3. Description générale du projet

Le but principal de ce projet est de réaliser un Atelier de Développement de Logiciels à Fonctionnalité Emergente (ADELFE) destiné à des concepteurs informatiques non spécialistes de ce type de logiciels. Ces logiciels sont fondés sur

**ADELFE : Atelier de Développement de Logiciels à Fonctionnalité Emergente
Projet précompétitif
12 / 2000**

1. Équipes IA :

IRIT

Adresse : Université Paul Sabatier,
118 route de Narbonne,
31062 Cedex Toulouse.

les techniques des systèmes multi-agents adaptatifs et auto-organiseurs (AMAS) développés au sein du laboratoire IRIT (<http://www.irit.fr/SMAC>) et de la société ARTAL Technologies. Par leurs capacités d'auto-organisation et d'adaptation, ces logiciels sont aptes à fonctionner dans un environnement à très forte dynamique et à faire face à des imprévus. Ils sont adaptés pour résoudre des problèmes complexes, distribués pour lesquels un contrôle global est impossible à mettre en œuvre. Avec l'essor des nouvelles technologies comme Internet, l'évolution des réseaux de machines, ces multi-agents apportent des solutions aux problèmes de conception de logiciels adaptés à des applications réparties de grande complexité. Le développement de plusieurs applications a montré qu'après la définition des composants du système que sont les agents, il faut les doter d'aptitude pour décider de leurs interactions. Ainsi, le concepteur de l'application implémente les agents et le système en fonctionnement se configure de manière automatique par auto-organisation sans l'intervention du concepteur. Actuellement, il n'existe pas de méthodologie et d'aide à la conception pour ces systèmes adaptatifs.

La réalisation de cet atelier de développement consiste à fournir :

- une notation basée sur UML,
- une méthodologie de conception, qui étend les recommandations pour l'analyse, la conception et le développement présentés dans le guide « The Unified Software Development Process » pour prendre en compte la modélisation et réalisation de systèmes multi-agents adaptatifs,
- une plate-forme comprenant un logiciel interactif de suivi du processus de développement, un outil de modélisation graphique, un outil de prototypage rapide et une bibliothèque de composants.

4. Description détaillée de la partie IA du projet

Le projet qui consiste en la réalisation d'une méthodologie orientée agent fait partie du domaine des systèmes multi-

agents. L'objectif d'ADELFE étant de considérer tous les enchaînements d'activités d'une conception d'un logiciel classique depuis les besoins jusqu'au déploiement, ADELFE se base sur la méthodologie du Rational Unified Process (RUP) et utilise les notations UML et AUML. L'outil de modélisation graphique utilisé et étendu pour ADELFE est le logiciel OpenTool de la société TNI-Valiosys. Notre but n'est pas d'ajouter une méthodologie de plus à celles déjà existantes mais de travailler sur les aspects tels que la complexité de l'environnement, les agents, la dynamique ou l'adaptation du logiciel qui n'ont pas encore été pris en compte dans les méthodologies classiques. ADELFE se fonde sur la vision d'un SMA en tant qu'organisation dynamique formée de différents agents coopérants qui de par leur comportement vont permettre au système global de s'adapter. Ce que nous avons ajouté à la méthode RUP sont les outils ou modèles suivants :

- un outil d'aide à la décision qui permet de savoir si la technologie AMAS est adéquate pour l'application,
- un modèle de l'environnement,
- un guide pour identifier les agents dans un système qui doit s'adapter,
- un modèle agent,
- un modèle d'interactions entre agents,
- un outil de prototypage rapide.

5. Etat d'avancement, résultat obtenu

Les principaux travaux réalisés concernent la définition de la méthodologie de conception des systèmes multi-agents à fonctionnalité émergente, la spécification et la conception de l'atelier ainsi que la définition d'un cas d'étude pour la partie validation.

La méthodologie est décrite et formalisée à l'aide du SPEM (Software Process Engineering Metamodel) et est donnée en terme de définitions de travaux (WorkDefinition), d'activités (activities) et d'étapes (steps). Les définitions de travaux (en gras) et les activités d'Adelfe sont les suivantes :

Les besoins préliminaires

Définition du cahier des charges du client

Validation du cahier des charges par l'utilisateur

Définition d'un cahier des charges consensuel

Etablissement d'un ensemble de motifs

Extraction des limites et des contraintes

Les besoins finals

Définition et modélisation de l'environnement

Détermination des cas d'utilisation

L'analyse

Analyse du domaine et étude de l'architecture du système

Adéquation de la théorie des AMAS

Identification des agents

Etude des interactions entre les différentes entités

La conception

Définition de l'architecture détaillée et du modèle multi-agent

Etude des interactions

Modélisation de l'agent

Prototypage rapide

Complétion de l'architecture détaillée

La notation UML doit permettre d'écrire tous les concepts manipulés dans la méthodologie ADELFE. Cette notation a donc été étendue selon deux axes :

- en utilisant AUML, développé par Odell et al., qui permet de décrire les protocoles d'interactions des agents à l'aide de diagrammes de séquences ;
- en ajoutant des stéréotypes liés au modèle de l'agent

Adelfe est une méthodologie dédiée à la conception de systèmes adaptatifs mais c'est aussi un outil interactif qui permet à l'utilisateur de gérer ses projets. Cet outil interactif est couplé à l'outil OpenTool qui permet de réaliser tous les diagrammes de la notation UML, AUML et ADELFE.

6. Bibliographie

Site Web du projet : <http://perso.univ-lr.fr/vcamps/Adelfe/Adelfe.html>

CASTERAN Jean-Christophe, GLEIZES Marie-Pierre, GLIZE Pierre – Des méthodologies orientées multi-agent – 8^{èmes} Journées Francophones Intelligence Artificielle Distribuée et Systèmes Multi-

Agents – Editions Hermès, Saint-Jean-la-Vêtre, octobre 2000.

BERNON Carole, CAMPS Valérie, GLEIZES Marie-Pierre, GLIZE Pierre – La conception de systèmes multi-agents adaptatifs : contraintes et spécificités – *Atelier de Méthodologie et Environnements pour les Systèmes Multi-Agents (SMA 2001)*, Plateforme AFIA, Grenoble du 25 au 28 juin 2001, Z. Guessoum et M. Occhetto, coords.

BERNON Carole, GLEIZES Marie-Pierre, PICARD Gauthier, GLIZE Pierre – The ADELFE Methodology for an Intranet System Design – *Agent Oriented Information Systems (AOIS'02)*, dans le cadre de CaiSE'02, Toronto, 27-28 mai 2002.

BERNON Carole, GLEIZES Marie-Pierre, PEYRUQUEOU Sylvain, PICARD Gauthier – ADELFE, a Methodology for Adaptive Multi-Agent Systems Engineering – *ESAW'02 Third International Workshop Engineering Societies in the Agents World*, Madrid, 16-17 September 2002.

**CLACIC : CaLAge de Codes
par Ingénierie des
Connaissances**
Gilles MOREL – CETMEF
Projet exploratoire
11 / 2002

1. Équipes IA :

LaRIA (Laboratoire de Recherche en Informatique d'Amiens)

Adresse : Faculté de Mathématiques et d'Informatique - CURJ –
33 rue Saint-Leu –
80039 Amiens Cedex
Tél. : 03 22 82 88 75
Mél : kassel@laria.u-picardie.fr
Personne à contacter : Gilles KASSEL

Contributeurs : Gilles KASSEL (Pr) – Sabine BRUAUX (doctorante) – Gilles MOREL (CETMEF – membre associé) – Yingjiu LIU (ingénieur de recherche)

2. Identification du reste du partenariat :

- CETMEF : Centre d'Études Techniques Maritimes et Fluviales (Ministère de l'Équipement), coordinateur du projet
- EDF-LNHE : Laboratoire National d'Hydraulique et Environnement

- CEA-DEN : Direction des Energies Nucléaires
- VERNE-ADER (association proche de l'UPJV)
- DIVERGENT / UTC (fait partie du « Groupe Université de Technologie de Compiègne »)
- DELTACAD (SSII implantée à Compiègne)

3. Description générale du projet

CLACIC est un projet RNTL labellisé en 2002 qui porte sur la réalisation d'un atelier d'assistance au calage des codes de simulation numérique (ou codes de calcul).

Les codes de calcul, qui simulent des phénomènes physiques dans des domaines très variés comme les écoulements hydrauliques ou les réactions nucléaires, doivent généralement subir une phase préalable dite « de calage » afin d'être rendus opérationnels sur un objet physique particulier ou une gamme d'objets. Ce processus de calage est généralement appliqué de manière empirique par les experts de chaque code, parfois avec l'aide d'outils d'optimisation.

Dans une première phase, le projet CLACIC vise à modéliser la tâche générique de calage à partir de l'analyse de trois codes de simulation se situant dans des domaines variés : la trajectographie des navires (code NAVMER du CETMEF), la simulation des réactions nucléaires (code FLICA du CEA) et la simulation des écoulements en rivière (code TELEMAR du LNHE).

Cette conceptualisation du processus de calage doit dans un second temps servir de support de référence pour faire émerger, d'une part, les besoins des utilisateurs en matière d'aide au calage et, d'autre part, les nouvelles méthodes qui pourraient être utilisées pour réaliser certaines tâches qui composent ce processus.

Ces besoins se traduiront ensuite par les spécifications et le développement d'un atelier logiciel qui devrait intégrer et faire collaborer différents types de méthodes : méthodes d'optimisation mathématiques, raisonnement à base de cas, raisonnement à base d'expertise.

La dernière grande étape du projet

consistera à adapter et valider cet outil générique sur les trois applications pilotes qui auront inspiré sa conception.

4. Description détaillée de la partie IA du projet

Le processus global de calage, bien que s'appuyant sur des connaissances propres à chaque code, suit *a priori* une trame récurrente et commune à tous les domaines. La première phase du projet CLACIC vise par conséquent à mettre en lumière et modéliser la tâche de calage comme une tâche générique (au sens de CommonKADS[1] et de Chandrasekaran[2]). Ce travail de généralisation et de modélisation n'a, à notre connaissance, jamais été réalisé de manière systématique et formelle malgré l'universalité du calage, probablement en partie à cause de la culture essentiellement numérique des experts en simulation. Il constitue une occasion d'évaluer la méthode CommonKADS sur un cas complexe et dans un contexte quasi industriel, d'identifier ses limites et d'avancer des propositions pour les dépasser en s'inspirant de travaux plus récents (cf. état d'avancement).

Un second objectif du projet CLACIC en rapport avec l'IA consiste à associer les approches à base de connaissances et les approches fondées sur l'optimisation mathématique pour améliorer l'aide au calage des codes. La modélisation du processus du calage évoquée précédemment permet de faire émerger et de rendre explicites des méthodes et des connaissances implicites utilisées par les experts, connaissances de nature empirique, heuristique ou mathématique. Ces différents types de connaissances pourront se traduire ultérieurement dans le cadre d'un atelier logiciel, sous des formes et technologies qui restent à définir, par la collaboration de méthodes de résolution et d'assistance hétérogènes à base de cas, de règles, de contraintes et d'optimisation.

5. Etat d'avancement, résultat(s) obtenus

La première phase de travaux concernant la modélisation de la tâche de calage

ge a débuté avec comme support initial la méthode CommonKADS (modèle d'expertise) et le langage CML [1]. La confrontation des premiers résultats avec le point de vue des experts a permis de mettre en lumière l'intérêt d'un tel cadre conceptuel mais aussi ses limites. Tout en conservant le cadre général de la méthode CommonKADS, les difficultés de modélisation et de communication rencontrées nous ont poussés à faire évoluer le langage de modélisation dans deux directions. Tout d'abord, la manière de définir et de différencier les concepts et les rôles pour la tâche de calage ne s'avérant pas assez précise et rigoureuse, nous avons été amenés à introduire une approche plus stricte fondée sur les ontologies formelles de Guarino (taxinomie de types et de rôles fondée sur une taxinomie de propriétés [3]). D'autre part, les limites du formalisme graphique proposé pour le langage CML nous ont conduits à faire appel au formalisme SADT pour rendre plus lisible et plus modulaire la structure d'inférence et le flux de données associés à la tâche de calage. Ce même formalisme SADT avait déjà été utilisé dans la méthode d'ingénierie des connaissances MACAO [4].

6. Bibliographie

- [1] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. Van de Velde, B. Wielinga : « Knowledge Engineering and Management, The CommonKADS Methodology ». MIT Press, Cambridge, Massachusetts, London, England, 1999, 455 p.
- [2] B. Chandrasekaran, T. Johnson : « Generic Tasks and Tasks Structures : History, Critique and New Directions ». Dans *Second Generation Expert Systems*, Springer-Verlag, 1993, p. 232-272.
- [3] N. Guarino, C. Welty : « A Formal Ontology of Properties. In R. Dieng and O. Corby (eds.) *Proceedings of the 12th Int. Conf. On Knowledge Engineering and Knowledge Management : EKAW2000*, Lecture Notes on Computer Science, Springer Verlag, p. 97-112. (article accessible à l'adresse : <http://ontology.ip.rm.cnrit>)
- [4] N. Matta : « Méthodes de Résolutions de Problèmes : leur explicitation et leur représentation dans MACAO-2 ». Thèse de l'Université Paul Sabatier de Toulouse, soutenue le 27 octobre 1995, 234 pages.

**CO2 : COnception par COntraintes
Projet précompétitif
labellisé en 2001
démarrage effectif :
15 février 2002**

1. Équipes I. A.

Dassault Aviation

Division de la Prospective -
Département Études Scientifiques
Avancées
Équipe Résolution de Contraintes et
Aide à la Décision
78, quai Marcel Dassault - Cedex 300
92552 SAINT CLOUD
Personne à contacter : Laurent Zimmer
Tél. : 01 47 11 52 95
Mél : laurent.zimmer@dassault-aviation.fr
Contributeurs : Alexis Anglada, Patricia Zablit

LIP6

Pôle Intelligence Artificielle - Groupe
OASIS
Équipe Contraintes
Personne à contacter : Philippe
Codognet
Tél. : 01 44 27 88 01
Mél : philippe.codognet@lip6.fr

IRIN

Équipe Algorithmes, Spécifications et
Contraintes
Groupe Contraintes
Personne à contacter : Laurent
Granvilliers
Tél. : 02 51 12 58 51
Mél : granvilliers@irin.univ-nantes.fr
Contributeurs : Frédéric Goulard, Marc
Christie

2. Autres équipes

CRIL TECHNOLOGY

Département Industrie
Pôle Informatique Scientifique et
Technique
Personne à contacter : Nicolas Fovet
Tél. : 01 40 83 52 24
Mél :
nicolas.fovet@criltechnology.com
Contributeurs : Anne Legendre

LEPT-ENSAM

Systèmes énergétiques et conception
Personne à contacter :
Jean-Pierre Nadeau
Tél. : 05 56 84 54 28
Mél :
nadeau@lept-ensam.u-bordeaux.fr
Contributeurs : Patrick Sabastian,
Jérôme Pailhes

LIPSI-ESTIA

Laboratoire en Ingénierie des Processus
et des Services Industriels
Personne à contacter : Xavier Fischer
Tél. : 05 59 43 84 69
Mél : x.fischer@estia.fr

3. Description générale du projet

Le projet CO2 (Conception par Contraintes) vise à proposer un nouvel outil d'Ingénierie Assistée par Ordinateur qui permette de mettre en œuvre des calculs dès la première étape de conception. A cette étape de conception préliminaire, dite « Avant-Projet », déterminante dans le cycle de développement d'un produit, les outils de l'ingénieur sont restés le papier et le crayon alors que dans les étapes ultérieures, le modèleur géométrique des outils de CAO a remplacé la planche à dessin.

L'enjeu du projet est donc de développer un nouveau type d'outil de calcul, interactif, adapté à la pratique de l'ingénieur, et qui adresse les deux points suivants :

- effectuer la recherche de solutions dans un espace de conception limité d'emblée par les contraintes du cahier des charges ;
- intégrer les calculs au plus tôt dans le cycle de conception alors que certains paramètres ne sont pas fixés.

Le logiciel développé dans le cadre de ce projet exploite les propriétés formelles du cadre CSP qui lui permettent :

- de gérer l'incertitude ou l'imprécision des grandeurs manipulées,
- séparer nettement dans un problème les aspects modélisation des aspects résolution,
- résoudre des problèmes inverses et

adresser la problématique du Design-to-X

4. Description détaillée de la partie I.A. du projet

Ce projet combine différentes méthodes et techniques de l'Intelligence Artificielle : raisonnement à base de modèle, résolution de contraintes et méthodes d'Aide à la Décision.

Raisonnement à base de modèle

L'approche « Model-based » adoptée dans l'outil devra garantir que la description des modèles de conception sur lesquels porteront les calculs sera suffisamment déclarative pour permettre aux ingénieurs de développer et partager des bases de modèles appropriables et réutilisables par d'autres.

Résolution de contraintes

La nature des problèmes rencontrés en conception impose aux méthodes de résolution de tenir compte des facteurs suivants :

- l'outil développé doit être interactif et la résolution la plus réactive possible,
- les problèmes ont une forte composante numérique même si elle n'est pas exclusive,
- beaucoup de problèmes n'ont pas de solution dès lors qu'on intègre toutes les contraintes du cahier des charges ;
- beaucoup de problèmes avec de nombreux degrés de liberté ont beaucoup de solutions.

Nous avons donc axé nos travaux sur :

- l'intégration dans l'outil d'algorithmes de résolution en Arithmétique d'Intervalle généraux et spécialisés, performants et garantissant des propriétés de correction ou de complétude ;
- le développement de méthodes de résolution de contraintes flexibles pour résoudre les problèmes sur-(ou sous) contraints ;
- le développement de méthodes de recherche locale qui permettent de donner la réactivité nécessaire à certains calculs pour assurer l'interactivité ;
- l'intégration de modèles qui ne peu-

vent s'exprimer directement sous des formes analytiques directement utilisables.

Aide à la Décision

Dans le cas des méthodes de résolution de contraintes flexibles, le modèle de conception est enrichi par des priorités, des préférences explicites sur des éléments du problème (contraintes, domaines ...) pour satisfaire au mieux un problème sur-contraint ou au contraire choisir une solution d'un problème sous-contraint. On entre dans une problématique décisionnelle. Différents cadres théoriques existent, le problème est de retenir celui ou ceux qui ont du sens en conception.

5. État d'avancement, résultats obtenus

Une première version de l'outil Constraint Explorer (CE) a été délivrée à tous les partenaires pour évaluation.

Une analyse critique a été effectuée sur :

- les algorithmes de résolution et l'arithmétique d'intervalle utilisés
- les études de cas menées sur des problèmes « métiers ».

Une deuxième version de CE avec un formalisme plus déclaratif et des algorithmes plus efficaces sera disponible début 2003. Un cadre théorique d'utilisation de contraintes flexibles adapté à CE a été défini. Le site Internet du projet est en cours de construction.

6. Bibliographie

- Benhamou F., Goualard F. (2000). Universally Quantified Interval Constraints. Sixth International Conference on Principles and Practice of Constraint Programming, Singapore. Lecture Notes in Computer Science 1894.
- S. Bistarelli, P. Codognet, Y. Georget, and F. Rossi, « Abstracting Soft Constraints, » *In: New Trends in Constraints*, K. Apt et al. (Eds.), Lecture Notes in AI 1865, Springer Verlag 2000
- Fischer X., Nadeau J.P., Sebastian P., Joyot P., Conception Inversée Intégrée (CII), IDMME 2000, CPI01, Fes (Maroc) octobre, (2001)
- L. Zimmer, P. Zablit « Global Aircraft pre-design based on constraint propagation and

interval analysis » in CEAS Conference on multidisciplinary Aircraft design and Optimization, 25-26 juin 2001

DraMachina : atelier d'aide à l'écriture de fictions interactives

Stéphane Donikian
IRISA/CNRS
Campus de Beaulieu
35042 Rennes Cedex
donikian@irisa.fr
Tél. : +33 2 99 84 72 57

La fiction interactive est une extension de la fiction classique, en ce sens qu'elle suppose une participation active du spectateur avec pour objectif d'influencer le déroulement de l'histoire [Szi99, Szi01, Mat00a, Mat00b, Rya97, Rya01]. La fiction est, par essence linéaire. Elle est une progression ordonnée d'événements, dont le mouvement est toujours ordonné vers l'avant, même si des flash-back ou des ellipses sont autorisés. Dans la fiction interactive, le spectateur n'est pas passif comme dans les œuvres littéraires, cinématographiques, théâtrales ou chorégraphiques classiques, mais actif et peut ainsi influencer sur l'évolution de l'histoire : il n'y a donc pas une histoire mais une infinité d'histoires pouvant se dérouler sur la base de la trame définie par l'auteur. Cela impose un niveau de complexité supplémentaire pour l'auteur, car il doit envisager l'ensemble des hypothèses requises pour qu'une action dramatique se produise et il doit aussi formuler l'impact de chaque action dramatique sur le déroulement de la suite de l'histoire. Les moyens dont le scénariste dispose aujourd'hui sont encore assez rudimentaires (essentiellement des outils de traitement de texte et de représentation graphique d'un scénario décomposé en une arborescence de scènes) au regard de l'évolution des techniques. Le projet DraMachina, objet d'un partenariat entre l'IRISA/CNRS (projet Siames) et la société Dæsign (anciennement Dramaera), contribuera à corriger cette situation. Ce projet RNTL précompétitif, démarré à la fin de 2000 et prévu pour s'achever en mars 2003, a pour but la mise au point d'un outil permettant

aux auteurs de concevoir des environnements narratifs en manipulant directement les éléments clefs du récit (lieux, personnages, rôles, relations entre personnages, actions dramatiques, etc.).

L'écriture d'une fiction interactive s'appuie sur la structure narrative d'un texte, le langage cinématographique ainsi que la psychologie et la notion de couple relationnel, c'est à dire l'état d'une relation entre deux personnages. Plusieurs chercheurs se sont intéressés à la structure narrative d'une œuvre de fiction [Pro70, Bre73, Dij73, Gre66, Tod69]. Des travaux sont actuellement pour utiliser ces travaux théoriques (sémiotique, structuralisme, post-structuralisme) dans les domaines du jeu et des nouveaux médias [Cos01, Kle01, Cav01, Mac01, Spi02, You01]. Une analyse de la méthodologie d'écriture qu'avait développée Dæsign a été effectuée, ce qui a permis de proposer un modèle de représentation plus ambitieux, capable de prendre en compte les divers types de fictions interactives ou de trames dramatiques définis par les théoriciens de la littérature. Nous avons fait le choix dans l'outil de laisser l'auteur choisir l'architecture de sa fiction et de ne pas lui imposer le cadre rigide des architectures toutes prêtes. Par contre, nous lui offrons les moyens de visualiser celle-ci et d'en analyser la cohérence à tout moment du processus d'écriture. L'auteur construit avant tout la protohistoire, c'est-à-dire une continuité narrative qui raconte l'une des histoires possibles pouvant se dérouler dans l'environnement. A partir de là, il en extrait des éléments particuliers comme les personnages, leur rôle, ou encore les protoscènes. Chaque personnage possède un rôle (fonction motrice dans l'histoire, par exemple « le méchant »), un objectif (état de l'histoire dans lequel il considérera sa fonction, comme accomplie), un système de valeurs (des valeurs, des interdits, les règles de transgression de ces interdits). Chaque personnage peut accomplir un certain nombre d'actions dramatiques, en cohérence avec son rôle (le méchant peut tuer). Les personnages sont également décrits dans leur relation avec les autres : ce sont les couples relationnels. On décrit la façon dont peut évoluer la rela-

tion entre deux personnages en fonction des actions qui se déroulent. Les briques de base de l'histoire sont appelées unités dramatiques. Elles sont représentées dans une carte, qui permet de repérer les différents chemins que peut prendre le récit, et les éventuels points critiques. Enfin, le « drama manager » représente l'auteur lors de l'exécution. Il arbitre les différentes demandes d'actions dramatiques, de façon à rester dans un arc dramatique fort et cohérent.

La méthode d'écriture que nous avons mise au point permet à l'auteur de manipuler des objets logiques (acteurs, rôles, unités dramatiques...) ayant chacun une fonction bien précise dans le scénario. Ces objets sont bien évidemment étroitement liés les uns aux autres. Nous proposons à l'auteur de manipuler directement ces objets via l'interface, d'en visualiser les liens, de repérer visuellement les conséquences de l'ajout ou du retrait d'un objet. L'application est fortement interactive et l'utilisateur a la possibilité d'ouvrir des vues multiples sur les éléments : il faut donc assurer la cohérence et la synchronisation des données. Pour cela l'architecture de l'application suit le modèle d'architecture PAC-amodeus. Nous avons développé pour DraMachina un éditeur de texte de type WYSIWYG adapté aux besoins du projet, en particulier l'intégration de liens hypertextes afin de simplifier l'accès aux différentes informations manipulées par l'auteur : personnages, lieux, objets, actions dramatiques, unités dramatiques. Un éditeur de graphes est apparu nécessaire pour la réalisation des dialogues et le traitement logique de la cohérence de l'architecture de la fiction. En complément de l'édition de graphes, l'édition de tableaux est nécessaire pour la saisie des dialogues. En effet, une des possibilités du joueur est de pouvoir décider du choix que fera un personnage en ce qui concerne les réponses à des questions posées, influant ainsi directement sur l'évolution de l'histoire. Le module d'extraction de données à partir du langage naturel permet de trouver automatiquement les actions, localisations et déplacements décrits dans un texte. Notamment, concernant, les actions, nous pouvons extraire diffé-

rentes informations : qui fait l'action, de quelle manière, dans quel lieu et quels objets ou autres personnages sont concernés par cette action. Destiné aussi à l'équipe de production de la fiction, l'outil permettra, grâce à la base de données que celui-ci génère, de réduire considérablement l'étape fastidieuse d'analyse et de synthèse des documents produits par le scénariste. Grâce à un export XML, l'outil sera connecté à un « moteur de fiction interactive » développé par la société Dæsign.

L'interface graphique est un point important de ce projet, car l'outil s'adresse à des utilisateurs de logiciels de bureautique. Ceci exige donc une ergonomie d'utilisation en adéquation avec les standards du marché. Pour ces raisons et après l'étude des solutions du marché, nous avons choisi d'utiliser Cordial de Synapse pour le traitement du langage naturel, Qt de Trolltech pour l'interface graphique et les différents éditeurs, Rational Rose comme atelier de génie logiciel pour l'étape de conception, XML pour le stockage des données et Visual C++ comme environnement de développement sous Windows.

Références

- [Bre73] Bremond (C.). – Logique du récit. – Editions du Seuil, 1973, collection Poétique.
- [Cav01] Cavazza (M.), Charles (F.) et Mead (S.J.). Agents' Interaction in Virtual Storytelling – IVA2001, LNAI 2190, pp 156-170, Springer-Verlag – 2001.
- [Cos01] Conférence COSIGN 2001, Computational Semiotics in Games and New Media. Amsterdam, septembre 2001.
- [Dij73] van Dijk (T.A.). – Grammaires textuelles et structures narratives. – Chapitre du livre <<Sémiotique narrative et textuelle>>, Larousse, Paris – 1973.
- [Gre66] Greimas (A.J.). – Sémantique structurale : recherche de méthode. – Paris, Larousse, 1966, Collection Langue et langage.
- [Kle01] Klesen (M.), Szatkowski (J.) et Lehmann (N.). A Dramatised Actant Model for Interactive Improvisational Plays – IVA2001, LNAI 2190, pp 181-194, Springer-Verlag – 2001.
- [Mac01] Machado (I.), Paiva (A.) et Prada (R.). Is the Wolf Angry or... Just Hungry? – International Conference on Autonomous Agents, Montréal, Canada, pp 370-376 – 2001
- [Mat00a] Mateas (M.) et Stern (A.). – Towards Integrating Plots and Characters for

Interactive Drama. – *AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop* – Nov. 2000.

[Mat00b] Mateas (M.). – A Neo-Aristotelian Theory of Interactive Drama. – *AAAI Spring Symposium on AI and Interactive Entertainment* – 2000.

[Pro70] Propp (V.). – Morphologie du conte. – Editions du Seuil, 1970, collection Points. (titre original : Morfologija skazki, Leningrad, Nauka, 1969).

[Rya01] Ryan (M.L.). – Narrative as Virtual Reality. – The John Hopkins University Press – 2001.

[Rya97] Ryan (M.L.). – Interactive Drama: Narrativity in a Highly Interactive Environment. – *Modern Fiction Studies* 43.3 pp 677-707, 1997.

[Spi02] Spierling (U.), Grasbon (D.), Braun (N.) et Iurgel (I.) – Setting the scene: playing digital director in interactive storytelling and creation. *Computer & Graphics* 26(2002) pp 31-44.

[Szi01] Szilas (N.). – A New Approach to Interactive Drama: From Intelligent Characters to an Intelligent Virtual Narrator. *Proc. AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, Technical Report SS-01-02, p. 72-76, AAAI Press – 2001.

[Szi99] Szilas (N.). – Interactive Drama on Computer: Beyond Linear Narrative – *Proc. AAAI Fall Symposium on Narrative Intelligence*, Technical Report FS-99-01, p. 150-156, AAAI Press – 1999.

[Tod69] Todorov (T.). – Grammaire du décaméron. – La Haie, Mouton – 1969.

[You01] Young (M.). – Notes on the Use of Plan Structures in the Creation of Interactive Plot. In *Proc. AAAI Fall Symposium on Narrative Intelligence*, Technical Report FS-99-01, p. 164-167, AAAI Press – 2001.

e.dot Entrepôts de Données Ouverts sur la Toile

Contact projet : Guy Ferran, Xyleme,
Tel. : 01.55 39 33 33 –
guy.ferran@xyleme.com

Objectifs

Développer un entrepôt de données XML intégrant de manière automatique toutes les informations liées au risque de contamination des aliments, qu'elles soient issues du Web, privées ou obtenues sous licence.

Innovation & Points forts

Un projet en rupture avec l'existant : vers le « tout-XML »

- Surveillance automatique du Web XML
- Extraction automatique de données multi sources
- Classification thématiques des informations récupérées
- Intégration sémantique de données hétérogènes
- Stockage des données dans un entrepôt natif XML
- Utilisation du langage de requêtes XML

Retombées

La création de **nouveaux logiciels et services** à partir des produits existants de Xyleme

La construction d'un **entrepôt de données « risque de contamination des aliments »** directement opérationnel pour les industriels de l'agro-alimentaire

Des **avancées technologiques** en matière d'intégration de données

Partenariat

Trois équipes de recherche en informatique :

- IASI/LRI (Université Paris-Sud), spécialisée dans la représentation de connaissances et l'intégration de données
- Gemo/INRIA, spécialisée dans les bases de données XML et le Web
- BIA/INRA, spécialisée dans l'utilisation des bases de données dans le cadre de la prévention des risques de contamination des aliments

Une entreprise experte des technologies liées à XML :

- Xyleme, spécialisée dans la recherche et l'intégration de larges volumes de contenus XML hétérogènes et multi sources. Xyleme a aussi développé un crawler puissant qui permet de surveiller le Web de façon très pointue.

Principe de base

Surveillance continue du Web :

- Crawling du Web HTML et XML
- Classification automatique des pages d'intérêt

Intégration simple dans l'entrepôt :

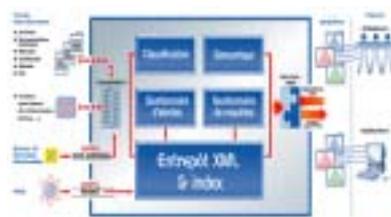
- Récupération automatique des données (en particulier XML) traitant du risque de contamination des aliments
- Possibilité d'enrichir les documents XML

Interrogation très fine des données stockées :

- Ecrans pointus, intuitifs et spécifiques
- Navigation dans le corpus via des hyperliens.
- Mise en place de thésaurus agro-alimentaire

Fourniture rapide de réponses précises :

- Liste triée de réponses
- Calcul original de la pertinence
- Accès direct aux éléments du document
- Granularité des réponses
- Gestion de liens vers des documents PDF...



Réalisations et résultats

Labellisation RNTL en Juin 2002
Début des travaux

EOLE : Environnement d'Optimisation temps réel en LignE dédié télécoms Projet exploratoire

Laurent Jeannin, Juliette Mattioli,
THALES Research & Technology
France
laurent.jeannin@thalesgroup.com

1. Description du projet et du consortium

L'objectif d'EOLE [EOLE 2000] est de fournir un environnement permettant de développer des systèmes d'optimisation combinatoire en ligne dans le domaine télécom.

La technologie développée permet aux mécanismes d'optimisation utilisés de fournir à tout moment une solution opérationnellement utilisable et de faire le meilleur usage du temps et des ressources mis à leur disposition, que ce temps soit connu à l'avance ou non. Elle a été validée sur trois applications pilotes télécoms nécessitant un traitement en-ligne du problème : l'optimisation des ressources réseau lors d'une reconfiguration, l'administration d'un réseau ATM (Asynchronous Transfer Mode) et l'optimisation de l'affectation des fréquences pour le téléphone cellulaire.

Le consortium à forte compétence IA (Intelligence Artificielle) et RO (Recherche Opérationnelle) est constitué des équipes suivantes : TRT – France / Platon, ONERA / DCSD, Cosytec, Bouygues / e-lab et l'école des mines de Nantes France-Télécom / R&D (eole@thalesgroup.com)

2. Description détaillée de la partie IA

EOLE repose en grande partie sur des technologies issue de l'IA car l'approche consiste à combiner la souplesse de modélisation de la programmation par contraintes, la puissance de résolution des algorithmes hybrides (méthodes constructives / approches méta-heuristiques), et l'adéquation aux contraintes opérationnelles temps réel des algorithmes anytime.

La programmation par contraintes permet de modéliser et de résoudre des problèmes d'optimisation combinatoire hors ligne. Elle offre une grande flexibilité en terme de modélisation (il est simple d'ajouter ou de retirer une contrainte à un modèle). De plus, la nature déclarative de ce paradigme permet au programmeur de se focaliser davantage sur la modélisation que sur la programmation. Les solveurs actuels fournissent de puissants algorithmes de résolution au travers des contraintes globales. Cependant ils ne sont pas adaptés à la problématique en ligne.

EOLE apporte une réponse à ce problème en définissant [EOLE 2001] une architecture générique pour un environnement de développement d'applica-

tions d'optimisation en-ligne. Sa finalité est d'avoir les mêmes propriétés en termes de génie logiciel pour la partie résolution que pour la partie modélisation. Un algorithme de recherche est alors la conjonction de quatre composants distincts. Un ensemble d'heuristiques pour ordonner les choix ; elles servent de guide pour le schéma de recherche. Un schéma de recherche est une expression structurée de l'espace de recherche. Une stratégie d'exploration est appliquée à un schéma de recherche pour en contrôler sa complexité. Enfin, une stratégie temporelle, permettant la gestion du temps, est définie par un schéma d'hybridation (i.e. une coopération entre plusieurs recherches paramétrées) fondé sur une répartition temporelle et une politique de réglage des paramètres (statique, itératif ou adaptatif). Le module de contrôle temporel constitue l'élément clé de l'architecture. Il fait office d'interface avec le monde extérieur. Il existe deux modes de fonctionnement de ce module : contractuel (la résolution s'arrête à l'issue du temps imparti initialement) et interruptible (la résolution est susceptible d'être interrompue à n'importe quel instant).

3. Résultats obtenus

Le projet EOLE s'est achevé en novembre 2002. Il a permis de spécifier l'architecture présentée précédemment. Cette architecture a été implantée dans différents environnements, et notamment dans ToOLS© au dessus de Eclair©, le solveur de contraintes dans les domaines finis développé par THALES Research & Technology. ToOLS© est un framework d'optimisation en ligne. Il a deux objectifs : donner un haut niveau d'abstraction pour décrire des algorithmes de recherche et offrir des moyens de gestion du temps permettant de contrôler facilement ces algorithmes. C'est une bibliothèque d'algorithmes de recherche prédéfinis basée sur la notion de patron de recherche (template). Un patron de recherche définit un composant abstrait d'un algorithme de recherche. Ces patrons sont primordiaux pour la réutilisabilité des algorithmes développés.

L'application pilote sur l'optimisation

des ressources réseau lors d'une reconfiguration a permis de valider opérationnellement l'outil ToOLS©. Les retours d'expérience ont montré que l'utilisation de ce framework facilite la conception de nouveaux algorithmes de recherche par combinaisons de stratégies d'exploration et de stratégies temporelles, ce qui en fait un bon outil de prototypage. Les primitives de recherche produisent un code concis et clair qui peut donc être facilement maintenu. Enfin les patrons de recherche permettent la capitalisation des algorithmes développés.

4. Bibliographie

[EOLE 2000] http://www.lcr.thomson-csf.fr/projects/www_eole/
[EOLE 2001] Eole consortium. Towards an on-line optimisation framework. CP-01 workshop On-Line combinatorial problem solving and Constraint Programming (OLCP'01), 2001.

**ExtraPloDocs : EXTRACTION de connaissance pour l'EXPLOitation de la DOCUMENTATION Scientifique
Projet exploratoire
labellisé en 2002
démarrage début 2003**

(<http://www-lipn.univ-paris13.fr/~extraplodocs/>).

ExtraPloDocs regroupe 3 partenaires :

2 laboratoires de recherche :

- Le LIPN assure la coordination du projet (équipe « Représentation des Connaissances et Langage Naturel ») Laboratoire d'Informatique de Paris-Nord (LIPN, <http://www-lipn.univ-paris13.fr>), UMR 7030, Institut Galilée, Université Paris-Nord, av. J.B. Clément, 93430 Villetaneuse. Contact : Adeline Nazarenko (nazarenko@lipn.univ-paris13.fr) ;
- Le laboratoire MIG de l'INRA Laboratoire Mathématique, Informatique et Génome (MIG), INRA, Domaine de Vilvert, 78352 F- Jouy-en-Josas cedex. Contact : Claire

Nédellec (*claire.nedellec@diamnd.jouy.inra.fr*);

1 industriel :

Hybrigenics est une société de protéomique fonctionnelle qui identifie et valide des nouvelles cibles et des molécules thérapeutiques grâce à l'étude des interactions protéine-protéine et au phénotypage dans les cellules humaines.

1. Objectif du projet

Le projet ExtraPloDocs vise à concevoir des outils logiciels permettant de donner accès au contenu d'une documentation technique. Il s'agit non seulement de sélectionner les documents pertinents au sein d'une documentation potentiellement volumineuse mais aussi d'extraire des informations factuelles de ces documents.

Aujourd'hui les moyens automatiques d'accès à l'information scientifique sont limités à la recherche documentaire à l'aide de requêtes, de mots-clés et d'index. L'imprécision des résultats rend la tâche d'exploitation des documents sélectionnés coûteuse voire infaisable. Pourtant l'exploitation de la documentation scientifique par les scientifiques eux-mêmes et par les techno-veilleurs est un facteur crucial de progrès et d'innovation. L'accès automatisé à l'information réclame de nouvelles technologies dans les domaines de la recherche documentaire et de l'extraction d'information.

Les approches développées dans ExtraPloDocs seront guidées et validées par les besoins identifiés en génomique fonctionnelle, tout en conservant leur applicabilité dans les domaines scientifiques présentant les mêmes spécificités (domaine scientifique expérimental, communication scientifique organisée en bases de connaissances distribuées et hétérogènes, tentatives de normalisation terminologique et ontologique...). Citons par exemple la médecine et l'astrophysique.

Du point de vue documentaire, la génomique fonctionnelle forme un domaine exemplaire à plusieurs titres. L'essor exponentiel des méthodes expérimentales en biologie moléculaire a profondément modifié les pratiques des

chercheurs au cours de la décennie précédente. Traditionnellement, un projet de recherche en biologie moléculaire s'intéresse à un petit nombre d'objets, gènes ou protéines, responsables d'une fonction biochimique ou physiologique précise. Auparavant, quelques centaines de publications scientifiques en moyenne se rapportaient au système étudié, et pouvaient être traitées manuellement. Cette manière de procéder n'est plus praticable aujourd'hui car le volume de la bibliographie concernée atteint des milliers, voire quelques dizaines de milliers de publications. Beaucoup des connaissances fines qui sont nécessaires pour interpréter et analyser les résultats expérimentaux ne sont décrites que dans ces publications, sous la forme de texte non structuré.

C'est pourquoi ExtraPloDocs vise à fournir des outils qui contribuent à faire le lien entre les données factuelles issues des expériences à grande échelle et structurées en bases de données et les connaissances de la littérature scientifique. Notons que les méthodes mises au point pour repérer, extraire et représenter l'information pertinente doivent posséder une grande précision et une grande sensibilité pour être compatibles avec le degré d'automatisation élevé requis par le volume des données textuelles à traiter.

Une collaboration active entre différents champs de l'IA

L'une des originalités du projet ExtraPloDocs repose sur la collaboration entre différents champs de l'IA : le traitement des langues naturelles (TAL), l'ingénierie des connaissances (IC) et l'apprentissage automatique (App). Le LIPN apporte son expertise en TAL et IC tandis que le laboratoire MIG prend en charge les questions d'apprentissage. La dimension applicative du projet repose sur l'expérience partagée de MIG et d'Hybrigenics en génomique.

Pour répondre à l'exigence de pertinence de la recherche documentaire comme de l'extraction d'information, il faut mettre en œuvre des moyens d'analyse du contenu reposant sur des ressources lexicales, syntaxiques et sémantiques spécifiques au domaine étudié.

Les méthodes et outils de traitement automatique des langues permettent de construire automatiquement des représentations normalisées et enrichies des documents sur lesquelles peuvent s'appliquer des moyens de sélection de documents et d'extraction d'information.

Les ressources sont spécifiques au domaine et les moyens de sélection et d'extraction sont spécifiques à la tâche. Les uns et les autres sont longs et coûteux à acquérir manuellement et l'acquisition manuelle pêche généralement par manque d'exhaustivité. L'application de l'apprentissage automatique à partir de corpus textuels dans les domaines spécifiques est une alternative reconnue mais encore débutante.

Il faut également prendre en compte les ressources existantes quand elles existent (une liste de noms de gènes par exemple). Ces connaissances permettent d'obtenir un meilleur traitement des données textuelles et facilitent donc l'apprentissage de ressources spécifiques à la tâche, que ces ressources soient d'ordre lexical, conceptuel ou bibliographique. Il faut par ailleurs interpréter les informations extraites au regard d'un modèle conceptuel, la langue introduisant toujours des déformations par rapport à l'ontologie du domaine. L'ingénierie des connaissances aide à articuler et intégrer de manière homogène ces différents niveaux de connaissances.

Les travaux récents en apprentissage de ressources ontologiques et terminologiques et de règles d'extraction vont dans ce sens, les partenaires d'ExtraPloDocs y ont contribué. Leur dispersion et leurs limitations empêchent encore une utilisation automatisée et efficace. Des étapes clés sont manquantes. Notamment, la prise en compte des ressources existantes et l'utilisation de l'ensemble des ressources, qu'elles soient acquises ou apprises, pour l'apprentissage de règles d'extraction, et de critères de sélection de documents, autres que thématiques.

ExtraPloDocs a pour objectif de concevoir et d'adapter des outils logiciels d'apprentissage automatique de ressources lexicales, syntaxiques et sémantiques à partir de textes en prenant en compte des ressources préexistantes. Ces textes sont préalablement normali-

sés par des techniques de traitement automatique des langues.

2. Etat d'avancement

Le projet ExtraPloDocs n'a pas encore réellement démarré. Les équipes participantes ont défini l'approche générale d'ExtraPloDocs et ont démarré deux tâches particulières.

L'approche générale est présentée dans la figure 1. Cette approche repose sur trois méthodes différentes d'exploitation d'une documentation scientifique : il s'agit de sélectionner les textes pertinents, de sélectionner les fragments les plus intéressants et enfin d'extraire de la connaissance véhiculée par les textes pour en donner une représentation formelle.

Ces trois méthodes sont complémentaires et reposent sur des techniques apparentées. Des traitements de complexité croissante sont effectués à chaque étape, mais les traitements intermédiaires apportent déjà une aide appréciable en matière d'exploration documentaire :

- La première étape « Sélection de corpus » consiste à constituer un corpus en sélectionnant au sein de la base documentaire les documents effectivement pertinents au regard de la requête formulée par l'utilisateur. Ces corpus sont directement exploitables par les utilisateurs comme source d'information et ils constituent le corpus approprié sur lequel devra porter l'extraction automatique d'information.

Cette sélection repose sur un classifieur de documents qui dépend lui-même des types de documents considérés et des critères de pertinence de l'utilisateur.

- La deuxième étape « Sélection de fragments » permet d'identifier dans les documents du corpus sélectionné à l'étape précédente, les portions de texte qui concernent effectivement le problème applicatif posé. Ils sont généralement très dispersés, par exemple les phrases sur les interactions géniques forment 3 % des résumés sur le sujet. Cette étape de filtrage est donc essentielle, que ce soit pour présenter directement les fragments sélectionnés à l'utilisateur sous la forme de texte surligné et comme une

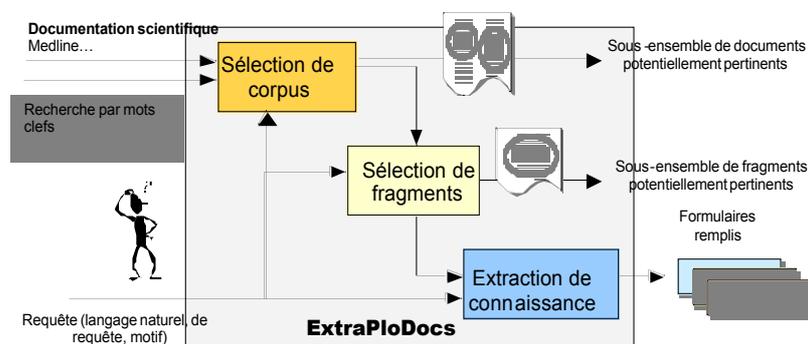


Figure 1 : Exploitation de documentation dans ExtraPloDocs

Fragment 1

Furthermore, HIPK2 and p53 cooperate in the activation of p53-dependant transcription and apoptotic pathways.

Fonctionnal role

Source : protein HIPK2
Target : apoptosis

Fonctionnal role

Source : protein p53
Target : apoptosis

Fragment 2

Here we demonstrate that homeodomain-interacting protein kinase-2, a member of a novel family of nuclear serine/threonine kinases, activates p53 by directly phosphorylating it at Ser 46by directly phosphorylating it at Ser 46

Membership

Element : protein kinase-2
Family : serine /threonine kinases

Interaction

Source : protein kinase-2
Target : p53
Type : Positive

Figure 2. Des fragments de texte, aux formulaires remplis.

aide à la lecture les documents, ou pour aborder le traitement lourd d'extraction de connaissance proprement dit.

Prenons comme exemple le domaine de la génomique fonctionnelle qui est l'application test de ExtraPloDocs. Pour identifier les interactions fonctionnelles entre les objets biologiques, problème clef de la génomique fonctionnelle, il est important de savoir repérer dans les résumés d'articles scientifiques de Medline des fragments tels que celui de la figure 2 (colonne de gauche).

De la même manière que pour la sélection de corpus, cette sélection de fragments repose sur un classifieur de phrases.

- La troisième étape « Extraction de connaissance » permet d'extraire et de donner une représentation formelle de l'information pertinente. La figure 2 montre quel type de connaissance le module d'extraction vise à extraire. Cette connaissance est représentée sous la forme de formulaires (colonne de droite sur la figure 2). Du fragment 2, on retient ainsi l'information que HIPK2 appartient à la famille des kinases sérine/threonine et que HIPK2 interagit positivement (activation) sur le gène p53.

Dans le cadre de cette architecture d'ensemble, les travaux ont débuté sur deux points spécifiques :

- Diverses techniques sont testées pour

la sélection de fragments. Les premières expériences menées montrent que des techniques de type « Naive Bayes » donnent de bons résultats.

- Le module d'analyse terminologique qui permet de repérer les expressions clefs d'un document est en cours de développement. Il s'appuie sur des outils existants en terminologie computationnelle. La contribution des diverses bases de connaissances existantes est actuellement à l'étude.

3. Références

Philippe Bessières, Adeline Nazarenko et Claire Nédellec. « Apport de l'apprentissage à l'extraction d'information - le problème de l'identification d'interactions géniques ». *Actes du 4ème Colloque International sur le Document Electronique (CIDE'01)*, Toulouse, oct. 2001.

Bisson G., Nédellec C. and Canamero D. « Designing clustering methods for ontology building - The Mo'K workbench » in *Proceedings of the workshop on Ontology Learning*, workshop of the European Conference on Artificial Intelligence (ECAI-2000), S. Staab, et al (Eds.). pp. 13-19, Berlin, August 2000.

Blaschke C., Andrade M. A., Ouzounis C. and Valencia A., « Automatic Extraction of biological information from scientific text: protein-protein interactions », in *Proceedings of International Symposium on Molecular Biology, (ISMB'99)*, 1999.

Califf M. E. and Mooney R. J., « Relational Learning of Pattern-Match Rules for Information Extraction. » In *Proceedings of AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, CA, March, 1998.

Craven M. and Kumlien J., « Constructing Biological Knowledge Bases by Extracting Information from Text Sources. » In *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology (ISMB-99)*, 1999.

Faure D. and Nédellec C., « Knowledge Acquisition of Predicate-Argument Structures from technical Texts using Machine Learning » in *Proceedings of Current Developments in Knowledge Acquisition: EKAW-99*, pp. 329-334, Fensel D. and Struder R. (Eds.), Karlsruhe, Germany, April 1999.

Freitag D., « Multistrategy learning for information Extraction. » In the *Proceedings of the Fifteenth Machine Learning Conference (ML-98)*, J. Shavlik (Ed.), Madison, USA, 1998.

Habert, B et Nazarenko, A., « La syntaxe comme marche-pied de l'acquisition des connaissances : bilan critique d'une expé-

rience ». *Actes de Journées sur l'acquisition des connaissances*, pp. 137-142, Sète, mai 1996.

Hamon T. and Nazarenko A., « Detection of synonymy links between terms: experiment and results », *Recent Advances in Computational Terminology*. John Benjamins, 2001.

Hobbs, J, Appelt, D., Bear, J. Israel, D., Kameyana, Stickel, M., M. Tyson, M., « FASTUS : a cascaded finite-state transducer for extracting information in natural language text ». in *Finite state Language Processing*, E. Roche a E. Schaves (eds), MIT Press, Cambridge, 1997.

MUC, *Proceedings of the Message Understanding Conference (MUC-4-7)*, Morgan Kaufman, San Mateo, USA, 1992-98.

Nédellec C. et Nazarenko A. « Ontology and Information Extraction », in *Handbook on Ontologies*, to appear, 2003.

Nédellec C., Ould Abdel Vetah M., et Bessières P., « Sentence Filtering for Information Extraction in Genomics, a Classification Problem ». In *Proceedings of the Conference on Practical Knowledge Discovery in Databases*, Springer Verlag, Freiburg, Septembre 2001.

Ono T., Hishigaki H., Tanigami A. and Takagi T., « Automated extraction of information on protein-protein interactions from the biological literature », in *Bioinformatics* vol 17, n° 2, pp. 155-161, Oxford Press, 2001.

Pillet V., *Méthodologie d'extraction automatique d'information à partir de la littérature scientifique en vue d'alimenter un nouveau système d'information*, thèse de l'Université de droit, d'économie et des sciences d'Aix-Marseille, 2000.

Poibeau T. et Nazarenko A., « L'extraction d'information, une nouvelle conception de la compréhension de texte ? » *Traitement Automatique des Langues*, 40(2), 1999.

Poibeau, T. *Extraction d'information à base de connaissances hybrides*. Thèse de doctorat en Informatique, Université Paris-Nord. Mars 2002.

Thomas, J., Milward, D., Ouzounis C., Pulman S. and Carroll M., « Automatic Extraction of Protein Interactions from Scientific Abstracts ». In *Proceedings of the Pacific Symposium on biocomputing (PSB'2000)*, vol.5, p. 502-513, Honolulu, 2000.

Nédellec C., Ould Abdel Vetah M., et Bessières P., « Sentence Filtering for Information Extraction in Genomics, a Classification Problem ». In *Proceedings of the Conference on Practical Knowledge Discovery in Databases*, Springer Verlag, Freiburg, Septembre 2001.

**INKA : Génération automatique de cas de tests assurant la couverture structurelle d'un logiciel
Projet pré-compétitif
Démarrage : 01 / 2001**

1. Équipes IA :

THALES SYSTEMES AEROPORTES

M. Bernard Botella
Adresse : 1 Bd Jean Moulin
78852 Elancourt Cedex, FRANCE
Tél : 01.34.59.78.23
Fax : 01.34.59.52.50
Mél :

bernard.botella@fr.thalesgroup.com

Université de Nice Sophia Antipolis – Laboratoire Informatique, Signaux et Systèmes (I3S)

M. Michel Rueher
Adresse : Université de Nice Sophia Antipolis ESSI Route des colles, BP 145, 06903 Sophia Antipolis
Tél : 04 92 96 51 54
Fax : 04 92 96 51 55
Mél : rueher@unice.fr

Université de Franche-Comté – Laboratoire d'Informatique (LIFC)

M. Bruno Legeard
Adresse : 16 route de Gray, 25030 Besançon cedex
Tél : 03 81 66 64 55
Fax : 03 81 66 64 50
Mél : bruno.legeard@univ-fcomte.fr

2. Identification du reste du partenariat :

Axlog Ingénierie

M. Bertrand Allo
Adresse : 19-21 rue du 8 mai 1945
94110 Arcueil
Tél : 01 41 24 31 16
Fax : 01 41 24 07 36
Mél : bertrand.allo@axlog.fr

Université Joseph Fourier – Laboratoire Logiciels Systèmes Réseaux (IMAG)

M. Ioannis Parissis
Adresse : BP 72, 38402 Saint Martin d'Hères Cedex
Tél : 04 76 82 72 89
Fax : 04 76 82 72 87
Mél : Ioannis.Parissis@imag.fr

3. Description générale du projet :

Le test structurel est une étape primordiale pour assurer une qualité minimale des logiciels. Les objectifs du test structurel ne proviennent pas des spécifications du logiciel comme pour les autres phases de tests, mais directement de directives de qualité associées au projet en termes de couverture de test des programmes. C'est pour cette raison que les tests fonctionnels ne peuvent assurer seuls une qualité optimale et doivent être complétés par une phase de test structurel qui vise explicitement cet objectif de couverture.

Cependant le test structurel est une tâche difficile, longue et coûteuse car trouver les cas de tests permettant d'exécuter les parties non encore couvertes par les tests fonctionnels requiert de compétences spécifiques et une analyse minutieuse de la structure du code source du logiciel.

Le projet InKa [1] vise à développer un outil permettant de déterminer automatiquement les cas de tests structurels, réduisant drastiquement le temps et le coût de mise en œuvre du test structurel.

4. Description détaillée de la partie IA du projet :

Le cœur d'InKa fait appel à la programmation par contraintes et repose sur une technique originale permettant de transformer le code source d'un programme en un CSP (Constraint Satisfaction Problem) équivalent.

Certaines variables de ce CSP représentent le passage ou non du flot de contrôle dans les branches du programme, les autres variables représentant l'état des variables du programme original.

En contraignant ces variables il est donc possible de choisir un point de passage, puis par résolution du système, d'obtenir les valeurs d'entrée permettant d'atteindre ce point.

Une couche de contraintes adaptée à la représentation d'un langage de programmation a donc été définie. Cette couche comporte différents composants qui collaborent au sein du même solveur. Tout d'abord un certain nombre d'opérateurs ont été définis pour modéliser les structures de contrôle du langage

(if, while, ...) [2]. Ces opérateurs modélisent à la fois le sens direct correspondant à l'exécution de la structure de contrôle mais également le sens inverse. Ensuite, les différentes données du programme et les opérations sur ces données sont modélisées via des solveurs adaptés. Les données assimilables à des entiers (int, long, char, enum, ...) sont modélisées à l'aide d'un solveur CLP (FD). Les données de type flottant ont fait l'objet d'une étude menée par l'IS3 et un solveur adapté a été défini [3]. La représentation des notions de pointage et d'allocation dynamique a été étudiée par le LIFC. Elle a nécessité la mise en place d'un niveau intermédiaire permettant d'exprimer des contraintes ensemblistes sur des couples (adresse-valeur), les valeurs étant des inconnues de type entier ou flottant. Ces contraintes permettent d'exprimer l'ajout, l'accès, la modification et la suppression d'une nouvelle donnée en mémoire.

L'ensemble de ces contraintes est implanté dans un solveur développé en prolog au dessus de la bibliothèque de contraintes de SICSTUS Prolog.

5. Etat d'avancement, résultats obtenus :

Le développement de l'outil InKa est réalisé par THALES SYSTEMES AEROPORTES et AXLOG. La pré-version (V0) est disponible, elle traite un sous-ensemble significatif du langage C. La version V1 qui complète le traitement du langage C et l'étend au C++ est en cours de développement.

6. Bibliographie :

Site Web du projet : http://www.axlog.fr/R_d/inka/inka_cadre.html
Articles disponibles à <http://www.essi.fr/~rueher>

[1] Génération automatique de cas de test structurels avec des techniques de programmation par contraintes, Techniques et Sciences Informatiques, 2002 (à paraître).

[2] A CLP Framework for Computing Structural Test Data. Arnaud Gotlieb, Bernard Botella, Michel Rueher. International Proceedings of Computational Logic (CL'2000), London, Jul. 2000. LNAI

1891, Springer Verlag, pp 399-413.

[3] *Solving constraints over floating point numbers*. Claude Michel, Michel Rueher, Yahia Lebbah. Proc of. 7th International Conference on Principles and Practice of Constraint Programming 2001 (CP'2001), LNCS 2239 (Toby Walsh Ed., Springer Verlag), pp 524-538

LUTIN : Outillage des patrons de conception et d'architecture avec prise en compte de leur sémantique
Projet exploratoire sur 36 mois
démarré le 01/01/2002

Page web : http://www.industrie.gouv.fr/rntl/AAP2001/Fiches_Resume/LUTIN.htm

Coordonnateur : Mikal Ziane

Mél : Mikal.Ziane@lip6.fr

Tél : 01 44 27 87 46

1. Équipe IA

Laboratoire d'Informatique de Paris 6 (LIP6)

Université Pierre et Marie Curie

Pole IA thème OASIS

8, rue du capitaine Scott

75015 Paris

2. Autres partenaires

Laboratoire PRISM

Université de Versailles St-Quentin en Yvelines

Equipe Systèmes d'Information et Architectures Logicielles

45, avenue des États-Unis

78035 Versailles Cedex

SOFTEAM

144 Avenue des Champs-Élysées

75008 PARIS

3. Objectifs

Le projet Lutin vise l'intégration des patrons de conception et d'architecture dans les ateliers UML et en particulier dans l'atelier Objecteering de la société

Softeam. L'objectif principal est d'améliorer la productivité des développeurs qui utilisent des patrons. Le point le plus original de notre démarche est la prise en compte de la partie problème des patrons. Ceci permettra la présélection automatique de solutions candidates aux problèmes exprimés par le développeur et la génération du code ou de modèles UML associés à la solution finalement choisie par ce dernier.

Partie IA

L'idée de base du projet Lutin est celle du génie logiciel à base de connaissances et consiste à intégrer une partie des connaissances des développeurs de logiciels dans des outils pour mieux les aider.

La première difficulté est d'identifier ces connaissances. Or les patrons de conception et d'architecture visent précisément à diffuser une partie des ces connaissances, celles qui permettent aux développeurs à objets de résoudre des problèmes récurrents dont les solutions ont été élaborées et validées (le plus souvent empiriquement) au fil des années.

Toutefois les patrons s'adressent à des personnes et ne sont pas écrits pour être intégrés à des outils. Les patrons sont décrits de façon informelle à l'aide notamment d'exemples. Ce format les rend très accessibles mais rend leur outillage particulièrement difficile.

De nombreux travaux ont abordé l'outillage ou la formalisation des patrons mais sans aborder la question de la représentation des connaissances que les patrons visent à disséminer. La plupart de ces travaux concerne la formalisation de la partie solution sans même effleurer la partie problème ! Enfin ces travaux n'évoquent pratiquement pas les mécanismes sous-jacents aux patrons qui permettent de retrouver les solutions à partir des problèmes.

Comment représenter les connaissances sous-jacentes aux patrons ?

Une partie de ces connaissances concerne justement ces mécanismes fondamentaux sous-jacents qui permettent de retrouver les solutions à partir du problème. Nous avons montré qu'au moins certains de ces mécanismes peuvent être bien représentés par un petit nombre de règles de réécriture très générales :

essentiellement des variantes de la notion de pliage [ZIA 00,01,02].

L'intérêt de ces diverses sortes de pliages est d'introduire différentes sortes d'indirections. Ceci permet de résoudre notamment des problèmes de découplage qui apparaissent très fréquemment dans les patrons et par là même de favoriser la maintenance (une modification sur une partie d'un logiciel à un impact plus localisé).

Nous avons pu ainsi « redécouvrir » les solutions de plusieurs patrons à partir de règles simples et très générales.

Comment représenter la partie problème des patrons ?

Un de nos premiers résultats dans le projet Lutin a été de clarifier la notion de problème que résout un patron et de montrer que ce problème a deux aspects : une intention fonctionnelle et une intention non fonctionnelle [ZIA 02]. L'intention fonctionnelle d'un patron ne pose typiquement pas de difficulté en tant que telle au développeur : il peut l'exprimer facilement dans son formalisme habituel (UML ou un langage de programmation). Ce qui pose problème au développeur c'est d'arriver à exprimer cette intention tout en respectant des règles de qualités. Ces contraintes de qualité correspondent à ce que nous appelons l'intention non fonctionnelle du patron. Nous travaillons actuellement sur le moyen d'exprimer ces contraintes en UML ou dans une variante d'OCL.

Ces contraintes contrôlent que les solutions candidates produites pas les règles de réécriture résolvent bien le problème du développeur. Autrement dit, le rôle d'un outil d'assistance est de proposer des solutions équivalentes à l'intention fonctionnelle "naïve" du développeur mais qui respectent les contraintes de qualité exigées.

Nous ouvrons donc la voie à une sélection semi-automatique des patrons en fonction du problème auquel fait face le développeur.

Mieux, il devient envisageable à terme, au moins dans certains cas, de détecter automatiquement le problème : lorsque les contraintes de qualité sont violées.

Il nous reste toutefois un travail considérable avant de disposer d'un véritable langage pour exprimer les contraintes de qualité.

4. Conclusion

Au bout d'un an nous avons pu clarifier la notion de problème d'un patron ce qui nous conduit à présent à développer un langage pour exprimer des contraintes non fonctionnelles en nous appuyant sur UML et OCL.

Dans certains cas il serait fastidieux pour le développeur d'exprimer son intention fonctionnelle complètement pour qu'elle soit finalement « corrigée ». Nous proposons donc qu'il s'exprime de façon un peu plus abstraite à l'aide de plusieurs sortes de méta-variables [ZIA 02] ce qui facilite par ailleurs la sélection des règles de réécritures applicables.

Nous avons développé en prolog un prototype qui comprend plusieurs règles de pliages et permet de retrouver, au moins en partie, les solutions de plusieurs patrons.

Par ailleurs nous avons développé les règles de traduction vers B [MAR 00a, 00b] qui nous permettront de valider le processus de transformation au moins pour certains patrons.

5. Bibliographie

- [MAR 00a] R. Marcano-Kamenoff, « Formalizing Pattern Applicability - an Approach Based on UML and B, » presented at ASE Doctoral Symposium, Grenoble, France, 2000.
- [MAR 00b] R. Marcano-Kamenoff, N. Levy, and F. Losavio, « Spécification et Spécialisation de Patterns en UML et B, » presented at LMO, Mont Saint-Hilaire, Québec, 2000.
- [ZIA 00] M. Ziane, « A Transformational Viewpoint on Design Patterns, » in Proc. IEEE Automated Software Engineering Conference (ASE), Grenoble, 2000.
- [ZIA 01] M. Ziane, « Towards Tool Support for Design Patterns Using Program Transformations, » in Proc. Langages et Modèles à Objets (LMO'01), Le Croisic, France, 2001.
- [ZIA 02] M. Ziane, « Redécouvrir les Solutions des Design Patterns, » rapport LIP6 2002/017, septembre 2002.

L'Interactivité dans des Mondes Virtuels Projet Exploratoire Année 2001

1. Partenaires

LAAS-CNRS, groupe Robotique et Intelligence Artificielle,
7 Av. du colonel Roche.
31077 Toulouse Cedex 4.
Contact: Malik Ghallab, malik@laas.fr,
Tel: 05 61 33 62 59.
Thierry Siméon, nic@laas.fr,
Tel: 05 61 33 63 49.

Virtools SA, 93 rue Vieille du Temple.
75003 Paris.
Contact: S. Dupasquier, dupasquier@virtools.com,
Tel: 01 42 71 46 86

Sous-traitance: **Kineo** Computer Aided Motion (spin-off du LAAS-CNRS créée pour valoriser la technologie Move3D),
Prologue - La Pyrénéenne,
31312 Labège Cedex.
Contact: J.P. Laumond,
jpl@kineocam.com,
Tel: 05 61 00 90 60.

2. Objectifs

Le projet *NeMove3D* porte sur l'augmentation des capacités d'entités virtuelles dans des environnements graphiques 3D. Il part de deux technologies maîtrisées respectivement par *Virtools* sur la programmation à base de comportements, et par le *LAAS-CNRS* sur la planification du mouvement. L'objectif principal réside dans la fusion de ces deux technologies : il s'agit de doter des objets des mondes virtuels 3D de plus d'autonomie en libérant l'utilisateur des contraintes liées à la gestion de l'espace et du mouvement. Le projet *NeMove3D* ouvre la voie à la conception d'interfaces de haut niveau apportant un gain de productivité aux développeurs de contenu graphique, et un saut qualitatif dans la richesse des scénarios dans des applications d'animation interactive ou les jeux vidéos.

3. Description et Résultats.

Synthèse de mouvement de personnages animés

L'animation d'objets graphiques repose aujourd'hui sur des comportements pré-programmés simples (*Building Blocks* dans *Virtools*) où la planification de mouvement n'apparaît qu'à un stade embryonnaire. L'autonomie de mouvement est cependant une composante essentielle pour le développement d'artefacts virtuels dotés de capacités décisionnelles plus évoluées. Le premier axe de recherche porte donc sur la synthèse automatique de mouvements pour des personnages animés évoluant dans des environnements encombrés d'obstacles.

La planification de mouvement constitue un thème de recherche très actif dans la communauté robotique depuis ces vingt dernières années. Le principe général de résolution consiste à transformer le problème du déplacement du système mécanique dans son environnement 3D en celui du déplacement d'un point dans un espace caractéristique des positions du système, appelé *espace des configurations*. Planifier un mouvement sans collision revient alors à construire et explorer les composantes connexes de cet espace. Alors que la planification de mouvement s'est longtemps heurtée à la complexité combinatoire inhérente au problème (en particulier lorsqu'il s'agit d'explorer des espaces de configuration hautement dimensionnés), de nouvelles approches de type probabiliste [1] permettent aujourd'hui de résoudre efficacement une large classe de problèmes pour des systèmes articulés décrits par des chaînes cinématiques complexes, tout en rendant compte des contraintes propres au mouvement du système. La technologie développée au LAAS et intégrée dans le moteur de mouvement *Move3D* [2] repose sur ces techniques récentes qui se démarquent par leur généralité et leur capacité à traiter des systèmes mécaniques complexes.

Les travaux menés dans *NeMove3D* visent à étendre ce moteur de mouvement aux problèmes nouveaux posés par le mouvement de personnages virtuels. Il s'agit en particulier d'intégrer des capacités de réactivité face aux environ-

nements dynamiques que l'on rencontre dans la plupart des applications pratiques, et aussi de prendre en compte des contraintes afin d'assurer le rendu réaliste des mouvements du personnage.



Synthèse automatique de la marche réaliste d'un personnage dans une scène encombrée d'obstacles

L'approche proposée dans [3] pour la **synthèse de mouvements de marche naturelle** d'un personnage animé repose sur l'utilisation de séquences de capture de mouvement (qui permettent de reproduire à l'identique le mouvement enregistré d'un acteur réel) et sur une méthode de composition permettant de recréer à partir de quelques séquences élémentaires (marche avant, virage à droite ou à gauche) de nouvelles séquences adaptées au mouvement à réaliser, tout en intégrant les contraintes d'évitement d'obstacle.

En parallèle au problème du déplacement autonome de personnages animés, des recherches plus amont portent également sur les problèmes posés par la **manipulation d'objets**, nécessaire à l'interaction de ces personnages avec leur environnement. Ces travaux [4] ont conduit à de nouvelles techniques permettant, à partir d'un raisonnement purement géométrique de produire automatiquement la décomposition d'une tâche de manipulation en sous-tâches élémentaires, en même temps que la séquence de mouvements sans collision solution de chacune de ces sous-tâches. Ces techniques permettent de résoudre des problèmes pour lesquels une simple tâche de *prise et pose* d'un objet néces-

site un plan de manipulation complexe d'autres objets afin de modifier la structure de l'environnement (déplacement d'un objet gênant, changement de prise).

Comportements Autonomes Planifiés.

Les personnages graphiques animés ont aujourd'hui des comportements entièrement préprogrammés correspondant à des automates finis souvent très simples, spécifiés généralement à l'aide des techniques de programmation traditionnelles (langages impératifs). En parallèle à l'automatisation de la planification de mouvement, qui fait l'objet de l'essentiel de *Nemove3D*, les travaux portent également sur le développement de techniques de planification de tâches, et plus généralement de comportement délibérés, afin de doter ces personnages de capacités décisionnelles permettant à terme un saut qualitatif au niveau de la richesse des scénarios et des fonctions qui pourraient être intégrés aux applications du type animation interactive et jeux vidéo. Deux directions de recherche sont explorées dans le projet.

D'une part, les systèmes dits de **raisonnement procédural** (tels que par exemple *PRS* [5] ou *Propice* [6] développés au *LAAS-CNRS*) ou ceux à base de règles (par exemple *RAP*, *Kheops*) ont été développés en IA pour permettre de spécifier à un haut niveau des comportements de systèmes complexes qui

réagissent à des événements perçus (par des réflexes) et à des buts assignés dynamiquement selon une librairie de plans connus. On étudie donc la mise en oeuvre de tels systèmes pour des personnages animés sur le plan des principes et des extensions algorithmiques éventuellement nécessaires.

Par ailleurs, les systèmes de *planification de tâches* s'appuient sur la modélisation préalable d'un environnement et des types d'actions réalisables par un personnage. Pour un but donné, ces systèmes effectuent la synthèse automatique d'un plan ou séquence d'actions permettant de réaliser ce but. Diverses techniques de planification de tâches sont connues. Celles un tant soit peu performantes mettent en oeuvre des représentations de l'environnement et des actions limitées. Elles admettent des hypothèses très restrictives : actions déterministes sans durée (transition instantanée d'états), environnement complètement connu et statique, pas d'adversité. Dans le cadre du projet *Nemove3D*, on s'intéresse au développement d'une planification de tâches plus adaptée aux applications de jeux et d'animation vidéo, en particulier à travers le couplage entre le planificateur temporel *IxTeT* [7] et le planificateur de mouvement *Move3D*.

4. Bibliographie

- [1] T. SIMEON. Planification de mouvement par échantillonnage aléatoire. 3ièmes Journées Nationales de la Recherche en Robotique (JNRR2001), Presqu'île de Giens (France), 15-17 Octobre 2001.
- [2] T. SIMEON, J.P. LAUMOND, F. LAMIRAUX. Move3D : a generic platform for path planning. 2001 IEEE International Symposium on Assembly and Task Planning (ISATP2001), Fukuoka (Japon), 28-29 Mai 2001.
- [3] J.PETTRE, T.SIMEON, J.P.LAUMOND. Planning human walk in virtual environments. 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2002), Lausanne (Suisse), 30 Septembre - 4 Octobre 2002.
- [4] T. SIMEON, J. CORTES, A. SAHBANI, J.P. LAUMOND. A general manipulation task planner. 5th International workshop on the Algorithmic Foundation of Robotics (WAFR2002), Nice (France), 16-18 Décembre 2002.
- [5] F. INGRAND, M.P. GEORGEFF, A.S. RAO. An architecture for real-time reasoning and system control. IEEE Expert Intelligent Systems and Their Applications, Vol 7, N° 6, December 1992.
- [6] O. DESPOUYS, F. INGRAND. Propice-plan : toward a unified framework for planning and execution. 5th European Conference on Planning (ECP99), Durham (GB), 8-10 September 1999.
- [7] P. LABORIE, M. GHALLAB. Planning with sharable resource constraints. International Joint Conference on Artificial Intelligence (IJCAI95), Montreal, 20-25 Août 1995.

Plate-Forme AFIA 2003

Laval 1er - 4 Juillet 2003

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

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

Organisée tous les deux ans, en alternance avec la conférence RFIA, la Plate-forme AFIA rassemble sur un même site plusieurs conférences spécifiques du domaine de l'Intelligence Artificielle. Cette manifestation est l'occasion de réunir des chercheurs ayant des thématiques voisines et complémentaires, avec des industriels. Ainsi, la précédente édition, organisée en 2001 à Grenoble par le laboratoire Leibniz a permis de rassembler 215 participants.

Ouverte aux conférences émanant des groupes de travail ou des collèges de l'AFIA, la plate-forme a également pour vocation d'accueillir plus largement l'ensemble des communautés qui sont scientifiquement proches de l'IA. L'IA est en effet une discipline « ouverte » qui, à côté de communautés scientifiques clairement identifiées comme faisant partie de ce domaine de recherche, est en collaboration avec d'autres communautés pour lesquelles l'IA n'est pas le domaine scientifique principal. Une manifestation comme la plate-forme est donc une occasion de dialogues et d'échanges avec ces groupes de chercheurs qui sont situés à la périphérie de l'IA.

La plate-forme 2003

Cette plate-forme, qui sera la troisième du nom, se déroule cette année du mardi 1er juillet au vendredi 4

juillet dans la ville de Laval. Elle est organisée par le centre de Recherche du groupe ESIEA. Elle regroupera trois conférences principales :

- **CAp** (Conférence d'Apprentissage)
Responsable du CP : Rémi Gilleron (Université de Lille, équipe Grappa)
- **IC** (Ingénierie des Connaissances)
Responsable du CP : Rose Dieng-Kuntz (Inria-Sophia, projet ACACIA)
- **RJCIA** (Rencontres Jeunes Chercheurs en IA)
Responsable du CP : Florence Dupin de Saint-Cyr (Université d'Angers, LERIA)

Mais la plate-forme accueillera également un ensemble d'ateliers («workshop») qui sont en cours de définition.

Enfin une journée commune de « posters et démonstrations », ouverte à tous, complètera le programme de la semaine. Cette journée a pour objectifs de permettre aux participants de présenter applications et travaux divers et surtout d'échanger librement sur leurs problématiques scientifiques.

Pour tout renseignement n'hésitez pas nous à contacter :

Gilles.Bisson@imag.fr

Anne-Marie.KEMPF@esiea.fr

Appel à communication Cap 2003 - Plate-forme AFIA

**Conférence d'Apprentissage
Laval les 1er - 2 Juillet 2003**

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

Président du comité de programme : **Remi Gilleron** (GRAPPA, Lille)

Dates importantes

- Date limite de soumission : 15 mars 2003
- Notification aux auteurs : 18 avril 2003
- Réception des articles : 16 mai 2003

Présentation

L'apprentissage automatique ou artificiel est essentiel pour la plupart des systèmes informatiques actuels. Son objectif est l'amélioration des performances des systèmes avec l'expérience. Il constitue donc une brique de base pour les systèmes d'extraction d'information et de connaissance à partir de tous types de données.

La Conférence d'Apprentissage (CAp) est, depuis plusieurs années, un rendez-vous annuel d'échange entre chercheurs, universitaires, industriels, utilisateurs avertis ou concepteurs de méthodes d'apprentissage. Elle est ouverte à toutes les approches de l'apprentissage et à tous les domaines d'application avec une priorité sur les méthodes émergentes et les nouveaux domaines d'application.

CAp doit être un événement incontournable pour discuter, confronter les points de vue, rechercher des solutions, initier des collaborations. Le comité scientifique désire encourager les soumissions et favoriser la venue du plus grand nombre de chercheurs concernés par l'apprentissage. Tout papier soumis ou en cours de soumission à une conférence internationale devrait être préalablement soumis à CAp ; les applications et projets réussis sont les bienvenus ; les échecs également s'ils contribuent à l'ouverture de nouvelles problématiques.

Thèmes

Pas d'exclusive sur les méthodes, modèles et objets

de l'apprentissage, ...

Les données hétérogènes, semi-structurées, séquentielles, géographiques, ...

Les applications en fouille de bases de données, fouille de textes et WEB, bio-informatique et biologie, ...

Format de soumission

La longueur des papiers est limitée à 12 pages. La soumission est électronique, les articles devant être envoyés au président du comité de programme gilleron@univ-lille3.fr. Les formats précis de soumission sont disponibles sur le site de la conférence. Le titre et le résumé du papier doivent être en langue française. Le comité se réserve d'accepter des papiers conditionnellement à une écriture dans un format court (4 pages) pour des travaux aux perspectives intéressantes mais qui seraient jugés non aboutis par le comité.

Composition du comité de programme

Président du comité de programme : Remi Gilleron (GRAPPA, Lille)

Florence d'Alché-Buc (LIP6, Paris),
Marc Bernard (EURISE, St Etienne),
Gilles Bisson (LEIBNIZ, Grenoble),
Jean-François Boulicaut (LISI, INSA Lyon),
Laurent Brehelin (LIRMM, Montpellier),
Antoine Cornuejols (LRI, Orsay),
François Coste (IRISA, Rennes),
François Denis (LIM, Marseille),
Pierre Dupont (UCL, Louvain-la-Neuve, Belgique),
Frédéric Garcia (INRA, Toulouse),
Yann Guermeur, (LORIA, Nancy),
Yves Kodratoff (LRI, Orsay),
Lionel Martin (LIFO, Orléans),

Stan Matwin (SITE, Ottawa, Canada),
Laurent Miclet (ENSSAT, Lannion),
Amedeo Napoli (LORIA, Nancy),
Claire Nédellec (INRA, Versailles),
Richard Nock (Antilles - Guyane),
Nicolas Pasquier (I3S, Nice),
Céline Rouveiroi (LRI, Orsay),
Lorenza Saitta (M&EC, Torino, Italie),

Michèle Sebag (LRI / Polytechnique-LMS, Orsay),
Marc Sebban (EURISE, St Etienne)
Henri Soldano (LIPN, Villetaneuse),
Fabien Torre (Grappa, Lille),
Christel Vrain (LIFO, Orléans),
Djamel Zighed (ERIC, Lyon2),
Jean-Daniel Zucker (LIP6, Paris),

Appel à communication

IC 2003 - Plate-forme AFIA

Ingénierie des Connaissances
Laval les 1er - 3 Juillet 2003

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

Présidente du comité de programme : **Rose Dieng-Kuntz** (INRIA-Sophia, ACACIA)

Dates importantes

- Date limite de soumission : 15 mars 2003
- Notification aux auteurs : 30 avril 2003
- Réception des articles : 23 mai 2003

Présentation

Placée sous l'égide du GRACQ (Groupe de Recherche en Acquisition des Connaissances) la conférence IC est un lieu d'échange et de réflexion de la communauté française sur les problématiques de l'ingénierie des connaissances. IC'2003 fait suite aux conférences :

- IC'02 (Insa-Rouen)
- IC'01 (Grenoble: dans le cadre de la deuxième plate-forme (AFIA);
- IC'00 (Toulouse);
- IC'99 (Ecole Polytechnique, Palaiseau, dans le cadre de la première plate-forme AFIA);
- IC'98 (Nancy) et IC'97 (Roscoff) et aux Journées d'Acquisition des Connaissances (JAC) qui se sont tenues
- tous les ans de 1990 à 1996 (Lannion, Sète, Dourdan, Saint-Raphaël, Strasbourg, Grenoble, Sète).

L'évolution actuelle de la « société de l'information » vers une « société de la connaissance », les besoins en gestion des connaissances dans les organisa-

tions ou communautés, l'engouement actuel de plusieurs communautés de recherche pour le Web sémantique peuvent permettre de renforcer l'impact des recherches menées depuis des années en ingénierie des connaissances.

En effet, l'ingénierie des connaissances propose des concepts, méthodes et techniques permettant d'acquérir, de modéliser et de formaliser des connaissances pour les mobiliser dans l'activité individuelle ou collective au sein d'une organisation ou d'une communauté. Structurées, formalisées et opérationnalisées pour être intégrées dans le fonctionnement d'un système basé sur des connaissances (SBC), ces connaissances deviennent des informations destinées à être, *in fine*, interprétées par un humain dans son interaction avec l'artefact construit. La réussite d'une démarche d'ingénierie des connaissances peut alors se mesurer par rapport à 3 objectifs :

- Construire un SBC ayant un comportement appréhendable et compréhensible par l'utilisateur. Cela passe par une modélisation à un niveau d'abstraction pertinent qui fait sens pour les différents acteurs impliqués (cognitivistes, experts métiers, utilisateurs, etc.) et leur permet de s'approprier le comportement du système et d'interagir avec lui. C'est le « niveau des connaissances ».
- Appréhender le fonctionnement et l'utilisation d'un SBC par rapport à son intégration dans une organisa-

tion. L'ingénierie des connaissances doit alors permettre de modéliser et d'expliciter le fonctionnement cognitif de l'utilisateur ou du groupe d'utilisateurs dans leur relation au SBC.

- Développer et mettre en place une réelle démarche d'ingénierie, reproductible et proposant des instrumentations techniques - i.e. des méthodes et des outils - fondées sur des approches rigoureuses de modélisation et de formalisation des connaissances.

L'IC s'associe par conséquent à de nombreuses disciplines :

D'une part, dans sa démarche d'ingénierie, l'IC mobilise les concepts et techniques de la représentation des connaissances, les méthodes d'analyse et de conception par objets, les techniques de Traitement Automatique de la Langue Naturelle et les réflexions sur la sémantique, le raisonnement à base de cas, l'ingénierie documentaire, l'ingénierie éducative, la conception de systèmes d'information, etc. ;

D'autre part, dans sa démarche de modélisation des connaissances, l'IC doit se rapprocher de disciplines permettant de caractériser et décrire les connaissances d'un domaine et d'évaluer leur mise en œuvre dans les SBC. Dans les sciences humaines et sociales, la sociologie, la psychologie et l'ergonomie cognitives, la terminologie et la linguistique de corpus, la gestion, peuvent ainsi concourir à une démarche d'explicitation de ce que sont les connaissances dans un contexte humain et organisationnel.

Thèmes

Ce rôle central et pluridisciplinaire de l'Ingénierie des connaissances se retrouve dans les thèmes privilégiés (mais non exhaustifs) dans l'édition 2003 de sa conférence annuelle :

Ontologies et modèles de connaissances
 Méthodologies et outils pour la construction, gestion, intégration, validation ou évolution d'ontologies ou de modèles de connaissances
 Langages semi-formels ou formels de représentation d'ontologies ou de modèles de connaissances
 Acquisition de connaissances à partir de textes
 Acquisition de connaissances à partir de données structurées ou semi-structurées ou de données multimedia

Intégration intelligente d'informations

...

Gestion des connaissances et coopération
 Mémoire d'entreprise, mémoire organisationnelle
 Approches cognitives, sociales, coopératives et organisationnelles de la gestion des connaissances
 Ingénierie des connaissances et théorie / modélisation des organisations
 Ingénierie des connaissances pour les systèmes de travail coopératif médiatisé

...

Web sémantique
 Langages de représentation de connaissances pour le Web sémantique
 Modélisation, annotation, gestion et évolution de connaissances sur le Web sémantique
 Recherche d'information sur Internet à l'aide de connaissances
 Architectures pour le Web sémantique
 Web sémantique et E-learning

...

Epistémologie et Cognition
 Fondements épistémologiques de l'Ingénierie des connaissances
 Approches cognitives pour l'Ingénierie des connaissances
 Apports pluridisciplinaires
 Documents, hypertextes, hypermédia et Ingénierie des connaissances documentaires
 Ingénierie des connaissances, Terminologie et Linguistique de corpus
 Ingénierie des connaissances et Ingénierie éducative
 Ingénierie des connaissances et Aide à la décision
 Ingénierie des connaissances et Conception
 Ingénierie des connaissances et Raisonnement à partir de cas
 Ingénierie des connaissances et Systèmes d'Information
 Ingénierie des connaissances et Modélisation par Objets
 Ingénierie des connaissances et Systèmes multi-agents

...

Applications
 Santé
 Biologie
 Télécommunications
 Transport
 Applications de conception et de gestion de production
 ...

Format de soumission

Les auteurs sont invités à soumettre des propositions d'article dans les 2 catégories suivantes :

Communications de recherche, présentant des travaux de recherche originaux, théoriques ou expérimentaux. Une rigueur devra être apportée à la rédaction des articles, non seulement au niveau du contenu scientifique mais aussi au niveau de la forme.

Communications appliquées, présentant des applications industrielles et opérationnelles significatives, des études de cas, des prototypes avancés. Le contexte de l'application devra être clairement décrit et les auteurs mettront l'accent sur les problèmes rencontrés, les enseignements tirés et les retours d'expériences. Ces communications appliquées devront inclure "Applications" à la liste de leurs mots-clés.

La soumission est exclusivement électronique, les articles (12 pages maximum en format A4) devront être envoyés à l'adresse ic2003@sophia.inria.fr. Les formats précis de soumission sont disponibles sur le site Web de la conférence.

Composition du comité de programme

Présidente du comité de programme :

Rose Dieng-Kuntz (INRIA-Sophia, ACACIA)

Membres :

P. Albert, ILOG, Paris

N. Aussenac-Gilles, IRIT, Toulouse
B. Bachimont, INA, Paris & UTC, Compiègne
C. Barry-Gréboval, LARIA, Amiens
J. Bézin, LRSG, Nantes
B. Biébow, LIPN, Villetaneuse
J.-F. Boujut, INP, Grenoble
D. Bourigault, ERSS, Toulouse
J. Charlet, AP-HP, Paris
O. Corby, INRIA-Sophia, Sophia Antipolis
F. Darses, CNAM, Paris
S. Després, Univ. Paris 5
J.-L. Ermine, Tech-CICO, Troyes
J. Euzenat, INRIA-Rhône-Alpes, Grenoble
J.-G. Ganascia, LIP6, Paris
N. Girard, INRA, Toulouse
G. Grosz, CGI-Europe, Paris
M.-C. Jaulent, SIM Broussais, Paris
G. Kassel, LaRIA, Amiens
J.-M. Labat, CRIP5, Paris 5
Ph. Laublet, LaLICC, Univ Paris 4, Paris
N. Matta, Tech-CICO, Troyes
A. Mille, LISI, Univ. Lyon 1
S. Moisan, INRIA-Sophia, Sophia Antipolis
A. Napoli, INRIA-Lorraine, Nancy
Y. Prié, LISI, Univ. Lyon 1
M. Revenu, GREYC, Caen
Ch. Reynaud, LRI, Orsay
P. Salembier, GRIC-IRIT, Toulouse
P. Tchounikine, LIUM, Le Mans
R. Teulier, GRID, ENS Cachan
F. Trichet, IRIN, Nantes
M. Zacklad, Tech-CICO, Troyes

Appel à communication

RJCIA 2003 - Plate-forme AFIA

*6ème Rencontres Jeunes Chercheurs en Intelligence Artificielle
Laval les 1er - 2 Juillet 2003*

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

Présidente du comité de programme : **Florence Dupin** de Saint-Cyr,
Université d'Angers

Dates importantes

- Titre et résumé : 8 Mars 2003
- Date limite de soumission : 15 mars 2003
- Notification aux auteurs : 18 avril 2003
- Réception des articles : 16 mai 2003

Présentation

La conférence RJCIA'2003 est une rencontre destinée aux chercheurs en IA qui sont en début de carrière : c'est-à-dire des doctorants ou des titulaires d'un doctorat depuis moins d'un an. L'objectif de cette manifestation

est triple :

1. Permettre aux étudiants préparant une thèse en Intelligence Artificielle ou l'ayant soutenue depuis peu, de se rencontrer et de présenter leurs travaux ;
2. Former ces mêmes étudiants à la préparation d'un article, à la révision de cet article pour tenir compte des observations émanant du Comité de programme et finalement à sa présentation devant un auditoire de spécialistes ;
3. Offrir un panorama de la recherche en IA en France.

Thèmes

Algorithmique de l'IA

- Algorithmes génétiques / évolutionnaires, métaheuristiques
- Complexité
- Contraintes
- Démonstration automatique
- Optimisation combinatoire
- Programmation logique
- Réseaux neuronaux
- SAT

Applications de l'IA

Apprentissage, Fouille et Extraction

- Classification
- Extraction de la connaissance,
- Fouilles de données

Architectures de système d'IA

Connaissances

- Ingénierie des connaissances
- Indexation
- Représentation de connaissances (graphes conceptuels, logique)

Interaction Homme-Machine

- Dialogue
- Modélisation de l'utilisateur,
- EIAH (Environnements informatiques pour l'apprentissage humain) et EIAO

Raisonnement

- Aide à la décision
- Diagnostic
- Fusion
- Planification
- Raisonnement à base de modèles
- Raisonnement temporel et spatial
- Raisonnement incertain (probabiliste, possibiliste, flou ...)

- Résolution de problèmes

Systèmes multi-agents et systèmes distribués

- Autonomie
- Collectifs
- Emergence,
- Réalité virtuelle

Traitement

- Traitement d'image et du signal
- Traitement de la parole
- Traitement du langage naturel

Web

- Ontologies - Web sémantique

Soumission

Des soumissions sont souhaitées sur des recherches substantielles, originales, et non encore publiées concernant tous les champs de l'intelligence artificielle, elle devront être envoyées par courrier électronique à la présidente du comité de programme: *Florence.Bannay@univ-angers.fr*

La soumission est un processus en deux étapes :

1. Les auteurs doivent d'abord soumettre un résumé de leur papier de recherche pour le 8 mars 2003, incluant le titre, les noms de tous les auteurs avec leurs affiliations complètes et leurs adresses email, l'auteur à contacter, le résumé du papier (au plus 200 mots), et également des mots clés dont au moins un sera extrait de la liste des thèmes de la conférence.
2. Le papier complet sera ensuite envoyé électroniquement à la même adresse pour le Samedi 15 mars 2003, au format PostScript ou PDF (ce dernier étant largement préféré).

Il est vivement recommandé de soumettre des papiers utilisant le style de formatage final prêt à l'impression (incluant les auteurs et les affiliations sur la première page). Les soumissions ne doivent pas excéder 10 pages en format définitif. Les consignes de formatage sont décrites sur le site de la conférence.

Composition du comité de programme

Présidente du comité de programme : Florence Dupin de Saint-Cyr (Université d'Angers)

La composition du CP est en cours de finalisation à la date d'édition du présent bulletin.

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/>

ELECTRONIC TRANSACTIONS ON ARTIFICIAL INTELLIGENCE (ETAI)

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

ARTIFICIAL INTELLIGENCE

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

ARTIFICIAL INTELLIGENCE V142 N°1 NOVEMBER 2002

- An information-based neural approach to generic constraint satisfaction, *H. JÖNSSON, B. SÖDERBERG*
- Process algebras for systems diagnosis, *L. CONSOLE, C. PICARDI, M. RIBAUDO*
- Complexity results for structure-based causality, *T. EITER, T. LUKASIEWICZ*

ARTIFICIAL INTELLIGENCE V142 N°2 DECEMBER 2002

- International Conference on MultiAgent Systems 2000, ed. by *E. DURFEE, S. KRAUS, H. NAKASHIMA, M. TAMBE*
- Multiagent control of self-reconfigurable robots, *H. BOJINOV, A. CASAL, T. HOGG*
- Plan coordination by revision in collective agent based systems, *H. TONINO, A. BOS, M. DE WEERDT, C. WITTEVEEN*
- The influence of social norms and social consciousness on intention reconciliation, *B. J. GROSZ, S. KRAUS, D. G. SULLIVAN, S. DAS*
- Believing others: Pros and cons, *S. SEN*
- Using similarity criteria to make issue trade-offs in automated negotiations, *P. FARATIN, C. SIERRA, N.R. JENNINGS*
- Surplus equivalence of leveled commitment contracts, *T. SANDHOLM, Y. ZHOU*
- Defection-free exchange mechanisms based on an entry fee imposition, *S. MATSUBARA, M. YOKOO*

ARTIFICIAL INTELLIGENCE V143 N°1 JANUARY 2003

- On the evaluation of agent behaviors, *A. DATTATRAYA MALI*
- Approximate inference in Boltzmann machines, *M. WEL-LING, Y. WHY TEH*
- Converting numerical classification into text classification,

S. A. MACSKASSY, H. HIRSH, A. BANERJEE, A. A. DAYANIK

- Uniform semantic treatment of default and autoepistemic logics, *M. DENECKER, V. W. MAREK, M. TRUSZCZYNSKI*
- Maximum likelihood bounded tree-width Markov networks, *N. SREBRO*
- Three perspectives of data mining, *Z.-H. ZHOU*

ARTIFICIAL INTELLIGENCE V143 N°2 FEBRUARY 2003

- Algorithms for propagating resource constraints in AI planning and scheduling: Existing approaches and new results, *P. LABORIE*
- Problem difficulty for tabu search in job-shop scheduling, *J.-P. WATSON, J. C. BECK, A. E. HOWE, L. DARRELL WHITLEY*
- Complexity results for standard benchmark domains in planning, *M. HELMERT*

AI MAGAZINE

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

AI MAGAZINE V23 N°3 2002

- Specifying Rules for Electronic Auctions, *P. R. WURMAN, M. P. WELLMAN, W. E. WALSH*
- vAI and Agents: State of the Art, *E. ALONSO*
- Support Vector Machines and Kernel Methods: The New Generation of Learning Machines, *N. CRISTIANINI, B. SCHÖLKOPF*
- AI and Music: From Composition to Expressive Performance, *R. LOPEZ DE MANTARAS, J. LLUIS ARCOS*
- An AI-Based Approach to Destination Control in Elevators, *J. KOEHLER, D. OTTIGER*
- The Timing of Bids in Internet Auctions: Market Design, Bidder Behavior, and Artificial Agents, *A. OCKENFELS, A. E. ROTH*
- Leveled-Commitment Contracting: A Backtracking Instrument for Multiagent Systems, *T. SANDHOLM, V. LES-SER*
- Strategic Design of Mobile Agents, *N. VULKAN*
- A Review of the Twenty-Second SOAR Workshop, *F. E. RITTER, I. G. COUNCILL*

AI MAGAZINE V23 N°4 2002

- Editorial Introduction: The Fifteenth Innovative Applications of Artificial Intelligence Conference (IAAI-2002), *S. CHIEN, J. RIEDL*
- MiTAP for Biosecurity: A Case Study, *L. DAMIANOS, J. PONTE, S. WOHLEVER, F. REEDER, D. DAY, G. WILSON, L. HIRSCHMAN*
- Staff Scheduling for Inbound Call and Customer Contact Centers, *A. FUKUNAGA, E. HAMILTON, J. FAMA, D. ANDRE, O. MATAN, I. NOURBAKHSH*
- Information Self-Service with a Knowledge Base That Learns, *S. D. DURBIN, D. WARNER, J. N. RICHTER, Z. GEDEON*

- Training and Using Disciple Agents: A Case Study in the Military Center of Gravity Analysis Domain, *G. TECUCI, M. BOICU, D. MARCU, B. STANESCU, C. BOICU, J. COMELLO*
- Applying Perceptually Driven Cognitive Mapping to Virtual Urban Environments, *R. W. HILL, JR., C. HAN, M. VAN LENT*
- Computational Vulnerability Analysis for Information Survivability, *H. SHROBE*
- In Memory of Ray Reiter (1939-2002), *F. PIRRI, G. HINTON, H. LEVESQUE*
- The AAI-02 and IAAI-02 Conferences, *S. HEDBERG*
- FLAIRS 2002 Conference Report, *R. SOORIAMURTHI AND T. REICHHERZER*
- The 2002 AAI Spring Symposium Series
- Report on the First International Conference on Knowledge Capture (K-CAP), *Y. GIL, M. MUSEN, J. SHAVLIK*
- Intelligent Integration of Information and Services on the Web, *E. STROULIA*
- AAI 2002 Workshops
- Consciousness Constrained, *P. DE PALMA*

COGNITIVE SCIENCE

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

COGNITIVE SCIENCE V26 N°1 JANUARY-FEBRUARY 2002

- Perceiving temporal regularity in music, *E. W. LARGE, C. PALMER*
- Memory for goals: An activation-based model, *E. M. ALTMANN, J. G. TRAFTON*
- Spanning seven orders of magnitude: A challenge for cognitive modeling, *J. A. ANDERSON*
- Learning words from sights and sounds: A computational model, *D. K. ROY, A. P. PENTLAND*

COGNITIVE SCIENCE V26 N°2 MARCH-APRIL 2002

- An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tour, *V. A.W.M.M. ALEVEN, K. R. KOEDINGER*
- Evidential diversity and premise probability in young children's inductive judgment, *Y. LO, A. SIDES, J. ROZELLE, D. OSHERSON*
- Eye scanpaths during visual imagery reenact those of perception of the same visual scene, *B. LAENG, D.-S. TEODORESCU*

COGNITIVE SCIENCE V26 N°3 MAY-JUNE 2002

Special section on linguistically apt statistical models

- Statistical models of syntax learning and use, *M. JOHNSON, S. RIEZLER*
- Discovering syntactic deep structure via Bayesian statistics, *J. EISNER*
- Statistical models for the induction and use of selectional preferences, *M. LIGHT, W. GREIFF*

- Modeling asynchrony in automatic speech recognition using loosely coupled hidden Markov models, *H.J. NOCK, S.J. YOUNG*

Other articles in this issue

- A simplicity principle in unsupervised human categorization, *E. M. POTHOS, N. CHATER*
- A non-representational approach to imagined action, *I. VAN ROOIJ, R. M. BONGERS, W. F.G. HASELAGER*
- Asymmetric interference in 3- to 4-month olds sequential category learning, *D. MARESCHAL, P. C. QUINN, R. M. FRENCH*

COGNITIVE SCIENCE V26 N°4 JULY-AUGUST 2002

- The distributional structure of grammatical categories in speech to young children, *T. H. MINTZ, E. L. NEWPORT, T. G. BEVER*
- Strategies in sentential reasoning, *J.-B. VAN DER HENST, Y. YANG, P.N. JOHNSON-LAIRD*
- Cognitively active externalization for situated reflection, *H. SHIROUZU, N. MIYAKE, H. MASUKAWA*
- Learning representations in a gated prefrontal cortex model of dynamic task switching, *N. P. ROUGIER, R. C. O'REILLY*

COGNITIVE SCIENCE V26 N°5 SEPTEMBER-OCTOBER 2002

- The misunderstood limits of folk science: An illusion of explanatory depth, *L. ROZENBLIT, F. KEIL*
- PROBABILITIES from EXemplars (PROBEX): A 'lazy' algorithm for probabilistic inference from generic knowledge, *P. JUSLIN, M. PERSSON*
- Symbolically speaking: A connectionist model of sentence production, *F. CHANG*
- Cultural preferences for formal versus intuitive reasoning, *A. NORENZAYAN, E. E. SMITH, B. JUN KIM, R. E. NISBETT*

COGNITIVE SCIENCE V26 N°6 NOVEMBER 2002

- Electrifying diagrams for learning: principles for complex representational systems, *P.C.-H. CHENG*
- Overtensing and the effect of regularity, *J.P. STEMBERGER*
- Circular reasoning, *L.J. RIPS*
- Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis, *N. FURL, P.J. PHILLIPS, A.J. O'TOOLE*
- Gesture offers insight into problem-solving in adults and children, *P. GARBER, S. GOLDIN-MEADOW*

COGNITIVE SCIENCE V27 N°1 JANUARY 2003

- Learning to divide the labor: an account of deficits in light and heavy verb production, *J.K. GORDON, G.S. DELL*
- Semantic grounding in models of analogy: an environmental approach, *M. RAMSCAR, D. YARLETT*
- Towards structural systematicity in distributed, statically bound visual representations, *S. EDELMAN, N. INTRATOR*
- One-year-old infants use teleological representations of actions productively, *G. CSIBRA, S. BRO, O. KOOS, G. GERGELY*
- The evolution of cognition-a hypothesis, *H. CRUSE*

APPLIED ARTIFICIAL INTELLIGENCE

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

APPLIED ARTIFICIAL INTELLIGENCE V16 N°9-10 OCTOBER-DECEMBER 2002

- Editorial Engineering Agent Systems: Best of « From Agent Theory to Agent Implementation (AT2AI)-3 », *P. PETTA, J. P. MÜLLER*
- What Agent Middleware Can (And Should) Do For You, *A. POGGI, G. RIMASSA, P. TURCI*
- Multi-Agent Corporate Memory Management System, *F. GANDON, A. POGGI, G. RIMASSA, ET AL.*
- Coordination Tools for MAS Development and Deployment, *E. DENTI, A. OMICINI, A. RICCI*
- The Design of Communicative Act Libraries: A Linguistic Perspective, *J. M. SERRANO, S. OSSOWSKI*
- An Observation Approach to the Semantics of Agent Communication Languages, *M. VIROLI, A. OMICINI*
- The Role Of Reflection In Simulating And Testing Agents: An Exploration Based On The Simulation System James, *A. M. UHRMACHER, M. RÖHL, B. KULLICK*
- Interfacing Indigolog and OAA: A Toolkit for Advanced Multiagent Applications, *A. LAPOUCHNIAN, Y. LESPÉRANCE*
- Heterogeneous Agents Situated in Heterogeneous Spaces, *S. BANDINI, S. MANZONI, C. SIMONE*
- Multi-Dimensional Matchmaking for Electronic Markets, *D. J. VEIT, C. WEINHARDT, J. P. MÜLLER*

APPLIED ARTIFICIAL INTELLIGENCE V17 N°1 JANUARY 2003

- The Virtual University's Faculty: An Overview Of Educational Agents, *S. PAYR*
- An Adaptive Social Network For Information Access: Theoretical And Experimental Results, *B. YU, M. VENKA-TRAMAN, M. P. SINGH*
- Emergent Computation Using A New Model Of Cellular Automata, *Y. HASSAN, E. TAZAKI*
- What Does Ai Contribute To Hydrology? Aerial Photos And Flood Levels, *D. RACLOT, C. PUECH*

COMPUTATIONAL INTELLIGENCE

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

COMPUTATIONAL INTELLIGENCE V18 N°4 NOVEMBER 2002

- Introduction to the Special Issue on Agent Technologies for Electronic Commerce, *B. SPENCER*
- Automated Negotiation from Declarative Contract Descriptions, *D. M. REEVES, M. P. WELLMAN, B. N. GRO-SOF*
- Contracting With Uncertain Level Of Trust, *S. BRAYNOV,*

T. SANDHOLM

- Reputation Formalization for an Information-Sharing Multi-Agent System, *J. CARTER, E. BITTING, A. A. GHORBANI*
- Distributed Reputation Management for Electronic Commerce, *B. YU, M. P. SINGH*
- A Reputation-Oriented Reinforcement Learning Strategy for Agents in Electronic Marketplaces, *T. TRAN, R. COHEN*
- Model Selection in an Information Economy: Choosing What to Learn, *C. H. BROOKS, R. S. GAZZALE, R. DAS, J. O. KEPHART, J. K. MACKIE-MASON, E. H. DURFEE*
- vAgent Reasoning Mechanism for Long-Term Coalitions Based on Decision Making and Trust, *J. VASSILEVA, S. BREBAN, M. HORSCH*
- Design of a Multi-Unit Double Auction E-Market, *P. HUANG, A. SCHELLER-WOLF, K. SYCARA*
- A Market-Driven Model for Designing Negotiation Agents, *K. MONG SIM*
- Architectural Components of Information-Sharing Societies, *J. CARTER, A. A. GHORBANI, S. MARSH*
- eMediator: A Next Generation Electronic Commerce Server, *T. SANDHOLM*

MINDS AND MACHINES JOURNAL FOR ARTIFICIAL INTELLIGENCE PHILOSOPHY AND COGNITIVE SCIENCE

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

MINDS AND MACHINES V12 N°4 NOVEMBER 2002

- Hypercomputation, *B. JACK COPELAND*
- Incompleteness, Complexity, Randomness and Beyond, *C. S. CALUDE*
- What is Church's Thesis? An Outline*, *J. DOYLE*
- Infinite Time Turing Machines, *J. D. HAMKINS*
- Quantum Hypercomputation, *T. D. KIEU*
- Computing Machines Can't Be Intelligent (...and Turing Said So), *P. KUGEL*

MINDS AND MACHINES V13 N°1 FEBRUARY 2003

- Transcending Turing Computability, *B.J. MACLENNAN*
- Alan Turing and the Mathematical Objection, *G. PICCINI-NI*
- Two Undecidable Problems of Analysis, *B. SCARPELLINI*
- Comments on "Two Undecidable Problems of Analysis", *B. SCARPELLINI*
- Physical Hypercomputation and the Church-Turing Thesis, *O. SHAGRIR, I. PITOWSKY*
- Neural and Super-Turing Computing, *H. T. SIEGELMANN*
- Computation and Hypercomputation, *M. STANNETT*
- Supermachines and Superminds, *E. STEINHART*

ARTIFICIAL INTELLIGENCE REVIEW

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

AI REVIEW V18 N°3-4 DECEMBER 2002

- Introduction, *P. CUNNINGHAM, R. BURKE*

- Acquiring Customers' Requirements in Electronic Commerce, *R. BERGMANN, P. CUNNINGHAM*
- simVar: A Similarity-Influenced Question Selection Criterion for e-Sales Dialogs, *S. SCHMITT*
- ExpertClerk: A Conversational Case-Based Reasoning Tool for Developing Salesclerk Agents in E-Commerce Webshops, *H. SHIMAZU*
- Interactive Critiquing for Catalog Navigation in E-Commerce, *R. BURKE*
- An Expressive Query Language for Product Recommender Systems, *D. BRIDGE, A. FERGUSON*
- A Generalised Approach to Similarity-Based Retrieval in Recommender Systems, *D. MCSHERRY*

AI REVIEW V19 N°1 MARCH 2003

- Agent-Based Simulation in the Study of Social Dilemmas, *N.M. GOTTS, J.G. POLHILL, A.N.R. LAW*
- Maximum Consistency of Incomplete Data via Non-Invasive Imputation, *G. GEDIGA, I. DÜNTSCH*

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

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

COMPUTATIONAL LINGUISTICS

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

NATURAL LANGUAGE ENGINEERING

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

NLE V8 N°4 DECEMBER 2002

- Dedication to William A. Gale, *K. CHURCH*
- Introduction to the special issue on evaluating word sense disambiguation systems, *P. EDMONDS, A. KILGARRIFF*
- Evaluating sense disambiguation across diverse parameter spaces, *D. YAROWSKY, R. FLORIAN*
- Parameter optimization for machine-learning of word sense disambiguation, *V. HOSTE, I. HENDRICKX, W. DAELEMANS, A. VAN DEN BOSCH*
- Combining Classifiers for word sense disambiguation, *R. FLORIAN, S. CUCERZAN, C. SCHAFER, D. YAROWSKY*
- Word sense disambiguation with pattern learning and automatic feature selection, *R. F. MIHALCEA*
- The role of domain information in Word Sense Disambiguation, *B. MAGNINI, C. STRAPPARAVA, G. PEZZULO, A. GLIOZZO*
- Evaluating lexical resources using SENSEVAL, *N. CALZOLARI, C. SORIA, F. BERTAGNA, F. BARSOTTI*

USER MODELING AND USER-ADAPTED INTERACTION

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

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

- Using Bayesian Networks to Manage Uncertainty in Student Modeling, *C. CONATI, A. GERTNER, K. VANLEHN*
- User Interfaces for All: Concepts, Methods and Tools, *B. A. NARDI*

COMPUTER SPEECH AND LANGUAGE

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

COMPUTER SPEECH & LANGUAGE V17 N°1 JANUARY 2003

- Classifier-based non-linear projection for adaptive endpointing of continuous speech, *B. RAJ, R. SINGH*
- Statistical language modeling based on variable-length sequences, *I. ZITOUNI, K. SMAÏLI, J.-P. HATON*
- Topics in decision tree based speech synthesis, *R. E. DONOVAN*
- Pronunciation modeling for ASR – knowledge-based and data-derived methods, *M. WESTER*
- Language modelling for Russian and English using words and classes, *E. W. D. WHITTAKER, P. C. WOODLAND*

MACHINE LEARNING

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

MACHINE LEARNING V50 N°3 MARCH 2003

- Combining Classifiers with Meta Decision Trees, *L. TODOROVSKI, S. D. EROSKI*
- Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results, *P. B. BRAZDIL, C. SOARES, J. PINTO DA COSTA*
- vLearning to Match the Schemas of Data Sources: A Multistrategy Approach, *A. DOAN, P. DOMINGOS, A. HALEVY*
- vClustered Partial Linear Regression, *L. TORGO, J. PINTO DA COSTA*
- Complete Mining of Frequent Patterns from Graphs: Mining Graph Data, *A. INOKUCHI, T. WASHIO, H. MOTODA*

MACHINE LEARNING V51 N°1 APRIL 2003

- PAC-Bayesian Stochastic Model Selection, *D. A. MCALLESTER*

- Relative Loss Bounds for Temporal-Difference Learning, *J. FORSTER, M. K. WARMUTH*
- Polynomial-Time Decomposition Algorithms for Support Vector Machines, *D. HUSH, C. SCOVEL*
- An Empirical Study of Two Approaches to Sequence Learning for Anomaly Detection, *T. LANE, C. E. BRODLEY*

NEURAL NETWORKS

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

NEURAL NETWORKS V15 N°8-9 OCTOBER-NOVEMBER 2002

- Introduction: new developments in self-organising maps, *N. ALLINSON, H. YIN, K. OBERMAYER*
- How to make large self-organizing maps for nonvectorial data, *T. KOHONEN, P. SOMERVUO*
- Analysis and visualization of gene expression data using Self-Organizing Maps, *J. NIKKILÄ ET AL.*
- vStatistical tools to assess the reliability of self-organizing maps, *E. de BODT, M. COTTRELL, M. VERLEYSEN*
- Recursive self-organizing maps, *T. VOEGTLIN*
- Self-organizing maps with recursive neighborhood adaptation, *J. A. LEE, M. VERLEYSEN*
- Data visualisation and manifold mapping using the ViSOM, *H. YIN*
- g-Observable neighbours for vector quantization, *M. AUPETIT, P. COUTURIER, P. MASSOTTE*
- Kernel-based topographic map formation achieved with an information-theoretic approach, *M. M. VAN HULLE*
- A self-organising network that grows when required, *S. MARSLAND, J. SHAPIRO, U. NEHMZOW*
- Generalized relevance learning vector quantization, *B. HAMMER, T. VILLMANN*
- Principles and networks for self-organization in space-time, *J. PRINCIPE, N. EULIANO, S. GARANI*
- Image denoising using self-organizing map-based nonlinear independent component analysis, *M. HARITPOULOS, H. YIN, N. M. ALLINSON*
- Integrating contextual information to enhance SOM-based text document clustering, *D. PULLWITT*
- Applications of the self-organising map to reinforcement learning, *A. J. SMITH*
- BYY harmony learning, structural RPCL, and topological self-organizing on mixture models, *L. XU*

NEURAL NETWORKS V15 N°10 DECEMBER 2002

Neural Networks Letter

- The V1-V2-V3 complex: quasiconformal dipole maps in primate striate and extra-striate cortex, *M. BALASUBRAMANIAN, J. POLIMENI, E. L. SCHWARTZ*
- New classification scheme of cortical sites with the neuronal spiking characteristics, *S. SHINOMOTO, K. SHIMA, J. TANJI*
- Stochastic resonance in the hippocampal CA3-CA1 model: a possible memory recall mechanism, *M. YOSHIDA, H. HAYASHI, K. TATENO, S. ISHIZUKA*

- Straight monotonic embedding of data sets in Euclidean spaces, *P. COURRIEU*
- The connections between the frustrated chaos and the intermittency chaos in small Hopfield networks, *H. BERSINI, P. SENER*
- Category regions as new geometrical concepts in Fuzzy-ART and Fuzzy-ARTMAP, *G. C. ANAGNOSTOPOULOS, M. GEORGIOPOULOS*
- Bayesian model search for mixture models based on optimizing variational bounds, *N. UEDA, Z. GHAHRAMANI*
- Threshold disorder as a source of diverse and complex behavior in random nets, *P. C. MCGUIRE ET AL.*
- Multi-objective cooperative coevolution of artificial neural networks (multi-objective cooperative networks), *N. GARCÍA-PEDRAJAS, C. HERVÁS-MARTÍNEZ, J. MUÑOZ-PÉREZ*
- Extracting regression rules from neural networks, *K. SAITO, R. NAKANO*
- Neural network based optimal routing algorithm for communication networks, *P. VENKATARAM, S. GHOSAL, B.P. VIJAY KUMAR*

Letter to the editor

- A comment on "Global stability analysis in delayed Hopfield neural network models", *C. SUN*
- A comment on "Global stability analysis in delayed Hopfield neural network models", *J. ZHANG*

NEURAL NETWORKS V16 N°1 JANUARY 2003

- Celebrating the Year with a Special Issue for IJCNN'03, *S. GROSSBERG, M. KAWATO, J.G. TAYLOR*
- Meta-learning in Reinforcement Learning, *N. SCHWEIGHOFER, K. DOYA*
- Learning to generate articulated behavior through the bottom-up and the top-down interaction processes, *J. TANI*
- Computational model for neural representation of multiple disparities, *O. WATANABE, M. IDESAWA*
- A synfire chain in layered coincidence detectors with random synaptic delays, *K. IKEDA*
- Mathematical analysis of a correlation-based model for orientation map formation, *T. YAMAZAKI*
- The functional localization of neural networks using genetic algorithms, *H. TSUKIMOTO, H. HATANO*
- A new EM-based training algorithm for RBF networks, *M. LAZARO, I. SANTAMARA, C. PANTALEON*
- Analysis of Tikhonov regularization for function approximation by neural networks, *M. BURGER, A. NEUBAUER*
- Predicting the behaviour of G-RAM networks, *G.G. LOCKWOOD, I. ALEKSANDER*
- On-line identification and reconstruction of finite automata with generalized recurrent neural networks [review article], *I. GABRIJEL, A. DOBNIKAR*
- Classification of clustered microcalcifications using a Shape Cognitron neural network, *S.-K. LEE, P.-C. CHUNG, C.-I. CHANG, C.-S. LO, T. LEE, G.-C. HSU, C.-W. YANG*
- On learning to estimate the block directional image of a fingerprint using a hierarchical neural network, *K.A. NAGATY*
- Self-organizing Neural Networks: Recent Advances and Applications - Studies in Fuzziness and Soft Computing; U.

Seiffert, L.C. Jain, (Eds.); Physica-Verlag, 2002, by *F. AZUAJE*

ARTIFICIAL INTELLIGENCE IN MEDICINE

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

AI IN MEDICINE V26 N°3 NOVEMBER 2002

- Survey on the use of smart and adaptive engineering systems in medicine, *M.F. ABBOD, D.A. LINKENS, M. MAHFOUF, G. DOUNIAS*
- Model-based interpretation of cardiac beats by evolutionary algorithms: signal and model interaction, *A. I. HERNÁNDEZ, G. CARRAULT, F. MORA, A. BARDOU*
- Case-based distance measurements for the selection of controls in case-matched studies: application in coronary interventions, *M. GYÖNGYÖSI ET AL.*
- Visualization and simulation techniques for surgical simulators using actual patient's data, *A. RADETZKY, A. NÜRNBERGER*
- Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles, *G. VALENTINI*

Book review

- Qualitative Methods for Reasoning under Uncertainty - Simon Parsons (Ed.), MIT Press, by *S. RENOUIJ*

AI IN MEDICINE V27 N°1 JANUARY 2003

- Evaluation of inherent performance of intelligent medical decision support systems: utilising neural networks as an example, *A.E. SMITH, C.D. NUGENT, S.I. MCCLEAN*
- A multiple classifier system for early melanoma diagnosis, *A. SBONER ET AL.*
- A combined neural network and decision trees model for prognosis of breast cancer relapse, *J. M. JEREZ-ARAGONÉS, J. A. GÓMEZ-RUIZ, G. RAMOS-JIMÉNEZ, J. MUÑOZ-PÉREZ, E. ALBA-CONEJO*
- Bagging tree classifiers for laser scanning images: a data and simulation-based strategy, *T. HOTHORN, B. LAUSEN*
- Midpoints for fuzzy sets and their application in medicine, *J. J. NIETO, A. TORRES*

Book review

- Safe and Sound: Artificial Intelligence in Hazardous Applications - John Fox, Subrata Das, AAAI Press, Menlo Park, CA, and MIT Press, by *M. MARCOS*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

<http://www.elsevier.com/inca/publications/store/5/0/5/7/8/7/>

IJAR V31 N°3 NOVEMBER 2002

- Synergies between evolutionary computation and probabilistic graphical models, *P. LARRAÑAGA, J. A. LOZANO*
- Evolutionary optimization and the estimation of search distributions with applications to graph bipartitioning, *H. MÜHLENBEIN, T. MAHNIG*
- Using a priori knowledge to create probabilistic models for optimization, *S. BALUJA*
- Scalability of the Bayesian optimization algorithm, *M. PELIKAN, K. SASTRY, D.E. GOLDBERG*
- Multi-objective optimization with diversity preserving mixture-based iterated density estimation evolutionary algorithms, *P. A.N. BOSMAN, D. THIERENS*
- Ant colony optimization for learning Bayesian networks, *L. M. DE CAMPOS, J. M. FERNÁNDEZ-LUNA, J. A. GÁMEZ, J. M. PUERTA*
- Mathematical modelling of UMDAc algorithm with tournament selection. Behaviour on linear and quadratic functions, *C. GONZÁLEZ, J.A. LOZANO, P. LARRAÑAGA*

IJAR V32 N°1 JANUARY 2003

- Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *J. ABONYI, J. A. ROUBOS, F. SZEIFERT*
- Epistemic independence in numerical possibility theory, *E. MIRANDA, G. DE COOMAN*
- Multi-interpretation operators and approximate classification, *J. ENGELFRIET, J. TREUR*

IJAR V32 N°2-3 FEBRUARY 2003

- Soft Computing in Information Mining, edited by *R. KRUSE, C. BORGELT*
- Information mining, *R. KRUSE, C. BORGELT*
- Mixed fuzzy rule formation, *M. R. BERTHOLD*
- Improved fuzzy partitions for fuzzy regression models, *F. HÖPPNER, F. KLAWONN*
- Fuzzy data analysis with NEFCLASS, *D. D. NAUCK*
- FS-FOIL: an inductive learning method for extracting interpretable fuzzy descriptions, *M. DROBICS, U. BODENHOFER, E. P. KLEMENT*
- A hierarchical recurrent neuro-fuzzy model for system identification, *A. NÜRNBERGER*
- Approximate symbolic pattern matching for protein sequence data, *B. C.H. CHANG, S. K. HALGAMUGE*
- Linguistic modeling with hierarchical systems of weighted linguistic rules, *R. ALCALÁ ET AL.*
- Web mining with relational clustering, *T.A. RUNKLER, J.C. BEZDEK*
- Symbolic state transducers and recurrent neural preference machines for text mining, *G. AREVIAN, S. WERMTER, C. PANCHEV*
- Enhancement of topology preservation and hierarchical

L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales

Jean Charlet

Habilitation à diriger des recherches de l'Université de Paris 6, Spécialité Informatique, soutenue le mardi 10 Décembre 2002

Résumé

Le travail présenté dans ce mémoire d'habilitation à diriger des recherches cherche à faire le point sur le domaine de l'Ingénierie des connaissances et à étudier ses apports, présents et futurs, à la gestion des connaissances médicales. L'ingénierie des connaissances ne peut pas résoudre à elle seule les problèmes de la gestion des connaissances médicales. Un certain nombre d'années passées à travailler dans ces domaines fondamentaux - pour l'Ingénierie des connaissances - et appliqués - pour la gestion des connaissances médicales - nous ont convaincu qu'il était nécessaire de mobiliser d'autres domaines dans une approche pluridisciplinaire. C'est que nous avons fait en entreprenant des recherches qui ressortissent à l'Ingénierie des connaissances mais aussi au domaine des systèmes d'information et à la gestion. C'est cette démarche que nous développons dans ce mémoire.

Pour ce faire, nous proposons d'abord une introduction à l'acquisition des connaissances et à la façon dont le domaine rencontre les applications médicales. Nous y abordons les concepts et les modèles de ce domaine et y rappelons rapidement les résultats que les travaux ayant concouru à ces recherches nous ont permis d'obtenir. Nous poursuivons par une analyse épistémologique de l'Ingénierie des connaissances pour en argumenter les apports et en définir un cadre de travail.

Nous abordons ensuite plus précisément des applications médicales et constatons que le caractère contextuel de l'information et ses rapports avec son support (ici informatique) nous amène à mettre l'accent sur l'intérêt du document textuel comme support et véhicule de la connaissance. Le projet de dossier médical informatisé HOSPITEXTE développé au sein du service est présenté dans ce paradigme et la problématique de l'indexation des documents médicaux est spécifiquement étudiée.

La notion d'ontologie est abordée de façon approfondie en justifiant et prenant une position affirmée sur ce que nous pensons que c'est et que ça n'est pas. Cette position nous semble nécessaire à la compréhension de ce que l'on peut faire

ou non avec les ontologies comme avec de nombreux « produits » terminologiques tels que les thésaurus ou les bases de données lexicales. Nous défendons, à l'occasion de la description de travaux développés en collaboration, l'utilisation d'outils du traitement du langage naturel comme mode d'accès privilégié aux concepts d'une ontologie. Des ontologies en médecine, construites durant ces travaux, servent d'exemples à nos propos.

Les systèmes d'information en médecine, sont abordés au regard du support de la mise en oeuvre de la gestion des connaissances médicales et vis-à-vis des questions d'interopérabilité. Ces systèmes dont le champ d'action grandit, de l'établissement de soins à la médecine de ville, impliquent l'ensemble des professionnels de santé pour une pathologie ou une situation donnée. Cette situation crée de nouvelles contraintes pour la mise en oeuvre de ces larges systèmes d'informations des réseaux, en particulier au niveau de l'interopérabilité qui doit obligatoirement passer par des standards qui se construisent et s'imposent lentement.

La dernière partie du mémoire s'appuie sur les précédentes pour, d'abord rechercher un certain nombre de points communs entre la gestion et l'Ingénierie des connaissances et, ensuite, proposer un certain nombre d'indications sur l'élaboration d'un système de gestion des connaissances en santé. À partir de là, nous repositionnons les enjeux de la gestion des connaissances médicales ainsi que leur spécificité au regard des problématiques développées dans les chapitres précédents. Les projets comme les développements normatifs sont resitués dans un contexte d'intégration. Enfin, nous essayons de tirer les conséquences de ces travaux dans le contexte des développements futurs attendus, en particulier au regard des nouvelles structures de soins qui se font jour.

Mots clés : Ingénierie des connaissances, épistémologie, théorie du support, ontologies, informatique médicale, gestion, thésaurus, dossier médical, interopérabilité, gestion des connaissances, Web sémantique.

Jean Charlet
STIM - DPA/DSI/AP-HP
91, bd de l'Hopital
75634 Paris Cedex 13
Mél : charlet@biomath.jussieu.fr

Dialogue entre agents naturels et artificiels ; une application aux communautés virtuelles

Guillaume Chicoisne

Thèse de l'INP Grenoble, soutenue le mercredi 11 Décembre 2002

Résumé

Cette thèse s'inscrit dans le cadre de la constitution de communautés mixtes rassemblant des agents naturels (humains) et des agents artificiels. Dans cet objectif, la thèse présentée traite des interactions pouvant exister entre ces types d'agents lorsqu'ils sont mis en présence au sein d'une communauté virtuelle.

Les modèles d'agent et d'interaction proposés reposent sur l'hypothèse que chaque message est porteur d'un potentiel de sens, négociable, qu'il est nécessaire d'interpréter et non pas porteur d'un sens, unique et précis, qu'il suffit de décoder. Ceci implique les deux points centraux suivants : l'enrichissement contextuel du message et le focus conversationnel. Le premier point, l'enrichissement contextuel, associe un message avec son contexte de façon à permettre l'interprétation de ce message plutôt qu'un simple décodage comme cela est le cas dans les applications informatiques classiques. De manière duale, un processus d'expression remplace l'encodage. Le deuxième point, le focus conversationnel, représente l'objet de la discussion, lui aussi négociable et construit de manière collective par l'ensemble des interactants, les différentes interventions entraînant l'enchaînement des tours de parole.

Un agent conversationnel dédié à la recommandation de films, dans le cadre d'une boutique virtuelle 3D, constitue le cadre technologique et applicatif de la thèse défendue qui se conclut sur l'analyse de plusieurs dialogues entre cet agent et un client humain.

Guillaume Chicoisne
Equipe MAGMA Laboratoire Leibniz
46, Avenue Félix Viallet
38031 GRENOBLE CEDEX

Représentation et apprentissage de concepts et d'ontologies pour le Web Sémantique

Alexandre Delteil

Thèse de l'INRIA Sophia-Antipolis, soutenue le mardi 26 Novembre 2002

Résumé

Le Web Sémantique se donne pour objectif de rendre compréhensible le contenu du Web à des programmes informatiques. La réalisation d'un tel but repose sur l'utilisation d'ontologies, permettant de définir un certain vocabulaire conceptuel utilisé pour exprimer des meta-données, et sur l'exploitation de langages de représentation des connaissances pour exprimer et raisonner sur ces ontologies et sur ces meta-données.

Le langage RDF(S), recommandation actuelle du W3C pour servir de fondement au Web Sémantique, n'ayant que peu d'expressivité, il est particulièrement important de proposer des langages possédant de plus grandes capacités d'expression, tout en restant simples et intuitifs d'utilisation. La notation graphique telle qu'utilisée dans le formalisme des Graphes Conceptuels remplissant à notre avis ces conditions, nous proposons trois langages de représentation de concepts et d'ontologies pour le Web Sémantique, qui possèdent tous cette nature graphique. DRDF(S), qui est notre proposition, étend RDF(S) par un mécanisme de définition de concepts et de propriétés, fondé sur le formalisme des Graphes Conceptuels Simples, i.e. par des formules positives, existentielles et conjonctives.

GDL est un langage intégrant Graphes Conceptuels et Logiques de Description, et possédant de fortes capacités d'expression telles que la négation et la disjonction. Il peut être vu comme la clôture booléenne des Graphes Conceptuels Simples ou comme la généralisation des Logiques de description permettant de représenter des motifs graphiques dans les descriptions de concepts, à l'aide de deux nouveaux constructeurs : les graphes existentiels et les règles de graphe. Enfin GDL+ est une généralisation de GDL permettant aussi la définition de rôles et l'utilisation de ces rôles définis dans les descriptions de concepts, et notamment dans les conclusions des règles de graphes. Tous ces langages sont dotés de procédures de décision correctes et complètes. La question de l'organisation des meta-données et de l'apprentissage de concepts est aussi importante pour gérer une grande quantité de données.

Nous proposons un algorithme permettant de regrouper systématiquement les objets décrits par un même motif. Cette

approche peut être vue comme une généralisation de l'Analyse Formelle de Concepts où les attributs sont des motifs relationnels. L'algorithme est incrémental dans le sens où la taille des motifs augmente graduellement d'une unité à chaque pas d'incrémentalité. Pour adapter l'algorithme à des situations réelles, des heuristiques sur la forme, l'expressivité ou la fréquence des motifs sont introduites.

Mots clés : représentation des connaissances, Web sémantique, logiques de description, graphes conceptuels, ontologies, apprentissage

Alexandre Delteil
INRIA – Equipe ACACIA
2004, route des Lucioles, BP 93
06902 Sophia Antipolis
Mél : Alexandre.Delteil@sophia.inria.fr
<http://www-sop.inria.fr/acacia/personnel/Alexandre.Delteil/>

Contribution à la conception de méthodes et d'outils pour la gestion de connaissances

Sylvie DESPRES

Habilitation à diriger des recherches de l'Université de Paris 5, soutenue le mardi 17 décembre

Résumé

Les travaux présentés constituent une contribution à la conception de méthodes et d'outils pour la gestion de connaissances. Il s'agit d'une recherche appliquée dans les domaines de l'accidentologie et du droit médical. Cette recherche est conduite dans un contexte pluridisciplinaire, en collaboration avec des psychologues cognitivistes, des ergonomes et plus récemment des linguistes. Dans un contexte de multi-expertise, après avoir identifié la trace de représentations mentales, j'ai proposé une méthode de construction collective de prototypes CoCA (Construire Collectivement pour Acquérir) les représentant et l'architecture d'un système coopératif d'aide à la construction ou la consultation de prototypes. Ensuite, la réutilisation de ressources expertes a été au centre de mes travaux et a permis, à partir du modèle de l'activité de différents acteurs, de définir les spécifications d'un module à insérer dans un serveur de connaissances et la réalisation d'un système de raisonnement à partir de cas d'aide à l'analyse d'accidents pour l'assurance. Ces différents résultats sont fondés sur une modélisation qui utilise une ontologie de l'accidentologie que j'ai développée à partir de corpus textuels.

Sylvie Despres,
UFR de mathématiques et informatique
Université René Descartes - Paris 5
45 rue des Saints Pères, F-75270
Paris Cedex 06, France
Mél : sd@math-info.univ-paris5.fr

Intelligence artificielle distribuée et gestion des connaissances: ontologies et systèmes multi-agents pour un web sémantique organisationnel.

Fabien GANDON

Thèse de l'Université de Nice Sophia Antipolis, soutenue le jeudi 7 Novembre 2002

Résumé

Ce travail concerne les systèmes multi-agents pour la gestion d'un web sémantique d'entreprise basé sur une ontologie. Il a été effectué dans le cadre du projet Européen CoMMA se focalisant sur deux scénarios d'application: l'assistance aux activités de veille technologique et l'aide à l'insertion d'un nouvel employé dans une organisation. Trois aspects ont essentiellement été développés dans ce travail :

- la conception d'une architecture multi-agents assistant les deux scénarios, et l'approche organisationnelle descendante adoptée pour identifier les sociétés, les rôles et les interactions des agents.
- la construction de l'ontologie O'CoMMA et la structuration de la mémoire organisationnelle en exploitant les technologies du Web sémantique
- la conception et l'implantation (a) des sous-sociétés d'agents chargées de la maintenance des annotations et de l'ontologie et (b) des protocoles supportant ces deux groupes d'agents, en particulier des techniques pour la distribution des annotations et des requêtes entre les agents.

Mots clés : gestion des connaissances – systèmes multi-agents – ontologies – mémoires d'entreprises – web sémantique

Fabien GANDON
INRIA - Equipe ACACIA
2004, rt des Lucioles, BP 93
06902 Sophia Antipolis

Mél : Fabien.Gandon@sophia.inria.fr
http://www-sop.inria.fr/acacia/personnel/Fabien.Gandon/
Tél: 04 92 38 77 88
Fax: 04 92 38 77 83

Adaptation du niveau d'analyse des interventions dans un dialogue – Application à un système de question – réponse

Laura Monceaux

Thèse de l'Université de Paris 11, soutenue le vendredi 13
décembre 2002

Résumé

Face à la diversité des types de dialogues mis en oeuvre de nos jours, nous avons étudié comment adapter, de manière générique, le niveau d'analyse des interventions d'un utilisateur. L'analyse par mots clés prônée par les systèmes de dialogues spécifiques (relatifs à une tâche particulière) est inappropriée pour traiter de telles interventions, car on ne peut vraisemblablement pas représenter toutes les connaissances nécessaires et résoudre les conflits liés à cette diversité.

Nous avons donc développé une analyse des interventions, indépendante de leur domaine et fondée sur leur forme syntaxique. Pour ce faire, nous avons été confrontée au problème du choix d'un analyseur syntaxique. Pour le résoudre, nous avons étudié les différents analyseurs syntaxiques existants en réalisant une classification de ceux-ci selon leurs propriétés. Puis, nous avons développé un protocole d'évaluation pour les analyseurs syntaxiques qui, à l'heure actuelle, fait défaut pour le français.

Suite à cette évaluation, il nous a paru intéressant de développer un algorithme de compromis entre plusieurs analyses afin de retourner l'analyse la plus probable. Celui-ci nous permettra non seulement d'utiliser les capacités de chaque analyseur mais également de quantifier chaque information retournée par un taux de confiance. Cette approche pourra notamment servir de base pour la constitution de corpus annotés.

A partir de la syntaxe des interventions et de connaissances sémantiques issues de la base lexicale WordNet (synonymie, hyperonymie...), nous avons développé un système d'analyse permettant d'étudier l'intention sous-jacente de l'intervention et son contenu propositionnel. Plus particulière-

ment, nous nous sommes intéressée aux interventions de demande d'informations dont le contenu propositionnel repose sur l'identification du type attendu de la réponse et de l'objet sur lequel porte la question.

Pour évaluer l'efficacité de ces critères, cette analyse a été intégrée dans le système de question-réponse QALC développé dans le groupe LIR qui participe depuis quatre ans à la campagne d'évaluation TREC pour la tâche question – réponse.

Mots clés : Traitement automatique des langues, Dialogue, Analyse syntaxique, Evaluation, Système de question – réponse

Decision Theoretic and Logic based Agents for Multi-Agent Systems

Pavlos Moraitis

Habilitation à diriger des recherches de l'Université de
Paris 9, soutenue le vendredi 6 Décembre 2002

Abstract

My research concerns both theoretical and application work on agents and multi-agent systems. The theoretical work is mainly based on the use of decision theory and logic. My research work could be summarized as follows:

- A point of view about agent architectures based on a modular structure, where each module is responsible for one of the possible capabilities an agent may have (e.g. problem solving, cooperation, communication, etc.). The agent's behavior is the result of the different modules interaction. This point of view also suggests the integration of another specific module in the agents' architectures, which is dedicated to the agent's personality and shows how this can have an influence on the other modules.
- The completion of the above point of view with the observation that decision-making is a common characteristic of several deliberation processes involved in the operation of the modules and the idea to propose a unified argumentation based model for their representation.
- The presentation of this argumentation model which is based on the extension of the Logic Programming without Negation as Failure (LPwNF) framework, the integration of the concepts of roles and context to it as well as its use to model personalities.
- A dynamic planning model based on graph representation, taking into account changes that are generated not only by the environment but also by the agent himself and its

exploitation to build a multi-criteria distributed planning framework.

- A multi-criteria negotiation approach, basically used in a distributed planning context, but also able to find a compromise between agents having work and private goals in a more general setting as well as an argumentation-based approach.
- A logical framework for modeling of complex dialogues, adopting the modular agent's architecture and associating each type of the possible dialogues (i.e. negotiation, persuasion, deliberation, etc.) to a specific module, ensuring the automated generation of dialogues and allowing the representation of embedded dialogues.
- Applications of the theoretical results in: e-commerce, marketing, information services, diagnostic and some experimental work on agent software engineering.

Key words : Agent architectures, Argumentation, Distributed Planning, Automated Decision Making, Negotiation, Agent Conversation

Dr. Pavlos Moraitis
Dept. of Computer Science
University of Cyprus
75, Kallipoleos Str.
Nicosia, Cyprus
Mél: moraitis@ucy.ac.cy

Des systèmes autonomes aux systèmes multi-agents : Interaction, émergence et systèmes complexes

Jean-Pierre Muller

Habilitation à diriger des recherches de Université Montpellier II, Ecole doctorale "Information, structures, systèmes", formation doctorale "Informatique", soutenue le vendredi 8 Novembre 2002

Résumé

L'ensemble des travaux a porté sur la production de comportements individuels ou collectifs à partir des interactions, soit entre un agent et l'environnement (systèmes autonomes), soit entre un ensemble d'agents et leur environnement (systèmes multi-agents). Ces travaux ont nécessité de poser la question du lien entre les comportements observés ou désirés et les mécanismes de leur production au sein des agents et en interaction. Pour ce faire nous invoquons la notion d'émer-

gence que nous utilisons à la fois comme grille de lecture pour la simulation de systèmes complexes et comme méthodologie de conception.

Mots clés : Systèmes autonomes, systèmes multi-agent, émergence, simulation, méthodologie

Jean-Pierre Muller
CIRAD TA 60/15
73, av. Jean-François Breton
34398 Montpellier cedex 5
Mél : jean-pierre.muller@cirad.fr
<http://cormas.cirad.fr/>

Étude de modèles de représentations, de requêtes et de raisonnement sur le fonctionnement des composants actifs pour l'interaction homme-machine

Nicolas Sabouret

Thèse de l'Université de Paris 11, soutenue le jeudi 19 décembre 2002

Résumé

Nous abordons dans cette thèse un problème émergent dans la problématique de l'interaction homme-machine : celui de la construction de réponses à des requêtes formelles modélisant des questions de bon sens posées par les utilisateurs ordinaires concernant le fonctionnement d'un composant logiciel actif.

Les recherches actuelles dans la communauté du raisonnement sur les actions proposent de travailler sur une modélisation du système en logique pour répondre à des requêtes bien formées et ne prennent pas en considération l'existence de « notions de bon sens » utilisées par les utilisateurs ordinaires dans la formulation des questions. Nous voulons au contraire pouvoir travailler directement sur le code du composant, et répondre à des requêtes faisant intervenir des notions de bon sens.

Nous présentons tout d'abord un langage de programmation et un modèle d'exécution spécifique pour les composants actifs, qui permet d'avoir accès en cours d'exécution à une description des actions effectuées par le composant et des données manipulées. Nous présentons ensuite un langage de requêtes qui permet de modéliser simplement une large classe de questions sur le fonctionnement issues de l'interaction en

langue naturelle. Nous n'abordons pas les problèmes spécifiques au traitement automatique du langage naturel (TALN), soulevés par la production de ces requêtes. Nous proposons des algorithmes de réponse qui permettent de traiter ces requêtes.

Nous montrons qu'il est possible de prendre en compte des notions de bon sens dans les requêtes et nous proposons un cadre formel pour le traitement de ces requêtes de bon sens sur le fonctionnement.

Nous étudions ensuite le problème de la construction de réponses portant sur l'exécution passée du composant. Nous montrons qu'il est possible et nécessaire d'utiliser une approche dynamique, pour extraire des régularités dans les actions du composant et mettre en évidence des comportements, individuels dans le cas d'une interaction avec l'utilisateur ou collectifs dans le cas d'une interaction dans un système multi-agents, nécessaires à la production d'une explication en réponse aux questions de l'utilisateur.

Nous montrons que nos travaux peuvent s'intégrer dans le cadre du Web sémantique pour définir des composants actifs sous la forme de pages Web dynamiques munies de capacité de raisonnement et d'interaction avec les utilisateurs ordinaires, par exemple pour des tâches d'assistance. Nous présentons enfin l'implémentation de notre modèle qui a été réalisée dans le cadre du projet InterViews et nous donnons des exemples de programmation et d'exécution de ces composants actifs

Mots clés : Représentation des connaissances, Raisonnement sur les actions et le fonctionnement, Interaction homme-machine, Agents conversationnels

Nicolas Sabouret
LIMSI-CNRS
BP 133
91403 ORSAY CEDEX
Mél : Nicolas.Sabouret@limsi.fr
<http://www.limsi.fr/Individu/nico>

Dossier « Web sémantique »

A paraître dans le bulletin numéro 54, Avril 2003

Coordonnateurs :

Jean Charlet (AP-HP, Charlet@biomath.jussieu.fr), Philippe Laublet (LaLICC, Philippe.Laublet@paris4.sorbonne.fr), Chantal Reynaud (LRI, Chantal.Reynaud@lri.fr)

Objectifs

L'expression Web sémantique, attribuée à Tim Berners-Lee au sein du W3C, fait référence à une certaine vision du Web de demain. Selon cette vision, les utilisateurs doivent pouvoir être déchargés d'une grande partie de leurs tâches de recherche, de construction et de combinaison de résultats, grâce aux capacités accrues des machines à accéder à la sémantique des contenus des ressources du Web et à effectuer des raisonnements sur ceux-ci. Les contributions attendues devront ainsi se situer dans le cadre d'une exploitation automatique des contenus des sources d'information et des fonctionnalités des services, accessibles par le Web.

Le Web sémantique, concrètement, est d'abord une infrastructure (architecture, protocoles de communication, nouveaux langages, etc.) pour permettre l'utilisation de connaissances formalisées, en plus du contenu informel actuel du Web. Cette infrastructure doit permettre de localiser, d'identifier et de transformer des ressources de manière robuste et saine tout en respectant l'esprit d'ouverture du Web avec sa diversité d'utilisateurs. Pour cela, elle doit s'appuyer sur un certain niveau de consensus portant, sur les langages de représentation et les ontologies utilisés. Elle doit contribuer à assurer, le plus automatiquement possible, l'interopérabilité et les transformations entre formalismes et entre ontologies. Elle doit réaliser et faciliter l'intégration d'informations provenant de sources d'information distribuées et hétérogènes. Elle doit faciliter la mise en œuvre de raisonnements complexes tout en offrant des garanties sur leur validité, offrir des mécanismes de protection et de qualification des connaissances. Elle doit permettre la composition automatique et l'exploitation de services Web. Ce dossier, coordonné par les animateurs de l'action spécifique Web sémantique du département STIC du CNRS, a pour but de faire le point sur les travaux en cours de la communauté francophone, portant sur l'ensemble de ces aspects.

Qui peut soumettre ?

Toute équipe de recherche de la communauté francophone appartenant à un établissement public ou industriel, impliquée dans des actions de recherche et de développement se situant clairement dans l'optique du Web sémantique.

Comment soumettre ?

Les contributions devront présenter succinctement l'équipe et les principaux travaux concernant le thème IA et Web Sémantique : concepts, techniques, méthodes, outils, expériences, applications mises en œuvre. Elles seront organisées selon le schéma suivant :

- Identification de l'équipe (ou du chercheur) : nom, adresse, site Web, personne à contacter (téléphone, mél) ;
- Membres de l'équipe concernés par le thème du Web sémantique ;
- Thème général de l'équipe ;
- Description des travaux ou projets en lien avec le thème du dossier ;
- Courte bibliographie ciblée (5 références maximum) et adresse d'un site Web où l'on peut trouver l'ensemble des références et articles.

Présentation

Les contributions feront 1400 mots au maximum et devront être envoyées par courrier électronique au format Word ou RTF. Aucune mise en forme particulière ne doit être faite sur les textes autre que gras, italique et la taille des caractères. Si vous êtes concernés par cet appel, envoyez votre contribution,

***** avant le 20 mars 2003 *****

à l'adresse suivante :

Chantal Reynaud
Université Paris-Sud LRI, Bâtiment 490
91405 Orsay cedex
Mél : chantal.reynaud@lri.fr

Pour tout renseignement, s'adresser par courrier électronique, à l'un des coordonnateurs.

Adhésion individuelle et abonnement		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
Nom : Prénom : Affiliation : Adresse postale : N° de téléphone : N° de télécopie : Adresse électronique : Activités (à titre professionnel / à titre privé (<i>rayez la mention inutile</i>)) :			
	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	
<input type="checkbox"/> Adhésion au collège <i>IAD-SMA</i> : ajouter 7,5 Euros pour les étudiants, 15 Euros pour les autres. <input type="checkbox"/> Adhésion au collège <i>Cafe (Apprentissage)</i> : gratuit.			

Adhésion Personne morale		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
Organisme : Adresse postale commune aux bénéficiaires couverts par cette adhésion : Nom et prénom du représentant : Fonction : Mél : Tél : Fax : Adresse postale :			
Le tarif d'adhésion comprend une partie fixe et une partie par bénéficiaire Coordonnées des bénéficiaires (10 maximum) :			
NOM, prénom	Mél.	Tél.	Fax
	Tarif de base fixe :	Tarif par bénéficiaire :	
<input type="checkbox"/> Laboratoires universitaires	100 Euros	30 Euros	
<input type="checkbox"/> Personnes morales non universitaires	300 Euros	30 Euros	
<input type="checkbox"/> Adhésion de soutien	600 Euros	Sans objet	
<input type="checkbox"/> j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA. <input type="checkbox"/> 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 : Marc AYEL, LIA-Université de Savoie, 73376 Le Bourget du Lac cedex. Mode d'adhésion : De préférence, en ligne via le site Internet de l'AFIA : http://www.afia-france.fr A défaut, cette page doit être envoyée au trésorier. Modes de paiement : 1) par chèque, à l'ordre de l'AFIA, envoyé au trésorier. 2) par bon de commande administratif, à l'ordre de l'AFIA, envoyé au trésorier. 3) Par virement bancaire sur le compte de l'AFIA : Société Générale, Résidence du Val de Seine, 78430 LOUVECIENNES. Code banque 30003, code guichet 01902, numéro de compte 00037283856 clef RIB 3. TVA non applicable, article 293B du CGI			

Les dossiers du *Bulletin de l'AFIA*

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

SOMMAIRE DU BULLETIN N° 53

La vie de l'A.F.I.A.	4
L'IA au Québec	6
Présentation de Laboratoires	9
Dossier IA et RNTL	25
Conférences	43
Sommaires des Revues	49
Résumés Habilitations et Thèses	55
Appels à dossiers	61

CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Été</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: 30 septembre
Sortie le 30 avril	Sortie le 31 octobre