

***BULLETIN***

***DE***

***L'AFIA***



**ASSOCIATION FRANCAISE POUR  
L'INTELLIGENCE ARTIFICIELLE**

***JUIN 2003***

***N° 55***

## Présentation du bulletin

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

### Pour contacter l'AFIA

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

**Contributions au bulletin**  
**Jean-Paul SANSONNET**  
voir ci-contre

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

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

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

### Membres d'honneur

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

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

### Personnes morales adhérentes à l'AFIA

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

### Bureau de l'AFIA

**Marc SCHOENAUER, président**

**Serge DUPUY, secrétaire**

**Marc AYEL, trésorier**

**Comité :**

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

## Comité de rédaction

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

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

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

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

**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

**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**  
LRI, bâtiment 490  
Université Paris-Sud  
91405 - ORSAY Cedex  
cornuejols@iie.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



Réa. Numérisation - 01 40 92 74 60

ISSN 1273-1323

Dépôt légal juin 2002

## Le bulletin de l'AFIA change de direction

Ce numéro du bulletin de l'AFIA est consacré essentiellement au dossier « Systèmes d'Information, Connaissances et Intelligence Artificielle ». Ce dossier a été édité par Danielle Boulanger, Eunika Mercier-Laurent et Manuel Zacklad. Son objectif est de faire le point sur les approches et les techniques de l'IA symbolique et numérique pouvant moderniser l'informatique au quotidien, celles d'un élève, d'un citoyen, d'une organisation et de l'entreprise. Il est aussi de dire à un public plus large que celui de la communauté des membres de l'AFIA que des idées géniales de chercheurs doivent être transformées en produits, utilisées plus largement, être intégrées dans l'informatique traditionnelle. Faire le pont entre les approches fondamentales de l'IA et les besoins plus directs en situation effective, voilà l'objet du « Knowledge Management » qui s'est développé au début des années 90 et qui connaît aujourd'hui avec l'explosion de la médiation Internet des connaissances une période critique alors que les modèles et outils n'ont pas été « dimensionnés » pour cette nouvelle donne.

Au mois de juin 2003, L'AFIA a tenu son assemblée générale annuelle qui fut l'occasion de changements tant sur le plan de la direction à donner à l'Association qu'au niveau des diverses responsabilités au bureau. Le Président de l'AFIA, Marc Schoenauer, donne son point de vue sur la situation actuelle de la communauté IA dans la rubrique « vie de l'AFIA ».

Mais le changement concerne aussi le bulletin de l'AFIA qui va évoluer dans la continuité. Après trois années de « bons et loyaux services », je passe la main à une nouvelle direction composée de deux rédacteurs en chefs : Jérôme Euzenat et Philippe Morignot. A cette occasion, je voudrais remercier les membres du comité de rédaction du Bulletin qui pendant toutes ces années ont consacré bénévolement beaucoup de leur temps pour faire vivre les rubriques et tout particulièrement Isabelle Robba, rédactrice adjointe, qui a eu le rôle difficile de finaliser les éditions.

Jérôme Euzenat est membre du bureau de l'AFIA depuis plusieurs années et il est responsable de la rubrique « Groupes de travail » du bulletin de l'AFIA. Il est chercheur à l'INRIA Rhône-Alpes Grenoble où il dirige l'équipe Exmo. Sa recherche concerne la représentation des connaissances distribuées, en particulier l'échange de connaissances formelles médiées par ordinateur : documents annotés, serveurs de connaissances, transformation de représentations, etc.

Philippe Morignot est responsable scientifique au sein de la société AXLOG à Paris. Après son doctorat à SupTélécom il a passé cinq ans dans plusieurs laboratoires de recherche : d'abord au KSL à Stanford puis au CRIN-CNRS à Nancy, et aussi à ICS-FORTH (Grèce) et au GMD (Allemagne). Sa recherche a porté sur l'architecture logicielle d'agents autonomes pour le contrôle de robots mobiles, avec en particulier les problèmes de génération embarquée de plans d'actions. Plus récemment, il a travaillé sur les systèmes d'information.

Jérôme et Philippe auront la tâche de faire évoluer le Bulletin, alors que nous avons envisagé de le rendre plus cohérent avec les autres médias de l'association comme le portail par exemple qui contient de plus en plus d'informations en ligne et qui doit s'équilibrer avec le Bulletin. Je suis persuadé que Philippe et Jérôme, qui a déjà travaillé sur le site de JeDAI, vont faire du bon travail avec le comité.

**Jean-Paul Sansonnet**

*Rédacteur en chef du Bulletin de l'AFIA*

## Message du Président

Je commencerai par remercier Gilles Bisson, président de la plate-forme AFIA 2003, ainsi que Anne-Marie Kempf et l'ESIEA, grands organisateurs, grâce auxquels la plate-forme a réuni plus de 200 personnes dans une ambiance chaleureuse et intellectuellement vivifiante.

Au delà des problèmes posés par la faible participation à la vie de l'association de membres par ailleurs enthousiastes, ainsi que nous avons pu le constater lors de l'Assemblée générale le 3 juillet, voici l'une des questions que se pose le bureau de l'AFIA, en cette rentrée 2003, et j'en profite pour vous la poser.

D'un côté, la fascination de l'IA reste entière : nombreux restent ceux qui rêvent d'une machine qui n'est jamais ni tout à fait la même ni tout à fait une autre, nous parle et nous comprend. D'un autre côté, nombreux sont ceux qui se lassent d'espérer - de nombreux industriels, hélas, en font partie - et rêvent à de nouveaux concepts : les systèmes complexes ou les méthodes émergentes, intelligence en essaim ou vie artificielle.

Au delà des modes, l'objectif est étrangement similaire : travailler à la frontière des problèmes faisables, là où les connaissances et/ou l'interaction des agents font la différence.

Où se situe donc la frontière entre IA et méthodes émergentes ? Quelles sont les spécificités de notre champ ? Longtemps, notre force est venue de l'attention portée à la représentation des problèmes, et de la conception d'opérateurs (symboliques ou numériques) dédiés aux représentations choisies. Mais les bénéfices de cette approche sont tels qu'elle fait désormais partie des principes de l'informatique.

Quel est donc le propre de l'IA ? Si cette question vous inspire, envoyez nous (comme au moins les membres du bureau se sont engagés à le faire) une page sur les grands dessins de l'IA selon vous. Rendez-vous dans un prochain bulletin pour le dépouillement et l'analyse de toutes les contributions !

**Marc Schoenauer,**  
*Président de l'AFIA*

## GRACQ

### Groupe « Ingénierie des connaissances » Groupe AFIA et structure d'animation du GdR-I3

#### *Mode de fonctionnement*

Le GRACQ, groupe « Ingénierie des connaissances », fédère, dans le cadre de l'Intelligence Artificielle, une communauté qui propose des concepts, méthodes et techniques permettant de modéliser, de formaliser, d'acquérir des connaissances dans les organisations, dans un but d'opérationnalisation, de structuration ou de gestion au sens large. L'intégration des systèmes à base de connaissances construits au sein du *système d'usage de l'utilisateur, ou, plus collectivement, au sein de l'organisation*, est un enjeu majeur qui va conditionner l'évaluation des propositions faites en matière de modélisa-

tion et les réflexions menées en Ingénierie de connaissances.

Cette problématique et d'autres ne sont abordables qu'en se rapprochant de disciplines abordant les mêmes objets de recherche et produisant des connaissances complémentaires aux siennes comme la sociologie, la gestion ou l'ergonomie qui peuvent concourir à une démarche d'explicitation de ce que sont les connaissances dans un contexte humain et organisationnel. C'est ainsi qu'en 2002 et à la suite des séminaires organisés au sein du GRACQ, des collaborations se sont réactivées entre Conception et Ingénierie des connaissances. C'est à la fois la preuve d'une démarche fonda-

mentalement pluridisciplinaire au sein de l'ingénierie des connaissances et du travail nécessaire de mise en commun des concepts pour activer cette pluridisciplinarité.

Cette communauté de travail se développe au sein du GRACQ, Groupe acquisition et ingénierie des connaissances depuis 1990. La communauté de recherche regroupée autour de ce thème s'est munie d'une liste de diffusion <info-ic@biomath.jussieu.fr> et d'un site Web <<http://www.irit.fr/GRACQ/>>, portail vers les travaux et l'actualité du domaine, toujours maintenu par N. Aussenac-Gilles qui a laissé sa place à F. Trichet pour l'animation de la liste info-ic au côté de J. Charlet.

D'un point de vue pratique, des directions de coopérations privilégiées ont amené la constitution de groupes de travail, développant des concepts et des outils sur le moyen et le long terme. C'est le cas du groupe COOP, au sein du GRACQ, et, en dehors, du groupe TIA.

## Groupe COOP

Le groupe COOP est un groupe restreint de recherche et d'animation de la communauté qui regroupe des chercheurs en ingénierie des connaissances et en sciences cognitives. Sa problématique initiale était celle de la modélisation de la coopération dans le contexte d'activités de résolution de problème ou interviennent des agents humains et logiciels. Il se positionne aujourd'hui clairement à l'intersection de l'ingénierie des connaissances et du CSCW (*Computer Supported Cooperative Work*) ce qui constitue une originalité dans le panorama scientifique européen. Il est le groupe fondateur et le comité de pilotage des conférences internationales COOP qui se tiennent tous les deux ans. Il assure l'organisation de ces conférences avec l'aide de L'INRIA et la cinquième édition, COOP 2004, se déroulera en juin 2004 à Cannes <<http://tech-web-n2.utt.fr/coop/>>

Le groupe est également impliqué dans la conférence interdisciplinaire CITE (Coopération Innovation Technologie) dont la deuxième édition, CITE 2003, se tiendra les 3 et 4 décembre à l'Université de Technologie de Troyes <<http://tech-cico.utt.fr/cite2003/>> en relation avec l'Action Spécifique du département STIC PCD-TC « Pratiques Collectives Distribuées et Technologies de la Coopération ».

## Séminaires du GRACQ

En dehors des groupes constitués décrits ici, les séminaires du GRACQ sont l'occasion de faire le point sur des directions de recherche ou d'entamer des débats sur des thèmes ou des coopérations qui nous semblent importantes. En 2003, 2 séminaires ont eu lieu :

- *Ingénierie des connaissances et santé* - 28 janvier 2002

Organisé par M.-C. Jaulent (INSERM), J. Lieber (INRIA) et J. Charlet (AP-HP) Paris, UFR Broussais-Hôtel-Dieu, 15, rue de l'école de médecine, 75006 PARIS. Le programme de cette journée est accessible sur le serveur du GRACQ à la rubrique activité 2002. Cette conférence a rassemblé plus de 40 personnes.

- *Ontologies et thésaurus* – 8 février 2002

Organisé par F. Sèdes (IRIT) dans le cadre du GT « modèles multi-média » du GdR-I3 à Toulouse.

## La conférence « Ingénierie des connaissances »

Placées sous l'égide du GRACQ, les conférences IC (Ingénierie des Connaissances) sont chaque année un lieu d'échange et de réflexion de la communauté française sur les problématiques de l'Ingénierie des connaissances. Elles rassemblent une centaine de participants.

Les journées IC-2002 ont été présidées par B. Bachimont de l'INA et organisées par le laboratoire PSI à Rouen sous la présidence de C. Gréboval-Barry. 46 propositions ont été reçues, 20 ont été acceptées sous forme de communications écrites ou orales, ainsi que 6 posters.

La conférence IC-2003 se déroule du 1 au 3 juillet dans le cadre de la plate-forme AFIA <<http://www.afia.polytechnique.fr/plateforme-2003/>> (plate-forme présidée par G. Bisson et organisée à Laval par A.-M. Kempf de l'ESIEA) et est présidée par R. Dieng-Kuntz (INRIA, Sophia Antipolis). Comme les années précédentes, des thèmes fléchés au sein de l'appel à communication ont été l'occasion d'apprécier la diversité des sujets constitutifs de l'Ingénierie des connaissances et des apports d'autres disciplines à sa réflexion. Sur les 35 articles reçus, 17 ont été retenus pour communication orale et publication.

Une sélection d'articles de plusieurs de ces conférences est par ailleurs opérée à partir des actes et diffusée sous forme de livre. Après un premier livre paru en 1996, qui reprenait des articles des années 1992-94,

puis le livre paru en 2001 (années 1995-98) avec la participation du CNET, un nouveau livre est en cours de finalisation qui reprend des articles des années 1999-2001. Il inclura un CD-Rom de consultation des articles via un index construit selon des techniques de traitement automatique des langues et de représentation de connaissances à partir de textes.

## Enseignement

L'Ingénierie des connaissances est enseignée en 3e cycle. Les jeunes chercheurs de la communauté sont ainsi principalement issus des DEA suivants : DEA en Informatique « Communication Homme-Machine & Ingénierie Educative » (Université du Maine – <http://ww-iupmime.univ-lemans.fr/depinfo/dea/index.html>), DEA « RACOR : Réseaux Avancés de Connaissances et Organisations » de l'Université Technologique de Troyes (UTT – [http://www-recherche.univ-troyes.fr/troisieme\\_cycle/frame\\_trois.htm#RACOR](http://www-recherche.univ-troyes.fr/troisieme_cycle/frame_trois.htm#RACOR)), DEA RCFR (Toulouse 3), DEA « I3 : Information, Interaction, Intelligence » (Univ. Paris XI - Centre Scientifique d'Orsay <http://www.lri.fr/DEA/I3>), DEA « Informatique : Systèmes Intelligents » (Paris Dauphine <http://L1.lamsade.dauphine.fr/dea127/>), DEA en Représentation des Connaissances et Formalisation du Raisonnement (Univ. Toulouse 2 – <>), DEA d'Intelligence Artificielle et d'optimisation combinatoire de l'université Paris XIII (LIPN), DEA d'Informatique de l'Université de Picardie Jules Verne à Amiens, <http://www.laria.u-picardie.fr/dea.html>. En outre des cours sont donnés dans des DESS comme le DESS d'IA de Paris 6 <http://www.dess-ia.infop6.jussieu.fr/>, des écoles d'ingénieurs comme l'Institut Galilée de l'université Paris XIII ou le DEA Dynamique des organisations (ESSEC, Mines, Polytechnique, Paris X).

## International

La communauté participe à et anime la communauté européenne, en particulier avec des communications dans la conférence bi-annuelle EKAW. À l'été 2002, Le thème de l'Ingénierie des Connaissances était bien représenté dans le comité de programme d'ECAI 2002 et dans les workshops organisés à cette occasion. Nathalie

Aussenac-Gilles et Alexander Maedche ont organisé

un *workshop* sur « Ontology Learning and NLP » durant la conférence ECAI 2002 qui s'est tenue à Lyon les 22 et 23 juillet 2002. Ce workshop a permis la présentation de 12 communications, de 2 conférences invitées (R. Basili et B. Bachimont) devant près de 40 participants. Des membres de la communauté ont organisé d'autres workshops à ECAI'2002 (R. Dieng-Kuntz et N.a Matta sur « Knowledge Management and Organizational Memories » le 22 juillet et R. Dieng-Kuntz, S. Handschuh et S. Staab sur « Semantic Authoring, Annotation and Knowledge Markup » le 23 juillet) ainsi qu'à EKAW 2002 (R. Dieng-Kuntz et F. Gandon sur « Knowledge Management and Corporate Semantic Webs » le 30 septembre). La communauté est représentée dans le réseau thématique OntoWeb.

## Orientations

La problématique de recherche de la conférence Ingénierie des connaissances est organisée en thèmes de recherche qui sont régulièrement remis à jour. L'animation de ces thèmes de recherche se fait à l'initiative de membres du GRACQ avec l'agrément du bureau. Depuis la fin 2002, les forces d'animation de la communauté se sont recentrées sur les actions spécifiques du CNRS dont un certain nombre touchent de près à l'Ingénierie des connaissances. Quelques unes sont bien avancées ou se terminent :

- AS « Indexation multimédia » (B. Bachimont, G. Vignaux – <<http://liris.cnrs.fr/as50/>>)
- AS « Web sémantique » (J. Charlet, Ph. Laublet, Ch. Reynaud – <<http://www.lalic.paris4.sorbonne.fr/stic/>>)
- AS « Construction de ressources terminologiques à partir de corpus » (A. Condamines, N. Aussenac-Gilles – <<http://www.irit.fr/ASSTICCOT/>>)

D'autres viennent de démarrer ou sont en cours de lancement :

- AS « Fondements théoriques et méthodologiques de la conception des EIAH » (P. Tchounikine)
- AS « Connaissances, activité, organisation » (R. Teulier, P. Salembier)
- AS « Ingénierie de l'information et des connaissances en santé » (M.-C. Jaulent, J. Charlet – <<http://www.spim.jussieu.fr/ASICS/>>)
- AS « Pratiques Collectives Distribuées et Technologies de la Coopération » (M. Zacklad, C. Henry, B. Turner)

- AS « Web sémantique et E-learning » (M. GrandBastien, D. Hérin et R. Dieng-Kuntz).

Les 2 AS, « Construction de ressources terminologiques à partir de corpus » et « Web sémantique » se déroulent au sein du réseau thématique pluridisciplinaire sur les documents (RTP-DOC – <<http://rtp-doc.ens-sib.fr/>>). Dans ce cadre, de nombreux contacts ont été pris avec les communautés Documents et Recherche d'information. Plus particulièrement, le GRACQ et les chercheurs de la structure d'animation 5.2 du GdR-I3 <<http://www.rfai.li.univ-tours.fr/gdri3/>> ont projeté de se rencontrer pour développer une réflexion commune

sur la question de la modélisation et des documents. Une journée est prévue le 6 octobre 2003, sans doute sur le thème des méta-données.

Les 2 AS « Connaissances, activité, organisation » et « Pratiques distribuées » se déroulent au sein du réseau thématique pluridisciplinaire Économie et STIC et collaborent de plusieurs façons. En outre plusieurs participants de ces deux AS et d'autres précitées (P. Salembier, B ; Bachimont, J. Charlet, M. Zacklad, R. Teulier) participeront à une réflexion commune lors de rencontres de CERISY (septembre 2003) qui feront l'objet d'un ouvrage commun.

# PRÉSENTATION DE LABORATOIRES

## Présentations de laboratoires dans le bulletin de l'AFIA

LIFIA, Grenoble .....	Bulletin n°1	LIA - Université de Savoie .....	Bulletin n°22
LRI, Orsay .....	Bulletin n°1	INRETS.....	Bulletin n°23
Service Systèmes Experts, Renault.....	Bulletin n°1	IRIN Nantes .....	Bulletin n°24
CEDIAG, .....	Bulletin n°2	CRIN - INRIA Lorraine.....	Bulletin n°25
CERT, ONERA, Toulouse.....	Bulletin n°2	DIRO - Université de Montréal .....	Bulletin n°26
IRIT, Toulouse .....	Bulletin n°2	IRIT - Toulouse (1) .....	Bulletin n°28
LAAS, Toulouse.....	Bulletin n°2	IRIT - Toulouse (2) .....	Bulletin n°29
HEUDIASYC, UTC.....	Bulletin n°3	LAAS - Toulouse (1) .....	Bulletin n°30
IFP, Rueil Malmaison.....	Bulletin n°3	Sony CSL .....	Bulletin n°31
DIAM, INSERM U194.....	Bulletin n°3	LAAS - Toulouse (2) .....	Bulletin n°32
Lab. Math. Info., Fac Médecine de Marseille..	Bulletin n°4	LIMSI - Département CHM .....	Bulletin n°33
GMD, St. Augustin (RFA) .....	Bulletin n°4	LAMSADE .....	Bulletin n°34
ONERA, Chatillon .....	Bulletin n°4	Institut autrichien de recherches en I.A. ....	Bulletin n°36
KSL, Université de Stanford (USA) .....	Bulletin n°5	LIP6 – Université Pierre et Marie Curie.....	Bulletins n°37 & 38
Dépt Applications de l'IA au CNET, Lannion.	Bulletin n°5	GREYC – Université de Caen .....	Bulletin n°40
LAFORIA, Univ. Pierre et Marie Curie.....	Bulletin n°6	LIFL – Université de Lille.....	Bulletin n°41
L'institut FAW, ULM (RFA) .....	Bulletin n°6	LRI (équipes IA et IAS).....	Bulletin n°43
Institut IIIA, Compiègne .....	Bulletin n°6	IMAG - Grenoble.....	Bulletin n°44
LAIR, OHIO State University (USA).....	Bulletin n°7	PSI (Perception, Système, Information - Rouen).....	Bulletin n°45
ARAMIHS, Labo mixte MATRA-CNRS,.....	Bulletin n°7	INRIA – Sophia Antipolis .....	Bulletin n°46/47
CEA, Service SERMA, Saclay .....	Bulletin n°8	LIH – Laboratoire d'Informatique du Havre .....	Bulletin n°46/47
Société ILOG.....	Bulletin n°8	Tech-CICO – Université de Technologie de Troyes .....	Bulletin n°51
LAIAC, Université de Caen .....	Bulletin n°9	LIFO – Université d'Orléans – Équipe Contraintes et Apprentissage.....	Bulletin n°52
Institut Français du Pétrole .....	Bulletin n°10	LIIA – Ecole Nationale des Arts et Industries de Strasbourg .....	Bulletin n°52
DFKI (Centre allemand de recherches en IA) .	Bulletin n°11	LRL – Laboratoire de recherche sur le langage – Université Blaise Pascal Clermont 2 .....	Bulletin n°53
GRTC, Marseille .....	Bulletin n°11	MIG – Mathématique, Informatique et Génome – INRA.....	Bulletin n°53
Inst. d'Analyse des Systèmes, Ac. Russe .....	Bulletin n°12	ESIEA Recherche, Laval et Paris .....	Bulletin n°54
Georges Mason Univ., Center for AI (USA)....	Bulletin n°13	Équipe Intelligence Artificielle et Applications (IAA) du Crip5, Paris 5.....	Bulletin n°55
IRISA, INRIA et Université de Rennes .....	Bulletin n°13	LAMIH – Univ. de Valenciennes et du Hainaut-Cambrésis.....	Bulletin n°55
Société INGENIA .....	Bulletin n°14		
LIPN, Université de Paris Nord .....	Bulletin n°14		
Institut EURISCO .....	Bulletin n°15		
LRDC, Université de Pittsburgh (USA).....	Bulletin n°15		
Société ISOFT .....	Bulletin n°16		
Dépt. d'Info de l'Université d'Ottawa .....	Bulletin n°16		
Equipe CHM, Université du Colorado (USA) .	Bulletin n°17		
LIRMM, Montpellier .....	Bulletin n°19		
Institut autrichien de recherches en I.A. ....	Bulletin n°20		
ENST Bretagne .....	Bulletin n°21		

## Équipe Intelligence Artificielle et Applications (IAA) du Centre de Recherche en Informatique de Paris 5 (Crip5)

<http://www.math-info.univ-paris5.fr/crip5>

**Responsable : Dominique Pastre**

### Le Crip5

Le Crip5 est une Equipe d'Accueil (EA 2517) créée en 1998 à l'Université René Descartes – Paris 5.

Il est issu du regroupement des anciennes équipes de recherche en informatique de Paris 5 :

- Équipe d'Accueil (EA 1499) Systèmes Intelligents de Perception (SIP) / Laboratoire d'Intelligence Artificielle de Paris 5 (LIAP5)
- Jeune Équipe (JE 348) Informatique et Communication

Il comporte actuellement (septembre 2003) 30 enseignants-chercheurs et 25 doctorants.

Il est depuis 2001 structuré en trois équipes :

- IAA : Intelligence Artificielle et Applications
- SIAR : Systèmes d'Information et Applications Réparties
- SPIR : Signal, Parole, Image, Réseaux

Le Crip5 est rattaché à l'école doctorale « Cognition, comportement et conduite humaine » de Paris 5.

### L'équipe IAA

*Mots-clefs* : Résolution de problèmes, Raisonnement, Représentation et Acquisition de Connaissances, EIAH, Connaissances et Applications distribuées, Evolution artificielle

L'équipe IAA regroupe actuellement tous les chercheurs en Intelligence Artificielle de Paris 5. Elle comporte 13 enseignants-chercheurs et 13 doctorants.

Pour des raisons pratiques et/ou historiques, cette équipe IAA est structurée

en trois groupes, ce qui ne signifie nullement que les thèmes soient complètement disjoints. D'ailleurs certains travaux sont partagés par plusieurs groupes.

- Le groupe SBC (issu du thème Systèmes à Base de Connaissances de l'ancienne Jeune Equipe Informatique et Communication) a pour vocation de travailler sur l'utilisation et l'acquisition de connaissances, sur le raisonnement, et de concevoir et réaliser des systèmes de résolution de problèmes ainsi que leur utilisation en EIAH.
- Le groupe SAFE (Systèmes d'Assistance à la Formation et à l'Education, créé en 2001, issu du groupe SBC) se veut spécialisé en EIAH et IHM, sans toutefois négliger l'aspect résolution de problèmes et utilisation de connaissances.
- Le groupe LIAP5 (issu du Laboratoire d'Intelligence Artificielle de Paris 5 de l'ancienne EA 1499) met désormais l'accent sur la recherche fondamentale en informatique et la vie artificielle.

### Groupe SBC : Systèmes à Base de Connaissances

*Enseignants-chercheurs* : Bruno Bouzy, Bernard Delforge, Sylvie Després, Dominique Pastre, Jean-Pierre Spagnol  
*Doctorants* : Rym Ben Haddada, Valentina Ceausu, Elise Fage  
*Post-doctorant* : Pierre Vinant

Le thème principal, et fédérateur, des projets de ce groupe est l'acquisition, la modélisation et/ou l'utilisation des connaissances. Différents types de rai-

sonnements (stratégique, mathématique, diagnostique) y ont une grande place. Les travaux réalisés peuvent être vus comme des résolutions de problèmes, le problème à résoudre pouvant être de gagner à un jeu de réflexion, de démontrer un théorème, d'aider un élève à apprendre, ou encore de comprendre des scénarios d'accidents.

### Jeux de réflexion

Bruno Bouzy

L'objectif du projet « Jeux de réflexion » est de découvrir des techniques de résolution de problèmes intéressantes pour l'IA à partir d'expériences informatiques sur les jeux de réflexion à deux joueurs avec une forte complexité (Go, Amazones, Hex, Echecs). La période 1999-2003 a été significative sur plusieurs plans : le raisonnement spatial et la modélisation d'abord, la théorie des jeux et la génération automatique de connaissances, entre autres, et enfin et surtout l'approche de Monte Carlo explorée récemment. L'ensemble de ce travail est guidé par le principe pratique de validation expérimentale. Le logiciel Indigo (<http://www.math-info.univ-paris5.fr/~bouzy/INDIGO.html>) est un programme jouant au Go développé au fil des années, implémentant des idées suivant les différents plans ci-dessus. Seules les idées donnant des résultats convenables en pratique dans le programme sont gardées. Indigo a participé à plusieurs compétitions internationales durant cette période : en 2000 à Londres aux olympiades des jeux de réflexion (5<sup>ème</sup> sur 6), à Edmonton à la seconde « 21st century cup » (10<sup>ème</sup> sur 14) et à Guyang en Chine (6<sup>ème</sup> sur 10) en 2002.

Concernant le raisonnement spatial et la modélisation, les nombreux concepts spatiaux intervenant dans une fonction d'évaluation au Go ont été explicités dans une première version puis affinés et retravaillés. L'apport de la modélisation du jeu de Go pour le raisonnement spatial a ainsi été montré. Toujours en suivant l'axe de la modélisation dépendante du domaine, une étude de fonctions d'évaluation pour le Go a également été présentée dans le cadre de la conférence « Computers and Games » en juillet 2002. Concernant la théorie des jeux appliquée au Go, un travail a montré les propriétés des « sous-jeux » intervenant au Go. Concernant la génération automatique de « patterns », un travail utilisant l'analyse rétrograde a été testé, mais n'a pas donné de résultats suffisants pour être intégré dans la version actuelle d'Indigo. En coopération avec le laboratoire d'IA de Paris 8, l'état de l'art, orienté IA, de la programmation du Go a été dressé afin de montrer pourquoi la programmation du go est un problème d'actualité pour l'IA. Le numéro spécial sur la morphologie mathématique de la revue IJPRAI inclut un article sur l'application de la morphologie mathématique à la programmation du jeu de Go. Le journal édité par l'ICGA (International Computer Game Association) a publié un numéro spécial sur la programmation du jeu de Go dans lequel est décrit le processus de décision original utilisé par Indigo. Ce processus repose d'une part sur de la recherche arborescente sur des coups calmes, et d'autre part, sur un mécanisme de calcul d'urgence des coups.

Enfin, une autre coopération a été explorée récemment avec B. Helmstetter (Université Paris 8) sur l'apport des techniques de Monte Carlo au go. Cette coopération a été très efficace suivant notre principe pratique. Suivant le travail initiateur de B. Brugmann datant de 1993, les mécanismes statistiques utiles pour la programmation du go ont été étendus, ou simplifiés, et testés expérimentalement. Cette approche est intéressante car elle

ne nécessite pas de connaissances du domaine, sauf celles des règles du jeu. La puissance des machines actuelles la rend possible sur des damiers 9x9 ou 13x13 et bientôt sur 19x19. En observant que l'approche statistique et l'approche fondée sur des connaissances donnent des résultats semblables sur 9x9, il était naturel de tester le niveau d'un programme combinant les deux approches. L'intégration d'une approche fondée sur des connaissances avec cette approche statistique a donc été prouvée expérimentalement en faisant jouer trois versions d'Indigo : la version fondée sur des connaissances, celle utilisant les statistiques seules et la meilleure, celle fondée à la fois sur les connaissances et sur les statistiques. Les « preprints » correspondant à la description de ces travaux sont disponibles en ligne (<http://www.math-info.univ-paris5.fr/~bouzy/publications.html>). Plusieurs stages de 2 à 4 mois ont été encadrés durant cette période : sur le jeu d'Hex, sur le jeu d'amazones et sur les Échecs ; enfin, en collaboration avec l'ENST, des projets de fin d'études sur Monte Carlo go ont été suivis. B. Bouzy a animé la session « Art et jeu » du colloque jeunes chercheurs en sciences cognitives en juin 2003 (CJC'03).

## Publications

- Bouzy B., *Programmation du jeu de Go et Raisonnement Spatial, Journées Nationales sur les Modèles de Raisonnement, JNMR'99*, Paris, 1999
- Bouzy B., *Complex Games in Practice*, 5th Game Programming Workshop in Japan, GPW'99, p. 53-60, Hakone, 1999
- Bouzy B., *Go patterns generated by retrograde analysis*, 6th Computer Olympiad, Computer-Games Workshop Proceedings, J.W.H.M. Uiterwijk (ed.), Report CS 01-04, 9 p., Maastricht, 2001
- Bouzy B., *Les concepts spatiaux dans la programmation du go*, Revue d'Intelligence Artificielle, vol. 15 n°2, 2001, pp.143-172
- Bouzy B., Cazenave T., *Computer Go : an AI oriented Survey*, Artificial Intelligence Journal, vol. 132 n°1, pp. 39-103, 2001
- Bouzy B., *A small go board study of dimensional and metric evaluation functions*, Proceedings Computers and Games 2002, Edmonton, 2002
- Bouzy B., *Mathematical morphology applied*

*to computer go*, International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), vol. 17 n°2, pp. 257-268, 2003

Bouzy B., *The move-decision strategy of Indigo*, International Computer Game Association Journal (ICGAJ), vol. 26 n°1, pp.14-27, 2003

Bouzy B., *Associating domain-dependent knowledge and Monte Carlo approaches within a go program*, Joint Conference on Information Systems (JCIS'03), Cary, septembre 2003

Bouzy B., Helmstetter B., *Developments on Monte Carlo go*, Advances in Computer Game 10 (ACG10), Graz, Autriche, novembre 2003

## Modélisation et automatisation du raisonnement mathématique

Dominique Pastre

et Jean-Pierre Spagnol

Le démonstrateur à base de connaissances MUSCADET, qui a été développé au LAFORIA jusqu'en 1992 et dont la seconde version, entièrement écrite en Prolog a été opérationnelle en 1998 continue à être améliorée et expérimentée. La version 2.4 est disponible sur Internet, ainsi qu'un manuel de l'utilisateur. Il participe à des compétitions internationales de démonstrateurs. D'autre part, le système ARGOS (thèse de J-P. Spagnol), développé à partir de MUSCADET 2.0, est capable d'une part de rédiger des preuves en français et d'autre part de proposer une interface fournissant une assistance aussi bien à un utilisateur enseignant qu'à un élève autonome.

– On rappelle que le démonstrateur MUSCADET est un système à base de connaissances fondé sur la déduction naturelle. Il a été expérimenté essentiellement dans des domaines mathématiques. Les connaissances sont constituées de données formelles, de stratégies de démonstration et de savoir-faire. L'utilisation de Prolog comme langage unique dans MUSCADET 2, a permis d'exprimer aussi bien des connaissances procédurales que des connaissances déclaratives, ainsi que tous les intermédiaires possibles et a augmenté les performances du nouveau

système. C'est la facilité avec laquelle l'expert peut donner les connaissances au système qui est maintenant privilégiée.

MUSCADET privilégie les notions d'hypothèse, de conclusion, de découpages d'un problème en plusieurs problèmes plus simples, de sous-théorèmes. Des règles sont construites par des métarègles à partir des définitions formelles. Les quantificateurs ne sont pas éliminés comme dans le principe de résolution mais, au contraire, restent le plus près possible des sous-formules sur lesquelles ils portent. MUSCADET travaille alors sur des sous-formules du 1er ordre, d'une manière analogue à ce que fait l'être humain, et non sur des ensembles de clauses.

Depuis 1998 un travail a été fait dans le cadre de la librairie TPTP (Thousands of Problems for Theorem Provers, <http://www.cs.miami.edu/~tptp>) destinée à tester et évaluer les systèmes de démonstration automatique de théorèmes. Ainsi MUSCADET 2 a pu être testé sur de nombreux problèmes autres que mathématiques. De nouvelles stratégies de démonstration ont dû être écrites, avec de nouveaux savoir-faire. D'autre part, plus d'une centaine de problèmes de théorie des ensembles ont été proposés pour enrichir la librairie. Les expérimentations faites ont montré que certains de ces théorèmes, démontrés par MUSCADET 2, ne le sont par aucun des autres démonstrateurs les ayant testés.

Enfin, MUSCADET 2 participe aux compétitions de démonstrateurs (<http://www.cs.miami.edu/~tptp/CASC>) organisées dans le cadre des conférences CADE (Conférence on Automated DEduction) depuis 1999. MUSCADET est le seul démonstrateur participant à ces compétitions n'utilisant pas le principe de résolution ni même la mise sous forme de clauses et ne peut ainsi concourir que dans une seule catégorie où les énoncés sont des formules du premier ordre (plusieurs autres catégories donnant les énoncés directement sous forme d'ensembles de clauses). Les

résultats obtenus montrent la complémentarité de MUSCADET par rapport aux démonstrateurs fondés sur le principe de résolution. MUSCADET est particulièrement efficace pour démontrer des théorèmes manipulant de nombreux concepts définis par d'autres énoncés à partir duquel il construit des règles, et particulièrement inefficace pour les énoncés uniques pouvant être énormes et/ou dénués de sémantique. Il est plus efficace sur des énoncés exprimés d'une manière naturelle, comme le fait le mathématicien dans une pratique courante, que sur les énoncés exprimés d'une manière plus axiomatique. Cela confirme la nécessité de faire collaborer plusieurs approches en résolution de problèmes.

MUSCADET 2.4 est disponible sur le Web à l'adresse <http://www.math-info.univ-paris5.fr/~pastre/muscadet/muscadet.html>.

– Le système ARGOS (<http://perso.club-internet.fr/jps0123/demonstrateurArgos.html>), conçu à partir du système MUSCADET 2.0, résout automatiquement des exercices de géométrie, proposés par un utilisateur, au niveau de l'enseignement secondaire. Il est ensuite capable d'expliquer sa solution, en français, en tenant compte des desiderata de l'utilisateur : niveau scolaire, filtrage de la preuve en fonction du contrat didactique en cours. Les connaissances sont fournies au système de façon déclarative et ce dernier se charge de les opérationnaliser en construisant automatiquement des règles modélisant le comportement du mathématicien en situation de résolution de problèmes. Le système dispose de stratégies de déductions et de création dynamique d'objets, apprises et mémorisées à partir de tous les exercices déjà résolus.

Ce système est maintenant doté d'une interface utilisable sur poste fixe, dans un Intranet ou sur Internet. Un enseignant ou un élève peut par exemple enregistrer une nouvelle propriété mathématique, la relier à tel ou tel

domaine et niveau scolaire, lui associer un modèle explicatif (instancié avec les bons objets lors de la construction de la preuve rédigée) et une phrase énonçant le savoir-faire associé (c'est-à-dire la méthode modélisée dans la propriété permettant de déduire la configuration associée).

Un utilisateur peut d'autre part ajouter un nouvel exercice dans la base qui est indexée en fonction d'un certain nombre de critères. Un enseignant peut décider de définir un nouveau domaine et d'y associer les exercices de son choix. Ceci lui permet donc de créer des séquences pédagogiques qu'il peut alors proposer à ses élèves.

Une liste d'exercices de plusieurs niveaux scolaires, cinquième, quatrième, troisième et seconde a été enregistrée permettant à un élève de sélectionner certains d'entre eux en fonction de différents critères. On peut lister tous les exercices d'un niveau scolaire donné, déduisant telle ou telle configuration, relevant de tel ou tel domaine ou savoir-faire.

L'élève choisit alors son exercice, en tenant compte du niveau de difficulté, ou cherche l'exercice ou la série d'exercices proposés par son professeur. Par un clic sur le lien une page est générée automatiquement mettant sous les yeux de l'élève une figure (construite sous GEOPLANW et modifiable : composant activX intégré) ainsi que l'énoncé correspondant. ARGOS intervient alors comme une aide à la résolution permettant à l'élève, à l'aide d'une barre de menu, d'extraire de l'énoncé les hypothèses et conclusions, de poser toute conjecture intermédiaire plausible. Il est possible de vérifier si l'on a pas oublié une hypothèse essentielle. On peut ensuite lancer la recherche en la paramétrant : niveau scolaire, connaissance prioritaire ou interdite, granularité explicative, temps limite. Lorsque le système a trouvé une solution, l'élève peut demander différentes aides. Il peut demander la liste des savoir-faire associés à ses conjectures hors contexte ou simplement ceux qu'a utilisés ARGOS

dans sa solution, instanciés avec les objets utiles. Il peut obtenir le dernier chaînon explicatif pour ses conjectures (pour le mettre sur la voie) ou la solution complète. Il peut aussi obtenir pour une configuration donnée toutes les instanciations déduites par ARGOS lors de sa recherche.

Le système a été utilisé dans plusieurs classes de collège et une classe de seconde. Plusieurs expérimentations sont en cours :

- Utilisation en seconde dans le cadre des modules et du suivi individualisé. Une étude statistique va permettre d'évaluer l'impact de son utilisation sur l'enseignement dispensé.
- Utilisation comme compagnon d'apprentissage. La meilleure méthode pour apprendre une notion est d'avoir à l'enseigner. L'élève va apprendre au système les connaissances du cours ainsi que la façon de les expliquer dans une preuve rédigée. En posant des exercices il pourra se confronter aux solutions trouvées par ARGOS et pourra voir en acte ses modèles explicatifs. Ceci ne pourra que conforter ses capacités de rédaction, principale difficulté pour les élèves.

## Publications

- Pastre D., *Le nouveau MUSCADET et la TPTP Problem Library*, Colloque Intelligence Artificielle, Berder, 1999, rapport LIP6 2000/002, p. 54-98
- Pastre D., *MUSCADET version 2.3* : Manuel de l'utilisateur, 2001, 16p
- Pastre D., *MUSCADET version 2.3* : User's Manual, 2001, 16p
- Pastre D., *MUSCADET2.3 : A Knowledge-based Theorem Prover based on Natural Deduction*, International Joint Conference on Automated Reasoning IJCAR 2001 (Conference on Automated Deduction CADE-JC), p. 685-689
- Pastre D., *Implementation of Knowledge Bases for Natural Deduction*, 8th International Conference on Logic for Programming, Artificial Intelligence and Reasoning, 2nd International Workshop on Implementation of Logics, Cuba, 2001, p. 49-68
- Pastre D., *Compétitions de démonstrateurs de théorèmes*, Rubrique « Jeux et compétition en IA », Revue d'intelligence artificielle, 15 n° 2/2001, p. 279-286
- Pastre D., *Strong and weak points of the*

- MUSCADET theorem prover*, AI Communications, 15 (2002), p. 147-160
- Spagnol J-P., *Automatisation du raisonnement et de la rédaction de preuves en géométrie de l'enseignement secondaire*, Thèse, Université René Descartes, 2001
- Spagnol J-P., *ARGOS, un démonstrateur de théorèmes en géométrie*, Sciences et techniques éducatives, vol. 8 n° 1-2, p. 113-125, EIAO'2001
- Spagnol J-P., Poster : *Modelisation and Automation of Reasoning in Geometry. The ARGOS System: a Learning Companion for High-School Pupils*, Intelligent Tutoring Systems, 2002, Biarritz,
- Spagnol J-P., *Modélisation et automatisation du raisonnement en géométrie, le système ARGOS*, Treizième Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Angers, 2002
- Hibou M., Labat J-M. et Spagnol J-P., *Génération de feuilles d'exercices de géométrie à l'aide d'énoncés indexés automatiquement*, conférence EIAH 2003 (Environnements Informatiques d'Apprentissage Humain), Strasbourg, avril 2003
- Spagnol J-P., *The ARGOS System: a Learning Companion for High-School Pupils*, workshop on Knowledge Representation and Automated Reasoning for E-Learning Systems (KRR-5), workshop of IJCAI2003, Acapulco, Mexique, August 2003

## Conception de méthodes et d'outils pour la gestion de connaissances

(Rym Ben Haddada, Valentina Ceausu, Bernard Delforge, Sylvie Després, Elise Fage, Pierre Vinant)

Les travaux se situent en ingénierie des connaissances et sont principalement organisés autour de la conception de méthodes et d'outils pour capitaliser, partager et créer des connaissances. Trois points clefs autour desquels s'articule la gestion des connaissances.

Il s'agit d'une recherche appliquée dans les domaines de l'accidentologie (accidentologie routière et accidentologie pour l'assurance) et du droit médical. L'enjeu d'une telle recherche est double : traiter de problèmes concrets posés par des acteurs attendant des réponses opérationnelles, tout en contribuant aux avancées de la discipline informatique.

Cette recherche est conduite dans un contexte pluridisciplinaire, en collabo-

ration avec des chercheurs en accidentologie routière, des psychologues cognitivistes, des ergonomes et plus récemment des linguistes, des chercheurs (juristes et médecins) en droit médical et des experts en assurance.

Les problématiques abordées sont l'utilisation de ressources expertes pour l'aide à la formation, la construction d'ontologies à partir de textes et la conception de systèmes coopératifs. Actuellement une réflexion est menée sur la modélisation des connaissances nécessaires à l'adaptation dans le contexte de la conception de système de raisonnement à base de cas.

## - Utilisation de ressources expertes pour l'aide à la formation

La réflexion sur l'exploitation d'une expertise en vue de sa diffusion et son partage a été initiée dans le cadre du développement de systèmes d'aide à la formation (serveur de connaissances à l'INRETS et système d'aide à l'analyse d'accidents). Elle s'est poursuivie dans le cadre d'un projet en accidentologie pour l'assurance.

En accidentologie routière, la forme de l'assistance a été imaginée à partir des modèles de l'activité des chercheurs et des enquêteurs (au sens de l'ergonomie cognitive) et du modèle de la tâche des chercheurs (au sens de l'ingénierie des connaissances). Les modèles de l'activité ont été construits pour comprendre ce que font les opérateurs en situation d'analyse d'accidents, afin de déterminer des propositions d'intervention sur le cours de l'action des enquêteurs et de définir les moments où le recours aux connaissances des chercheurs pourrait intervenir. En accidentologie pour l'assurance, la finalité du travail était l'élaboration d'un compagnon S3A (Système d'Aide à l'Analyse d'Accidents) qui, dans une première étape, permet aux rédacteurs novices de s'affranchir du recours systématique aux experts pour confirmer leurs diagnostics. S3A est un système de raisonnement à base de cas. Les modèles de la tâche et de l'activité des rédacteurs

expérimentés ont permis de dégager les éléments indispensables à la modélisation des cas de S3A.

La démarche adoptée dans le contexte de ces deux applications, encore peu expérimentée, contribue à une avancée dans le domaine de la gestion des connaissances. Elle exploite les connaissances de l'organisation dans les formes adéquates pour aider à leur appropriation par les différents acteurs de l'organisation.

## - Construction d'ontologie à partir de textes

La problématique peut être ainsi formulée : « Comment construire, en collaboration avec les experts, une ontologie d'un domaine à partir de textes et du modèle de l'activité des experts ? ». Le succès d'une telle entreprise dépend en partie de la capacité à surmonter les difficultés liées à la complexité de la langue et des textes. Par conséquent une méthodologie de construction fondée sur une approche linguistique a été choisie. La construction d'une telle ressource est réalisée dans deux domaines d'application ceux de l'accidentologie et du droit médical.

En accidentologie routière, la particularité des scénarios – qui relatent le déroulement d'un événement, résultat d'un enchaînement d'actions – conduit à considérer dans l'étude terminologique, la représentation des actions qui sont décrites par des formes verbales plutôt que des syntagmes nominaux. Cette démarche est reproductible dans le domaine du droit médical, si le modèle est construit à partir des actions menées en justice.

Les connaissances disponibles sur les scénarios se trouvaient dans des textes ou émanaient d'experts. Les textes, le modèle de l'activité et le recours aux experts ont donc joué un rôle central dans ce travail dont l'enjeu était la construction d'un modèle explicite et lisible des scénarios. Aucune ontologie de l'accidentologie n'était alors disponible et donc susceptible d'être réutilisée. A l'heure actuelle la question de

l'enrichissement de l'ontologie est étudiée dans le cadre d'un stage de DEA qui est centré sur l'élaboration d'une ressource terminologique établie à partir des procès-verbaux d'accidents piétons. Le logiciel TERMINAE (<http://www-lipn.univ-paris13.fr/~szulman/TERMINAE.html>) et des algorithmes de fouilles de données sont utilisés pour réaliser ce travail.

Dans le domaine du droit médical, une ressource terminologique est construite à partir d'un ensemble de textes de doctrine rédigés par les experts du domaine. Cette ressource est élaborée à des fins de recherche d'information sur le Web selon des profils utilisateurs (juristes, médecins ou novices). L'implémentation au moyen du logiciel TERMINAE est en cours de réalisation dans le cadre d'un stage de magistère. L'ontologie du droit de Guiraudé Lame (<http://ontologie.w3sites.net>) est exploitée pour établir cette ressource terminologique du droit médical.

## - Vers un méta modèle de représentation fondé sur une ontologie

La nécessité d'élaborer une représentation générique des scénarios était primordiale car cette notion est centrale en accidentologie. Une fois le modèle de la tâche d'analyse d'accident construit, la terminologie utilisée en accidentologie identifiée et l'ontologie élaborée, il devenait possible de construire un méta modèle formel du concept scénario. Le terme méta modèle indique qu'il s'agit de la construction d'un modèle de représentation des scénarios qui inclut les modèles du domaine de l'accidentologie. En outre, ce méta modèle est conçu dans un but d'implémentation et doit par conséquent répondre aux exigences inhérentes à l'exploitation des données par un ordinateur. Un langage formel est en cours d'élaboration. Il s'appuie sur une ontologie des relations, permet de décrire les relations spatiales, de transcrire la logique temporelle et d'expliquer la logique causale du déroulement de l'accident.

## Publications

- Després S., *Modélisation de représentations mentales pour une aide aux enquêteurs en accidentologie*, RFIA, Paris, 2000
- Després S., Delforge B., *Designing medical law ontology from technical texts and core ontology*, Workshop de la conférence EKAW Ontology and Texts, 2000
- Després S., Wolff M., *Modélisation et formalisation de l'activité cognitive de l'expertise des accidents : Vers un développement d'un Système d'Aide à l'Analyse des Accidents (S3A)*, rapport final du projet cognitique, rapport intermédiaire du projet cognitique n°21, 2001.
- Després S., *Une comparaison raisonnée des apports de la terminologie et de l'intelligence artificielle pour servir et améliorer la construction d'ontologies*, TIA-2001, inist, Nancy, 2001
- Després S., *Exploitation d'une expertise en accidentologie pour l'aide à l'Etude Détaillée d'Accidents*, rapport interne, 2002
- Després S., Wolff M., *Modélisation et formalisation de l'activité cognitive de l'expertise des accidents : Vers un développement d'un Système d'Aide à l'Analyse des Accidents (S3A)*, rapport final du projet cognitique n°21, 2002
- Després S., *Construction d'ontologies à partir de textes*, rapport interne, 2002
- Després S., *Contribution à la conception de méthodes et d'outils pour la gestion des connaissances*, Habilitation à Diriger des Recherches en Informatique, Université René Descartes, 2002
- Wolff M., Després S., *Raisonnement à partir de cas : Modélisation et formalisation de l'activité cognitive de l'expertise des accidents*, In J.C. Spérandio & M. Wolff (eds.), *Formalismes de modélisation pour l'analyse du travail et l'ergonomie*, Paris, PUF 2003

## Etude cognitive du raisonnement mathématique, découverte et représentations

Dominique Pastre et Sylvie Després

Découvrir une solution, avoir de bonnes heuristiques, utiliser de bonnes représentations, avoir les bonnes idées, créer de bons intermédiaires (objets ou problèmes) sont des points cruciaux en raisonnement mathématique ou plus généralement en résolution de problèmes. Depuis longtemps, des observations ou des auto-observations ont été faites pour essayer d'analyser le raisonnement humain. Outre le fait de mieux comprendre le raisonnement humain, l'intérêt de tels travaux est de dégager

des techniques et des heuristiques qui pourraient améliorer le comportement des systèmes d'intelligence artificielle.

Deux travaux ont été effectués récemment dans cet esprit :

- Des auto-observations ont été effectuées pour la résolution d'une quinzaine de problèmes mathématiques ou de type « casse-tête ». On a en particulier observé qu'il est rare d'aller droit au but, mais que plusieurs problèmes ont été résolus par des chemins détournés, après de nombreux essais, tâtonnements, dessins, cas particuliers, parfois même manque de rigueur, la solution rigoureuse et élégante n'étant déduite qu'après. Ceci est loin du comportement des systèmes actuels. On a aussi observé que la manipulation des divers objets intervenant dans un problème est essentielle, et qu'une idée, même fautive, permet de manipuler, donc de progresser et de trouver d'autres idées qui s'avéreront bonnes.
- Une étude a été faite sur l'utilisation des dessins par l'homme et par quelques systèmes d'intelligence artificielle, dans le cadre d'un hommage à Herbert Simon. Ce travail a été complété par l'analyse d'une expérimentation proposée aux participants d'un colloque d'Intelligence Artificielle. Divers problèmes étaient présentés, soit uniquement par du texte pour un groupe, soit avec l'aide de dessins pour l'autre groupe. Les résultats de cette expérience ont montré l'aide que les dessins peuvent apporter, mais aussi leurs limites et une grande variabilité selon les individus.

D'autre part, en accidentologie routière, une étude a été réalisée sur le rôle des dessins (schémas) associés aux représentations des scénarios d'accidents. La sémantique de ces dessins est centrée sur la manœuvre effectuée par l'utilisateur et contrairement aux représentations textuelles la variabilité entre les experts est faible. Le dessin est complémentaire des représentations textuelles et joue un rôle essentiel dans le raisonnement de celui qui analyse l'accident.

## Publications

Pastre D., *Chemins détournés, idées fausses et bonnes idées*, Colloque Intelligence Artificielle, Berder, 2000, rapport LIP6 2001/014, p. 77-114

Pastre D., *Utilisation de dessins en résolution de problèmes*, Revue d'intelligence artificielle, 16/2002, p. 123-164

Després S., *Contribution à la conception de méthodes et d'outils pour la gestion des connaissances*, Habilitation à Diriger des Recherches en Informatique (Chapitre 2), Université René Descartes, 2002

## Groupe SAFE : Système d'Assistance à la Formation et l'Éducation

*Enseignants-chercheurs* : Michel Futersack, Jean-Marc Labat, Françoise Le Calvez, Yannick Parchemal, Marie Urtasun et, à partir de septembre 2003, Elisabeth Delozanne

*Doctorants* : Younès Abdelouahed, Nikolay Georgiev, Mathieu Hibou, Lahcen Oubahssi, Issam Rebaï, Ossian Rogé

Les systèmes informatisés de formation, ou plutôt d'aide à la formation tant la place de l'enseignant humain reste indispensable et même prépondérante, nécessitent de concevoir des systèmes interactifs qui prennent en compte l'humain dans toute sa complexité sur le plan cognitif. Ce thème est donc fortement pluridisciplinaire par nature. Sur le plan de l'Informatique, IA et IHM sont les deux domaines au cœur des problématiques de ce groupe. En IA, ce sont essentiellement les approches symboliques qui sont étudiées car les modèles doivent être intelligibles par les humains, enseignants et/ou apprenants. Par ailleurs, l'expérimentation dans le domaine des EIAH (Environnement Informatiques d'Apprentissage Humain) nécessite de réaliser des artefacts qui soient, au moins sur un domaine restreint, des produits finis. Cela nécessite que les équipes aient une taille suffisante. C'est pourquoi, au sein du Réseau Thématique Pluridisciplinaire « Ap-

prentissage, Education et Formation » (RTP39 du département STIC du CNRS), une équipe-projet a été créée sur la région parisienne (AIDA : Approche Interdisciplinaire pour les Dispositifs informatisés d'Apprentissage, <http://www.math-info.univ-paris5.fr/AIDA>). Cette équipe est constituée de chercheurs en Informatique de Paris 6 et Paris 5, en psychologie cognitive de Paris 8 et en Didactique des Mathématiques de Paris 7. Enfin, l'équipe AIDA est membre du réseau d'excellence européen Kaléidoscope.

Les membres de ce groupe s'intéressent plus particulièrement aux problématiques suivantes :

### *Outils pour l'utilisation de ressources numériques*

Lahcen Oubahssi, Jean-Marc Labat et Issam Rebaï

La mise en commun et l'échange de composants logiciels libres et réutilisables est de nature à faciliter la mise en synergie des travaux de l'ensemble de la communauté française de recherche. Le groupe contribue à la conception d'une plate-forme de mutualisation permettant aux concepteurs d'outils et d'environnements de déposer les composants qu'ils ont conçus (thèse en cours d'I. Rebaï). Le premier objectif est de construire un entrepôt de composants accessible via le Net avec des outils d'indexation et de recherches personnalisable. La définition d'un jeu de métadonnées, comportant à la fois des informations générales, fonctionnelles et techniques, permet d'effectuer des recherches fines sur tous les aspects du composant et plus particulièrement sur les aspects fonctionnel et sémantique. Ce jeu de métadonnées est inspiré des standards actuels comme la LOM (Learning Object Metadata) ou l'OSD (Open Software Description). Un second objectif est de construire une plate-forme d'assemblage de composant suivant l'architecture MDA (Model Driven Architecture).

## *Nouvelles fonctionnalités pour les environnements d'apprentissage informatisés*

Michel Futersack, Jean-Marc Labat et Ossian Rogé

Les situations de formation qui font appel aux TIC combinent souvent un travail autonome effectué en présence ou à distance et un travail, individuel ou collectif, partiellement encadré. Dans ces situations, il convient de fournir au tuteur humain les informations nécessaires à un soutien et à une intervention efficaces. Mais ces informations doivent être, pour une large part, de nature qualitative et être fournies de manière synthétique afin de ne pas noyer l'enseignant sous une masse d'informations (thèse de O. Rogé). L'objectif est de concevoir un composant domaine indépendant et respectant les standards internationaux en cours de définition.

Dans le cadre du projet SPEP (Système Pédagogique d'Entraînement de Peloton, projet soutenu par TT&S - Thomson Training & Simulation- et dirigé par M. Joab), la problématique à laquelle on s'attache est celle de la formation d'une équipe (« team training ») : il s'agit non seulement d'évaluer les compétences individuelles de chaque membre de l'équipe, mais aussi les compétences inter-individuelles (qualité de la communication et de la collaboration). La finalité du projet est de réaliser un ensemble d'outils pour aider les instructeurs conduisant une formation sur simulateur (simulateurs de char Leclerc). Un système d'aide à l'évaluation des compétences techniques et tactiques d'un peloton de chars Leclerc a été réalisé et validé par les instructeurs militaires. Ce système se greffe sur un environnement de simulation distribuée et utilise un système à base de connaissances implanté en JESS (Java Expert System Shell). En aval de ce système a été réalisé un logiciel de génération de rapport de débriefing interactif qui permet à l'instructeur de présenter à ses stagiaires le résultat de l'exercice suivant différentes perspectives : l'ana-

lyse du comportement de sauvegarde, de l'agressivité, etc.

## **Publications**

Joab M., Auzende O., Futersack M., Le Leydour P., *Structuring a simulation exercise database using STREX*, ITEC 2000, Amsterdam

Joab M., Auzende O., Futersack M., Bonnet B., Le Leydour P., *Computer Aided Evaluation of Trainee Skills on a Simulator Network*, Intelligent Tutoring Systems 2002, ITS 2002, p. 521-530, Biarritz, 2002.

## *Architecture de tuteurs intelligents incluant des activités de résolution de problèmes*

Mathieu Hibou, Michel Futersack, Jean-Marc Labat, Françoise Le Calvez et Marie Urtasun

Dans l'approche constructiviste de l'apprentissage que le groupe souhaite privilégier, la modélisation d'activités de résolution de problèmes est un aspect essentiel pour développer l'interactivité entre apprenants et systèmes logiciels. Le groupe poursuit son travail sur l'utilisation de la résolution de problèmes dans les EIAH. Deux projets sont en cours : le projet Combien? (voir paragraphe suivant) et le projet Simplus (appel d'offre RIAM).

Ce dernier projet est réalisé en collaboration avec la PME Exosim et le centre de l'Ecole Nationale Supérieure des Mines de Saint-Etienne. Au sein du projet (développement d'un environnement pédagogique s'appuyant sur un simulateur), l'objectif est d'élaborer une méthode pour concevoir des joueurs virtuels et des explications à partir de l'expertise embarquée dans le simulateur. Cette méthode s'appuie sur le développement de joueurs virtuels et d'explications sous forme de systèmes à base de connaissances construits à partir de l'expertise embarquée dans le moteur de simulation des jeux.

Enfin, en collaboration avec J-P. Spagnol (du groupe SBC), un générateur de feuilles d'exercices a été réalisé en s'appuyant sur ARGOS, son résolveur de problèmes de géométrie, basé

sur la déduction naturelle et l'utilisation de connaissances et métaconnaissances. Les exercices résolus par ARGOS sont indexés automatiquement par les théorèmes utilisés et l'apprenant ou le professeur peut demander au système de lui fournir des exercices sur un thème donné.

## **Publications**

Futersack M., Labat J-M.: *Quelle planification pédagogique dans les EIAH ?*, Revue STE, vol 7, numéro spécial « Education et Informatique, hommage à Martial Vivet » Ed. Hermes, 2000, p. 165-177

Labat J-M., EIAH : *Quel retour d'informations pour le tuteur ?*, Conférence TICE 2002, p. 81-87, 2002

Hibou M., Labat J-M., Spagnol J-P., *Automatically generated exercise pages*, AIED 03 workshop, Sydney, 2003

Hibou M., Labat J-M., Spagnol J-P., *Génération de feuilles d'exercices de géométrie à l'aide d'énoncés indexés automatiquement*, congrès EIAH 2003, p. 247-258

## *Méthodologie de conception*

Françoise Le Calvez et Marie Urtasun

Le groupe Combien? comprend trois autres enseignants-chercheurs de Paris 6 et/ou professeurs de mathématiques. Une description complète de ses travaux et publications se trouve à l'adresse : <http://www.math-info.univ-paris5.fr/combien>.

Le projet Combien? a pour but de définir une méthodologie de conception de différents composants d'un EIAH. Pour valider ses réflexions, il réalise un système pédagogique d'aide à l'apprentissage humain dans le domaine mathématique des dénombrements. L'objectif pédagogique n'est pas de former des experts en combinatoire mais d'entraîner les étudiants à la modélisation et de les rendre capables de représenter une situation par une structure complexe. Des réflexions théoriques et des expériences d'implémentation ont été menées de front.

Les travaux antérieurs à 2000 ont permis de définir un modèle conceptuel objet du domaine sur lequel ont été créées des machines à construire des

configurations (éléments à dénombrer, répondant aux contraintes demandées). Ces machines permettent aux élèves de construire, en utilisant la méthode constructive qui a été définie, une configuration solution du problème et à partir de cette construction de compter le nombre d'éléments de l'ensemble-solution.

Les travaux récents du groupe Combien? ont concerné le diagnostic et les explications. Les réflexions ont porté sur la gestion des erreurs dans une interface pédagogique. Les questions qui se posent sont : quelles sont les erreurs « intéressantes » pour l'apprentissage et qui doivent être permises à l'élève, quelles sont celles que l'interface ne permettra pas à l'élève de faire car elles ne relèvent pas d'un processus pédagogique, à quel moment les erreurs détectées doivent-elles être présentées à l'élève, sous quelle forme et avec quels enseignements. A partir de ces réflexions, une méthode de conception des interfaces à partir des erreurs permises a été proposée. Un modèle de détection incrémentale d'erreurs, la « détection ciblée » a été conçu. La détection ciblée utilise le fait que l'environnement de recherche des erreurs est une représentation conceptuelle structurée en arborescence englobant solution de l'exercice et construction en cours de l'élève. Ce mécanisme de détection d'erreurs a été implémenté dans les machines à construire les configurations. Les connaissances nécessaires à la détection étant organisées sous forme de base de schémas d'erreur.

Pour pouvoir introduire des connaissances de natures diverses dans le système, un langage de description de ressources, DESCRIPT, a été conçu et implémenté. Un stage a été réalisé pour définir une meta-grammaire permettant de décrire les grammaires définissant chaque ressource et ainsi avoir un outil très général.

Actuellement quatre machines sont opérationnelles. Elles sont conçues pour être un monde « complet » ; elles apportent tous types d'aide (utilisateur,

conceptuelle, explications) de façon contextuelle. En particulier le « cours » traditionnel se trouve distribué aux endroits qui semblent les plus propices pour l'élève. Ces machines ont été validées avec des élèves de terminale (Lycée Carnot, Paris ; Lycée Félix Faure, Beauvais), avec des étudiants en préparation au CAPES de mathématiques (Paris 6), avec des étudiants en DEUG MIAS 2ème année option mathématiques discrètes (Paris 6) et avec des collègues de mathématiques.

Une réflexion sur la conception des EIAH à la lumière de ces travaux a été menée. Un certain nombre de leçons pour l'ingénierie des EIAH en ont été tirées.

Plusieurs voies sont ouvertes pour la poursuite de ces travaux :

- réaliser de nouvelles machines pour qu'elles couvrent les différentes classes de problèmes de dénombrement généralement enseignées ;
- transposer ces réflexions dans un autre domaine en utilisant les mêmes techniques, le travail est avancé concernant les isométries planes conservant une figure ;
- proposer de nouvelles activités à l'élève pour qu'il puisse tirer profit de l'activité de résolution de problèmes en prenant conscience de ce qu'il a acquis comme concepts mathématiques.

## Publications

Tisseau G., Giroire H., Le Calvez F., Urtasun M., Duma J., *Design of a learning Environment in combinatorics : Nondeterministic machines to improve modelling skills*, in proceedings of AI-ED 99, Le Mans, 1999

Tisseau G., Giroire H., Le Calvez F., Urtasun M., Duma J., *Spécification du dialogue et génération d'interfaces à l'aide d'interacteurs à réseau de contrôle*, IHM 99, Montpellier, 1999

Tisseau G., Giroire H., Le Calvez F., Urtasun M., Duma J., *Principes de conception d'un système pour enseigner la résolution des problèmes par la modélisation*, RFIA'2000 p. 121-130, Paris, 2000

Duma J., Giroire H., Le Calvez F., Tisseau G., Urtasun M., *Spécification de dialogues et construction d'interfaces modulaires et réutilisables*, Actes du colloque d'IA :

Apprentissage et Acquisition des connaissances, Berder 1999, Kocik F. et Pannérec T. Eds, Rapport interne LIP600, n°2 p.117-129

Tisseau G., Giroire H., Le Calvez F., Urtasun M., and Duma J., 2000, *Design principles for a system to teach problem solving by modelling*, ITS'2000, Montréal, 2000, Lecture Notes in Computer Science N°1839, p. 393-402

Duma J., Giroire H., Le Calvez F., Tisseau G., Urtasun M., *Gestion des erreurs dans une interface pédagogique*, Colloque Intelligence Artificielle, Berder, p. 22-37, 2000, rapport interne LIP6, 2001/014

Giroire H., Le Calvez F., Tisseau G., Duma J., Urtasun M., *Un mécanisme de détection incrémentale d'erreurs et son application à un logiciel pédagogique*, RFIA'2002, Angers, 2002, p. 1063-1072

Giroire H., Le Calvez F., Tisseau G., Duma J., Urtasun M., *Targeted Detection: Application to Error Detection in a Pedagogical System*, ITS'2002, Biarritz, 2002, p. 998

Giroire H., Le Calvez F., Tisseau G., Duma J., Urtasun M., *Un logiciel pour apprendre à résoudre des exercices de dénombrement*, TICE 2002, Lyon, 2002, p. 369

Tisseau G., Giroire H., Le Calvez F., Duma J., Urtasun M., *Combien?, un EIAH pour les dénombrements : expérimentations et leçons pour l'ingénierie*, EIAH 2003, Strasbourg, Avril 2003, p. 509-516, et <http://www.archive.eiah.org/EIAH2003>

*Article interactif EIAH 2003* : <http://www.archive.eiah.org/EIAH2003/DemoEx06/DebutArticleCombien.html>

Le Calvez F., Giroire H., Duma J., Tisseau G., Urtasun M., *Combien? a Software to Teach Students How to Solve Combinatorics Exercises*, Workshop « Advanced Technologies for Mathematics Education » AIED 2003, In Supplementary proceedings of the 11th International Conference on Artificial in Education, p. 447-454, Sydney, July 2003

## Suivi et modèles des connaissances de l'apprenant

Mathieu Hibou et Jean-Marc Labat

La modélisation de l'apprenant, est un thème de recherche particulièrement difficile sur lequel la communauté internationale bute depuis longtemps. Il n'en reste pas moins que cette modélisation est indispensable, au moins sur certains aspects si on veut mettre au point des environnements prenant réellement en compte les apprenants qui les utilisent. Un travail a été commencé sur des modèles utilisant les réseaux bayésiens dans le cadre de la thèse de M. Hibou,

avec comme objectifs d'étudier les spécificités et la validité de tels formalismes pour la représentation des connaissances ou de l'activité cognitive d'un sujet. En effet, la construction du réseau soulève de nombreuses questions. Dans le cas où ce sont des experts qui le déterminent, leurs choix et leurs décisions ne sont pas toujours clairement explicités ou justifiés. Dans les approches utilisant des données pour élaborer le réseau, et notamment sa structure, on peut s'interroger sur la pertinence d'une modélisation se limitant à des observables directs. De plus, si des techniques d'apprentissage machine sont parfois utilisées pour modifier les probabilités a priori du réseau en fonction du comportement de l'apprenant, il n'est jamais envisagé que la structure des connaissances puisse varier d'un individu à l'autre ou au cours du temps lors d'une phase d'apprentissage.

## **Ingénierie des connaissances**

Younès Abdelouahed, Nikolay Georgiev et Jean-Marc Labat

En partenariat avec une PME, 12-Planet, une recherche portant sur la conception et le développement d'un service interactif de réponses à des questions posées en langue naturelle a démarré (thèse de Y. Abdelouahed). Il ne s'agit pas de travailler sur la compréhension des langues naturelles mais d'intégrer l'agent logiciel ALICEBOT (le gagnant en 2000 de la compétition portant sur le fameux test d'intelligence de Turing) au sein d'un système innovant.

Dans le cadre de l'étude du comportement des internautes, il devient important de pouvoir analyser finement les contenus lus et les actions effectuées sur le Web. Le Web étant à la fois un mode de réception d'information et un moyen d'action (communication, achats, etc.), les contenus donnent accès non seulement aux centres d'intérêt, mais aussi au comportement. Couplées à la mesure du média Internet, ces analyses permettront

de tirer des conclusions qui intéresseront à la fois le monde de la recherche et l'industrie. Cette recherche se fait en partenariat avec une entreprise A.C. Nielsen Netratings, et J-L. Minel de l'université Paris 4, (bourse CIFRE de N. Georgiev).

## **Publication**

Georgiev N., Labat J-M., Minel J.L., Nicolas L., *Extraction de données à partir de pages HTML par création semi-automatique de règles XSLT*, congrès IC 2003

## **Enseignement à distance**

Yannick Parchemal

Les recherches effectuées, appliquées à l'enseignement à distance sont décrites et accessibles à l'adresse <http://des-cartes.math-info.univ-paris5.fr:8080/plenadis>.

Le premier objectif est la réalisation d'une plate-forme ergonomique de travail collaboratif avec cours et tutorat en ligne. Plenadis est une plateforme d'enseignement à distance, offrant aux étudiants et aux tuteurs l'environnement souhaitable avec des cours en ligne, des forums, des salons de discussions et des tableaux interactifs.

Le deuxième objectif est le suivi des étudiants et l'évaluation automatisée des enseignements. Le logiciel Questel permet l'édition en ligne des questionnaires, un envoi et un traitement des résultats automatisés.

Questel est utilisé depuis 2001 à l'université René Descartes et une expérimentation de Plenadis est prévue en 2003/2004 dans le cadre d'enseignements à distance de l'université.

## **Publication**

Parchemal Y., Ycart B. *Evaluation en ligne des enseignements : le logiciel Questel, Sciences et techniques éducatives*, rubrique « Evaluation en ligne », volume 9 n°3-4/2002, p. 491-499

## **Groupe LIAP5 : ex Laboratoire d'Intelligence Artificielle de Paris 5**

*Enseignants-chercheurs* : Norbert Cot, Claude Lattaud

*Doctorants* : Vincent Airault, Elias O'Regan, Marc Métivier, Stéphane Sikora

Les recherches menées au LIAP5 s'orientent suivant deux axes focalisés sur l'évolution artificielle : les aspects théoriques d'une part et les aspects expérimentaux d'autre part. La théorie couvre en particulier l'étude des automates cellulaires, et plus spécialement celle du « Jeu de la Vie » de John Conway, de leur généralisation dans un environnement continu et des similarités entre plusieurs règles.

D'un point de vue expérimental, l'étude porte principalement sur les phénomènes de co-évolution dans des écosystèmes artificiels ainsi que sur l'apprentissage par imitation via des techniques évolutionnaires.

L'axe central du premier thème est fondé sur l'évolution de communautés de plantes artificielles et les interactions entre elles, tant du point de vue compétition que coopération, et collectif qu'individuel. L'objectif est de pouvoir déterminer par simulation l'évolution de ces communautés de plantes sur plusieurs générations. Ce projet se fonde essentiellement sur des données issues de la biologie et de la bioclimatologie et ses expérimentations sont validées par comparaison avec des résultats obtenus dans la nature. L'étude des dynamiques d'évolution de populations d'agents mobiles est également un pôle de recherche de l'équipe, et ceci en collaboration avec l'Institut International du Multimédia.

Le second thème concerne l'élaboration de méthodes d'apprentissage par imitation dans des environnements virtuels en utilisant conjointement des algorithmes issus de l'apprentissage et de l'évolution artificielle.

## Publications

Magnier M., Lattaud C. & Heudin J.C., *Complexity classes in the Two-dimensional Life Cellular Automata subspace*, in *Complex Systems* vol. 11:6, 2000, p. 419-436

Sikora S., Steinberg D. & Lattaud C., *Integration of simulation tools in virtual worlds: towards virtual ecosystems*, in *Proceedings of the 2nd International Conference on Virtual Worlds*, 2000, p. 19-25

Métivier M. & Lattaud C., *Modélisation du comportement à base de systèmes de classifieurs durant un processus d'avatarisation*, in *Extraction des connaissances et apprentissage* vol. 1:3, Hermès, 2001, p. 61-85

Métivier M. & Lattaud C., *Anticipatory Classifier System using Behavioral Sequences in Non-Markov Environments*, in *Proceedings of the 5th International Workshop on Learning Classifier Systems*, Stolzman W. & Lanzi P.L. ed., 2002

Métivier M., Lattaud C. & Heudin J.C., *A stress based speciation model in LifeDrop*, in *Proceedings of the 8th International Conference on Artificial Life*, Sydney, Australia, 2002

Lattaud C., *Long-term competition for light in plant simulation*, in the 6th Genetic and Evolutionary Computation Conference GECCO-2003 (à paraître)

Lattaud C., *Co-Evolution in Artificial Ecosystems : Competition and Cooperation using Allelopathy*, in the 6th International Conference on Artificial Evolution, 2003 (à paraître).

## Les formations de troisième cycle

L'équipe IAA participe à l'encadrement de deux DEA et d'un DESS :

- le **DEA IARFA** (Intelligence Artificielle et Reconnaissance des Formes, école doctorale EDITE, cohabilité Paris 6, Paris 5, ENPC,

ENSTA, ENST Paris, Paris 5 est responsable des filières « Vie Artificielle » et « Reconnaissance des Formes »)

- le **DEA MIASH** (Mathématiques et Informatique Appliquées aux Sciences de l'Homme, école doctorale « Cognition, comportement et conduite humaine », co-habilité Paris 1, Paris 4, Paris 5, EHESS, ENST Bretagne, Paris 5 est responsable de la filière « Interaction Homme-machine, Intelligence Artificielle et Internet », <http://www.math-info.univ-paris5.fr/Enseignements/cycle3/filieresiii.html>)
- le **DESS IASV** (« Informatique et Applications aux Sciences de la Vie », <http://www.math-info.univ-paris5.fr/DESS>)

## Contrats

Bourses CIFRE : 12 Planet, A.C. Nielsen NetRatings, A6 Média Guide  
 Projet RIAM avec Exosim  
 Projet Cognitique (1999) et valorisation industrielle (2002)  
 Projet PREDIT-GO3 (2003)

## Collaborations

Lip6 (Paris 6), Didirem (Paris 7), Cognition et Actions finalisées (Paris 8), INRP  
 Réseau d'excellence européen Kaleidoscope  
 LEI (Laboratoire d'Ergonomie Informatique, Paris 5)  
 INRETS, département MA

(Mécanismes d'accidents) de Salon de Provence

Accord de collaboration avec Cybésosphère

## Logiciels

*Disponibles sur Internet :*  
 Indigo (<http://www.math-info.univ-paris5.fr/~bouzy/INDIGO.html>)  
 MUSCADET (<http://www.math-info.univ-paris5.fr/~pastre/muscadet>)  
 S3A (mise en ligne en attente de l'autorisation de la société KAIDARA)  
 Plenadis et Questel (<http://descartes.math-info.univ-paris5.fr:8080/plenadis>)

*En démonstrations sur Internet :*  
 ARGOS ([http://www.ac-amiens.fr/lycee60/faure\\_beauvais/fr/dossiers/math/Argos/sf136ex1.htm](http://www.ac-amiens.fr/lycee60/faure_beauvais/fr/dossiers/math/Argos/sf136ex1.htm))

SWI-prolog n'étant pas installé sur le serveur du rectorat d'Amiens qui héberge ces pages la solution n'est pas obtenue dynamiquement mais est pré-enregistrée pour un seul exercice.

*Combien?* (<http://www.math-info.univ-paris5.fr/~lecalvez/combien/machineCE.html>)

*Participation à des compétitions internationales :*

Indigo (21<sup>st</sup> century cup 2002, <http://www.intelligentgo.org/en/igf/21cc2002>)

MUSCADET (CADE ATP System Competitions, <http://www.cs.miami.edu/~tptp/CASC>)

## LAMIH : Laboratoire d'Automatique, de Mécanique et d'Informatique, Industrielles et Humaines

*Université de Valenciennes et du Hainaut-Cambrésis,*

**Contacts : C. Kolski, R. Mandiau, S. Piechowiak**

Nos activités de recherche se déroulent au sein du Laboratoire d'Automatique, de Mécanique et d'Informatique, Industrielles et Humaines - LAMIH, de l'Université de Valenciennes et du Hainaut-Cambrésis, dirigé par le Prof. Patrick Millot (Directeur) et le Prof. Jérôme Oudin (Directeur Adjoint).

À partir de 1998, suite à une réorganisation du LAMIH, nos travaux se sont inscrits dans le groupe de recherche RAIHM – Raisonnement Automatique et Interaction Homme-Machine, dirigé par le Prof. C. Kolski. Les recherches portent essentiellement sur les deux thèmes suivants : d'une part, méthodes formelles et semi-formelles en conception et évaluation de systèmes interactifs, et, d'autre part, le raisonnement distribué ou centralisé.

Dans le cadre de cette présentation, nous développerons les deux derniers thèmes. Les permanents associés à ce thème sont plus particulièrement : René Mandiau (Pr.), Sylvain Piechowiak (Pr.), Emmanuel Adam (Mdc), Véronique Delcroix-Froidure (Mdc), Emmanuelle Grislin-Le Strugeon (Mdc).

### **1. Intégration d'agents dans les processus administratifs**

Cette recherche consiste à proposer des principes d'aide à la gestion commune d'informations (le plus souvent des documents) au sein d'une organisation de type administrative. L'intérêt de cette recherche se justifie par des organisations possédant une structure pyramidale et mettant en jeu un nombre

important d'acteurs qui ont des rôles différents et des degrés de responsabilité propres. Dans un système administratif, les activités consistent essentiellement en la création, la gestion et la distribution d'informations effectuées par l'ensemble des acteurs. L'état de l'art du domaine fait aussi ressortir un manque évident de méthodes dans la conception de systèmes multi-agents, dans un contexte industriel. Une méthode fondée sur des approches du génie logiciel, appelée MAMOSACO pour « Méthode Adaptable de MODélisation de Systèmes Administratifs Complexes », est proposée pour de telles applications. Son originalité est de permettre la représentation des notions de responsabilité ainsi que de coopération entre les acteurs de l'organisation. Ce travail en liaison avec la société Solvay (Bruxelles) a fait l'objet de la thèse d'E. Adam (soutenue en septembre 2000).

Afin de poursuivre ces travaux, une validation de MAMOSACO est proposée dans un autre contexte applicatif, la veille technologique. À ce sujet, un outil fondé sur l'intégration d'agents permettant d'améliorer la recherche d'informations est en phase de validation.

### **2. Approche multi-agents pour la Régulation de trafic de bus**

L'objectif de cette étude vise à améliorer la qualité des transports en fournissant une aide à la régulation de trafic de bus. Elle modélise l'application sous forme d'une organisation multi-agents. Ce processus permet de compenser les incidents du réseau (se traduisant par des retards) en effectuant certaines manœuvres sur les autobus. La régula-

tion étant complexe, elle est réalisée par un opérateur humain qui travaille essentiellement par expérience (le processus de régulation n'est pas formalisé). Afin de répondre aux besoins du régulateur et à la satisfaction des usagers, un outil d'aide a été développé.

Une approche originale de modélisation du réseau a été envisagée en considérant que tout l'environnement est constitué d'agents simples regroupés dans une organisation en fonction des flux d'informations. Les principes décrits précédemment ont fait l'objet d'une application liée à la CTPM (Compagnie des Transports du Pays de Montbéliard). Ce travail a conduit à une étude similaire pour la région valencienne (Sté Semurval). La thèse soutenue en décembre 2002 de H. Laichour (codirigée par Pr. S. Maouche, Lille1). L'originalité de la thèse réside dans la prise de décision des agents et dans la généralisation des travaux de régulation sur les deux sites étudiés (la CTPM à Montbéliard et la Semurval à Valenciennes).

### **3. Approche théorie des jeux appliquée au trafic routier**

Une étude en cours vise à modéliser des interactions conflictuelles sous forme d'un modèle mathématique inspiré de la théorie des jeux. Cette modélisation est appliquée à la simulation de trafic routier. Dans ce contexte, un agent humain est plongé dans un environnement de simulation (routier en l'occurrence) mettant en jeu des interactions entre agents artificiels d'une part, agents artificiels/agents humains d'autre part. La modélisation de chaque agent et de

ses comportements s'avère une entreprise délicate, puisque les comportements sont différents (voire totalement égoïstes). Des solutions dans des situations critiques (par exemple des carrefours) ont été proposées et validées. La thèse en cours d'A. Champion (prévue fin 2003) en liaison avec CS-SRILOG et l'INRETS-Arcueil, contribue à cette étude.

#### 4. Comportement émergent de piétons virtuels

La thèse de D. Hanon (débutée en 2002) porte sur une architecture hybride d'agent, dans un but d'animation comportementale de personnages virtuels. Le contexte est différent puisqu'il s'agit d'un nouveau domaine applicatif, faisant l'objet d'un projet PREDIT (projet RESPECT) en collaboration avec différents partenaires dont la société CORYS TESS. Cependant les questions posées restent fondamentalement identiques (architecture, contrôle, interactions). L'objectif global du projet est la création d'un logiciel éducatif à destination des enfants-piétons. Il s'agit d'apprendre aux enfants à se déplacer en sécurité en ville. Le logiciel est donc fondé sur une simulation d'un environnement urbain, comprenant des véhicules et d'autres piétons. Notre équipe est chargée de la modélisation des aspects décisionnels inhérents à ces piétons virtuels "peuplant" la simulation. Chaque piéton virtuel est ainsi piloté par un agent. Le modèle décisionnel des agents-piétons est fondé sur la combinaison de comportements élémentaires.

#### 5. Approche basée sur les CSP Distribués

La modélisation des organisations multi-agents a conduit à la gestion de nombreuses informations pouvant être conflictuelles et distribuées entre les différents agents. La nécessité de conserver une cohérence informationnelle entre ceux-ci est évidemment cruciale. Dans ce cadre, une application est décrite

te sous formes de contraintes à relâcher afin d'obtenir une solution. L'originalité consiste à décrire un mécanisme original de contrôle des relaxations de contraintes. Ceci fait l'objet de la thèse de J. Ma, depuis septembre 2001.

#### 6. Architecture logicielle appliquée au système d'information

La thèse de C. Petit-Rozé (en collaboration avec l'INRETS et la région Nord-Pas de Calais ; soutenance prévue fin 2003) propose une architecture de système d'information à base d'agents, et une méthode de personnalisation de l'information délivrée. Le système multi-agents proposé comporte quatre équipes d'agents aux compétences et fonctions distinctes. Les agents interagissent en vue de répondre au mieux aux requêtes des utilisateurs, à la fois en terme d'efficacité (temps de réponse) et de qualité (adéquation à la demande de l'utilisateur). Le résultat fourni est personnalisé en fonction des connaissances que les agents possèdent sur l'utilisateur (habitudes, préférences, handicaps, etc.). Le cadre applicatif est celui de l'information aux usagers des transports terrestres : préparation et suivi d'itinéraire multi-modes de transports.

#### 7. Architecture multi-agents pour la conception et l'évaluation de systèmes interactifs

Dans le cas d'un système interactif, une conception à base d'agents conduit à s'interroger sur la façon dont le système sera perçu et utilisé par ses utilisateurs. En effet, l'interaction homme-système multi-agents présente des caractéristiques spécifiques qui méritent une attention particulière de la part du concepteur.

Par ailleurs, les agents intelligents semblent constituer une base intéressante dans la conception d'interfaces homme-ordinateur capables de s'adapter à l'utilisateur du système. Dans un contexte multi-utilisateurs, l'adaptation

du système est déterminée par un ensemble d'éléments supplémentaires : cohérence des niveaux d'adaptation entre les différents outils (dont des agents) à disposition ; partage des modèles utilisateurs entre les agents intelligents (ce qui implique un partage des connaissances et des mécanismes de modélisation) ; prise en compte de l'autonomie des agents dans le partage de tâches. Par conséquent, notre objectif est de composer un système d'agents intelligents dont la tâche globale est de fournir une interface adaptative à l'utilisateur, qui respecte à la fois : a) les critères globaux usuels d'adaptativité des interfaces ergonomiques et les buts globaux du système, b) les critères locaux d'adaptativité et les buts locaux de chaque agent intelligent.

Une des approches que nous proposons, consiste à implémenter le composant de dialogue de l'architecture d'IHM ARCH par un agent intelligent. L'interface est ainsi contrôlée par un agent et la représentation globale du système hommes-ordinateurs est fondée sur un modèle orienté agents.

#### 8. Diagnostic à bases de modèles

Le problème est celui du diagnostic à base de modèles en représentant chaque modèle par un réseau de contraintes. Chaque contrainte est décrite par un ensemble de règles de résolutions qui définissent des propagations possibles en fonction de conditions sur les variables en jeu. En appliquant différents algorithmes de renforcement de la consistance on recherche les conflits. Une relaxation d'un ensemble de contrainte qui conduit à résorber l'inconsistance constitue un diagnostic. Cette technique est étendue au cas des intervalles de valeurs ainsi qu'aux valeurs datées permettant ainsi le traitement de dispositifs évoluant dans le temps.

Une des difficultés réside dans l'élaboration des modèles du fait de leur

taille. Une contribution a été apportée sur le thème de l'aide à l'élaboration et à la validation des modèles dans le cadre de la thèse F. Lefebvre (soutenue en juin 2000) en collaboration avec l'INRETS et CS-Transport.

Cette contribution concerne la méthodologie d'élaboration d'un modèle. L'objectif à atteindre est d'élaborer un modèle traduisant fidèlement les propriétés du dispositif réel et permettant de calculer des diagnostics compatibles avec le niveau d'abstraction des observations du comportement réel. Le travail de F. Lefebvre a montré que l'existence d'un modèle de conception est bénéfique pour l'élaboration du modèle pour le diagnostic et que ce dernier s'obtient au travers un cycle complexe qui s'opère en quatre phases : la spécification des hypothèses sur le problème du diagnostic, la recherche des informations sur le dispositif à diagnostiquer, l'utilisation du savoir-faire en modélisation et l'exploitation du savoir-faire en diagnostic. Cette démarche a été suivie pour concevoir le modèle d'un dispositif complexe de contrôle/commande ferroviaire : le système sol de la signalisation des lignes à grande vitesse produit par CS-Transport (TVM 430).

## 9. Diagnostic coopératif

Face à la difficulté à obtenir des modèles complets ou à obtenir des observations directement sur les dispositifs diagnostiqués, nous avons également exploré une voie de diagnostic interactif et coopératif. Cette recherche a été menée dans le cadre de la thèse de D. Jouglet en collaboration avec le CNET et France Telecom.

Dans certaines situations, le dispositif à diagnostiquer et le modèle utilisé pour rechercher les diagnostics ne sont pas directs. Par exemple, dans les configurations où la prise d'observations est effectuée par un opérateur humain en vue d'alimenter un outil de recherche des diagnostics. Dans ces cas, on ne dis-

pose pas d'observations directes sur le dispositif défaillant mais d'un ensemble d'informations généralement fournies par une personne plaignante. La qualité des échanges entre l'opérateur et ces personnes est primordiale car elle influence directement la qualité des informations fournies. D'autre part, les personnes qui fournissent les informations ne sont pas objectives.

En réponse à ce type de problème, nous avons défini la notion d'Espace Commun de Travail. Nous avons proposé une architecture supportant cet espace commun, prenant la forme d'un système interactif d'aide à la décision. Cette proposition a été validée dans le contexte du diagnostic. Ainsi, l'architecture de l'outil repose sur une base de connaissances organisée selon différents points de vue. Chaque point de vue correspond à un arbre de décision et permet de guider la recherche des diagnostics. Différentes fonctions ont été définies pour « naviguer » dans ces arbres ; elles permettent à l'opérateur de gérer efficacement les hypothèses qu'il formule lors des différentes étapes de son raisonnement. Ce problème est connu dans le domaine de la recherche d'explications, par exemple en enseignement interactif assisté par ordinateur, où les situations d'échec contribuent à un objectif plus global d'acquisition du savoir.

## 10. Raisonnement incertain et diagnostic

Depuis 1998 nous avons amorcé une étude sur la prise en compte d'informations incertaines dans le raisonnement. L'approche suivie exploite un réseau bayésien. Nous avons défini un algorithme de recherche des diagnostics les plus probables. Nous poursuivons nos travaux d'une part pour caractériser l'influence de la topologie d'un réseau bayésien sur les temps de calcul et, d'autre part, pour manipuler des modèles hiérarchiques. De plus, nous souhaitons intégrer ce raisonnement de nature abductive dans une structure plus

globale en complément du raisonnement déductif principalement traité jusqu'ici. L'exploration de cette voie a démarré dans le cadre d'un projet GRRT, et se poursuit actuellement dans le cadre de la thèse de M.A. Maalej.

## Quelques Références

- [1] ADAM Emmanuel (2000)  
Modèle d'Organisation multi-agent pour l'aide au travail coopératif dans les processus d'entreprise : application aux systèmes administratifs complexes, Thèse de Doctorat, Université de Valenciennes, 28 Septembre 2000
- [2] E. ADAM, R. MANDIAU et C. KOLSKI (2001)  
Holonc Multi-Agent for Cooperative Work : Application of a HOMASCOW to administrative processes., *Journal of Applied Systems Studies : Methodologies and Applications for Systems Approaches*, Cambridge International Science Publishing, Cambridge, England (special issue Holonc and Multi-agent Systems), vol. 2, n° 1, 2001
- [3] E. ADAM, R. MANDIAU (2003)  
Bringing Multi-Agent Systems into human organizations: application to a Multi-Agent Information System, accepted in the Fifth International Bi-Conference Workshop on AGENT-ORIENTED INFORMATION SYSTEMS (AOIS-2003), 13 October 2003, Chicago, Illinois, at ER'03
- [4] E. ADAM, C. KOLSKI, R. MANDIAU & E. VERGISON (2003)  
A software engineering workbench for modeling groupware activities", *HCI' 2003 : Proceedings of the 10th International Conference on Human - Computer Interaction*, J. Jacko & C. Stephanidis (Ed.), vol. 4, June 22-27, 2003, Crete, GREECE, pp. 1499-1503
- [5] E. ADAM, M. LECOMTE (2003)  
A software engineering workbench for modeling groupware activities, *HCI' 2003 : Proceedings of the 10th International Conference on Human - Computer Interaction*, J. Jacko & C. Stephanidis (Ed.), vol. 3, June 22-27, 2003, Crete, GREECE, pp. 629-633
- [6] A. CHAMPION, S. ESPIE, R. MANDIAU et C. KOLSKI (2001)  
Multi-agent Road Traffic Simulation : The Coordination Issue, *Proceedings of the 13th European Simulation Symposium : Multi-Agent Modelling and Simulation in Industry and the Environment*, Marseille, France, October 18-20th, 2001

## PRÉSENTATION DE LABORATOIRES

- [7] A. CHAMPION, S. ESPIE, R. MANDIAU et C. KOLSKI (2001)  
Multi-Agent Road Traffic Simulation : Towards Coordination by Game Theory Based Mechanisms, *Proceedings of the Agents in Traffic and Transportation : Workshop in Conjunction with ITS – Intelligent Transport Systems - World Congress 2001*, Sydney, Australia, 30th Sept. / Oct. 5, 2001
- [8] V. DELCROIX, S. PIECHOWIAK, J. RODRIGUEZ (2002)  
Computing diagnoses with higher posterior probability using bayesian networks, *International Conference on Information Processing and Management of Uncertainty in Knowledge-based systems (IMPU'2002)*, Annecy, july 1-5, 2002
- [9] V. DELCROIX-FROIDURE, S. PIECHOWIAK, M.A. MAALEJ (2003)  
Calcul des diagnostics les plus probables a posteriori. *Revue d'Intelligence Artificielle (RIA)* (acceptée en février 2003)
- [10] E. GRISLIN-LE STRUGEON, E. ADAM, C. KOLSKI (2001).  
Agents intelligents en interaction homme-machine dans les systèmes d'information, Environnements évolués et évaluation de l'IHM, *Interaction homme-machine pour les SI*, volume 2. (pp. 209-248). Kolski C. (Ed.), Paris: Hermès.

- [11] D. HANON, E. GRISLIN-LE STRUGEON et R. MANDIAU (2003)  
A Behavior Based Architecture for the Control of Virtual Pedestrians, *Proc. 2<sup>nd</sup> Conf on Computation Intelligence, Robotics and Autonomous Systems*, Singapore, 15-18 Dec. 2003
- [12] D. JOUGLET (2000)  
Coopération homme-machine pour le diagnostic technique : Application aux dérangements téléphoniques. Thèse de Doctorat, Université de Valenciennes, 26 janvier 2000
- [13] LAICHOURE H. (2002)  
Modélisation multi-agent et Aide à la Décision : Application à la Régulation des Correspondances dans les Réseaux de Transport Urbain., Thèse de Doctorat, Lille 1, 20 Décembre 2002
- [14] F. LEFEBVRE (2000)  
Contribution à la modélisation pour le diagnostic des systèmes complexes : application à la signalisation des lignes à grande vitesse. Thèse de Doctorat, Université de Valenciennes, 17 juin 2000
- [15] J. MA, S. PIECHOWIAK, R. MANDIAU (2003)  
Towards an Influence Factors guided algorithm on Constraint Relaxation in Distributed Constraint Satisfaction Problems, *Proc. 2<sup>nd</sup> Conf on Computation Intelligence, Robotics and Autonomous*

*Systems*, Singapore, 15-18 Dec. 2003

- [16] R. MANDIAU, E. GRISLIN-LE STRUGEON et A. PENINOU (2002)  
Organisation et applications des SMA, ISBN 2-7462-0439-8, Hermes-Lavoisier, Paris, 2002
- [17] C. PETIT-ROZE, E. GRISLIN-LE STRUGEON (2003).  
Intelligent Agents to Structure and to Process Personalized Information Systems. In M. Mohammadian (Ed.), *Proceedings International Conference on Intelligent Agents, Web Technologies and Internet Commerce - IAWTIC'2003* (12-14 February 2003, Vienna, Austria). (pp. 165-174). Canberra, Australia: DiskTech Pty Ltd.
- [18] S. PIECHOWIAK, J. RODRIGUEZ (2001)  
Compilation des contraintes booléennes naires pour le traitement des CSPs dynamiques, *Revue de l'Intelligence Artificielle*, Hermes, vol 15 n°2, pp. 219-245, 2001
- [19] S. PIECHOWIAK (2004)  
The localization and correction of errors in models: a constraint based approach, à paraître dans *Journal of Applied Intelligence*, Kluwer Academics Publishers.

## Systemes d'Information, Connaissances et Intelligence Artificielle

Coordonnateurs :

**Danielle Boulanger** (*db@sunlyon3.univ-lyon3.fr*), **Eunika Mercier-Laurent** (*eml@wanadoo.fr*)

**Manuel Zacklad** (*manuel.zacklad@utt.fr*)

### Objectif

L'objectif de ce dossier est de faire un point sur les approches et les techniques de l'IA symbolique et numérique. Elles peuvent contribuer à moderniser l'informatique au quotidien, que ce soit pour un élève, un citoyen, une organisation ou une entreprise. Il est aussi d'indiquer à un public plus large que la communauté des membres de l'AFIA, que des idées géniales de chercheurs doivent être transformées en produits. Ceux-ci, utilisés plus largement, peuvent être intégrés dans l'informatique traditionnelle. Certains programmes de recherche sont également développables à partir des besoins des utilisateurs. Parmi les besoins potentiels des utilisateurs, citons le « retour d'expérience » utilisant des approches et des techniques IA, intéressant pour comprendre et reproduire les succès et tirer des leçons des erreurs.

### Aspects historiques

Depuis l'invention de l'ordinateur, les technologies de l'électronique et celles de la programmation n'ont cessé d'évoluer.

L'intelligence artificielle depuis plus de 50 ans propose des méthodes et des techniques pour acquérir, découvrir et traiter des connaissances.

Dans des années 80, des approches et des techniques comme les systèmes experts, la programmation par contraintes, les réseaux neuromimétiques, la modélisation de connaissances ont été utilisées. Elles ont contribué à accompagner des activités d'entreprise, comme : la conception, la prise de décision en situation complexe (Sachem) ou en situation de crise (Jeux Olympiques d'Albertville), l'optimisation de la production ou la logistique, et bien d'autres. L'enthousiasme initial est retombé, suite à des échecs dont jamais le retour d'ex-

périence n'a été analysé. Une des raisons de cet échec a été le manque de prise en compte de l'humain, de sa nature, de ses motivations, de ses habitudes et de sa résistance au changement.

Au début des années 90 apparaissent les outils du raisonnement à partir des cas. Ils intègrent le raisonnement par analogie, très proche du raisonnement humain. Ils sont utilisés en entreprise essentiellement en diagnostic industriel et plus tard pour collecter et exploiter le retour d'expérience.

### L'ordinateur et l'innovation

Dans notre vision, celle de l'Innovation à Partir des Connaissances (Knowledge Innovation ®), l'ordinateur connecté est capable de traiter des connaissances et travaille en symbiose avec l'utilisateur. Il est son assistant intelligent qui le renseigne sur les événements de sa ville, comme : réserver un billet de train, acheter des produits et services, visiter un musée... Il fait partie d'un système d'information de l'organisation étendue<sup>1</sup> qui permet d'apprendre en permanence grâce aux échanges dans ces réseaux. Cependant, les systèmes actuels, bien qu'automatisés intègrent peu d'intelligence et parfois même perturbent l'utilisateur au lieu de l'aider rapidement et efficacement. Les approches et les outils d'IA ont donc un rôle à jouer.

### Quelques approches IA

Le retour d'expérience en utilisation des approches et des techniques IA est recommandé pour comprendre, reproduire les succès, et tirer des leçons des erreurs.

Des approches et des techniques comme les systèmes experts, la pro-

1. Organisation, clients, partenaires, fournisseurs, distributeurs...

grammation par contraintes, les réseaux neuromimétiques, la modélisation de connaissances sont intéressantes. Elles ont été utilisées pour accompagner des activités en entreprise, comme : la conception, la prise de décision en situation complexe, en situation de crise, et l'optimisation de la production ou la logistique, et bien d'autres.

Les techniques et outils du raisonnement à partir des cas, intégrant le raisonnement par analogie, très proche du raisonnement humain, sont utilisés en entreprise. Leurs destinations ont été essentiellement le diagnostic industriel et plus tard la collecte et l'exploitation du retour d'expérience.

Le mouvement du Knowledge Management (KM), a été lancé d'abord par les spécialistes en management comme Peter Drucker, à la fin des années 80. Il a donné des idées à la communauté IA, qui était la seconde à rejoindre ce mouvement (Karl Wiig, CEDIAG...). Il s'est ensuite propagé dans d'autres domaines comme les ressources humaines (Leif Edvinson), la formation, le marketing, et les documentalistes.

L'arrivée de l'Internet a été une opportunité pour réutiliser des techniques comme le Groupware et le Workflow (INRIA, Kayak 1981), l'hypertexte et la programmation objet. En revanche, tout ce que l'IA avait proposé a été passé sous silence. On peut citer : la construction de sites Web à l'aide de modèles KADS, les techniques de traitement du langage naturel, les possibilités du raisonnement à partir de cas, les agents intelligents, les techniques de modélisation pour réutiliser les mêmes connaissances, les techniques d'apprentissage. Ce dernier domaine a proposé des techniques de découverte des connaissances dans les bases de données. Plus tard, en collaboration avec les spécialistes du traitement de la langue et ceux des réseaux neuromimétiques, cela a

conduit aux techniques du text mining. En revanche, l'expérience a prouvé qu'opérer du data mining sur des données brutes d'entreprise, au lieu de connaissances, ne donne pas de résultats intéressants. On est conduit à « nettoyer » les données, c'est à dire sélectionner des enregistrements pertinents, ce qui demande beaucoup d'énergie et de temps.

## Constats et conclusions

En observant les cours d'informatique on peut constater que les approches et les « techniques de la connaissance » sont très peu enseignées. L'enseignement du KM, dans la plupart des cas se limite au travail collaboratif et à la construction d'Intranets, sans structurer des connaissances.

Dans certaines entreprises le tout puissant Microsoft fait tout. Dans cette approche l'utilisateur est bel et bien oublié. Il doit s'adapter à ce que la Direction des Systèmes d'information, sollicitée par une armée de vendeurs des outils et de services, lui propose.

Dans les entreprises, toutes les approches et les techniques de l'IA sont concernées. Souvent elles doivent être associées pour une meilleure efficacité et convivialité, ce qui constitue en excellent défi pour la recherche collaborative multi-domaines.

Eunika Mercier-Laurent  
et Serge Dupuy

**INTELLIGENCE  
ARTIFICIELLE DANS LES  
ANNÉES 80 ET GESTION DES  
CONNAISSANCES DANS  
LES ANNÉES 90 :**

**POURQUOI SI PEU  
DE PROMESSES TENUES ?**

**COMMENT MIEUX FAIRE  
DANS LA PROCHAINE  
DÉCENNIE ?**

**Auteur :** Jean Rohmer, vice-président fondateur de l'AFIA,  
jean.rohmer@fr.thalesgroup.com

2. La plaquette de l'AFIA mentionne tous (ou presque) les domaines de l'IA.

## Activités

*Jean Rohmer travaille pour Thales Communications, à concevoir des architectures combinant les réseaux sémantiques, le datamining et le text mining.*

*Un de ses centres d'intérêt est d'étudier l'acceptabilité des techniques « intelligentes » et « symboliques » par les utilisateurs. Il s'exprime ici à titre personnel.*

*Avant de rejoindre Thales, Jean Rohmer a dirigé successivement Le Centre de Recherche en IA Groupe Bull, le CEDIAG centre de développement de produits et de services en IA du Groupe Bull, les sociétés Dallas et Idéliance qui ont développé le logiciel de gestion de connaissances Idéliance, utilisé en particulier par l'Armée Française, actuellement propriété de la société Thales.*

*Il est membre de l'Institut F.R. Bull où il anime le groupe multidisciplinaire de réflexion « L'Informatique vue par l'Entreprise ».*

En 2003, nous en sommes, dans le domaine de la gestion des connaissances, à peu près dans l'état où nous en étions en 1993 à propos de l'intelligence artificielle : dans un contexte de crise du secteur des NTIC, bien peu d'entreprises ont mis en place de manière généralisée la pratique de la gestion des connaissances.

En tant qu'acteur depuis plus de vingt ans de tout ce domaine que l'on pourrait appeler le « traitement symbolique et intelligent de l'information », nous voulons ici contribuer à un retour d'expérience, en essayant de comprendre pourquoi ces technologies par ailleurs si séduisantes pour les utilisateurs et si excitantes pour les chercheurs, ne se sont pas plus développées. Ce texte est volontairement sans complaisance - et essentiellement autocritique ! Notre but est d'inviter la communauté scientifique à trouver des manières de procéder autrement dans la décennie à venir.

**Pourquoi l'IA des années 80 n'était déjà plus florissante au début des années 90 ?**

A la fin des années 80, quasiment un chercheur sur deux travaillait en

Intelligence Artificielle. Mais vers 1994, presque toutes les équipes de recherche ou d'études avancées des grands groupes avaient disparu. Nous entrevoyons la liste partielle suivante d'explications :

- les machines étaient beaucoup trop chères et pas assez performantes : une station de travail UNIX pouvaient coûter jusqu'à 500 000 francs. Le rapport performance / prix était en gros de mille à dix mille fois plus mauvais qu'aujourd'hui.
- les chercheurs ne s'intéressaient pas aux applications simples de l'intelligence artificielle. Plutôt que de s'appliquer à confronter ces nouvelles techniques aux réalités du terrain, ils s'empressaient de sophistiquer sans fin des outils qui n'avaient jamais connu un début d'application. C'était une fuite sans fin des réalités (à laquelle l'auteur a participé).
- le reste de la communauté scientifique « non IA » (Génie Logiciel, Bases de Données, ...) était agacée, mais aussi isolée, par la mode ambiante et la prétention de ses adeptes. Les tenants de l'IA ont cru bon devoir se passer de son savoir-faire, et, en réaction, la communauté non IA n'a pas ménagé l'autre camp, en le soumettant à des exigences et des critiques qui n'avaient d'égales que l'arrogance des autres.

**Pourquoi la Gestion des Connaissances est si peu répandue au début des années 2000 ?**

C'est l'économie qui a eu raison de l'intelligence artificielle au début des années 90, qui ont vu la disparition de tous les grands acteurs traditionnels de l'informatique de l'époque, IBM excepté.

En parallèle, une réflexion critique avait tout de même eu lieu, qui a donné naissance vers 1990 au domaine de la Gestion des Connaissances. Elle a été amplifiée par la sortie commerciale du logiciel de travail en groupe Lotus Notes racheté depuis par... IBM. Plutôt que de raisonner à la place des gens, pourquoi ne pas utiliser l'ordinateur simplement pour les aider à organiser et partager leur connaissance ?

Ce mouvement ouvrait le champ à de nouveaux acteurs dans les sciences et dans l'entreprise : sciences cognitives, sciences humaines, directions des ressources humaines. La technique s'effaçait, l'humain repassait au premier plan.

On s'est trouvé de ce fait face à un « vide technologique » : finis les langages Lisp et Prolog, les bases de données déductives ... La nature a vite rempli ce vide par deux énormes phénomènes : la montée en puissance des logiciels de Microsoft, et bien sûr l'arrivée d'Internet .

Les « technologies subtiles » de la gestion qualitative des connaissances (modélisation des connaissances, agents intelligents, ontologies, règles métiers...) se sont trouvées balayées par les « technologies basiques » consistant à maîtriser l'invasion quantitative de « trop d'information ».

Dans ce paysage, s'affrontent ainsi deux écoles :

- l'une est centrée sur l'effort *explicite* de modélisation des connaissances, effort individuel ou collectif
- l'autre privilégie l'automatisation *implicite* dans la gestion des documents
- la première a développé des outils de modélisation et de gestion *explicite* de connaissances, reposant pour l'essentiel sur un effort humain (éditeurs de graphes de type The Brain, Mindmanager, outils de gestion de réseaux sémantiques plus sophistiqués comme Ideliance ou Mondeca)
- la seconde s'est concentrée sur des outils automatiques d'extraction, découverte, ou organisation *implicite* de connaissances à partir de textes, en réduisant au minimum l'intervention humaine (les outils de text mining, les outils de classification automatique de documents comme Arisem, outils d'extraction de concepts et de terminologies comme ceux de Lingway, etc ...).

Les déboires récents de la plupart des sociétés actives dans ce domaine montrent que ni les unes ni les autres de ces approches technologiques de la gestion des connaissances n'ont connu un grand succès.

Nous voudrions inviter la communauté scientifique à réfléchir au « pourquoi » de ce qu'il faut bien appeler les échecs successifs des technologies de traitement symbolique intelligent de l'information depuis vingt ans.

– côté gestion explicite des connaissances par effort individuel, la situation est très claire : à part quelque pourcent d'individus, personne n'a envie de consacrer du temps à organiser, gérer, publier ses connaissances. Nous préférons tout refaire, réécrire pour la dixième fois une présentation Powerpoint, plutôt que de passer une demi heure par jour –c'est le « tarif » - à utiliser un éditeur sémantique. D'accord pour interroger Google pendant des heures, mais pas pour le quart d'heure de synthèse structurée des résultats. **L'obstacle à la Gestion des Connaissances est notre paresse et notre manque de rigueur.**

(Personnellement, nous tenons à jour un réseau sémantique depuis 5 ans, et ne pourrions plus travailler autrement, mais avons pour l'instant renoncé à « vendre » aux autres l'intérêt de ce « sport »).

Aujourd'hui l'outil numéro 1 de gestion structurée de connaissances dans de grandes et performantes organisations, c'est Excel !

Même les systèmes les plus modestes dans leurs ambitions de partage explicite de connaissance (intranets de publication) ont un mal fou à s'imposer. C'est tellement plus facile d'envoyer par mail à chaque fois et à chacun la nouvelle version du document ...

– côté gestion implicite des connaissances (via le text mining ou la classification automatique), tout repose sur la capacité de la machine à comprendre « un peu » le sens des documents. Or c'est là un des défis non relevés des années 80 ...

La compréhension du langage naturel semble rester hors de portée. L'état de l'art ne sait pas résumer un texte, extraire les bons concepts. *Trop intelligents pour être compris par une machine, nous ne le sommes pas assez pour l'éduquer ...*

Et si l'on tente de lui mâcher le travail au moyen d' « ontologies » censées aider à cette compréhension du texte, on retrouve l'autre échec : personne n'a envie de constituer et surtout mettre à jour ces ontologies ...

En résumé, nous avons le sentiment de nous trouver face à une double contradiction :

- **Ce que l'on saurait faire (gérer nous mêmes explicitement nos connaissances), on ne veut pas le faire**
- **Ce que l'on voudrait faire (laisser gérer implicitement nos connaissances par la machine), on ne sait pas le faire**

Plutôt que de proposer nos propres pistes, nous souhaiterions, à travers cet article un peu provocateur, inviter la communauté scientifique à réfléchir aux moyens de « trouver le passage » pour dépasser ces contradictions. Un point clé de cette réflexion sera bien sûr d'imaginer « quoi faire » du Web Sémantique et de ses standards émergents (RDF, OWL, ...).

## Équipe Condillac « Ingénierie des Connaissances »

### La gestion ontologique des connaissances

Université de Savoie - Campus Scientifique  
73 376 Le Bourget du Lac cedex - France

**Contact :** Pr. Christophe Roche  
tél: 04 79 75 87 79  
mél : christophe.roche@univ-savoie.fr  
<http://ontologie.univ-savoie.fr>

### Membres de l'équipe concernés par le thème

Christophe ROCHE, Professeur, Daniel BEAUCHÊNE, Maître de Conférences, Françoise DELOULE, Maître de Conférences, Christine RIEU, Maître de Conférences, Christophe TRICOT, Doctorant, Cécile MILLION-ROUSSEAU, Directeur d'Ontologos, société

de valorisation et de transfert technologique de l'équipe Condillac, Samah REGUIGUI, Ingénieur R&D Ontologos.

## Contexte global de recherche / application

L'évolution des marchés et des technologies a conduit à une profonde modification de nos sociétés que l'on qualifie volontiers de « sociétés de l'information ». Le nouvel enjeu économique est devenu la maîtrise des informations, des connaissances et des savoir-faire.

Face aux nombreux problèmes que soulèvent la gestion des connaissances (modélisation, (ré)utilisation, partage) et la gestion des communications (langages métier différents, acteurs hétérogènes, multilinguisme) les ontologies sont souvent présentées comme la panacée dans cette « quête du sens » pour une meilleure maîtrise des informations. Les référentiels métier, définis en termes d'ontologies d'entreprise et de terminologies, constituent la pierre angulaire de nos recherches.

Enfin, recherche d'informations et documentaire, veille technologique, mémoire d'entreprise, gestion de la relation client et des ressources humaines, constituent le cadre applicatif de nos recherches. Ces applications sont menées en relation directe avec la société Ontologos, société de valorisation et de transfert technologique de l'Équipe Condillac créée dans le cadre de la loi sur l'innovation et la recherche de 1999.

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

Nos travaux se répartissent selon trois grands axes de recherche :

1. Référentiels Métier :
  - Analyse linguistique ;
  - Terminologies ;
  - Ontologies.
2. Cartographie Sémantique
3. Capitalisation des Connaissances : CEM (Corporate Enterprise Memory)

Nous présentons ci-dessous un exemple de nos travaux de recherche menés au sein de l'Équipe Condillac et une des applications réalisées en collaboration avec la société Ontologos.

*Exemple de travaux de recherche :*

## Construction de Référentiels Métier (Terminologies & Ontologies)

Notre objectif est la construction de référentiels métier qui soit : consensuelles, c'est-à-dire pour lesquelles les experts s'accordent quant à la signification des termes ; cohérentes, au sens de la consistance logique de l'ensemble des significations ; partageables et réutilisables.

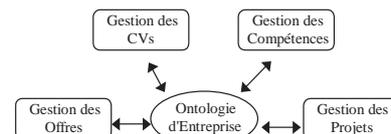
La construction d'un référentiel métier, lexique de spécialité propre à une pratique sociale, pose deux problèmes : celui de la construction du lexique, c'est-à-dire de la détermination des termes qui le composent ; et celui de la signification des termes définie en fonction des concepts métier de l'entreprise. Il n'existe pas à l'heure actuelle de consensus quant aux méthodes de construction de référentiels métier. Le type d'applications et les objectifs visés permettent de choisir entre une « terminologie textuelle », une « terminologie conceptuelle » ou une approche intermédiaire, la « terminologie ontologique ».

Dans tous les cas, il semble judicieux d'initier le processus de construction d'une terminologie par l'étude des textes produits au sein du domaine d'application dans la mesure où ils constituent le support principal des expressions métier. Cependant, tout n'est pas dans le texte et la méthodologie de construction doit tenir compte des propriétés épistémologiques du langage de représentation des signifiés. C'est pourquoi le modèle que nous proposons tente dans la mesure du possible d'unifier principes linguistiques, épistémologiques et logique.

### Exemple d'application : Gestion des Ressources Humaines

La gestion des ressources humaines, qu'elle soit interne à l'entreprise, ou externe pour la recherche de personnes qualifiées, repose d'une part sur la connaissance des individus et de leurs compétences et d'autre part sur la connaissance de l'entreprise et de son métier. C'est par la « mise en correspondance » de ces connaissances que nous pourrions améliorer l'emploi et valoriser les connaissances individuelles et que nous pourrions au mieux gérer l'entreprise.

La modélisation de ces connaissances repose sur les **référentiels métier** de l'entreprise, c'est-à-dire sur les concepts métier et les termes nécessaires à leur désignation (terminologie). Ces référentiels constituent ce qui est communément appelé **l'Ontologie d'Entreprise**.



L'analyse du contenu des informations, qu'elles soient structurées ou non : cv, compétences, offres d'emplois, projets, permet de marquer sémantiquement les informations par leur sens métier ; c'est-à-dire en référence à l'ontologie de l'entreprise. Il devient alors possible de les indexer, de les classer, de les rechercher ou de les mettre en correspondance selon les objectifs visés : recherche de cv, rapprochement des compétences, des offres et des projets.

La construction des référentiels métier est donc le premier problème à résoudre, et de leurs « qualités » dépendra la pertinence de la gestion des ressources humaines. C'est la raison pour laquelle nous avons adopté l'approche ontologique

## Courte Bibliographie

(<http://ontology.univ-savoie.fr/>)

- Conférence invitée. C. Roche « Ontology: a technical roadmap from theory to applications » ICEIS 2003, 5th International Conference on Enterprise Information Systems. 23-26, April 2003 Angers – France
- C. Roche « The Differentia Principle: a Cornerstone for Ontology », « Knowledge Management and Philosophy » Workshop in WM 2003 Conference, Luzern, 2-4 april, 2003
- C. Roche, C. Million-Rousseau « Construction d'Ontologies : l'Importance du Modèle Ontologique » Journées Francophones d'Extraction et de Gestion des Connaissances EGC 2003, Lyon, 22-24 janvier 2003
- C. Millon-Rousseau, C. Roche. « L'Ontologie au Service des Ressources Humaines ». Troisièmes Journées Francophones d'Extraction et de Gestion des Connaissances EGC 2003, Lyon, 22-24 janvier 2003
- D. Beauchêne, C. Millon-Rousseau, C. Rieu. « Détection automatique de l'insatisfaction du client dans un contexte de com-

## Équipe "Intelligence Economique"

Département des Sciences Sociales pour l'Ingénieur (DSSI), Université Toulouse I

**Responsable du DSSI :** Anne-Marie ALQUIER-BLANC

**Contact** pour l'équipe « Intelligence Economique » : Maryse SALLES, Maryse.Salles@univ-tlse1.fr

**Domaine global de recherche du DSSI :** ingénierie de méthodes pour les systèmes d'aide à la décision.

### Thèmes traités par l'équipe Intelligence Economique

Méthodes d'ingénierie du besoin en intelligence économique (thème principal), méthode de conception de systèmes d'intelligence économique.

### Contexte de la recherche

#### *L'intelligence économique*

L'intelligence économique (I.E.) n'a pas de définition consensuelle, et en termes de système d'information, n'a pas de réels référentiels.

L'équipe a retenu la définition proposée par G. Colletis dans le numéro 1 de la *Revue d'Intelligence Economique* : l'intelligence économique, pour une entreprise donnée, est « sa capacité à combiner efficacement des savoir-faire et compétences internes et externes, en vue de résoudre un problème productif inédit ».

Une *capacité* de l'entreprise renvoie à un ensemble de connaissances.

Le système d'information d'intelligence économique (S.I.E.) de l'entreprise sera considéré comme le système de représentation des connaissances de l'entreprise concernant les relations, réelles ou potentielles, qu'elle entretient avec son environnement.

Un tel système d'I.E. a pour vocation première d'apporter une **aide à la déci-**

**sion**, pour toutes les décisions qui affectent les relations de l'entreprise avec son environnement, de quelque manière que ce soit et avec quelque intensité que ce soit. Les décisions concernées sont pour l'essentiel des décisions complexes, non répétitives, peu ou mal structurées.

La perspective dans laquelle se situe l'équipe « Intelligence Economique » est celle de la mise au point de méthodes adaptées pour la conception de SIE.

#### *L'ingénierie des systèmes d'information d'intelligence économique*

Les axes de travail de l'équipe se situent donc dans le domaine général de l'ingénierie de méthode, et plus particulièrement dans celui de l'**ingénierie du besoin**.

Les produits de la recherche couvrent théorie, techniques et systèmes, avec une volonté majeure de confrontation au réel et de méthodologie adaptée à l'**action effective**.

Le contexte de l'intelligence économique, comme celui, plus généralement, des systèmes d'aide aux décisions peu ou non structurées, présente des problèmes particuliers. Le plus déterminant de ces problèmes réside dans le fait que les structures classificatoires ou opérations logico-mathématiques **ne préexistent pas à la spécification du SIE**, et ce, ni dans les processus cognitifs individuels, ni comme référentiel dans l'entreprise, ni même dans la littérature du domaine.

L'ingénierie du besoin en intelligence économique requiert donc une méthodologie spécifique d'acquisition des connaissances.

La méthodologie développée par l'équipe « Intelligence Economique » s'organise en trois composants principaux : modèles, démarches et supports à la mise en œuvre.

Cette méthodologie a permis de développer une première méthode complète d'analyse du besoin en I.E. (voir plus bas « Projet MEDESIIE »).

#### *Réalisations de l'équipe « Intelligence Economique »*

Les productions de l'équipe répondent à deux objectifs distincts :

- la constitution d'un corps de connaissances sur l'ingénierie de méthodes

spécifiques à l'intelligence économique

- la réalisation d'outils méthodologiques complets utilisables par les professionnels de l'intelligence économique

#### *Modèles*

L'ingénierie de besoins intègre dans la modélisation :

- le fonctionnement cognitif du décideur face à la situation de décision, qui s'étudie en tant que processus mental personnel du décideur.
- la situation de décision elle-même, qui s'étudie en rapport à l'entreprise, son fonctionnement et son référentiel d'organisation.

La modélisation de la connaissance adressée dans les SIE intéresse la connaissance produite ou à produire par l'organisation :

- sur elle-même et son fonctionnement (objets, processus, structures internes)
- sur son environnement (définition des objets externes, relations existantes ou potentielles avec ces objets, ...)

Les modèles développés s'expriment à trois niveaux d'abstraction : méta-modèles, modèles conceptuels et modèles conceptuels partiellement instantiés (voir plus bas « Supports à la mise en œuvre »).

Les modèles concernent :

- l'entreprise et sa stratégie
- l'environnement de l'entreprise (proche et lointain)
- le besoin en intelligence économique (construit sur 3 dimensions principales, renvoyant au fonctionnement cognitif du décideur, à la situation de décision dans l'entreprise, et aux objets de l'environnement concernés)
- le produit d'I.E.

Ces modèles ont été construits en classifiant les connaissances issues de différentes sources : un corpus de 70 entretiens réalisés auprès de dirigeants de PME (voir plus loin « Le projet MEDESIIE »), des éléments de méthodes fournis par des professionnels du conseil en I.E., la littérature du domaine et de domaines connexes (économie industrielle, gestion, systèmes d'information, ...).

## Démarches

Au niveau de l'ingénierie de méthode en I.E., les démarches développées concernent l'extraction de connaissance pour la définition des méta-modèles et modèles conceptuels.

La méthode d'analyse du besoin produite propose une démarche détaillée, qui s'articule en quatre phases. Les trois premières phases sont organisées selon une logique dite « en cascade », la dernière suivant une démarche par prototype, de type « spirale ».

## Supports à la mise en œuvre

Dans l'ingénierie des systèmes d'information, les aides à la mise en œuvre s'inscrivent dans quatre grandes perspectives, qui ne s'excluent pas l'une l'autre :

1. La transmission de l'expertise nécessaire aux futurs concepteurs, essentiellement par la formation
2. Le guidage de la démarche de conception, qui s'appuie sur des documents types, des check lists, des exemples, etc.
3. La mise à disposition d'outils logiciels de type AGL
4. La fourniture d'éléments déjà constitués, à utiliser tels quels, qui peuvent être d'importance et de niveau d'abstraction très divers.

Les travaux de l'équipe ont porté sur les 4 types de supports, avec un effort tout particulier sur les 3<sup>ème</sup> type (AGL) et 4<sup>ème</sup> type (fourniture d'éléments déjà constitués).

Un ensemble de modèles pré-instanciés a été constitué au sein de la méthode d'analyse du besoin. Ces modèles pré-instanciés concernent tous les modèles évoqués plus haut.

## Le projet MEDESIIE

Le projet MEDESIIE (*Méthode de définition de système d'information pour l'intelligence économique*) est un projet d'ingénierie de besoin pour l'I.E., financé par la Région Midi-Pyrénées. Ce projet, consacré entièrement à l'analyse du besoin en I.E. des PME, s'est déroulé sur trois ans. Il a eu pour résultat principal une méthode structurée pour l'analyse du besoin des décideurs en matière d'I.E.. La méthode MEDESIIE est accompagnée d'un atelier logiciel desti-

né à aider à sa mise en œuvre ainsi qu'à faciliter la capitalisation des connaissances.

Ses principales phases ont permis de réaliser :

- un large état de l'art
- une enquête terrain, réalisée sous forme d'entretiens auprès des responsables de 70 PME de la région Midi-Pyrénées, qui a permis de recueillir leur besoin en intelligence économique tel qu'ils l'ont exprimé, ainsi que les caractéristiques de leurs entreprises
- la construction de divers modèles et modèles pré-instanciés utilisés dans la méthode
- le développement d'un atelier logiciel d'aide à la mise en œuvre de la méthode.

Le projet a réuni des chercheurs de diverses disciplines (intelligence économique, économie industrielle, systèmes d'information, gestion, bibliométrie).

## Une méthode centrée sur les PME

La méthode MEDESIIE a été principalement conçue pour l'analyse **du besoin en I.E. des PME**. Notons qu'elle est cependant, dans plusieurs de ses composants, transposable pour d'autres types d'organisations.

La méthode considère les PME comme dotées de caractéristiques propres. Des modèles et des typologies ont été développés dans la méthode pour rendre pleinement compte des spécificités des PME.

La méthode est principalement destinée à des professionnels intermédiaires en I.E., amenés à effectuer des analyses du besoin en I.E. au sein de multiples organisations. L'atelier logiciel propose des outils facilitant aussi bien la capitalisation de connaissances que l'adaptation de la méthode par ses utilisateurs.

## L'analyse du besoin en I.E.

MEDESIIE et les outils de support associés aident à réaliser :

- le recueil du besoin en I.E., tel qu'exprimé par les dirigeants de la PME, à l'aide de guides d'entretien exemplifiés, et personnalisés en fonction du type de la PME
- l'analyse du besoin ainsi recueilli, avec le support de « besoins cibles »,

et des aides à l'identification et l'interprétation des écarts entre le besoin exprimé et le « besoin cible »

- l'entreprise ayant, en coopération avec le consultant, défini le besoin qu'elle considère comme pertinent, MEDESIIE permet alors d'affiner ce besoin, pour chaque décideur concerné, avec le support d'un outil de prototypage de produit d'I.E.

La méthode a été testée une première fois, au sein du projet MEDESIIE, de façon limitée. Elle est actuellement en cours d'utilisation dans deux sociétés de service en I.E.

## Perspectives

Des projets sont actuellement à l'étude pour d'une part étendre la méthode à d'autres types d'organisations, d'autre part spécifier la méthode pour une population particulière (PME d'un secteur industriel spécifique par ex.).

## Courte bibliographie

Salles, M. *Stratégies des PME et intelligence économique. Une méthode d'analyse du besoin*. Ouvrage à paraître, éditions Economica, septembre 2003.

Salles, M. *Modélisation des situations de décision dans une méthode d'ingénierie du besoin en I.E.* Communication, conférence IERA (Intelligence Economique : Recherches et Applications), Nancy, France, 14-15 avril 2003.

Zid, T., Salles, M. *Projet MEDESIIE (rapport final tâche WP6) : Spécification de l'atelier logiciel d'aide à définition du besoin*. Université Toulouse I, décembre 2002.

Salles, M., *Projet MEDESIIE (rapport final tâche WP5) : Méthode MEDESIIE de définition du besoin en Intelligence Economique des PME*. Université Toulouse I, décembre 2002.

Salles, M., *Problématique de la conception de méthodes pour la définition de Systèmes d'Intelligence Economique*. *Revue d'Intelligence Economique*. n° 6-7, Avril-Octobre 2000.

Site Web (en re-construction) : <http://atlas.irit.fr/MEDESIIE>

## Équipe MODEME

Centre de recherche de l'IAE  
Université Jean Moulin - Lyon3  
15 Quai Claude Bernard

69007 Lyon  
tél : 04 78 78 71 58  
fax : 04 78 78 77 50  
mél : criae@univ-lyon3.fr

## Responsable :

Pr. Danielle BOULANGER  
db@univ-lyon3.fr

## Membres concernés

Danielle BOULANGER, Professeur,  
Responsable scientifique  
Gilles DUBOIS, Maître de conférences  
Guilaine TALENS, Maître de conférences  
Caroline WINTERGERST, Maître de conférences  
Eric DISSON, chercheur docteur  
Vincent COUTURIER, ATER  
Magali SEGURAN, ATER

## Contexte global de recherche

Les travaux de l'équipe concernent essentiellement l'intégration de concepts et de techniques d'intelligence artificielle dans les systèmes d'information coopératifs. Plus précisément, nous travaillons sur la mise en œuvre de processus coopératifs via des systèmes multi-agents, ce qui nous amène à définir des structures d'agents coopératifs, des protocoles de coopération et des architectures fédératives assurant l'interopérabilité entre sources d'information hétérogènes.

Dans ce cadre, les thèmes de recherche plus particulièrement développés sont :

1. La représentation de connaissances (modèles de connaissance interne des agents et connaissances globales d'un domaine pour la résolution de conflits sémantiques, application des travaux sur les ontologies),
2. La définition et la validation de protocoles de coopération, de résolution de conflits et d'apprentissage adaptés au contexte des systèmes d'information coopératifs,
3. La définition et la validation de structures d'agents informationnels et de systèmes multi-agents coopératifs, la capitalisation et la réutilisation de ces structures via la définition de patterns,

4. La sécurité des systèmes d'information coopératifs, interopérabilité de modèles de sécurité et protocoles de sécurité entre agents informationnels.

Les axes ci-dessus sont détaillés dans la section suivante.

## Description des travaux

1. La représentation de connaissances dans nos travaux exploite un modèle orienté objet développé dans le cadre d'une thèse de l'équipe permettant de représenter le corpus de connaissance sous forme de graphes d'objets. Chaque nœud est un objet descriptif fournissant une représentation d'un élément d'information issu d'une source de données ou de l'environnement. Chaque arc est un objet-lien qui permet de représenter différents types de relations entre les objets descriptifs. Ces objets à la structure « légère » permettent la génération dynamique de graphes de représentation sur l'environnement et sur l'ensemble des sources d'information locales. Ces graphes sont exploités dans les protocoles de traitement de requêtes globales et de résolution de conflits sémantiques.

Une extension de ce modèle descriptif orienté objet a été développée dans le cadre d'une application de raisonnement à partir de cas. Nous explorons actuellement les liens de notre approche avec les travaux sur les ontologies.

2. Nous avons défini une structure d'agent capable de raisonner sur les connaissances orientées objet décrites ci-dessus et mettant en œuvre plusieurs protocoles en fonction des phases de fonctionnement du système multi-agents assurant la coopération des sources de données locales. Le protocole global de coopération est composé de différentes phases exploitant un protocole adapté. La première phase utilise le protocole de passation de contrat (contract net) pour détecter et mobiliser les agents susceptibles de répondre de façon pertinente à une requête, puis les réponses des différents agents sont collectées de façon asynchrone (request until satisfaction) et enfin les bases de connaissances des agents sont mises à jour grâce à un protocole d'apprentissage dérivé du protocole de Sian. Les

modèles d'architecture d'agent et de système multi-agents développés sont documentés et présentés sous forme de patterns dans le but de constituer un langage de patterns adapté à la construction d'architectures coopératives orientées agent. La plate-forme développée en C++ est actuellement en phase de réécriture et d'extension en environnement Java.

3. Nous disposons d'une architecture fondée sur un système multi-agents, dont les fonctionnalités ont été validées par un premier prototype. Nous évoluons vers une démarche de capitalisation des connaissances issues de l'architecture coopérative et du prototype (rétro-conception) afin de proposer une bibliothèque de patterns et de composants réutilisables dédiés à la coopération de systèmes d'information. Il s'agit d'éviter la redécouverte coûteuse des abstractions et des concepts logiciels lors du développement de processus coopératifs entre applications.

Au sein de l'architecture proposée, deux types d'éléments sont susceptibles d'être capitalisés et réutilisés : le prototype et les 'savoir et savoir-faire' mis en œuvre dans la conception de l'architecture. Les patterns s'appliquant à toutes les phases du développement logiciel (analyse, conception et implantation), nous en proposons l'utilisation afin de capitaliser l'expertise mise en œuvre dans la construction de processus coopératifs entre applications. Quel que soit son niveau d'application, un pattern décrit un problème récurrent intervenant dans un contexte et sa solution. Parallèlement, les patterns conçus doivent permettre la génération de composants réutilisables lors du développement de toute application coopérante.

La démarche sera constituée de trois phases :

- **Découverte des patterns de coopération** elle repose sur un processus itératif d'expertise basé sur l'existant logiciel et conceptuel défini au sein de l'architecture. La découverte des patterns est tributaire des modèles, du code ou de tout autre produit logiciel existant qui contiennent, par essence, un savoir-faire lié au domaine d'application.

- **Généralisation et optimisation des patterns précédents** à partir de l'existant logiciel et conceptuel. La phase de vérification et d'optimisation des patterns peut être favorisée par l'utilisation de guides de rédaction.
- **Implémentation des composants réutilisables** qui permettront de résoudre les problèmes de coopération.

Notre approche permettrait donc une rétro-conception de l'architecture et du prototype existant, mais aussi de compléter et généraliser les connaissances extraites à l'aide d'autres méthodes ou techniques (protocoles...) nécessaires afin de résoudre de façon optimale les divers problèmes liés à la coopération.

4. Dans l'élaboration d'une fédération de sources d'information, il est important de traiter la sécurité des données. La modélisation proposée doit prendre en compte l'interopération de systèmes d'information ouverts et évolutifs et, de plus, garantir le respect des différentes politiques de sécurité locales. Le contrôle d'accès dans un tel contexte est rendu d'autant plus complexe que ses objectifs vont à l'encontre de ceux de la fédération. La solution adoptée pour le contrôle des accès dans une fédération ne peut être qu'un 'compromis' entre l'ouverture au niveau fédéré et l'autonomie des politiques de sécurité locales (hypothèse forte des fédérations à couplage lâche).

Pour résoudre cette double difficulté, nous construisons un processus d'intégration d'un système local à la fédération se déroulant en trois étapes. Dans un premier temps, les administrateurs locaux définissent un schéma d'exportation de données. Un modèle de rôle original appelé "modèle de rôle étendu" permet alors la représentation unifiée des schémas d'accès locaux tout en conservant les propriétés du contrôle de flux des trois principales politiques d'accès (discrétionnaire, modèle de rôles et modèles multi-niveaux). Lors de la troisième phase du processus, les schémas d'accès décrits sont enrichis pour établir les autorisations d'accès inter-schémas.

L'objectif principal de la politique de contrôle des flux du niveau global est de respecter les profils utilisateurs locaux : un utilisateur local ne peut lire ou modi-

fier des informations fédérées que si elles sont équivalentes aux informations locales sur lesquelles il a de telles autorisations.

Dans ce but nous définissons une session globale sécurisée permettant de respecter une politique d'importation et d'exportation des données entre la fédération et ses divers membres.

L'ensemble de ces propositions s'appuie et complète le modèle de coopération développé au point 1.

## Bibliographie

- Objets-Agents : continuum ou différences ?*  
Danielle Boulanger, Joël Colloc, Gilles Dubois, Caroline Wintergerst  
Journée SMA du PRC IA, Paris, 16 décembre 1994
- A Knowledge Based System : An Object Case Approach*  
Guilaine Talens, Danielle Boulanger, Isabelle Dedun  
12th International Conference on Tools with Artificial Intelligence, ICTAI'2000, Vancouver, 13-15th November, 2000.
- A Multi-Agent System using Semantic Metadata for the Cooperation among Multiple Information Sources*  
Gilles Dubois, Danielle Boulanger,  
Workshop « Knowledge Management Theory and Applications », KMTA'00, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases, Lyon, France, 12-16th September, 2000
- A Role-Based Model for Access Control in Database Federations*  
Eric Disson, Danielle Boulanger, Gilles Dubois  
3rd International Conference on Information and Communications Security, ICICS'01, Xian, China, 13-16 November 2001, LNCS 2229, Springer Verlag
- Advanced Object Approach for Information System Cooperation*  
Danielle Boulanger, Gilles Dubois  
Encyclopedia of Microcomputers Vol 27 supplement 6, Allen Kent & James Williams Executive Editors, Marcel Dekker Inc., New-York, 2001

## Équipe Orpailleur

UMR LORIA 7503 - Bâtiment LORIA  
- Campus Scientifique - B.P. 239 - F-54506 Vandoeuvre-lès-Nancy Cedex  
URL : <http://www.loria.fr/equipes/orpailleur/accueil.php>

**Responsable :** Amedeo Napoli  
(Amedeo.Napoli@loria.fr,  
tél : 03.83.59.20.68).

## Membres de l'équipe concernés par la thématique

Rim Al Hulou, Mathieu d'Aquin, Sébastien Brachais, Jean Lieber, Amedeo Napoli, Emmanuel Nauer, Laszlo Szathmary.

## Contexte général

L'objectif du projet Orpailleur est de concevoir des « systèmes intelligents », c'est-à-dire capables d'exploiter des connaissances pour résoudre des problèmes dans un domaine donné. Les connaissances sont multifformes et peuvent s'acquérir auprès d'experts - expertise, expériences, explications, stratégies - ou bien provenir de documents ou de données brutes. Nous décrivons dans la suite quatre systèmes qui exploitent des connaissances dans des cadres particuliers d'application.

## 1 - IntoBib : système hypertexte de fouille de données et de recherche d'information intelligente

Contact : Emmanuel Nauer,  
Emmanuel.Nauer@loria.fr

Le besoin en information est primordial dans de nombreux domaines, comme celui de la recherche ou celui de la veille scientifique et technique. Les données relatives à un domaine sont de plus en plus facilement accessibles ; toutefois cette quantité croissante de données disponibles nécessite de mettre en œuvre des moyens particuliers pour les exploiter. Le but du système IntoBib est de fournir au chercheur, ou au spécialiste de l'information scientifique et technique, un environnement dans lequel il puisse exploiter les données de son domaine (références bibliographiques et Web en particulier) pour des besoins de recherches bibliographiques ou d'analyses du domaine. IntoBib est un système de fouille de données bibliographiques. Il repose sur l'utilisation de l'hypertexte pour accéder de façon exploratoire aux données. Des fonctionnalités de fouille (dénombrements, classifications, extractions de règles, etc.) peuvent être déclenchées à la demande

pour analyser plus précisément certains sous-ensembles de données.

L'idée principale qui a conduit à la mise en œuvre du système IntoBib est que la fouille de données et la recherche d'information sont deux approches extrêmement complémentaires pour appréhender des données : la fouille de données permet de guider la recherche d'information à partir des connaissances extraites des données, et inversement, la recherche d'information permet de guider la fouille de données par l'exploitation des connaissances issues de la fouille de données. Le système tire également profit de la fouille des données bibliographiques pour favoriser l'accès à l'information sur le Web. Concrètement, une aide à la formulation de requêtes, à partir d'éléments extraits des données bibliographiques préalablement identifiées par l'utilisateur, est proposée pour améliorer la précision des réponses obtenues lors de l'interrogation de moteurs de recherche.

## 2 – Raisonnement pour l'exploitation de données semi-structurées

Contact : Rim Al Hulou ,  
Rim.Al-Hulou@loria.fr

Le monde informatique regorge de données aux formats très hétérogènes : SQL, HTML, textes, images, etc. qu'il est nécessaire d'intégrer pour construire des applications telles que des intranets ou des bibliothèques électroniques. Ces données dites semi-structurées peuvent être structurées, peu structurées ou non structurées. Leur traitement nécessite de mettre en œuvre des moyens adaptés pour les exploiter. La conception d'un « système intelligent » de traitement de données semi-structurées est en cours de réalisation ; ce système exploite à la fois les données et les connaissances correspondant au domaine des données.

Les formalismes de représentation de connaissances sont de bons candidats pour la construction d'un environnement permettant d'exploiter des données semi-structurées textuelles pour des besoins de recherche documentaires ou d'analyse de données. Dans ce cadre, une approche générale pour le traitement de telles données, fondée sur l'utilisation du langage XML et sur l'utilisation des logiques de descriptions est

proposée. Concrètement, le langage XML est utilisé afin de décrire et annoter les données selon une ontologie représentant les connaissances du domaine des données. Cette ontologie est stockée dans une base de connaissances représentée avec la logique de descriptions RACER. Dans cette base de connaissances, des mécanismes de raisonnement par classification et par subsomption sont mis en œuvre pour raisonner sur le domaine des données et sur les requêtes. Enfin, ces connaissances peuvent être utilisées pour raisonner sur les documents : par généralisation, par similarité, etc.

## 3 – Le projet Kasimir : gestion des connaissances en cancérologie

Contact : Jean Lieber,  
Jean.Lieber@loria.fr

Le projet Kasimir réunit Orpailleur, l'association Oncolor (réseau de soins en cancérologie de Lorraine), le centre régional de lutte contre le cancer Alexis Vautrin (Vandoeuvre-lès-Nancy) et le laboratoire d'ergonomie du CNAM (Paris). L'approche de gestion des connaissances de ce projet est issue de travaux en ergonomie sur les activités méta-fonctionnelles. Elle s'appuie sur une confrontation de faits « du monde réel » (les cas médicaux) avec les bases de connaissances : ces dernières doivent évoluer pour tenir compte de ces faits.

La plate-forme informatique que nous construisons implante d'ores et déjà des bases de connaissances, dans un formalisme de représentation des connaissances par objets, pour la prise de décision thérapeutique. Elle est destinée à être utilisée par les praticiens. Une version s'appuyant sur une déduction floue, pour la prise en compte des effets de seuil, a aussi été développée. Depuis peu, la construction de bases de connaissances se fait grâce à un éditeur de connaissances s'appuyant sur le système Protégé-2000.

Actuellement, les adaptations de la base pour les cas particuliers sont étudiées dans l'optique du raisonnement à partir de cas (RàPC). L'acquisition et la représentation des connaissances pour l'adaptation sont au centre de ce travail. À plus long terme, l'évolution des connaissances par les confrontations

aux cas particuliers doit être étudiée : à partir des adaptations effectivement réalisées par les experts cancérologues, des évolutions des bases de connaissances doivent être proposées. La plate-forme est destinée, à terme, à devenir un portail sémantique manipulant des connaissances représentées dans les langages du web sémantique (RDF(S), OWL). Une extension de ces langages à la déduction floue et au RàPC (qui est non déductif) sera nécessaire.

## 4 – KVM, Knowledge Valorisation Matrix : un outil de gestion de connaissances

Contact : Amedeo Napoli,  
Amedeo.Napoli@loria.fr

L'objectif principal du système KVM est d'aider une entreprise à mettre en place une gestion de ses connaissances, en lui proposant une méthodologie pour structurer, codifier et représenter son portefeuille de connaissances. KVM a pour but de fournir une aide aux décisions stratégiques et une gestion dynamique des références du portefeuille de connaissances ; c'est un système configurable, modulaire et évolutif.

Dans le système KVM, une connaissance est considérée comme un élément dans le portefeuille de connaissances de l'entreprise, qui est constitué d'un ensemble d'objets – appelés « zones de connaissance » – qui contiennent la connaissance. Le système KVM caractérise les objets du portefeuille et leurs attributs à travers une matrice tri-dimensionnelle ; chaque dimension constitue un référentiel particulier de représentation. Les trois axes de représentation sont les suivants : WBS (Work Breakdown Structure) permet une décomposition technique de l'entreprise, OBS (Organization Breakdown Structure) correspond à la structure des ressources de l'entreprise, MBS (Model Breakdown Structure) s'appuie sur un cadre de codification de la connaissance de référence.

Une matrice (Knowledge Valorisation Matrix) est construite avec des éléments de la connaissance, considérés comme des briques de base, qui sont agrégées et intégrées dans une hiérarchie à trois dimensions. La gestion du portefeuille de connaissances s'appuie sur la structu-

ration des éléments de connaissance dans ce référentiel tri-dimensionnel et sur les attributs dynamiques qui donnent une codification unique à chaque zone de connaissance. KVM offre une codification structurée des processus de connaissance, de l'organisation (OBS), du modèle (MBS) et de la définition technique du système (WBS) de toute connaissance observable dans une entreprise. Cette définition du modèle par diverses variables structurées est générale et indépendante des industries, des commerces, de la technologie et des qualifications. Elle provient pour partie de l'épistémologie des sciences, des sciences économiques de transaction et de la théorie d'organisation et des systèmes.

**Bibliographie** (documents accessibles à partir du serveur Web) :

- [1] E. Nauer. *Principes de conception de systèmes hypertextes pour la fouille de données bibliographiques multibases*. Thèse en Informatique, Université Henri Poincaré Nancy 1, Janvier 2001.
- [2] R. Al Hulou and A. Napoli. Combining XML and DL for describing and querying documents. In *International Workshop on Description Logics - DL'2002*, Toulouse, France, April 2002.
- [3] J. Lieber, M. d'Aquin, P. Bey, B. Bresson, O. Croissant, P. Falzon, A. Lesur, J. Lévêque, V. Mollo, A. Napoli, M. Rios and C. Sauvagnac. The Kasimir Project: Knowledge Management in Cancerology. In *Proceedings of the 4th International Workshop on Enterprise Networking and Computing in Health Care Industry (HealthComm 2002)*, pages 125--127, 2002.

## Projet RAMCESH

**Contact** : RM. Faure, CETu, 25 Avenue F. Mitterrand 69500 BRON, 04 72 14 34 81.  
rm.faure@equipement.gouv.fr

**Équipe** : N. Faure, D. Ballouard, I. Provillard, S. Kunian, F. Danjon, RM. Faure

**Environnement** : le Centre d'Études des Tunnels

Le CETu (Centre d'Études des

Tunnels) est en Europe le seul organisme gouvernemental à s'intéresser exclusivement aux tunnels. Il rassemble donc des experts reconnus dans le domaine, et a constitué avec le temps une solide expérience acquise sur le terrain. Cette expérience se conjugue avec le dynamisme de la production scientifique et réglementaire à ce sujet, favorisé par les nombreux congrès internationaux concernant les ouvrages souterrains.

La masse de connaissances induites par cette participation internationale ajoutée à l'expérience propre du CETu nécessitait donc un programme de capitalisation et gestion de connaissances, adapté à sa mise en forme pour transmission et restitution.

A cette fin, le CETu participe activement au projet RAMCESH (*Recueil Assisté pour la Maîtrise de la Conception des Espaces Souterrains Habités*), coordonné par l'entreprise SOLEM Expertises S.A., dont les problématiques sont similaires. Le projet RAMCESH est une collaboration à l'échelle européenne rassemblant une dizaine d'organismes et entreprises, et visant à produire un système d'aide à la décision dans le domaine de la géotechnique (dans un premier temps) reposant sur la connaissance accumulée au cours des dernières années.

Le CETu a parallèlement un certain nombre de recherches orientées vers la gestion des connaissances, comme les projets Didactu (didacticiel en ligne), KBT (base de données de tunnels), et Base de référence (qui regroupe les textes à utiliser pour les projets de tunnels)

Le projet fondateur de l'expérience de la gestion des connaissances au CETu est le projet Xpent, dont la problématique concernant la stabilité des pentes reste proche du domaine propre aux tunnels. Xpent est un projet de système expert qui remonte à une bonne dizaine d'années et dont les enseignements profitent encore aujourd'hui au CETu.

## Structure de RAMCESH

Le projet RAMCESH est un outil col-

laboratif subdivisé en trois parties, numérotées de 1 à 3, qui sont autant de modules composant une application.

RAMCESH 1 est le module de constitution de l'ontologie, qui repose sur une analyse lexicale et taxonomique du contenu textuel d'un document proposé par un expert du domaine. L'analyse est systématique, et consiste majoritairement en l'utilisation de filtres informatiques qui permettent à l'expert de recenser, définir et classer les mots et les locutions du texte qui recouvrent des concepts-clés du domaine. Ces concepts sont répertoriés au sein d'une ontologie simplifiée (i.e. qui ne comporte qu'un nombre très limité de relations entre les concepts).

Cette ontologie permet donc le repérage sémantique de tous les concepts et l'indexation des documents. Ce repérage est d'autant plus accessible que RAMCESH propose une interface graphique où l'arborescence est directement visible et modifiable par l'expert. Il a été constaté à de nombreuses reprises l'avantage maïeutique indéniablen de cette dimension graphique lors des premiers essais de complétion du lexique par des experts géotechniciens totalement étrangers au projet.

RAMCESH 2 est le cœur du système (au sens recueil des connaissances) et permet à l'expert de constituer à partir des concepts ontologiques des « granules de connaissances » tirés de l'analyse systématique de textes. Un granule de connaissance est un assemblage argumenté de concepts ontologiques servant à décrire une situation. Écrits sous forme de règles de production (si A alors B), les granules de connaissance incluent des éléments de logique floue afin de permettre une recherche plus pertinente des connaissances. La variation des concepts propres au granule le long de l'arborescence ontologique générée dans RAMCESH 1 permet également le dégagement d'un « champ sémantique », un ensemble de granules proches qui peuvent avantageusement alimenter la réflexion de l'utilisateur.

Le granule peut en outre être instancié par une « détermination », un ensemble de valeurs variables selon le type de granule, afin d'obtenir un descriptif de cas spécifique.

Tris et inférences permettent une utilisation dite encyclopédique de l'outil ou, au choix de l'utilisateur, une utilisation méthodique qui lui permet de constituer intégralement un assemblage de granules, un raisonnement qui soit propre au projet sur lequel il travaille. C'est également là qu'intervient la notion d'analogie et les liens éventuels vers des projets internes au CETu comme KBT (Knowledge Base of Tunnels), dans la mesure où les choix proches de ceux de l'utilisateur effectués pour d'autres projets sont présentés en temps réel à l'écran.

Ces deux modules permettent aux experts, dont le temps est souvent précieux, de n'être mobilisés qu'un minimum de temps pour la phase de validation. Cependant, dans la mesure où le débat peut naître entre plusieurs experts à partir de la définition d'un élément (concept, granule ...), il est nécessaire d'entretenir un workflow précis avec des modérateurs (eux-mêmes experts du domaine) pour guider les discussions et éventuellement apporter des solutions idoines au sein des connaissances stockées par le système pendant la durée du débat.

RAMCESH 3 apporte la notion de suivi de chantier aux deux modules précédents, liant les granules de RAMCESH 2 aux problèmes de financement, de compétences, de matériels et de tâches qui leur sont afférents. Par ce biais, il est possible de composer automatiquement un diagramme de gestion de projet qui découle directement des choix effectués par l'utilisateur au sein de RAMCESH 2. RAMCESH 3 permet donc de noter sur une base quotidienne les progrès effectués dans le chantier et de les comparer automatiquement aux progrès prévus. Les différences relevées entre les objectifs et les progrès réels sont donc très vite accessibles, ce qui permet aux experts attachés au chantier d'apporter des correctifs et de prendre les mesures qui s'imposent selon les cas.

## Références bibliographiques :

- Faure R.M., Leroueil S., Rajot J.P., Larochelle P., Sève G. et Tavenas F., 1988, XPENT, système expert en stabilité des pentes. *Proceedings 5th International*

*Symposium on Landslides*, Bonnard ed., Lausanne, p.625-630

- Faure R.M., Mascarelli D., Vaunat J., Leroueil S. et Tavenas F., 1992, Present state and development of XPENT, expert system for slopes stability problems. *Proceedings 6th International Symposium on Landslides*, Bell ed., Christchurch, p.1671-1678
- Ballouard D., Provillard I., Ramcesh1, recherche d'une ontologie des tunnels, *mémoire de DESS SIA*, Lyon3, 2003
- Mazzoleni G., Présentation du thème de recherche prioritaire n°12, Recueil des connaissances. Rapport interne, CETu 2003.
- Faure N., Modélisation et gestion des connaissances, le cas des géosciences, *Thèse Lyon 3*, en préparation

## Department of Artificial Intelligence Systems

(<http://sun.ae.wroc.pl>)

The Department of Artificial Intelligence Systems was established in 1996 as a part of the Institute of Economic Computer Science at Wrocław University of Economics, Poland; head of the department Mieczyslaw L. Owoc Ph.D. Eng. e-mail: [mieczyslaw.owoc@ae.wroc.pl](mailto:mieczyslaw.owoc@ae.wroc.pl), Ph. No. : ++4871 3680 503.

**Staff:** Violetta Galant Ph.D., Janina Jakubczyc Ph.D., Krzysztof Hauke Ph.D., Maria Mach Ph.D., Malgorzata Nycz Ph.D. Eng., Barbara Smok Ph.D., Artur Kotwica M.Sc. Eng. Szymon Kosciow M.Sc. Eng., Maciej Pondel M.Sc. Eng., Maciej Wasiaak M.Sc.

## Departmental interests:

Knowledge acquisition and management (leading topic), temporal and fuzzy reasoning, knowledge validation and verification, expert system applications in the management context, distance and open learning, machine learning.

## Research description

From the very beginning the research and didactic fields of the department included modern technology of data processing. For example, research referring: database systems, information system development and later, expert systems and artificial intelligence domain

as a whole were performed. Some of the projects were granted by the Polish Scientific Research Committee. The department manages yearly seminar: « Knowledge Acquisition from Databases » recently changed onto « Knowledge Acquisition and Management » where people from different university centres from Poland and representing foreign institutions discuss theoretical and practical aspects of knowledge acquisition. The department took part in the international projects : the TEMPUS project (co-operation with Limburg University – Belgium, University of Limerick - Ireland, Superiore Ecole de Commerce Marseille - France and Economische Hochschule Aachen – Germany) and “Knowledge Acquisition and Intelligent Distributed Learning in Resolving Managerial Issues” – with Limburg University and Catholique University Leuven (Belgium) and “Machine Learning 2000” – with Malardalens University (Sweden) and Riga University (Latvia).

## Publications

- Bonner R.F., Galant V., Owoc M.L.: On Features of Decision Trees as a Technique of Knowledge Modelling. *Proceedings of the Workshop on Computer Science and Information Technologies CSIT'99 Moscow, Russia, 1999*
- Owoc M.L., Galant V.: Validation of Rule-Based Systems Generated by Classification Algorithms. *Proceed. of the Information Systems Development Conference, Bled'98 (Slovenia) Kluwer Academic/Plenum Publishers, New York*
- **Owoc M.L.**, Ochma\_ska M., G\_adysz T.: On Principles of Knowledge Validation: Validation and Verification of Knowledge Based Systems: Theory, Tools and Practice, A. Varmesan, F. Coenen (red). *Proc. of the 5th European Symposium on Knowledge Based Systems and Components - EURO-VAV'99, Oslo (Norway) 1999 Kluwer Academic/Plenum Publishers, New York*
- F. Coenen, A. Varmesan, ... **Owoc M.L.** et al. *Validation and verification of knowledge-based systems: Report on EUROVAV99. The knowledge engineering review*, vol. 15, no. 2, June 2000. Cambridge University Press
- Mach M., Owoc M.L.: *Validation as the Integral Part of a Knowledge Management Process*. Proc. of the Informing Science '2001 Conference – Challenges for Informing Clients. A Transdisciplinary Approach, Kraków

## Cyberenergy and cyberentropy : How to create value with the cybertechnologies

Henry Samier<sup>1</sup>, Simon Richir<sup>2</sup>, Victor Sandoval<sup>3</sup>

1&2 ISTIA Innovation, Université d'Angers, 62 avenue notre dame du lac, 49000 ANGERS

3 ENPC, 6 et 8, avenue Blaise Pascal, Cité Descartes, 77455 MARNE-LA-VALLEE

samier@istia.univ-angers.fr

richir@istia.univ-angers.fr

sandoval@mail.enpc.fr

**Abstract:** We present a new approach for a better comprehension of what is going on with the diffusion and implementation of the information technologies at larger scales. This is one support for creating a new space. This approach has at first many theoretical and many practical implications. We simply call it: Five Dimensional approach (5D). It appears as a new attribute of space (spatial and temporal) added to classical 4 space-time dimensions. Under those conditions, a new challenge is emerging for the human race: to live and work in 5D and not only in the well-known 4D. This paper illustrates this approach with some theoretical developments linked to Virtual Manufacturing (VM). VM uses cyberCAD/CAM and, to optimise the way it works, VM needs to know and use the main parameters of this new dimension. VM is a new type of manufacturing, entirely existing on this dimension. So this situation leads us to entirely new problems. Particular attention is paid to virtual spaces, manufacturing and collaboration, skills matrix and manufacturability and virtual prototyping and virtual partnerships.

**Keywords:** cyber technologies, cyberspace, model, measurement, concurrent engineering, agile manufacturing.

### Introduction

Currently we speak about digital world, cyberspace, hyperworld or hyperspace. All this refers to a new reality emerging from the increasing imple-

mentation of ICT at a wider scale in companies and countries. So we are adding a new dimension to our daily activities. We can consider digital as a largest space, cyberspace being a subset, and hyperspaces as a space bigger than our own. Hyperspace is a human creation having its own nature, principles and laws and its own development. We need to consider its internal dynamics and its external relations. Nevertheless all this is different from our own world.

For example, Hyperspace is a learning tool: we need to learn it and we need to use it as a working tool. Learning is short or long according to the field involved. Before we learn to apply it we receive « information »: we were in a situation of « information ». The problem actually is different: we need to learn to apply knowledge, we need to « learn to learn » and even more we need to learn by structures. We need to learn to discover knowledge from data and information contained in structures (see supracognition theory [CAS 00]). More generally, we must learn to live and work in this new space that we call for commodity 5D [SAM 01].

### Characteristics of the 5D worlds

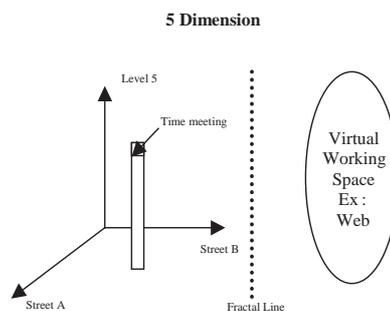


Figure 1:  
From 2D to 5D for living and working

Reading this figure we follow the dimension progressing from 2D to 5D. A simple look enables us to position the different dimensions. Immediately we have in front of us differences between 4D and 5D. Just some comments from this:

First: spaces have different nature and laws

Second: an interface between the two spaces represented by the vertical line

Third: we call this line a fractal line or border

Four: this is a fundamental problem for optimising systems in the future. We indicate only one example: classical algorithms of operation research are not directly applicable to this field. Here we are facing a world full of feedback or non-linearities are common bread for every one living and working in those dimensions.

	3D analogical	4D analogical	5D digital
(xyz)	X	X	0
(x,y,z,t)	N	X <sub>p</sub> t	0
Speed (V)	0	V=f(xyz,t)	V=Cte
Accel. (a)	0	a=f(xyz,t)	a=Cte
Theory*	0	(0)	(X)

Table 2: Point, speed and acceleration in different spaces  
Tools Functions 4D/5D

Table n°2 shows a comparison between spaces using only points, speed, acceleration and the need for a theory. Table n°2 compares also analogical and digital character of the spaces 3D, 4D and 5D. We can see 5D as an entirely digital space.

Another exercise to understand the above differences is that of networks followed by flow. We compare old flow, essentially material flows, with new flows: essentially, information flows. Table 3 illustrates this comparison.

Flow exchange	Material	Support	Transport
Material Flow	Mass (Kg)	Carrier ; Cost(\$)= f(time, mass)	Road,air, cost(\$)= f(distance, speed)
Information Flow	No Mass (0Kg)	ICT Cost(\$)=Cte Annual Cost	Networks Cost(\$)=Cte ( < 1 s)

Table 3:  
Material and Information flows

We can observe some fundamental differences, which can be, compared with the preceding tables, especially 1 and 2. As the result, in this new world, the scales are changing drastically. Time, distances, speed, mass are all different. Let us only say that we need here

a new thermodynamic. Only one thing to understand: mass are null and one important question is "How can one build densities in that case?"

## Laws of the hyperspace

There are many laws operating in hyperspace. All of them must be arranged into two categories. The first contains the general laws. The second concerns the binary laws. These last means the laws existing only because the nature of the binary strings. We give some examples of these: the bigger the objects are the more they are distant. The longer the distance to be crossed by one object, the more it is enveloped (addressing). The bigger the object is, the more it is fragile. All objects must be represented by simple line (Berger's line or Bit's line). This line is starting at zero point and can reach the infinite. Near zero the bits are stables, and they become less and less stables when we leave this point. That line, for example, can represent all objects on the Internet, and we can compare them comparing lines. All Web sites can be represented by an URL equation reduced to the bit's line and then compare them.

The next comments concern some general laws. These are related to the different layers of hyperspace. It is easy to comment it following the hyperspace geology from the physical to the abstract layers.

## A. The Moore Law

The Moore law operates at the physical layer or lower layers of the hyperspace: ships and computers. Thus the Moore law operates firstly in the computer manufacturing sector. Then its effects are exported to other sectors using computers and ships. A French publisher P. Berger, has developed an historical extension of this law to the past, to very long periods of time starting with the origin of humankind and presenting different steps lived by humankind as the Moore law and its factor of growth. It is possible to postulate a generalisation of the Moore law as follows:

- 1) the law can then be applied to other sectors (for example, to CAD/CAM, to mixing set technologies), and
- 2) the law can be applied to other per-

iods of history of our societies than that particular concerning the electronic devices spreading in our days.

## B. The Metcalfe law

Following Bob Metcalfe, the value of one network is proportional to the square of the user number. The higher the number of users of one network is, the more it has a value because of the potential communication each other has. Thus connectivity underlying them is a generator of the new values. This is true if users are connected to the network in order to share knowledge, or make contributions and calls. In fact, if a new user entering the network permits the creation of an amount of the new connections, which can be equivalent to the number of users already existing, then the possibility for sharing potential knowledge (as a common background) is also increasing. And this operates in CAD field, for example. But optimisation here is quite different from that of the traditional classical spaces (see OR): each user must enter in an optimal way for their own interest and for the set of users.

## C. The Amdahl Law

Amdahl observed that computer processing power was increasing at a rate several times greater than the time required to input and output data. From this observation he predicted that a significant percentage of the overall throughput would be consumed by the requirements to manage input and output. Reasoning so, he developed a formula, called the Amdahl Law to determine overall throughput: the time it takes to gather, analyse and output data. Amdahl's Law operates at the simple processor level as the computer and networked computer. So this law can be studied and applied to the case of an Intranet or simply on the Internet.

## D. The Gilder Law

The capacity of bandwidth telecom triplicates every year. This law is operating everywhere and affecting everybody using Net technologies.

## E. Information nature's law.

- 1) Information has an owner or seller but he or she doesn't lose the ownership after selling it;

- 2) The user or the buyer receives information and appropriate it as he or she buys in the market or picks up this information on the DBs or Web servers. He or she can use this information to generate more and more information following the main result of the Brooke's equation of the information sciences

- 3) The owner (seller) and user can do copies of this information and the owner cannot forbid the reproduction of this, but he or she cannot resell that information without any modification. A good example is that of the software editors: they sell software copies, and buyers can use it at their own convenience.

Top growth law or asymptotic growth's law (Web server's laws as an example)

There is also a limit for the growing of the Internet. We call it the laws of the top in digital worlds or asymptotical rational and logical laws. So these worlds are limited or relative, there is a digital relativity at this point. Asymptotic growth was so for a large variety of modern products including radio, TV set, VCRs, cars or freezers. In the particular case of telecommunications, telephone diffusion can be described by a rate of installation by  $n$  households or  $k$  enterprises and by the renewal rate.

## F. Three Click Rule (or 3CR).

Good web design respect 3CR rule, that means, reaching pertinent information we need at almost three clicks, no more. In VM this is a very important law, even more because we have a very big volume of information to handle. This means to optimise web design it is necessary a set of skills, including those of CAD/CAM designers.

## G. Virtual community creating value's law

In the LAN, WAN and Intranets case, we can extend the Metcalfe's law operating here. But this extension goes beyond physical layers and reaches the content of the interchange between members of one particular community. This last exists by the network and it is called « a virtual community ». One point is to study the sources of value

added and how members of the community use them. To better use the Metcalfe's law, we can use the advantages of the connexion:

- The number of associated increases,
- There is a facility to diffuse information,
- There is a facility to set up contacts,
- people are near because distances are reducing (see below),
- time is optimised because the possibility of parallel connection and concurrent thanks to the network architecture,
- time contraction plays fully (see below).

### 5D, hypertime and hyperdistance's laws

In hyperspace, time has the new characteristics. It is sequential, parallel and a concurrent at once. This process takes place in computer network, which permits a combination of many kinds of time (network time, computer time, biological time, universal time, sociological time, psychological time, etc). These characteristics constitute the inner nature of the new time, which we call hypertime, or « spatial time ». Spatial time exists only because the digital spaces permit the realisation of these characteristics. Then time is an endogenous factor that we can manage and master following our needs. We must know the nature of the spatial time or hypertime. For example, an engineer team working on a design will have some local constraints but the members work following to their own time. The time is then at the same time local and general. The question is to know what the local and the general elements are and to solve it according to each situation. The time pockets are in ourselves, as participants in the productive process, inside company before outside. In 5D time is spatial, a hypertime. Then the solution must integrate the computer nets but starting from people nets and combining those optimally. Then we must study first how to contract time, where to contract time, identify factors acting here, negatives and positives. The main question is that to understand time is a personal creation, to get out a right approach of time. Time is relative because the nature of the networks and the nature of time

created by every one as a user of the networks.

We need to create spaces within digital worlds, layer by layer. Time and space are closely related. We create time and space and every action on the time can interact within the space more, in competition for resources, people can use computer and do many things with space and time. All this in order to reach a determined objective. So they play with computer, hyperspace and hypertime, integrating fully 5D.s

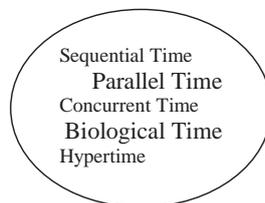


Figure 4: Different kinds of time

As time, distances in hyperspace and in natural real world are different. As a simple understanding exercise we propose some kinds of distances existing in hyperspace. For doing it, we use the classification proposed by P. Berger [BER 01].

Organic distance. « Physical » or « organic » distance: if the objects are in touch we can consider the distance between them is that of its organic centres. If those objects are not in touch it is necessary to add other intermediary objects. This last point is deduced directly from a direction computation. The « soft » distance is done by direct addressing (distance in the addressing tree). The length of that distance is drawn according to the steps Yes or No in the tree. The organisational « structure » of the bits recreates one localisation form, the « matter », having low level, including the case working with purely information. For a processor, we can say that the mass of one object makes the distance between inputs and outputs. The length of this distance is an appreciation of the depth of the treatment.

Hamming distance: it is more functional. It is the content distance. We can also get out a distance by morphing two images (this is used in a image recogni-

tion). If each message has only one bit, there are only two values for this distance: they are identical or they are opposite. If the messages are longer, the problem is more complex. We can have identity or opposition. But if the messages have a different length, we have more difficulties and we must know what to do with them. According to Hamming two strongly complementary messages are a maximal distance. But this point is not evident. So P. Berger proposes to define the distance as the complexity: the number of necessary bits to describe the second message starting from the first (and reciprocally).

Intermediary distance: the language difference is one of the forms of distance. On the contrary, the language normalisation reduces the distances. But the chifrement (case of the security exchange between systems) introduces one more element of distance. It creates a very specific format factor. As the distance increases, the seller of value added must do more and more.

« Sociometric » distance. This distance is bigger if objects are different in communicating. This is evident by the spatial congestion but also by the length of the direction. We must consider the metric of the distance. In this case the Moreno's sociometry can play a part. It permits the definition of the measures between participants within a given network. But even more interesting are the cognitive distances: on the web we have many things! This last distance asks for a normalisation but this even at the preliminary steps.

Cyberenergy and cyberentropy  
In hyperspace there are multiple interactions between being and doing agents and/or objects of this space. As one possible way to formalise this world, we can propose to consider the dynamic of these objects with a model integrating both cyberenergy and cyberentropy. Cyberenergy is the capacity to perform an informational work, one object conveys correctly in the cyberspace one information between persons participating on a given project. That's some kept pint in Virtual manufacturing, for example, for the teams working on CAD design.

Let us set up one important assumption: the higher is the cyberenergy of a given object in the cyberspace, the more the information should be diffused widely, and understood easily. Thus we need to model the process of production and consumption of cyberenergy in the cyberspace [SAM 01].

To begin, cyberenergy of an object is directly linked to:

- the size of the network on which the object is. For example, an object on the Internet must have a higher cyberenergy. On the contrary, an object on a PC alone its cyberenergy is normally null.
- the format of the object. Information stored on a known format has no problems to be understood to the other intelligent agents. On the contrary, if the object is stored in a proprietary format, unknown by the others or not diffused, the information will be lost, and its cyberenergy will be null.

### (1) Cyberenergy=NxF

where: N represents the network size and F represents the format involved.

Take now cyberentropy. It represents the trend of the object to be degraded over the time, that means the trend of the object to lose cyberenergy. Thus cyberentropy says object cannot perform its informational work at the same level of performance as they did before.

Let us consider cyberentropy as a derivative of the cyberenergy, this means it reflects the slope of the curve (cyberenergy vs time) of a given object when it has abandoned it. So the more this object is cyberenergetic the higher the slope will be.

Problem: What is conditioning this slope? In a nutshell: a bad network choice. Putting an object on the Intranet rather than in the Internet can limit the number of persons potentially interested in, and the persons who can be reached thanks to the connectivity, from whom information conveyed by the object should result pertinent.

To preserve the cyberenergy from any decrease over the time, we can survey, for example:

- a) a frequency of updating the object. If the object needs to be updated regu-

larly and it is not the case, that object will quickly become useless. It cannot provide pertinent information on time. A simple example is that of updating web sites: if it is a web site addressed to the electronic commerce, it needs to be updated frequently;

b) a bad choice of the data storage format. If the format becomes obsolete, the probability increases for the object to become obsolete too. This problem is particularly strong in CAD/CAM systems.

At a first approximation the formula can operate to perform that work. But N, taken as the net size on which the information is being, if it depends on time, it must also take into consideration the accessibility and liability aspects. Thus the size of the net is important but at the same time it becomes limited. One particular problem is the efficiency of access to the nets: when this efficiency decreases, the risk of receiving erroneous information rises too. Finally, the cyberenergy handling information must contain a term for degradation of information as a function of the time. In fact, the same information, dated a week ago, has not the same value a couple of weeks after. In CAD/CAM systems, and by this means in Virtual Manufacturing, this may be a dramatic impact. It depends narrowly on the nature of that information. It also exists other characteristics for qualifying the value of information such as, for example, the exclusivity, the scarcity, and the potentiality.

Then, we can improve the formula (1) for expressing better the cyberenergy, using this new formula:

$$(2) \text{ cyberenergy} = N = R(t, a, f) \times F(p(t)) \times K(t)$$

Where:

- N=R size of the net depending on the time t, accessibility (function of the architecture and the size of the nets) a; and liability of information transmission f;
- F coefficient of format of the object depends on p(t), the weights of user units of F with respect to the total units
- K coefficient of degradation function of time

### Cyberentropy

Consider now cyberentropy as a derivative of the cyberenergy: under the original formulation, we have Cyberenergy=NxF and then

### (3) Cyberentropy=NxF/dt+FxN/dt

Under this form the cyberentropy take into account a bad choice of the network across the term dN/dt. Also this takes into account the bad choice of the storage by the term dF/dt. On the contrary, this formulation of the cyberentropy does not permit to express the frequency of updating of an object, or in other words, its obsolescence. By this reason, the use of the complete formulation of the cyberentropy gives Cyberentropy=NxFxK. This formulation represents in a better way the cyberentropy because it permits the introduction of the information updating effect by the term dK/dt. Then the final formula becomes:

$$(4) \text{ Cyberentropy} = NFxdK(t)/dt + FKxdN(t, a, f)/dt + NKxF(p(t))/dt$$

Another point of this Cyberenergy=NxF is related to the lack of the notion of access priority. This priority concerns the decisions that can be strategic, for example, in organizing collaboration between several agents doing Virtual Manufacturing. In this kind of situation, it is necessary to know:

- to define precisely the objective of the project and the sharing of the tasks,
- to perform the fellow and the updates of the project,
- to make a planning of virtual meetings,
- to define what the first decisions are (priority order).

For considering these parameters, we can modify the original formula (1) adding logistic components. Under these conditions, the formula is as follows:

### Virtual Manufacturing (VM) and Virtual Spaces

VM may be defined as an integrated, synthetic manufacturing environment to

enhance all levels of decision and control. The vision of VM is to provide a capability to « Manufacture in the Computer » [LIN 95]. We update this vision « to manufacture in networked computers » which allows us to VM by Internet, if having appropriated tools and software. Also this means to interact VM and 5D or hyperspaces nature, characteristics and laws. That means VM will provide a modelling and simulation environment so powerful that the fabrication/assembly of any product, including the associated manufacturing process can be simulated in the networked computers. VM affects all industrial company activities, starting from the manner in which market signals are taken into account, the relations with customers and providers until the internal reconfiguration. Here are concerned the impact on the company memory, capital investment, design of production, cost estimation, risks management, customer interfaces, functionality interfaces, workshop... This means that an entire networked activity of company is needed. Once again, we insist on the fact that: companies thinking in traditional ways use only partially the potentialities of new spaces, they must think all activity integrating real (or natural) space dimensions and hyperspace dimension as only one environment.

The scope of VM ranges from workshop to overall company. But at the age of Internet, this concerns all companies and not only the bigger ones. In addition, this needs to set up a theoretical approach based on the best practices data basis provided by consortium working on manufacturing projects. Under these conditions the future workshop will be a turning point from the "virtualisation of the real to the realization of the virtual", a new productive reality [CAL 97], [SAA 99].

In the next paragraphs, we present several cases showing an only partial approach of integrated worlds thanks to the networked computers. These are punctual cases that must be incorporated in a vision of the all space dimensions: hyperspace.

### Virtual shared space

The distributed and shared infrastruc-

tures are composed of virtual spaces constituted of virtual reality models that could be manipulated by complementary expertise of trading partners. Those virtual creations and manipulations by multidisciplinary teams enable people to reach a common understanding and to elaborate shared concepts in a better and faster way than any long discussion during meetings. Virtual spaces shared by disciplines provide a collective sense of identity and fundamental purpose as requested by the knowledge-creating company. One practical problem is to identify the rationale of any interaction need and interdependencies between entities manipulated by complementary disciplines and, to be able to manage the consistency of generated constraints. The virtual project office example of virtual shared space will be demonstrated in the project office case, where several industrial trading partners forming a consortium will use the virtual project office as a distributed infrastructure [PAL 98] (see figure 5).

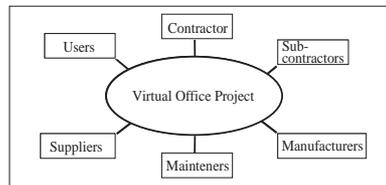


Figure 5 - Example of virtual shared space: the virtual project office (Source EPICE ESPRIT project).

### Virtual Prototyping

Virtual prototyping models offer an entirely new perspective in particular for the Internet trading partners and provide a strong sense of reality and the means to test the model of a product through meaningful interaction. An Internet virtual salesroom concept has to be based on a virtual three-dimensional photo realistic salesroom environment featuring a selection of virtual product models (figure 8). Virtual prototyping is the application of advanced modelling and simulation techniques, user interface techniques and virtual reality techniques to support electronic commerce for customers and also business-to-business electronic commerce in the value chain - that is product design, manufacturing, and sales [SAR 95], [GUP 95], [HOF 95].

Virtual prototyping means the process of using virtual prototypes instead of, or in combination with, physical prototypes, for innovating, testing and evaluating of specific characteristics of a candidate design. The following aspects are dominant:

- Functionality, created virtually, clearly defined and realistically simulated (for example product functionality and dynamic behaviour);
- Human Interaction: the human functions involved must be realistically simulated, or the human must be included in the simulation (that is

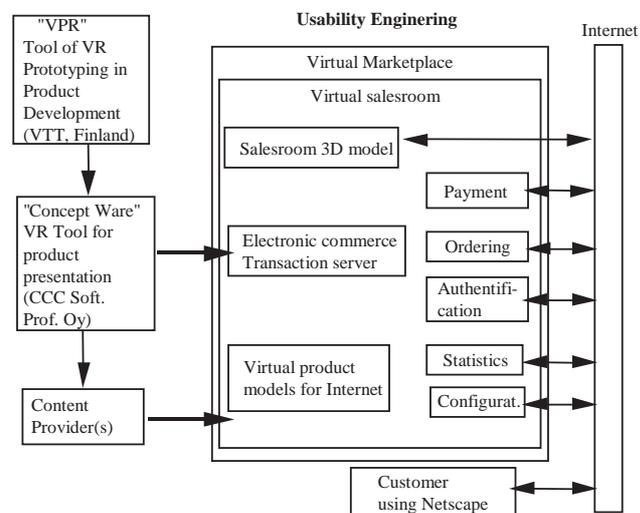


Figure 6 - The proposed concept for Internet electronic commerce using digital virtual product models in marketing and sales [PAL 98]

- real-time operator-in-the-loop simulation);
- c) Environment: an off-line (non-real-time) computer simulation of the functions can be carried out, or a combination of computer (non-real-time) and hardware-in-the-loop (real-time) simulation can be carried out.

In virtual prototyping, a virtual world comprises both the VR model of a product under development (that is virtual prototype), and the model of the product's target environment, called virtual target environment [SVA 98]. Virtual worlds, the VR models, are highly advanced human-computer interfaces based on virtual reality techniques, providing a strong intuitive sense of reality and the means to change the world through meaningful interaction. Virtual worlds may include various levels of reality, thus enabling us to concentrate on any desired aspects of the product.

## Manufacturability and Matrix of Competencies

Industrial companies must have a virtual presentation of its kernel competencies using, for example, appropriate Internet servers as brokers of these competencies. Each company has its own competency portfolio, developing and enriching it continuously, and supplying then virtually. At this point, Manufacturability Analysis is an essential step in defining VM. This is a measure of the effort required to manufacture the part according to the design specifications (estimate the manufacturing time and cost). Given a computerized representation of the design (i.e. a solid model) and a set of manufacturing resources, the automated manufacturability analysis problem can be defined as follows: Determine whether or not the design attributes (e.g., shape, dimensions, tolerances, surface finishes) can be achieved. Two alternatives appear:

- the design is found to be manufacturable, and then determine a manufacturability rating, to reflect the ease (or difficulty) with which the design can be manufactured;
- the design is not manufacturable, then identify the design attributes that cause manufacturability problems.

Linked to the Manufacturability Analysis, there is a Manufacturing Resource Matrix. In this case the system uses a tool for identification of potential participants to supply manufacturing parts of virtual enterprise it identifies industrial that are able to manufacture automatically a part and have the corresponding capacity to do it.

Finally, we have the Parts Characterization Matrix, another VM tool, which must operate automatically across the distributed manufacturing network.

## Virtual Partnerships, enterprises

VM as a powerful capability would take into account all of the variables in the production environment from shop floor processes to enterprise transactions. In other words, VM will accommodate the visualization of interacting production processes, process planning, scheduling, assembly planning, logistics from the line to the enterprise, and related impacting processes such as accounting, purchasing and management.

VM is linked to the Virtual Prototyping, the Virtual CAD and Virtual CAM made most of the time by simulation (Stefanescu, et al., 1998). But these technical solutions need to be completed by a special attention on people working on them. The first problem here is the building of the appropriated engineering teams. The second

problem is that of the building of virtual partnerships forming Virtual Enterprises (VE). In both cases the manufacturability and skill matrix are the essential constituents. The constitution of VE, for example, needs an electronic presentation and qualification partners. Here are intervening virtual spaces because all this process exists only in that space.

But another double challenge is appearing: using some techniques of virtual reality to build, evaluate and test VE, and to combine it with the ad hoc techniques developed within the framework of VE building such as, for example, the VE algebra that we can produce according to every project. We think VE exists only by project. New research problems emerge, one of them being that considering communities as virtual reality entities. The separated fields must be then integrated, creating a virtual lifecycle. Figure 4 illustrates the relations between all these notions.

Note. Within the framework of the VEGA project, researchers made a model of virtuality with three independent dimensions : location, virtuality of the interface, and the virtuality of structure. This question is studied in the case of P. Sieber study the computer services in Germany [SIE 97]. Later the study includes Switzerland companies. Figure 7 illustrates one case of the computer service industry. Three axes represent market experience, work configuration and competence leverage. For each the

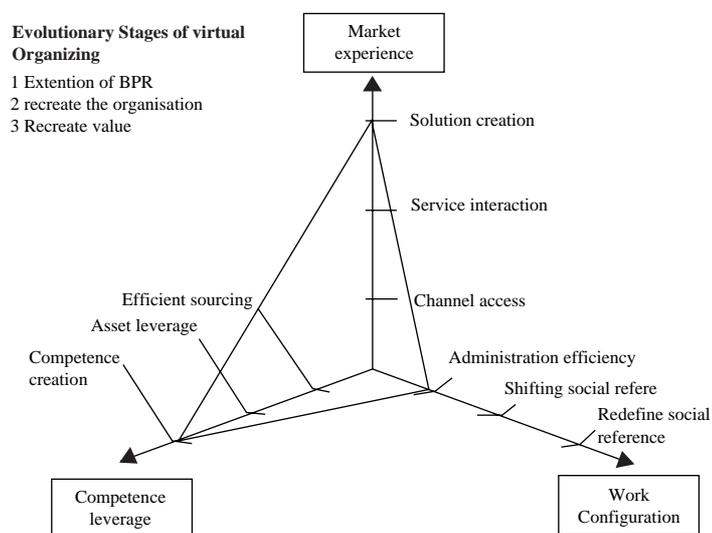


Figure 7: Virtualisation degree of company following three axes, [SIE 97].

author distinguishes three levels as indicated in figure 8. Here virtualisation has priorities such minimal level of experience (market experience) and competence leverage. The hyperplan must evolve to reach extreme values on each axis. This is optimal virtualisation.

## Conclusion

As we see, the partial vision is determined by separated works done by different communities taking the VM direction. To advance we propose integrate these solutions into the hyperspace dimensions. Figure 8 illustrates a widening of virtual towards hyperspace dimensions. We go from virtual prototyping to virtual enterprise [RIC 03].

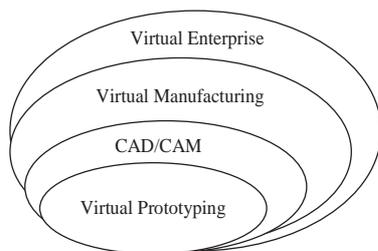


Figure 8: Widening virtual towards hyperspace dimensions.

We cross now this with higher dimension hyperspaces and we will see new problems and envision new solutions. Let us think about only one case: CAD and Cyberenergy and Cyberentropy. We call it CyberCAD. Classical designers and intelligent agent designers perform CyberCAD. The question here is to design the working conditions of both in orders to get a higher cyberenergy possible. Designers will create higher cyberenergy if they have more performing intelligent agents to help them. Then they will work using a common format, say STEP or another one, which must optimise the existing CAD/CAM Exchange formats in order to reduce the cyberentropy. And doing so we can take prototyping, enterprise, etc. In practical manner, we can do a matrix with just two columns. One containing elements belonging to the hyperspace and 5D. This is Hyperspace or 5D column. The other contains the elements concerning VM. This is the VM column. Then we can compare easily to understand where

we are really in using potentialities and capabilities of the hyperspaces for doing VM.

## References

- [BER 01] BERGER P., *Traité de systématique digitale*, Paris, 2001.
- [CAL 97] CALLET P., *L'atelier du virtuel ou le projet ARTESIEN*. Internal Report. Laboratoire Productique et Logistique – Ecole Centrale de Paris, 1997.
- [CAS 00] CASTRO A., *Supragnition theory*. Working paper, Universidad de Concepcion, Chili, 2000.
- [GUP 95] GUPTA S., HERMES : A Research Project on the Commercial Uses of the Web. In : *Journal of computer-Mediated Communication*, Vol. 1 N° 3, 1995.
- [HOF 95] HOFFMAN D., T. Novack, and P. Chatterjee : *Commercial Scenarios for the Web : Opportunities and Challenges*. In : *Journal of Computer-Mediated Communication* Vol. 1 N° 3, 1995.
- [LIN 95] LIN E., MINIS I., NAU D.S., REGLI W.C., *Contribution to Virtual Manufacturing Background Research*, [www.isr.umd.edu/Labs/CIM/vm/report/report.html].
- [PAL 98] PALLOT M., SANDOVAL V., *Concurrent Enterprizing*, Kluwer Academic Press, Boston, 1998.
- [RIC 03] RICHIR S., SAMIER H., SANDOVAL V., *Les cybertechnologies dans les entreprises industrielles*, Hermes-Science, Paris, 2003.
- [SAA 99] SAADOUN M., SANDOVAL V., *Virtual Manufacturing Communication to VRIC '99*, Proceeding of VRIC, Laval, France, June 1999.
- [SAR 95] SARKER R., KARIM M., HAQUE A., "An Optimum Batch Size for a Production System Operating Under a Continuous Supply/Demand", *International Journal of Industrial Engineering*, vol. 2, no. 3, pp189-198, USA, 1995.
- [SIE 97] SIEBER P., GRIESE J., *Virtual organizing as a strategy for the "Big Six" to stay competitive in the global market* in *Proceedings of Thirtieth Hawaii International Conference on System Sciences*, 7-10 January, 1997.
- [SAM 01] SAMIER H., SANDOVAL V., *Five imension and digital relativity theory and practice*, *Communication on 5D to the Special Professional Session, VRIC 2001*, Laval, France, May, 2001.
- [SAM 01] SAMIER H., SANDOVAL V., TARAVEL B., *5 Dimension, theorie et pratique*, Conference CONFERE, Marrakech, Maroc, July, 2001.
- [SVA 98] SVASTA and all *Computer aided design of electronic modules*, Technical Publ. House, Bucharest, 1998.

## VERS UN MOTEUR DE RECHERCHE ÉVOLUÉE DE DOCUMENTS MULTIMÉDIA PAR LE CONTENU TOWARDS A SOPHISTICATED MULTIMEDIA DOCUMENTS SEARCH ENGINE

C. Golbreich  
 Université Rennes 2  
 6 av. Gaston Berger,  
 35043 Rennes Cedex  
 Christine.Golbreich@uhb.fr

## Résumé

Il existe deux approches principales pour l'indexation et la recherche de documents audiovisuels. La première, dite « content-based search », est une approche numérique qui repose sur des caractéristiques visuelles des images. La seconde, est une approche symbolique qui utilise généralement des mots clés. Le problème est que ni l'une ni l'autre n'est satisfaisante pour un utilisateur qui cherche des documents correspondant à un contenu sémantique de plus haut niveau. Cet article expose les principes et premières spécifications d'une plateforme de recherche évoluée de documents multimedia par le contenu qui vise à dépasser ces limitations. L'idée est de réaliser un moteur de recherche *intelligent hybride*, c'est-à-dire qui repose d'une part sur une indexation et une recherche conceptuelle à partir d'une ontologie de concepts, et qui combine d'autre part les techniques propres à chaque média. Un premier outil de recherche « hybride », « par concepts », et « similarité » a permis de commencer à tester ces idées. Les résultats de ces premières expérimentations sont présentés.

## Mots Clef

Indexation et recherche par le contenu, Multimédia Recherche documentaire, Indexation conceptuelle, Ontologie, Moteur hybride, Logique de description, Raisonnement à partir de cas.

## Abstract

There are two main approaches for multimedia documents indexing and search : a numerical approach based on

the visual features of the pictures, and a symbolic approach based on keywords. None of them provide the user with capacities for finding documents corresponding to a high level semantics content. This paper describes the main principles and first specifications of a framework dedicated to an audiovisual documents sophisticated content-based search. The idea is to build an intelligent and hybrid search engine, based on conceptual indexing and search, from a concepts ontology, and which makes possible to combine the specific techniques of each media. A first hybrid, concept and similarity-based tool has been developed in order to test these ideas. The first results of the experiments done are described.

### Keywords

Content-based indexing and search, Multimedia, Document search, Hybrid engine, Description logics, Case based reasoning.

### 1 Introduction

Il existe deux approches pour l'indexation et la recherche de documents audiovisuels. L'approche « visuelle » porte sur le contenu physique de l'image. Elle est fondée sur des caractéristiques objectives d'apparence des images : couleur, forme, texture, etc. [17]. L'avantage est que les descripteurs du contenu sont calculés automatiquement à partir des images. L'inconvénient est le décalage entre le bas niveau des descripteurs et le haut niveau sémantique des requêtes des utilisateurs. La seconde approche concerne le contenu sémantique. Elle utilise en général des *mots-clé* et/ou des *textes* libres pour décrire le contenu des documents audiovisuels. L'avantage est que ces descripteurs, sont porteurs de « sens ». Par contre, dans la plupart de ces systèmes les mots-clés, sont définis dans un dictionnaire à plat, ou au mieux organisés dans un thésaurus hiérarchisé, d'où les problèmes de bruit ou silence des moteurs de recherche classiques liés au caractère fortuit des mots-clé, à leur imprécision, à un pouvoir d'expression limité, à la rigidité du modèle du dictionnaire / thesaurus, à la difficulté de sa

maintenancet de garantie de sa cohérence.

Les moteurs de recherche existants [1][6][7][11][26][20] sont généralement basés sur l'une ou l'autre de ces approches et leurs améliorations. En bref, pour trouver un document sur une tomate rouge, les uns font une recherche d'images d'un cercle rouge, les autres une recherche par mot-clé ou plein texte avec le mot « tomate ». Le principal problème est que dans de nombreuses applications, l'utilisateur recherche de documents (séquences, vidéos, images) qui correspondent à un sujet donné. Par exemple, un documentaliste de la vidéothèque de TF1 recherchera des sujets sur « le Président chinois en visite en France en octobre 1999 », un professeur d'Histoire cherchera dans la vidéothèque de son université des séquences vidéo sur « l'exploitation du lin en Bretagne au 16<sup>ème</sup> siècle » etc. Il est donc nécessaire d'offrir des outils et techniques qui permettent, pour de telles requêtes de haut niveau d'abstraction, de trouver les documents souhaités.

J'ai proposé pour cela de développer une plate-forme de recherche « évoluée » de documents audiovisuels par le contenu. Son but est d'offrir un moteur plus puissant que les moteurs de recherche de documents audiovisuels existants. L'idée clé est de réaliser un moteur de recherche *intelligent hybride* qui repose d'une part (i) sur une indexation et une recherche *conceptuelle*, c'est-à-dire fondée sur des concepts, et d'autre part, (ii) sur une indexation et une recherche hybride, c'est-à-dire qui combine les techniques propres à chaque média :

(i) Indexation et recherche « conceptuelle ». Pour dépasser les limitations de moteurs classiques, les documents multimédia (images et textes ou vidéos) ne seront pas annotés par des mots-clé mais par des descriptions formelles. Le moteur sera donc basé sur un index de concepts représentés formellement, et non des mots, reliés par des relations qui ne seront pas qu'arborescentes. La prise en compte d'une telle sémantique oblige à revoir le fonctionnement des moteurs de recherche, qui ne peuvent plus se contenter de simples recherches de termes, mais doivent être dotés de capacités d'inférence grâce à des méca-

nismes d'exploitation de « connaissances ».

(ii) Indexation et recherche « hybride ». La complémentarité des approches visuelle et sémantique invite à les combiner afin de profiter des résultats acquis dans chacun des domaines pour augmenter la puissance du moteur, en utilisant à la fois les informations textuelles, et/ou conceptuelles précisant le contenu sémantique des vidéos, personnes, lieu, action etc. et les caractéristiques d'apparence visuelle des images, couleur, forme, texture de l'image, ce tant pour l'indexation des documents que pour la formulation des requêtes et la recherche. La prise en compte d'une recherche basée sur des descripteurs et des techniques à la fois symbolique et numérique oblige à revoir le fonctionnement des moteurs de recherche.

Dans une première partie, cet article présente les deux idées clé du moteur de recherche *intelligent hybride* : l'indexation et la recherche « conceptuelle » reposent sur un modèle formel du domaine (ontologie) (§2.1), l'intégration intelligente d'informations nécessaire à un traitement multimédia repose sur une approche de type médiateur (§2.2). La représentation employée pour cela est décrite (§3). La deuxième partie (§4) expose les résultats de premiers travaux réalisés dans le cadre d'un stage de DEA [25] pour tester la faisabilité de ces idées.

### 2 Moteur de recherche intelligent hybride

Pour augmenter la puissance du moteur, deux fonctionnalités principales sont attendues : pouvoir faire une recherche (i) plus « intelligente » et (ii) multimédia. Cette section décrit les techniques proposées pour y répondre : le moteur sera plus « intelligent » grâce à une technique d'indexation et de recherche conceptuelle inspirée de l'intelligence artificielle qui rendra possible des inférences automatiques. Le moteur sera « hybride » grâce à une architecture de type médiateur qui permettra d'intégrer les différents types de données et d'utiliser conjointement les techniques propres à chaque média.

## 2.1 Indexation et recherche « conceptuelle »

Un moteur plus intelligent doit satisfaire deux exigences principales : pouvoir faire des requêtes plus fines et des inférences automatiques.

### 2.1.1 Requêtes et inférences

Le premier besoin est de pouvoir répondre à des requêtes plus fines, qu'elles portent sur le contenu sémantique et/ou visuel des vidéos, par exemple une requête du type :

- trouver « un gros plan en noir et blanc du Président chinois en visite en France en octobre 1999 »

c'est-à-dire, schématiquement représentée en logique du 1<sup>er</sup> ordre par :

*video* (V), *plan* (P), *part-of* (V, P), *résumé* (P, R), *contenu* (R, C), *agent* (C, Pr), *président* (Pr), *nationalité* (Pr, chinois), *action* (C, visiter), *date* (C, octobre 1999), *lieu* (C, France), *gros-plan* (P), *noir&blanc* (P)

La deuxième nécessité est de pouvoir faire des inférences automatiques, par exemple :

- soit la requête ci-dessus, sachant que Jiang Zemin est Président chinois en 1999 le moteur retrouvera les plans indexés par Jiang Zemin
- soit la requête : trouver une vidéo sur « Xavière Tibéri avec son époux » : grâce à la connaissance que Xavière Tibéri est l'épouse de Jean Tibéri (epouse (Xavière Tiberi, Jean Tibéri), ou que la relation epouse est inverse de epoux (epouse(x,y) epoux (y,x)), le moteur retrouvera les vidéos indexées par Xavière Tibéri et Jean Tibéri, ou Jean Tibéri et son épouse.
- soit la requête : trouver une vidéo sur « Yves MONTAND âgé » : grâce à la base de connaissances lui indiquant que Yves MONTAND est né en 1921 et définissant une personne âgée comme ayant plus de 60 ans, le moteur limitera ses recherches aux bases de vidéos réalisées à partir des années 1981

C'est ce qui m'a conduit à proposer une plate-forme générique d'indexation et de recherche à base de « connaissances ».

### 2.1.2 Plate-forme d'indexation et de recherche basée sur une ontologie

La plate-forme visée est un environ-

nement générique (*framework*) dédié à la construction de serveurs de documents pour des vidéothèques numériques (Figure 1)

L'approche proposée s'inspire du principe fondamental des systèmes à base de connaissances : séparer les connaissances (la définition des concepts, des individus, des relations, la description des vidéos) et les mécanismes d'exploitation de ces connaissances pour répondre aux requêtes. L'idée clé est d'utiliser pour l'indexation et la recherche des documents non des mots mais une ontologie de concepts du domaine.

L'ontologie du domaine, souvent appelée modèle du domaine, est un modèle formel du domaine décrivant les concepts du domaine et leurs relations. Les vidéos sont annotées à l'aide de marqueurs « ontologiques » pertinents issues de cette ontologie. Grâce à ces annotations leur contenu est rendu accessible en tant que faits. L'utilisation d'ontologies pour la recherche d'information présente certains avantages par rapport aux méthodes basées sur de simples mots-clé. Une ontologie fournit un vocabulaire commun pour exprimer des informations sur les contenus des documents. Cette même ontologie peut être utilisée pour formuler les requêtes. En plus, elle permet de spécifier des relations entre concepts exprimant des implications comme : *adroite* (Obj1, Obj2) => *agauche* (Obj2, Obj1), ou *epouse* (X,Y) <=> *epoux* (Y,X), ou encore *personneagee* (X) => *age* (X) > 60. Ces connaissances fournissent un moyen de déduire des informations qui n'ont été spécifiées qu'implicitement.

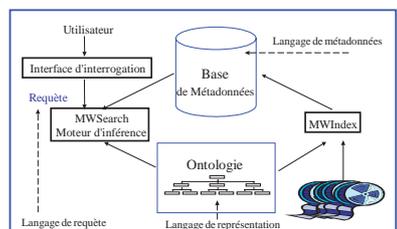


Figure 1 - Indexation et recherche « conceptuelle »

L'autre avantage important d'une telle approche déclarative, est que pour construire un nouveau serveur de vidéos il suffit de définir deux composants spé-

cifiques à l'application : la base de métadonnées c'est-à-dire les données décrivant le contenu des vidéos, et l'ontologie, c'est-à-dire le modèle du domaine décrivant la classification hiérarchique des concepts du domaine et leurs relations (§3.2.1). Les algorithmes de réponse aux requêtes nécessaires au moteur d'inférence (MWSearch) et les autres outils, interfaces et outil d'indexation (MWIndex), sont généraux. Ils ne dépendent que des langages de représentation. Ils seront donc définis une fois pour toutes et réutilisés pour chaque serveur à construire.

## 2.2 Intégration Intelligente d'Informations

Un moteur hybride doit permettre à la fois d'intégrer différents types de données pour pouvoir faire des requêtes multimédia et multimodales et d'utiliser conjointement les techniques d'indexation et recherche propres à chaque média.

### 2.2.1 Requêtes multimédia et multimodales

Le premier besoin est de pouvoir répondre à des requêtes plurimédia qu'elles portent sur les données réelles (i.e. image, plan, texte) ou sur les métadonnées associées, par exemple une requête du type :

- Trouver un gros plan noir et blanc de Jiang Zemin parlant des droits de l'homme schématiquement représentée par :

*video* (V) et *part-of* (V, P) et *grosplan* (P) et *noir&blanc* (P) et *image-clé* (P, I) et *figure* (I, Jiang Zemin) et *résumé* (P, R) et *contenu* (R, C) et *agent* (C, Jiang Zemin) et *action* (C, parler) et objet (C, droits de l'homme).

Contrairement à l'exemple §2.1.1 les prédicats en gras ici désignent des prédicats qui font appel à des outils spécifiques de traitement d'image ou vidéo pour retrouver un gros plan, un plan noir et blanc, une image-clé.

Le deuxième type de requête auxquelles il faut pouvoir aussi répondre sont des requêtes « bimodales » (ou plurimodales) sur image et métadonnées comme par exemple :

- Trouver une vidéo où il y a un gros plan de parlant des droits de l'homme, qui est for-



malisée comme ci-dessus mais ici le prédicat « figure » fera appel à un moteur de recherche d'images *figure* ( P,



)  
La solution que j'ai proposé de retenir pour réaliser un moteur de recherche hybride permettant de traiter ces requêtes est une architecture de type médiateur (Figure 2).

### 2.2.2 Architecture médiateur

Ce type d'architecture [22] issu de travaux récents à la frontière des bases de données et de l'intelligence artificielle est une des architectures qui a été proposée pour traiter le problème d'intégration d'informations. Un médiateur est une façon d'intégrer différentes sources d'information qui traitent d'entités du même monde réel.

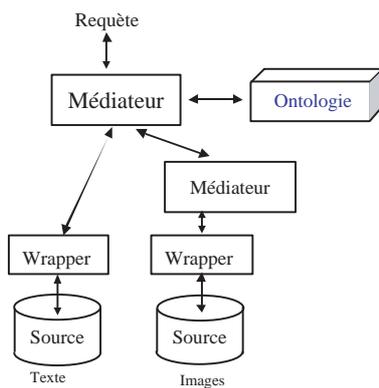


Figure 2 - Architecture médiateur

Le médiateur utilise les sources d'information brutes correctement interfacées par un « wrapper » et /ou d'autres médiateurs pour répondre à ses requêtes. Il fournit à l'utilisateur un accès intégré à des sources d'informations hétérogènes et distribuées, grâce à une interface uniforme, qui lui donne l'impression qu'il utilise un système centralisé. Un tel système (e.g. Information Manifold [16], TSIMMIS [8], Occam [23], Internet Softboat [9] InfoSleuth [2] etc.) se charge, pour une requête donnée, de filtrer automatiquement les sources d'informations pertinentes et de combiner leurs informations. En général, de tels outils engendrent pour une requête de l'utilisateur, à partir de l'ontologie du domaine et de la description des sources d'infor-

mation, un plan de requêtes grâce à un algorithme de planification.

La Figure 2 présente un médiateur pour accéder à différentes sources de données : Video, Image, Texte, Document XML, qui peuvent être stockées sous forme de bases de données relationnelles (RDB), objet (OODB), entrepôts XML, pages HTML, fichiers structurés etc Les « ontologies » jouent un rôle central dans un tel système.

### 2.2.3 Le médiateur MW

Le système MW qui a été spécifié comprend différents composants.

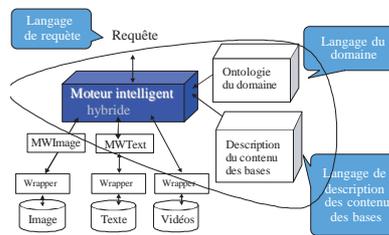


Figure 3 - Système MW

- **MWInter** : un agent d'interface MWInter (MW-MultiMedia-Query Interface) pour spécifier les requêtes multi-média et/ou multi-modales, afficher les résultats
- **MWImage** : un agent de recherche image MWImage (MW-Image-Query) pour réaliser les tâches de traitement d'images. Il aura plusieurs fonctionnalités notamment :

1. Traitement les requêtes d'image par similarité visuelle
2. Identification de régions dans une image (homogènes par rapport à différents critères visuels (signatures))
3. Identification d'objets

- **MWText** : un agent de recherche MWText (MW-Text-Query) sert à faire des recherches sur les textes d'annotations ou d'incrustations.

#### • MWtal

L'agent de traitement linguistique MWtal (MW-Linguistic management agent) sert à gérer les relations entre les mots et les concepts de l'ontologie. Ces relations comme par exemple, la synonymie, la polysémie seront exploitées pour reformuler les requêtes.

### • MMultisearch

L'agent de recherche intelligente hybride MMultisearch (MM-search agent) sert à traiter les requêtes hybrides portant sur le texte, les images, les vidéos, les mots-clé, les concepts. Les mécanismes de ce module sont ceux d'un moteur d'inférence complété d'un contrôleur basé sur un algorithme de planification.

## 3 Représentation des connaissances

Deux points sont essentiels. Premièrement, il faut disposer de langages adaptés pour représenter (i) les requêtes (ii) l'ontologie du domaine, (iii) le contenu des documents, (iv) les contenus des bases de documents (vues), (v) les compétences des spécialistes de chaque média :

- le *langage des requêtes* servira à spécifier les requêtes qui peuvent être exprimées au serveur de documents.
- le *langage du domaine* servira à décrire les ontologies du domaine
- le *langage des métadonnées* servira à représenter les données décrivant le contenu des vidéos.
- le *langage de ressources* servira à donner une description des bases, c'est-à-dire pour chaque base de documents un modèle conceptuel des données qu'elle contient.

Deuxièmement, il faut définir des mécanismes de planification qui, pour une requête, permette de sélectionner les spécialistes des média concernés et de combiner leurs résultats.

### 3.1 Langage de représentation

#### 3.1.1 Quel langage médiateur ?

La plupart des systèmes de type médiateur recommandent un formalisme déclaratif associé à une sémantique claire, c'est pourquoi les langages de représentation fondés sur la logique sont très utilisés. Il existe toutefois deux tendances parmi les systèmes existants :

- les systèmes qui utilisent des langages relationnels à la Datalog, bien adaptés à représenter les relations d'arité quelconque et exprimer des requêtes fines. En particulier, la liaison de variables permet par exemple de faire la distinction entre deux requêtes comme : *femme(x)*, *agent-de-police(x)* : pour trouver une vidéo dont le

sujet porte sur le métier d'agent-de police pour une femme et d'autre part, femme(x), agent-de-police (y), mari(x,y) : pour trouver un sujet sur une femme dont le mari est agent de policere contrairement à une recherche sur les mots-clé « femme, agent-de-police ».

- les systèmes qui utilisent des langages de type objet ou réseaux, bien adaptés à la structuration des connaissances.

L'inconvénient de ces langages est qu'ils n'ont pas de sémantique claire et rigoureuse, et pas toujours de sémantique opérationnelle, permettant de leur associer des algorithmes d'inférence (de classification ou d'héritage). Les logiques de description, dérivées de la logique des prédicats, des réseaux sémantiques et des frames, font exception puisqu'elles offrent une sémantique formelle associée à des algorithmes de classification

Certains systèmes comme Tsimmis, Garlic, Information Manifold, Infosleuth ont des langages mixtes : Datalog pour les requêtes et objet pour le domaine, ou SQL pour les requêtes et KIF pour le domaine. Le problème qui se pose alors est l'articulation entre les deux langages ([15]).

### 3.1.2 Logique de description

On a choisi d'utiliser un formalisme logique de la famille des logiques de description pour représenter les ontologies, les contenus des vidéos, les requêtes, les compétences des agents. Un moteur d'inférence exploitera la sémantique formelle du langage de représentation et permettra alors un raisonnement automatique. Eventuellement, le formalisme des graphes conceptuels pourra être utilisé pour les composants dédiés au traitement du langage naturel.

Les logiques de description sont des formalismes de représentation ayant pour origine KL-ONE [4] lui-même issu des réseaux sémantiques et des frames. Ce sont des formalismes logiques intéressants pour la représentation de connaissances structurées. Une logique de description comprend plusieurs types d'entités : des concepts, des rôles, des individus. Les concepts représentent des classes d'individus d'un domaine et les rôles des relations binaires entre les

C, D ← A	// A est un concept primitif
TOP   BOTTOM	// concept le plus général
	// le plus spécifique
C & D	// conjonction de concepts
(# r : C)	// co-domaine du rôle r
ONE-OF {I1 .. In}	// liste ex/haustive des instances
MIN μ	// μ nombre réel
MAX μ	
r FILLS {I1 .. In}	// valeurs associées au rôle r
r AT-LEAST n	// restriction de cardinalité n
r AT-MOST n	

Figure 4 - Syntaxe de C-Classic [3]

<b>Concepts</b>
ENTITE, ANIME, INANIME, ANIMAL, PERSONNE, METIER ...
PRESIDENCEREP FONCTIONPOLITIQUE
PERSONNEPOLITIQUE := PERSONNE
& (# a-pour-fonction FONCTIONPOLITIQUE)
& (#debut-fonction DATE) & (# fin-fonction DATE)
PSDT-REPUBLIQUE := PERSONNE & (# a-pour-fonction
PRESIDENCE-REP)
HOMME := PERSONNE & (# sexe FILLS « masculin »)
HOMME-AGE := HOMME & (# age ENTIER MIN 60)
<b>Rôles</b>
sexe, age, a-pour-fonction, debut-fonction ...
<b>Individus :</b>
Tutu = PERSONNE & (FILLS age 12) & (FILLS sexe "masculin")

Figure 5 - Exemples de concepts, rôles, individus

individus. Concepts et rôles sont soit primitifs soit complexes. Les concepts et rôles complexes sont définis à l'aide de « descriptions » construites à partir d'un certain nombre de constructeurs. Par exemple, la syntaxe du langage C-Classic est la suivante où A dénote un concept primitif, les Ri dénotent des rôles primitifs, C et D des concepts arbitraires et R un rôle arbitraire.

Un système développé avec une logique de description comprend deux composantes : une composante terminologique, la T-Box, qui définit les concepts du domaine, et leurs liens de subsomption, et une composante assertive ou A-Box, qui contient la définition des individus et des règles (Figure 5).

Les logiques de description ont été choisies en raison de leurs services de raisonnement puissants, intéressants à la fois pour la construction, la maintenance des ontologies et pour le catalogage. Les deux mécanismes d'inférence principaux des logiques de description sont (1) la classification et (2) la reconnais-

sance d'instances, tous deux basés sur les calculs des relations de subsomption. La relation de subsomption exprime qu'un concept est plus général qu'un autre : un concept A subsume B si A est plus général que B (on note  $B \subseteq A$ ). La classification automatique permet d'insérer automatiquement un concept dans la hiérarchie. La reconnaissance d'instances permet d'identifier pour un individu les concepts les plus spécifiques dont il est instance.

(1) La classification est un service clé crucial pour la construction automatique de hiérarchies, et la mise à jour des ontologies de concepts nécessaires à l'indexation. En effet, dans de nombreux domaines documentaires, par exemple pour la vidéothèque des journaux télévisés de TF1, il est indispensable de disposer d'une ontologie « ouverte » que les documentalistes peuvent facilement enrichir et faire évoluer. En effet, si lors de l'indexation, un concept manque dans la terminologie, ce qui se produit inévitablement dans des domaines vastes (tel par exemple

celui de la vidéothèque de l'information télévisée), il faut pouvoir le rajouter et que la mise à jour de l'ontologie se fasse alors automatiquement. Ainsi par exemple, si un nouveau concept est créé pour président de la république : PSDT-REPUBLIQUE := PERSONNE & (# a-pour-fonction PRESIDENCE-REP) est créé (Figure 5), du fait que PRESIDENCE-REP  $\subseteq$  FONCTIONPOLITIQUE, l'algorithme de calcul des subsomptions déduit automatiquement d'après leurs définitions que le concept PSDT-REPUBLIQUE est subsumé par le concept PERSONNEPOLITIQUE : PSDT-REPUBLIQUE  $\subseteq$  PERSONNEPOLITIQUE.

(2) La reconnaissance d'instances est un service clé pour l'indexation et le « catalogage » basés sur les ontologies.

En effet, de nouveaux documents doivent être stockés en permanence (e.g. chaque jour tous les sujets de trois journaux télévisés). En raison de l'importance du nombre de documents, le catalogage automatique des documents est précieux. La possibilité de construire automatiquement des hiérarchies de concepts et d'y rattacher des individus sont donc des services précieux pour l'organisation d'une base contenant des masses de documents vidéos (e.g. 250 000 fiches de données existantes de la vidéothèque de TF1).

### 3.2 Ontologie, base de documents, requêtes

Toutes les connaissances, modèle du domaine, base de documents, requêtes, sont représentées dans la logique de description *C-Classic*. Nous avons choisi d'utiliser *C-Classic* en raison des avantages suivants : son pouvoir d'expression apparaît suffisant, l'algorithme pour le calcul de subsomption est polynomial. De plus, des algorithmes intéressants basés sur les formes normales ont été proposés par [21] pour le raisonnement à partir de cas et l'apprentissage de concepts. *C-Classic* est un langage relativement simple et facile à utiliser. Conçu pour des applications nécessitant rapidité de réponse aux requêtes, le langage de description de concepts est moins riche que dans d'autres langages de logique de description, mais en contrepartie, les tests de subsomption, la classification et l'instanciation sont rapides. Par rapport aux

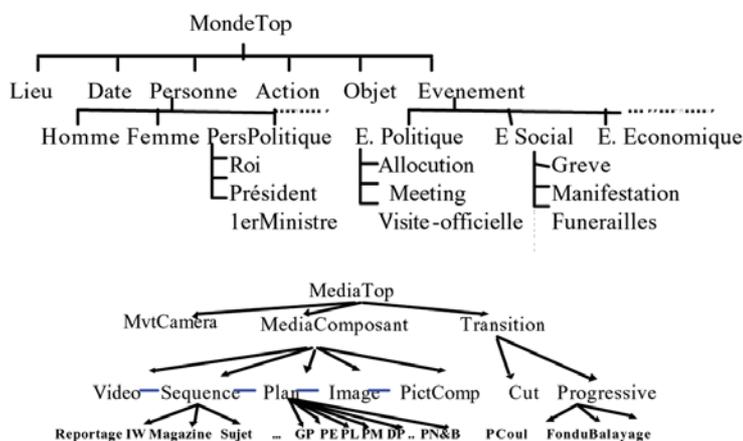


Figure 6 - Ontologies du domaine

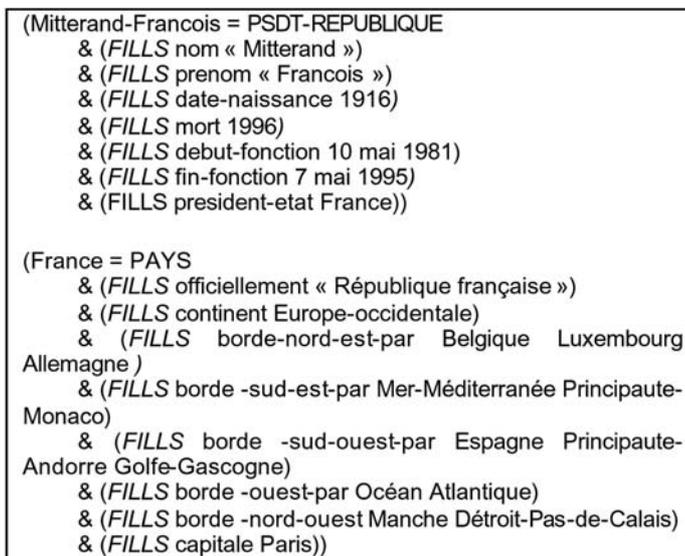


Figure 7 - Exemples d'individu en C-Classic

langages de SGBD traditionnels relationnels, il présente des possibilités intéressantes, puisqu'il permet de construire des descriptions génériques (concepts primitifs ou définis) ou individuelles (instance), d'associer une sémantique bien définie à chaque construction, d'organiser les descriptions par niveau de généralité, et de procéder à des inférences, tout en offrant un processus de requête qui reste tractable.

#### 3.2.1 Connaissances du domaine

L'ontologie du domaine, souvent appelée modèle du domaine est un modèle formel du domaine décrivant les concepts du domaine et leurs relations. Ils forment la terminologie. Cette terminologie définit le vocabulaire utilisé

pour décrire les données et les requêtes. Pour un serveur de documents multimédia, il y a deux ontologies principales (Figure 6). La première « World Model » concerne les objets du monde réel propres à l'application, par exemple pour le système d'indexation et de recherche pour la vidéothèque des journaux télévisés de TF1, les objets du domaine sont les concepts liés à l'actualité télévisée organisés en catégories pertinentes pour la recherche documentaire (LIEU, DATE, PERSONNE, ACTION, OBJET, EVENEMENT, HOMME POLITIQUE, PRESIDENT, PAYS, VILLE etc.). La seconde « Media Model » définit les objets du monde des media (e.g. VIDEO, PLAN, SEQUENCE, CUT etc.).

La Figure 5 donne des exemples de représentation de concepts de la T-Box, et la Figure 7 des exemples d'individus de la A-Box.

### 3.2.2 Base de documents multimédia

Les documents à indexer sont des documents multimédia : vidéos, segments de vidéo (e.g. sujets de journaux télévisés), images et textes (e.g. documents AFP). De façon générale, une vidéo comprend plusieurs séquences, elles-mêmes composée de plusieurs plans. Un plan est défini comme « un segment de film compris entre deux coupures successives ». Ainsi, par exemple un journal télévisé JT est composé d'un ensemble de segments audiovisuels traitant chacun d'un sujet Si d'actualité appelé « sujet » et stocké dans la vidéothèque sur une cassette vidéo ou sous forme numérique. Chaque sujet Si est composé d'un ensemble {P<sub>1</sub>, P<sub>2</sub>, P<sub>j</sub>, ... P<sub>n</sub>} de plans.

A chaque plan P<sub>j</sub> est associé (Figure 8) (1) une image fixe index (imagerie ou mosaïque), (2) un texte, descriptif informatif du plan (3) une liste de mots-clé.

#### ▪ Résumé « conceptuel »

Nous proposons d'associer en outre aux contenus de chaque plan vidéo (ou document image & texte) (4) une description formelle structurée, sorte de « résumé conceptuel » du plan qui décrit les personnes, lieu, date, événement journalistique, objets, action, impliqués. Cette description est exprimée en termes d'individus particuliers appartenant aux concepts PERSONNE, LIEU, DATE, EVENEMENT, OBJET, ACTION de l'ontologie ou leurs sous-concepts (Figure 6). Un concept particulier DESCR a donc été défini en C-Classic pour représenter ces informations formellement :

```
DESCR = MEDIATOP
& (#a-lieu LIEU)
& (#a-action ACTION)
& (#a-evenement EVENEMENT)
& (#a-date DATE)
& (#a-agent E-VIVANTE)
```

#### ▪ Contenus visuel et informationnel (Figure 9)

On distingue deux types de contenu selon le point de vue : le contenu « visuel » qui se rapporte aux données que l'on voit et le contenu informationnel qui se rapporte aux données relatives au sujet traité (e.g. sujet d'actualité pour le JT). Ces deux contenus sont liés, parfois identiques, le plus souvent le conte-

nu visuel illustre le contenu informationnel (Figure 9). On note DI<sub>j</sub>, la description du contenu informationnel du plan j et DV<sub>j</sub> celle de son contenu visuel. A chaque plan est donc associée une description structurée représentée par un individu particulier du concept C-Classic DESCRIPTION-DOC en deux parties décrivant l'une le contenu visuel DV<sub>i</sub> et l'autre le contenu informationnel DI<sub>i</sub>, par un individu particulier du concept DESCR.

```
DESCRIPTION-DOC = MEDIATOP
& (# a-description-visuelle DESCR)
& a-description-visuelle AT-MOST 1
& (# a-description-informationnelle DESCR)
& a-description-informationnelle AT-MOST
```

Voici par exemple, la représentation de l'individu PLAN1 indexant le plan de la Figure 8.

Les principaux éléments qui apparaissent dans la description du plan Figure 8, sont pour les données visuelles Ehud Barak et Kofi Annan individus appartenant au concept PERSONNE-POLITIQUE, et pour les données informationnelles, par exemple Jerusalem

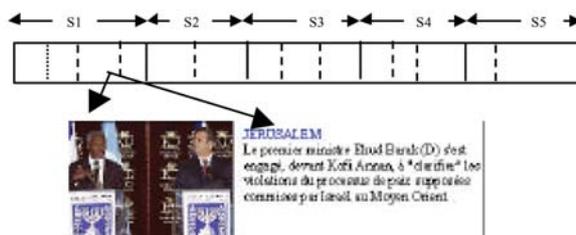


Figure 8 - Segmentation en sujets, plans avec image clé et texte descriptif associés



Figure 9 - Exemple de contenu visuel et informationnel

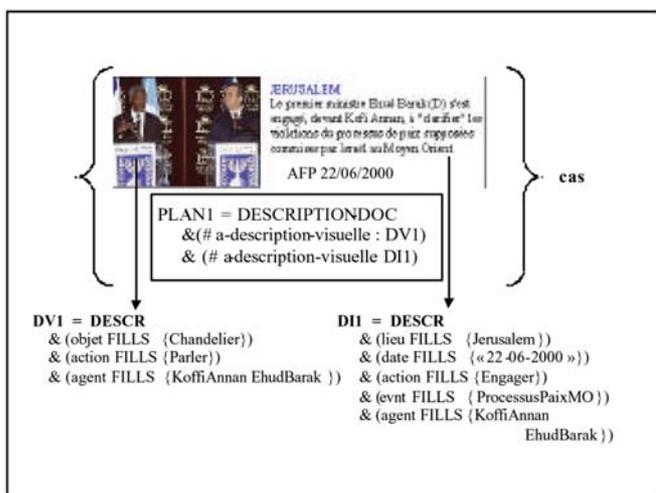


Figure 10 : Individu représentant les contenus du plan

individu appartenant au concept VILLE MOYENORIENT, ProcessuspaixMO individu appartenant au concept EVENEMENTPOLITIQUE, Engager individu appartenant au concept ACTION etc.

Les requêtes sur les contenus des plans sont représentées à l'aide d'un individu particulier du concept DESCRIPTION-DOC définissant les critères de recherche, par exemple un plan « sur le Président chinois en visite en France en octobre 1999 », un plan sur « la défaite de l'OM », un plan où l'on voit « un prince au Moyen-Orient » etc. (§ 4.1.1.2)

#### 4 Méthodes de raisonnement par similarité pour l'indexation et la recherche

Un premiers outil réalisé par Haifa ZargAyouna dans le cadre de son stage de DEA [25] a permis de commencer à tester ces idées et à faire de premières expérimentations. Il utilise une version préliminaire de l'implémentation de *C-Classic* développée au LIPN [13][14] qui fournit un analyseur lexical et syntaxique, réalise les calculs de forme normale, de subsomption par l'algorithme SOLLIN [21][14], et la classification de concepts. Ce travail a permis de mettre en œuvre une indexation et une recherche (i) hybride, (ii) par concepts, et (iii) par similarité :

(i) méthodes numériques et symboliques sont utilisées conjointement pour l'indexation et la recherche de docu-

ments multimédia. Des scénarios de recherche hybride combinant recherche visuelle par le contenu et recherche symbolique par les concepts ainsi qu'un scénario d'aide à l'indexation ont été testés.

(ii) l'indexation et la recherche symbolique reposent sur une approche par concepts, non par mots-clé : les concepts sont représentés dans la logique de description *C-Classic*, ce qui permet de bénéficier de mécanismes d'inférence puissants.

(iii) les méthodes de recherche utilisées reposent non sur un appariement exact mais sur un raisonnement par similarité de type numérique (visuelle) et/ou symbolique (raisonnement à partir de cas).

Les méthodes proposées ont été testées sur une base « jouet » constituée de 120 images et des textes associés, du domaine des actualités. La description complète d'une trentaine de cas a été réalisée en *C-Classic* pour constituer la base de cas nécessaire aux expérimentations. Les différents algorithmes utilisés (PPG, DISSIM, [19], ELENA+ [21] et les scénarios d'expérimentations ont été codés en Java.

#### 4.1 Recherche

Les techniques de recherche utilisent une méthode de recherche par « similarité conceptuelle », plus puissante qu'une recherche par mots clé. Plusieurs scénarios ont été testés : un scénario de recherche par similarité conceptuelle seule, et deux scénarios de recherche

hybride combinant similarité numérique et similarité symbolique.

##### 4.1.1 Recherche par « similarité conceptuelle »

La recherche par « similarité conceptuelle » désigne une recherche par concepts et par similarité, plus précisément par un raisonnement à partir de cas. La recherche de documents par concepts est plus puissante qu'une recherche par mots-clés car les inférences rendues possibles par la formalisation permettent de retrouver plus de documents satisfaisants grâce aux connaissances définies dans l'ontologie. La recherche par similarité est pertinente dans le domaine du multimédia car elle est plus souple qu'une recherche classique par appariement strict entre description des documents et requête. Elle permet de trouver plus de cas, puisqu'elle retourne non seulement les cas correspondants exactement à la requête mais aussi des cas « proches » de la requête. Les résultats obtenus par similarité peuvent être intéressants dans différentes situations. Par exemple, lorsqu'un utilisateur recherche un contenu visuel pour illustrer une actualité, il peut se contenter d'une image sur le même thème (e.g. une image d'un tremblement de terre au Pérou et pas forcément l'image précise du tremblement de terre du 25 juin 2001. Il est souvent difficile de décrire une image ou une vidéo. Un utilisateur qui a une idée vague du document qu'il cherche peut formuler sa requête de façon approximative et être donc satisfait par des solutions approchantes, etc.

##### 4.1.1.1 Méthode

L'outil d'indexation et de recherche conceptuelle réalisé pour les expérimentations repose sur un raisonnement à partir de cas et la formalisation en logique de description qui en est donnée dans [19]. Les concepts de l'ontologie ont été décrits en *C-Classic*. Chaque description de cas est composé de plusieurs parties : (1) une image index du cas (image clé du plan ou mosaïque, ensemble d'images pour un sujet) sur laquelle se fait la recherche d'images similaires (2) une texte descriptif (description informelle en plusieurs parties) sur lequel peut se faire une recherche

basée sur des techniques de langage naturel, (3) une *liste de mots-clé*, et (4) une *représentation formelle* du cas par un individu C-Classic (§ 3.2.1) sur laquelle se fait la recherche de cas similaires du point de vue des concepts<sup>1</sup>. L'ensemble des cas ainsi formés des documents et de leurs annotations forment la base de cas. La requête est également représentée comme un individu C-Classic. Les grandes étapes du raisonnement pour la recherche sont les suivantes :

- Recherche des anciens cas Ci de la base similaires à la requête Co. : grâce au mécanisme de classification, on retrouve les concepts les plus spécifiques auxquels Co appartient et on forme l'ensemble des cas qui appartiennent à au moins un de ces concepts<sup>2</sup>

- Sélection des cas les plus similaires à Co

La recherche des cas les plus similaires repose sur la notion clé du Plus Petit Généralisant de deux concepts (PPG) en logique de description due à [9]. Le PPG est un concept qui représente formellement la similarité de deux concepts par l'ensemble de leurs propriétés les plus spécifiques communes. Calculer le PPGi (Co, Ci) revient donc à trouver le concept le plus spécifique subsumant les deux concepts. L'implémentation du calcul du PPG en C-Classic repose sur l'algorithme de [21] qui s'effectue à partir des formes normales des concepts.

La recherche est faite en plusieurs temps, selon la procédure proposée dans [19]: 1° pour chaque ancien cas Ci calcul du Plus Petit Généralisant PPGi (Co, Ci) avec la requête Co, 2° classement et sélection des PPGi les plus spécifiques, 3° discrimination par DISS en cas de similarité identique.

1. Les parties texte et mot-clés n'ont pas été prises en compte dans ces expérimentations mais sont prévues en particulier pour permettre la compatibilité avec les descriptions par mot-clé existantes des outils antérieurs.

2. Dans ses expérimentations, H. Zargayouna prend tous les cas de la base parce qu'elle n'a pas défini à partir de DESCRIPTION-DOC une hiérarchie de sous-concepts correspondant à une organisation en grandes catégories de documents.

#### 4.1.1.2 Exemples et résultats

Les exemples suivants, issus des expérimentations faites par H. Zargayouna décrites plus en détail dans son mémoire [25] montrent que, dans de nombreuses situations, notre moteur de recherche par concepts est plus puissant qu'un moteur de recherche par mots-clé. Les résultats obtenus avec notre outil sont comparés à ceux du moteur par mot-clés QBIC [20] où les mot-clés choisis pour annoter les images sont les noms des concepts utilisés dans la représentation C-Classic des cas par des individus

- Traitement de requêtes plus fines et prise en compte d'une sémantique plus riche

D'abord notre outil est au moins aussi puissant qu'un moteur classique de requêtes de base de données, puisque lorsqu'une correspondance exacte avec la requête existe dans la base, elle est trouvée. Ensuite, il permet de traiter des requêtes plus fines et de prendre en compte une sémantique plus riche, qu'un moteur par mot-clé, puisqu'il permet de distinguer contenu informationnel et visuel et d'autre part permet d'exprimer une conjonction de critères sur le lieu, l'agent, les personnages etc.

#### Exemples

Soit la recherche d'un plan où l'on voit des données précises e.g. les personnages Ehud Barak et Yasser Arafat. Notre outil, pour la requête Co = DESCRIPTION-DOC & (#a-description-visuelle : (a-agent FILLS { Ehud Barak, Yasser Arafat })), retourne les deux images de la base satisfaisant exactement la requête, tandis que QBIC pour les mots-clé Ehud Barak, Yasser Arafat retourne 6 images dont 4 ne sont pas satisfaisantes, parce que QBIC ne prend pas en compte la conjonction et ne distingue pas contenu visuel du contenu informationnel.

De même pour la requête Co = DESCRIPTION-DOC & (#a-description-visuelle : (a-agent FILLS { Kofi-Annan Arafat })) & (#a-description-info : (a-lieu FILLS {Syrie})) & (#a-action : RESTAURER-PAIX)) on obtient comme unique résultat un plan où on voit Koffi Annan et qui porte sur sa visite le 24 juin 2000

en Syrie pour restaurer la paix, tandis que QBIC fournit plusieurs résultats inadéquats.

- Apports d'inférences automatiques

Soit une recherche générale de documents d'hommes politiques au Moyen Orient, ou des différents Présidents français depuis 1960. Grâce aux mécanismes de reconnaissance d'instances et de classification, notre outil, par exemple pour la requête Co = DESCRIPTION-DOC & (#a-description-visuelle : (# a-agent : PERSONNE-POLITIQUE)) & (#a-description-info : (# a-lieu : MOYEN-ORIENT)) retourne tous les documents d'un homme politique quels que soient la ville, le pays, du Moyen Orient où se passe la scène, soient 6 images de la base. QBIC ne retourne aucun document cars ils sont annotés non par « Moyen Orient » mais par des lieux précis tels Syrie, Jerusalem etc. Il sera possible d'étendre les fonctionnalités du moteur pour affiner la recherche en parcourant l'ontologie, par exemple on pourra préciser qu'on souhaite uniquement des documents d'hommes politiques dans une ville d'un pays précis du Moyen Orient.

- Apports du raisonnement à partir de cas

Lorsqu'il n'y a pas de document dans la base correspondant exactement à la requête, la recherche par similarité renvoie les documents les plus similaires grâce au raisonnement à partir de cas basé sur le calcul du PPG.

#### Exemple

Soit la requête d'un plan où on voit un personnage d'une catégorie donnée dans une certaine région e.g. un prince au Moyen Orient : Co = DESCRIPTION-DOC & (#a-description-visuelle : (# a-agent : PRINCE) & (# a-lieu : MOYEN-ORIENT)). Comme il n'y a pas une telle image dans la base, les résultats retournés sont les images les plus similaires à la requête, c'est-à-dire 6 images Figure 11 où l'on voit un personnage d'une catégorie proche « personnage politique », roi, président, premier ministre, secrétaire ONU au Moyen Orient. On note que la recherche par mots-clé avec le moteur de recherche Qbic ne retourne aucune

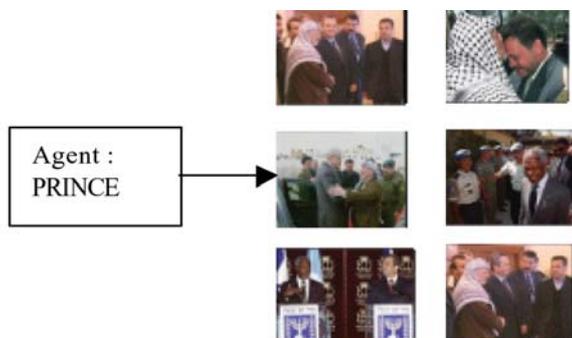


Figure 11 - Recherche de « prince au Moyen Orient »

image, car aucune n'est décrite avec le mot-clé « Prince ». Or, par exemple pour illustrer une séquence politique au Moyen Orient, un utilisateur pourrait finalement se satisfaire de ce type d'images, sans que ce soit précisément un Prince.

En conclusion notre moteur de recherche est plus performant car il permet, dans de nombreuses situations, d'éliminer le bruit ou le silence de moteurs classiques par mots-clé.

#### 4.1.2 Recherches « hybride » et « multimodale »

Il est intéressant d'utiliser conjointement moteurs de recherche visuelle et conceptuelle pour bénéficier de leurs avantages complémentaires. Deux scénarios de recherche dite « hybride », où sont exécutées séquentiellement dans un ordre ou l'autre, recherche numérique avec un moteur de recherche visuel et recherche symbolique avec un moteur par similarité conceptuelle, ont été expérimentés.

- Le premier scénario consiste à faire d'abord la recherche conceptuelle pour restreindre l'espace de recherche à un ensemble de documents pertinents par rapport au sujet voulu puis utiliser le moteur de recherche visuelle sur cet ensemble pour affiner alors la recherche sur des critères visuels.

#### Exemple

Supposons qu'un l'utilisateur, par exemple un journaliste, recherche un document de match de foot de l'OM, et choisisse pour image exemple l'image (Figure 12) de joueurs de cette équipe. Comme ce type de recherche repose sur des critères visuels (signatures) les résultats trouvés sont des documents de

personnages (Figure 12) et non de match, ce qui ne correspond pas du tout aux documents voulus. Une recherche directe par un moteur de recherche visuelle peut ainsi conduire à des résultats très éloignés des documents recherchés. Il est préférable dans un tel cas de faire d'abord une recherche conceptuelle (étape (1) Figure 12) par la requête Co = DESCRIPTION-DOC & (#a-description-visuelle : (# a-evenement : MATCH-FOOT)), ce qui permet d'obtenir les 11 documents de match, puis de choisir une de ces images comme image exemple pour à l'étape

(2) affiner la recherche avec le moteur de recherche visuel jusqu'à un document satisfaisant sur l'équipe voulue.

- Le deuxième scénario exécute ces recherches en ordre inverse : une recherche visuelle sert d'abord à restreindre la base de documents sur laquelle opère ensuite le moteur de recherche par concepts.

#### Exemple

Au lieu de faire une recherche conceptuelle en parcourant toute la base de cas pour la requête requête Co = DESCRIPTION-DOC & (#a-description-visuelle : (# a-evenement : DEFAITE-FOOT-OM )), on limite (Figure 13) la recherche conceptuelle (2) aux 6 documents retrouvés par le moteur visuel à l'étape (1), ce qui réduit l'espace de recherche de 4/5 dans le cas de cet exemple.

- Un dernier scénario hybride a été expérimenté dans le cas d'une requête mixte image et concepts dite « multimodale ». En fait, la procédure a consisté non pas à utiliser les deux moteurs en séquence mais à faire l'intersection des

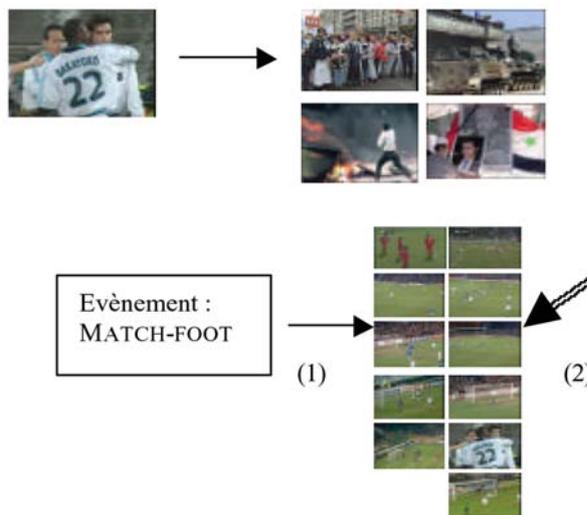


Figure 12 - Haut : recherche visuelle seul, Bas : (1) recherche conceptuelle suivie de (2) recherche visuelle sur le résultat

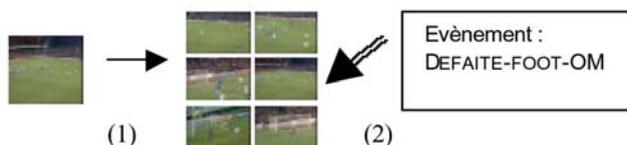


Figure 13 - Recherche visuelle suivie de recherche conceptuelle

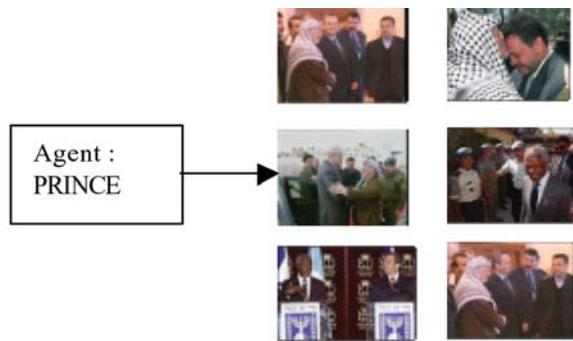


Figure 14 - Requêtes multimodales

résultats obtenus séparément par chacun des moteurs pour les requêtes mixtes suivantes des deux exemples (figure 14).

Sur ces exemples, et en raison notamment de la trop petite taille de la base de documents, les résultats n'ont pas été concluants : pour le premier l'intersection est vide, pour le second on obtient les mêmes résultats que pour la recherche conceptuelle seule. Le scénario de recherche multimodale devra toutefois être approfondi en particulier une perspective intéressante est d'étendre le langage de requête pour pouvoir exprimer une requête multimodale par exemple comme :

Co = DESCRIPTION-DOC & (#a-description-visuelle :

(# a-agent : ) & (# a-lieu : MOYEN-ORIENT))

#### 4.2 Indexation

Une méthode semi-automatique est proposée pour l'indexation. Le but est d'aider à indexer un nouveau document

sans description (le nouveau cas) en se servant des indexations des documents qui existent déjà dans la base (anciens cas de la base). La méthode proposée repose sur un raisonnement à partir de cas. La partie problème du cas est la partie connue du nouveau document. Dans le cadre de nos expérimentations, c'est une image. La partie solution est la partie manquante qu'on cherche à inférer, dans nos expérimentations c'est la description conceptuelle (ce pourrait être les mots-clés etc.).

Les étapes de raisonnement sont celles du cycle classique du raisonnement à partir de cas (Figure 15) : (1) une recherche par le contenu à partir des caractéristiques numériques (couleur, texture etc.) des images permet d'abord de trouver des cas « visuellement » similaires. Deux moteurs d'image, QBIC [20] et Surfimage [18] ont été utilisés pour cela ; (2) l'utilisateur classe les images ainsi trouvées en pertinentes et non pertinentes d'où deux catégories de cas qui serviront d'exemple et contre-

exemples à l'algorithme d'apprentissage Elena+. Cet algorithme d'apprentissage proposé dans [21] permet l'apprentissage d'un concept en C-Classic à partir d'instances positives et négatives de ce concept Il repose sur l'algorithme Eléna de [5] et sur le calcul du PPG[19]. Il est utilisé ici pour apprendre une ou plusieurs descriptions symboliques pour le nouveau cas ; (3) les descriptions proposées doivent être validées par l'utilisateur qui les confirme ou les modifie, par exemple en les complétant ; (4) le nouveau cas ainsi indexé est rajouté à la base.

#### Exemple

La Figure 16 présente le résultat de l'indexation semi-automatique du nouveau document dont on connaît une image index (à gauche) mais pas la description formelle qu'on cherche à inférer. La procédure décrite permet de trouver d'abord à l'aide du moteur d'images 6 images (à droite), classées par l'utilisateur en pertinentes (les 3 qui ont un rapport avec l'image à indexer) et non pertinentes.

A chacune d'elles est attachée sa description formelle. A partir de ces descriptions d'instances positives et négatives, l'algorithme Elena+ construit l'indexation formelle en C-Classic du nouveau cas (donnée de façon informelle en bas). Celle-ci donne les caractéristiques commune des images sélectionnées, c'est-à-dire exprime schématiquement qu'on voit sur l'image Yasser Arafat qui fait un geste diplomatique avec un homme politique lors d'un événement politique dans une ville du Moyen Orient. A l'étape sui-

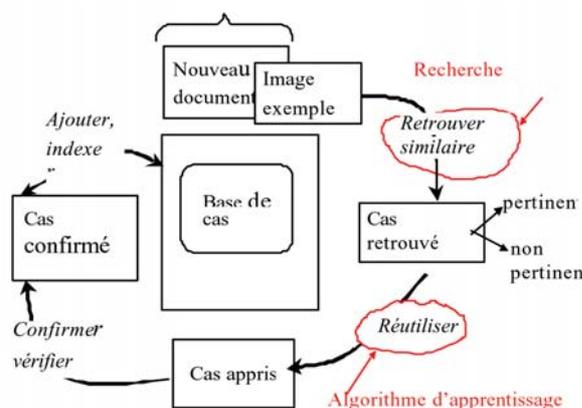


Figure 15 - Cycle du raisonnement à partir de cas

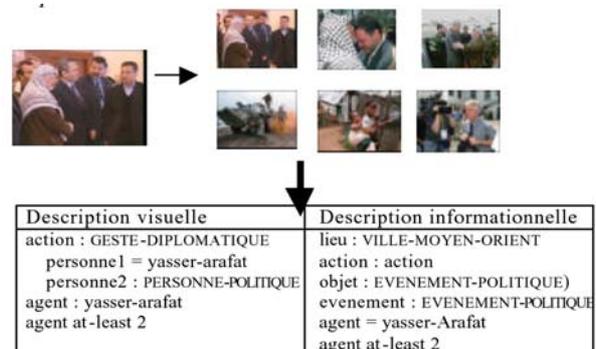


Figure 16 - Indexation semi-automatique

vante, l'utilisateur peut accepter cette description ou l'affiner par exemple en précisant, s'il les reconnaît, qu'il s'agit de Ehud Barak et du lieu Ramallah.

D'autres exemples sont décrits en détail dans [25]. En conclusion, cette méthode d'indexation présente un double avantage. D'une part, par ce procédé, la description formelle des documents en C-Classic est, générée semi-automatiquement, ce qui du point de vue de l'ingénierie des connaissances est un apport important puisque la tâche d'écriture formelle est ainsi considérablement facilitée. D'autre part, l'indexation semi-automatique est une nécessité pour l'indexation de masses de documents comme par exemple dans le domaine de l'actualité télévisée. Il est sûrement souvent préférable d'avoir une indexation plus grossière mais plus automatisée.

## 5 Conclusion

Cet article a présenté les principes et premières spécifications d'une plateforme de recherche évoluée de documents multimédia par le contenu. Le premier outil et les premières expérimentations qui ont été faites mettent en évidence qu'il est possible de développer un moteur de recherche plus intelligent et multimédia, sur la base des principes et techniques décrites. L'outil de recherche réalisé est plus performant que les moteurs classiques car il permet dans de nombreuses situations d'éliminer le bruit ou le silence d'une recherche par mots-clé. Les scénarios de recherche hybride et multimodale ainsi que l'indexation formelle semi-automatique de documents multimédia à partir d'un raisonnement à partir de cas sont tout-à fait originales et ouvrent des voies intéressantes à approfondir. Il n'existe pas à ma connaissance à ce jour d'outil permettant une indexation et une recherche hybride reposant sur un index de concepts formels. Mais il s'agit d'un projet ambitieux qui nécessitera des travaux de recherche et de développement de plusieurs années pour la réalisation d'un véritable prototype et son utilisation à la réalisation d'un système pour un site en vrai grandeur.

## Remerciements

Je remercie le LIPN et ses membres de nous avoir permis d'utiliser leur implémentation de C-Classic, ainsi que Jérôme Nobécourt et H. ZargAyouna pour leur contribution à son « débogage » indispensable à nos expérimentations.

## 6 Références

- [1] Bach J. R., et al, Virage image search engine: an open framework for image management, Symposium on Electr. Imaging: Science Tech - Storage & Retrieval for Image and Video Databases IV, IS&T/SPIE, Feb. 1996.
- [2] Bayardo, R., W. Bohrer, R. Brice, A. Cichocki, G. Fowler, A. Helal, V. Kashyap, T. Ksiezyk, G. Martin, M. Nodine, M. Rashid, M. Rusinkiewicz, R. Shea, C. Unnikris hnan, A. Unruh, D. Woelk Semantic Integration of Information in Open and Dynamic Environments In Proceedings of the ACM SIGMOD International Conference on Management of Data, 1997, pp. 195-206
- [3] Borgida A. Patel-Schneider. A semantics and complete algorithm for subsumption in the Classic description logic. Journal of Artificial Intelligence Research, 1:278-308, 1994.
- [4] Brachman R. J. and. Schmolze J. G. An overview of the KL-ONE knowledge representation system. Cognitive Science, 9(1):171--216, January--March 1985.
- [5] Brezellec P. And Soldano H. Elena : a bottom-up algorithm, Morgan Kaufmann Ed., 10 th Int. Conf. On Machine Learning, pp.9-16, 1993.
- [6] Chang, S.-F. Smith, J. R Meng H. J., Wang H., and. Zhong D, Finding Images/Video in Large Archives-Columbia's, Content-Based Visual Query Projects," CNRI Digital Library Magazine, Feb. 1997. (<http://www.dlib.org/dlib/february97/columbia/02chang.html>)
- [7] Chang, S.-F., Chen, W., Meng H.J., Sundaram, H. and D. Zhong, VideoQ- A Fully Automatic Object-Oriented Content-Based Video Search System, <http://www.ctr.columbia.edu/videoq>
- [8] Chawathe S, Garcia-Molina H, Hammer J, Ireland K, Papakonstantinou Y, Ullman J, Windom J. The TSIMMIS project : Integregation of heterogeneous information sources paru dans Proceedings of IPSJ Japan, 1994
- [9] Cohen W. , Borgida A., Hirsh, H., Computing Least Common Subsumers in Description Logics, Proceedings of the Tenth National Conference on Artificial Intelligence, 1992S
- [10] Etzioni and Weld. A softbot-based interface to the internet. paru dans CACM 37(7):72-76. 1994. <http://www.cs.washington.edu/research/softbots>
- [11] Excaliber Image Search System, <http://www.excalib.com/rev2/products/vrw/vrw.html>
- [12] Garcia-Molina H, Papakonstantinou Y., Widom, J. Object exchange across heterogeneous information sources. In Proc. of the International Conference on Data Engineering, 1995.
- [13] Gianfelici, S. Progetto e realizzazione di un sistema basato sulla logica terminologica per l'integrazione di schme concetuali. Universita degli studi di ancona, 2000
- [14] Ghomari D. Mémoire de DEA, Université Paris Nord, 1998
- [15] Lattes, V. Rousset MC, The use of CARIN language and algorithms for Information Integration: the PICSEL project 1998
- [16] Levy, A., Rajaraman, Ordille. Query-Answering Algorithms for Information Agents, AAAI 1996
- [17] Nastar, C., Indexation d'images par le contenu : un état de l'art, CORESA'97, 1997.
- [18] Nastar, C., Mitschke, M., Meilhac, C., Boujemaa N., Surfimage: a Flexible Content-Based Image Retrieval System, [http://www.acm.org/sigmm/MM98/electronic\\_proceedings/nastar/](http://www.acm.org/sigmm/MM98/electronic_proceedings/nastar/)
- [19] Salotti, S., Ventos, V., Une approche formelle du raisonnement à partir de cas. Revue d'Intelligence Artificielle, 1999.
- [20] Steele D., and Yanker, P. Query by Image and Video Content: The QBIC System," IEEE Computer Magazine, Sep. 1995, Vol.28, No.9, pp. 23-32.
- [21] Ventos V., C-CLASSICde : une logique de description pour la définition et l'apprentissage de concepts avec défauts et exceptions. Thèse, Université Paris Nord, 1997.
- [22] Wiederhold, G. Mediators in the architecture of future information systems, IEE Computer 25:3, pp. 33-49, 1992.
- [23] Weld, D., Kwok, C. Planning to gather information, In Proc. AAAI 1996, 1996.
- [24] Wooldridge M., Nicholas R., Jennings. Intelligent Agents : Theory and Practice, Knowledge Engineering Review, October 1994. Revised January 1995.
- [25] ZargAyouna h. Indexation et recherche multimédia. Rapport de DEA, Université Paris 9 Dauphine, 2000
- [26] Zhong D., Chang S.-F., Video Object Model and Segmentation for Content-Based Video Indexing, IEEE Intern. Conf. on Circuits and Systems, June, 1997, Hong Kong. (special session on Networked Multimedia Technology & Application).

## Vers une indexation et recherche multimédia

C. Golbreich  
 Laboratoire d'Informatique Médicale  
 Faculté de Médecine,  
 Université Rennes 1  
 Av du Pr. Léon Bernard,  
 35043 Rennes France  
 Christine.Golbreich@uhb.fr

Cet article présente une approche innovante pour une indexation et recherche multimédia de documents, les premiers travaux réalisés dans cette optique, et les perspectives actuelles. Le but est d'offrir un moteur plus puissant que les moteurs existants, afin de traiter des requêtes complexes sur le contenu de documents multimedia. L'idée est d'améliorer les performances par une recherche basée sur des « connaissances » au lieu de mots-clé ou du texte, et de combiner recherche « sémantique » et « image » pour allier les avantages respectifs des deux types de recherche.

Il existe deux approches principales pour l'indexation et la recherche de documents multimédia. La première, « content-based search », est une approche numérique qui repose sur des caractéristiques objectives d'apparence visuelle des images : couleur, forme, texture, etc. L'avantage est que les descripteurs du contenu sont calculés automatiquement à partir des images. L'inconvénient est le décalage entre le bas niveau de ces descripteurs et le contenu sémantique des requêtes des utilisateurs. La seconde est une approche symbolique qui repose sur des *mots-clé* et/ou des *textes* servant à résumer le contenu sémantique des documents audiovisuels. L'avantage est que ces descripteurs sont porteurs de plus de « sens ». Par contre, les mots-clés, sont en général définis dans un dictionnaire ou un thésaurus hiérarchisé, d'où les problèmes de bruit ou silence des moteurs de recherche par mot-clés, liés au caractère fortuit de ces annotations, à l'imprécision des mots-clé, les difficultés de maintenance et de vérification de cohérence des thesauri. De plus, l'indexation est la plupart du temps manuel-

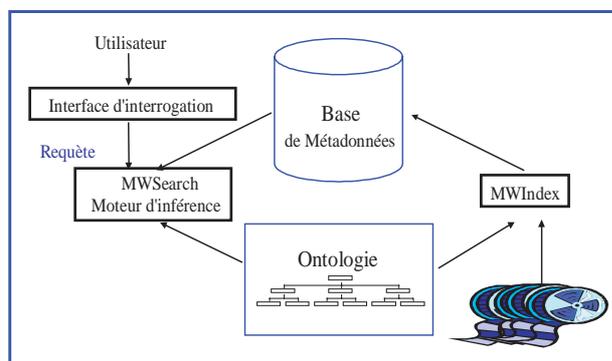


Figure 1 - Indexation et recherche basée ontologie

le, ce qui peut être très coûteux. Les moteurs de recherche existants QBIC, Webseek, Virage, Excalibur, Blobworld, ImageRover [1][8][17] sont généralement basés sur l'une ou l'autre de ces approches ou leurs améliorations. Mais, ni l'une ni l'autre n'est vraiment satisfaisante, d'où l'idée de réaliser un moteur de recherche *intelligent hybride*<sup>1</sup> qui repose sur une indexation par « concepts » au lieu de mots-clés (ou allie les deux), et une indexation « multimedia » qui combine description « sémantique » et « content-based ».

#### Moteur de recherche intelligent hybride

L'indexation et la recherche reposent sur une ontologie formelle, l'intégration multimédia sur un « médiateur ».

#### Recherche par concepts

L'idée clé est d'utiliser pour l'indexation et la recherche des documents non des mots, mais une ontologie de concepts du domaine (Figure 1) L'ontologie, ou modèle du domaine, est un modèle formel décrivant les concepts du domaine et leurs relations. Les documents sont annotés à l'aide de marqueurs « ontologiques » pertinents issus de cette ontologie. Grâce à ces annotations leur contenu est rendu accessible en tant que faits. L'utilisation d'ontologies pour la recherche d'information présente plusieurs avantages par rapport

aux méthodes par mots-clé. L'ontologie fournit un vocabulaire précis et partagé pour décrire des informations sur les contenus des documents. Elle peut être utilisée pour formuler les requêtes. Elle permet de spécifier des connaissances, qui peuvent être exploitées pour des déductions utilisées pour les réponses. Des raisonnements automatiques sont possible (classification, vérification)

L'autre avantage est qu'avec une telle approche déclarative, pour construire un nouveau serveur de documents, il suffit de définir deux composants: la base de métadonnées, c'est-à-dire les données décrivant le contenu des documents, et l'ontologie spécifiques à l'application. Le moteur de recherche (MWSearch) et les autres outils, interfaces et outil d'indexation (MWIndex), sont généraux. Ils ne dépendent que des langages de représentation et sont donc définis une fois pour toutes.

#### Recherche multimédia

Le moteur de recherche hybride repose sur une architecture de type médiateur [19]. Un médiateur fournit à l'utilisateur un accès intégré à des sources d'informations hétérogènes, grâce à une interface uniforme, qui lui donne l'impression qu'il utilise un système centralisé. Un tel système (e.g. Information Manifold [12], TSIMMIS [5], Internet Softboat [6] InfoSleuth [2], PICSEL [15] etc.) se charge, pour une requête donnée de l'utilisateur, de filtrer automatiquement les sources d'informations pertinentes et de combiner leurs informations, à partir de l'ontologie du domaine et de la description des sources. Pour un moteur de recherche

1. Technique que j'avais définie à l'origine en vue du projet PRIAMM MediaWorks plate-forme générique pour l'indexation et la recherche de documents audiovisuels par le contenu (CNRS, INRIA, TF1, AEGIS, EML) pendant mon détachement au LIMSI CNRS

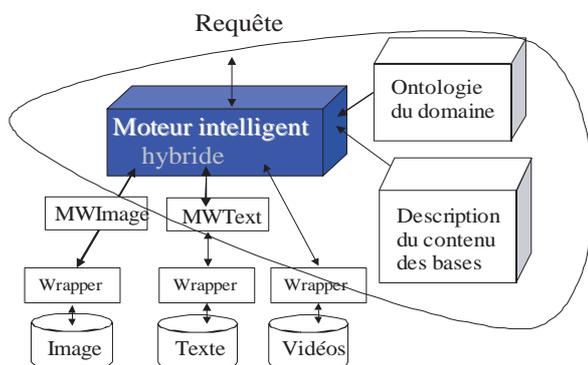


Figure 1 - Moteur Hybride

hybride, le médiateur comprend différents composants spécialisés pour le texte, les images, les mots-clé, les concepts.

### Indexation et recherche par similarité et hybride

Un premiers outil [20] a permis de mettre en oeuvre cette approche. Son originalité est d'offrir une indexation et une recherche (i) par concepts et similarité, (ii) multimédia.

### Recherche

Un scénario de recherche par similarité conceptuelle seule, et deux scénarios de recherche hybride combinant similarité numérique et symbolique ont été testés. Les expérimentations ont été faites sur une base « jouet » de 120 images du domaine des actualités. La description complète d'une trentaine de cas a été faite en *C-Classic*.

### - Recherche par « similarité conceptuelle »

La recherche sémantique repose sur une indexation non par mots-clé, mais par concepts, ce qui permet de bénéficier de mécanismes d'inférence puissants. L'ontologie, la base de documents, les requêtes, sont représentées dans la logique de description *C Classic*. Chaque document de la base (cas) est représenté par un individu composé de plusieurs parties : (1) une image (e.g. image clé du plan pour une vidéo) (2) un texte descriptif, (3) une liste de mots-clé, et (4) une description conceptuelle du cas représenté formellement par un individu *C-Classic*. La requête est représentée comme un indi-

vidu *C-Classic*. Le moteur de recherche repose non sur un appariement exact, mais sur un appariement par similarité numérique (visuelle) et/ou symbolique. La recherche utilise un raisonnement à partir de cas et sa formalisation en logique de description [16]. La recherche des cas les plus similaires repose sur la notion de Plus Petit Généralisant de deux concepts (PPG) [6] concept qui représente formellement la similarité de deux concepts par l'ensemble de leurs propriétés les plus spécifiques communes. Les différents algorithmes utilisés (PPG, DISSIM, ELENA+) pour l'implémentation ont été repris de [16][18] et les scénarios d'expérimentations codés en Java. Les résultats obtenus avec cet outil ont été comparés à ceux du moteur QBIC [17]. Les exemples, issus des expérimentations (cf. [20]) montrent que dans la plupart des cas, un moteur de recherche par concepts est plus puissant qu'un moteur de recherche par mots-clé car il permet :

- de traiter des requêtes plus fines, par exemple des conjonctions de critères, de prendre en compte une sémantique plus riche, de faire des inférences
- de traiter, grâce aux mécanismes de reconnaissance d'instances et de classification, des requêtes générales (e.g. trouver des documents d'hommes politiques au Moyen Orient, ou des

Présidents français depuis 1960), tandis que QBIC ne retourne aucun document, car ils sont annotés non par les concepts mais pas les individus.

- si une correspondance exacte avec la requête existe, il la trouve, mais s'il n'y a pas de document dans la base correspondant exactement à la requête, il retourne les documents les plus similaires.

### - Recherches « multimédia »

Plusieurs scénarios de recherche « hybride », ont été expérimentés. Les moteurs de recherche visuel et par concepts ont été utilisés conjointement. L'un sert à restreindre l'espace de recherche à un ensemble de documents, l'autre sert à affiner la recherche. Ainsi, au lieu de faire une recherche conceptuelle sur toute la base, on limite la recherche par concepts aux documents retrouvés par le moteur visuel, ce qui réduit l'espace de recherche de 4/5 dans le cas de cet exemple (Figure 2).

Le scénario inverse a aussi été testé. Un troisième scénario hybride a été expérimenté. Il permet de répondre à une requête multimodale, image et concepts, en faisant l'intersection des résultats obtenus par chacun des moteurs.

### Indexation

Une méthode semi-automatique est utilisée pour l'indexation. Le but est d'aider à indexer un nouveau document sans description en se servant des indexations des documents qui existent déjà dans la base (anciens cas). La méthode repose sur un raisonnement à partir de cas. Dans nos expérimentations, il s'agissait d'inférer pour une image, sa description conceptuelle. Les étapes de raisonnement sont celles du cycle classique du raisonnement à partir de cas (1) une recherche par le contenu à partir des caractéristiques numériques des images permet d'abord de trouver

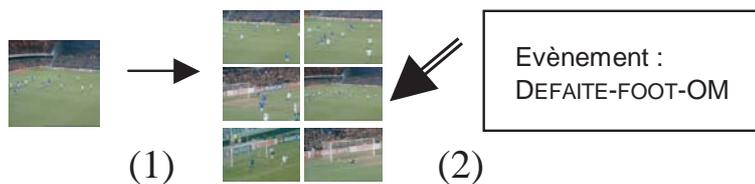


Figure 2 - Recherche visuelle suivie de recherche conceptuelle (documents du match de défaite OM)

des cas « visuellement » similaires. Deux moteurs d'image, QBIC [17] et Surfimage [14] ont été utilisés pour cela (2) l'utilisateur classe les images trouvées en deux catégories pertinentes et non pertinentes, qui serviront d'exemple et contre-exemples à l'algorithme d'apprentissage Elena+. Cet algorithme [3] [18] permet l'apprentissage d'un concept en C-Classic à partir d'instances positives et négatives de ce concept. Il est utilisé ici pour apprendre une ou plusieurs descriptions symboliques pour le nouveau cas. (3) les descriptions proposées doivent être validées par l'utilisateur qui les confirme ou les modifie, par exemple en les complétant (4) le nouveau cas ainsi indexé est rajouté à la base (cf. [20]). Cette méthode d'indexation présente un double avantage. D'une part, la description formelle des documents en C-Classic est, générée semi-automatiquement, ce qui du point de vue de l'ingénierie des connaissances est un apport important puisque la tâche d'écriture formelle est considérablement facilitée. D'autre part, automatiser l'indexation et déléguer le plus possible les mécanismes de bas niveau au système [11] est indispensable pour traiter des masses de documents comme c'est le cas pour des archives audiovisuelles ou en imagerie médicale. Il est souvent préférable d'avoir une indexation plus grossière mais plus automatisée.

### Conclusion et perspectives

L'outil et les premières expérimentations [20] montrent qu'il est possible de développer un moteur de recherche intelligent hybride, sur la base des techniques décrites. et qu'un tel moteur est plus performant que les moteurs classiques car il permet dans de nombreuses situations d'éliminer le bruit ou le silence d'une recherche simple par mots-clé ou par image-exemple. La recherche multimédia, par concepts, par similarité, ainsi que l'indexation formelle semi-automatique de documents à partir d'un raisonnement à partir de cas sont tout-à-fait originales et ouvrent des perspectives prometteuses. Ces techniques d'indexation et de recherche, ont été initialement définies pour des archives audiovisuelles, mais sont applicables à

toute base de documents multimédia. Au Laboratoire d'Informatique Médical de Rennes, nous nous intéressons à des outils basés sur les ontologies pour l'intégration et l'analyse d'informations biomédicales hétérogènes (média ou sémantique). Différentes applications sont privilégiées dans le domaine de la radiologie et de l'imagerie médicale (automatiser le plus possible l'annotation des compte-rendu radiologiques), de l'anatomie cérébrale [10], de la génomique (intégrer des informations médicales et biologiques pour l'annotation des gènes), de l'insuffisance rénale terminale etc., en collaboration avec plusieurs équipes médicales ou de biologie. Nous entendons poursuivre les travaux sur la recherche hybride, en combinant plusieurs modes de recherche, texte/image/mots-clés/descripteurs de terminologie (e.g. UMLS [13], Gene Ontology) avec une indexation basée sur une ontologie formelle.

### Remerciements

Je remercie le LIPN de m'avoir permis d'utiliser leur implémentation de C-Classic, J. Nobécourt et H. ZargAyouna pour l'outil et les expérimentations réalisés.

### Références

- [1] Bach J. R., et al, Virage image search engine: an open framework for image management, Symposium on Electr. Imaging: Science Tech - Storage & Retrieval for Image and Video Databases IV, IS&T/SPIE, Feb. 1996.
- [2] Bayardo, R., W. Bohrer, R. Brice, A. Cichocki, G. Fowler, A. Helal, V. Kashyap, T. Ksiezyk, G. Martin, M. Nodine, M. Rashid, M. Rusinkiewicz, R. Shea, C. Unnikrishnan, A. Unruh, D. Woelk Semantic Integration of Information in Open and Dynamic Environments In Proceedings of the ACM SIGMOD International Conference on Management of Data, 1997, pp. 195-206
- [3] Brezellec P. And Soldano H. Elena : a bottom-up algorithm, Morgan Kaufmann Ed., 10th Int. Conf. On Machine Learning, pp.9-16, 1993.
- [4] Chang, S.-F., Chen, W., Meng H.J., Sundaram, H. and D. Zhong, VideoQ- A Fully Automatic Object-Oriented Content-Based Video Search System
- [5] Chawathe S, Garcia-Molina H, Hammer J, Ireland K, Papakonstantinou Y, Ullman J, Windom J. The TSIMMIS project :

Intergregation of heterogeneous information sources paru dans Proceedings of IPSJ Japan, 1994

- [6] Cohen W. , Borgida A., Hirsh, H., Computing Least Common Subsumers in Description Logics, Proceedings of the Tenth National Conf. on Artificial Intelligence, 1992
- [7] Etzioni and Weld. A softbot-based interface to the internet. paru dans CACM 37(7):72-76. 1994. <http://www.cs.washington.edu/research/softbots>
- [8] Excalibur Image Search System, <http://www.excalib.com/rev2/products/vrw/vrw.html>
- [9] Garcia-Molina H, Papakonstantinou Y., Widom, J. Object exchange across heterogeneous information sources. Proc. of the International Conference on Data Engineering, 1995.
- [10] Golbreich , Dameron O., Gibaud B., Burgun A. How to represent medical ontologies in view of a Medical Semantic Web ? In Proc. of 9th Conf. on Artificial Intelligence in Medicine in Europe, AIME 2003, Springer LNCS.
- [11] Golbreich, C. Explorer les documents de bibliothèques numériques, un besoin croissant des entreprises. NimeStic'01, Nîmes, 2001
- [12] Levy, A., Rajaraman, Ordille. Query-Answering Algorithms for Information Agents, AAAI 1996
- [13] Lindberg D.A, Humphreys, B.L. McCray AT. The Unified Medical Language System. Meth. Inf Med Aug; 32(4) (1993) 281-91
- [14] Nastar, C., Mitschke, M., Meilhac, C., Boujemaa N., Surfimage: a Flexible Content-Based Image Retrieval System, In ACM-Multimedia 1998
- [15] Rousset M-C et al.. Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes : le projet PICSEL, Revue I3, 2002.
- [16] Salotti, S., Ventos, V., Une approche formelle du raisonnement à partir de cas. Revue d'Intelligence Artificielle, 1999.
- [17] Steele D., and Yanker, P. Query by Image and Video Content: The QBIC System," IEEE Computer Magazine, Sep. 1995, Vol.28, No.9, pp. 23-32.
- [18] Ventos V., C-CLASSICde : une logique de description pour la définition et l'apprentissage de concepts avec défauts et exceptions. Thèse, Université Paris Nord, 1997.
- [19] Wiederhold, G. Mediators in the architecture of future information systems, IEE Computer 25:3, pp. 33-49,1992.
- [20] ZargAyouna h. Indexation et recherche multimédia. Rapport de DEA, Université Paris 9 Dauphine, 2000.

## A theory of computer semiotics

Peter Bøgh Andersen

Cambridge university press, Cambridge (UK),  
1997 (0-521-44868-9 ; première édition 0-521-39366-1, 1990)

### Introduction

Pourquoi écrire un compte-rendu sur un livre vieux de plus d'une décennie ? Parce qu'au fur et à mesure de l'augmentation de la communication à travers l'outil informatique, les défauts de prise en compte du sens des « données » manipulées sont de plus en plus apparents. Ils sont aussi vite corrigés que détectés, mais de manière ad hoc, sans vraiment s'intéresser à la construction du sens accomplie par les utilisateurs. Le besoin d'un cadre unificateur se fait sentir et ce livre pourrait en être la base.

Son but est de proposer une théorie de sémiologie informatique, c'est-à-dire de la nature des signes informatiques et de la manière dont les ordinateurs communiquent à l'aide de signes. Il adopte une double perspective par l'approche théorique de son objet et par l'observation, dans le contexte d'utilisation, des impacts de l'informatique. Bien qu'écrit en 1990, et le résultat de travaux de la décennie 80, il n'a, pour l'essentiel pas vieilli (on y reviendra).

Enfin, il s'agit d'un ouvrage de science du langage qui s'intéresse profondément à l'outil informatique, ce n'est pas si fréquent et cela mérite donc toute notre attention.

L'ouvrage est décomposé en trois grandes parties consacrées à la théorie, l'appréhension des ordinateurs comme implémentant des systèmes de signes et l'étude des pratiques et de l'impact de l'ordinateur dans un contexte de travail. Il contient une bibliographie, un index et a été augmenté d'une introduction lors de sa dernière édition.

### I. Théorie

La première partie propose les grands principes d'une approche sémiologique des applications informatiques.

Partant d'un cadre structuraliste dans lequel on analyse les productions afin d'isoler les signes et de les interpréter, l'auteur signale à partir de quelques exemples les limites de l'approche. Il propose donc d'étendre le cadre inspiré de Ferdinand de Saussure et

de Louis Hjelmslev afin de surmonter ces problèmes. Les extensions portent sur :

- la prise en compte d'actes non symboliques, c'est-à-dire d'actes n'apparaissant pas sur deux plans (signifiant/signifié ou, pour reprendre les termes de l'ouvrage, expression/contenu) et plus spécialement d'actes n'étant pas exprimés mais étant visibles de l'interlocuteur;
- la prise en compte du contexte social, et concernant cette étude la prise en compte de l'environnement de travail et de la tâche à accomplir ;
- l'étude d'un sous-langage (langage de travail) lié à une pratique particulière.

Le premier point est à relier à la théorie des actes de langage. Les deux points suivants étendent le cadre à l'aide des travaux de sociolinguistique dont le but est de réintroduire le contexte social dans la linguistique, en particulier ceux de Michael Halliday.

De ces diverses extensions, il appert qu'il est nécessaire d'étendre les méthodes structuralistes comme le test de commutation : remplaçant un élément sur l'un des deux plans

– signifiant ou signifié – et observant la modification sur l'autre.

L'auteur ajoute à ces extensions la nécessité de traiter l'évolution du système de signes et de son contexte. Cette préoccupation est à rapprocher de l'aspect social : les rapports sociaux évoluent, les systèmes de signes doivent aussi pouvoir évoluer.

Cette partie ne s'en tient pas à une discussion de salon telle que ce bref compte-rendu pourrait le laisser penser : l'auteur y détaille chacun des concepts utilisés d'un point de vue théorique et à l'aide d'exemples présentés au début du chapitre. Quelques choix peuvent être discutés comme, la restriction à un seul langage de travail au lieu de considérer des sous-langages emboîtés (évoqué cependant), l'affirmation de l'absence de distinction signifiant/signifié en informatique (p142) qui ne tient pas lorsque l'on s'intéresse à la sémantique des langages (de programmation ou de représentation de connaissance), ou des considérations peu structuralistes sur l'intelligence des systèmes informatiques.

Cependant, cette partie est vraiment passionnante. En particulier, la vision des ordinateurs comme des media n'a fait que se développer depuis l'écriture du livre grâce au développement d'Internet et du Web. Depuis que l'ouvrage a été rédigé, ce rôle de l'ordinateur comme medium a été considérablement renforcé. Mais comme c'est toujours avec son poste de travail que l'on interagit, ce livre garde toute son actualité.

Ma principale critique concerne d'ailleurs les vues proposées dans la section finale de cette première partie (§I.2.5) : « computers as media ». En fait, les ordinateurs devraient être considérés comme des media particuliers : des média capables de calcul. Il serait donc nécessaire d'étendre la vue proposée par l'auteur en permettant à l'ordinateur la manipulation de signes en tant que signes (on peut considérer que la sémantique des langages informatiques est une première réponse). Une partie de cet aspect, la versatilité peu usuelle des signes informatiques et leur potentiel impact sur le monde, sera prise en compte dans la suite.

## II. Ordinateurs

La partie suivante a pour but de proposer une analyse des types de signes utilisés par les ordinateurs. Bien entendu, le plan de l'expression (ou du signifiant) du système de signes, ce qui est analysable, est l'interface entre la machine et l'utilisateur. C'est donc à partir de l'interface que l'analyse peut commencer.

Quels aspects sont particuliers aux signes informatiques ? La question est ambitieuse. Si la plupart des exemples donnés par l'auteur correspondent à l'utilisation d'une interface graphique, la réponse est aussi abstraite que la question. L'auteur caractérise un signe informatique par des caractéristiques permanentes lui permettant d'être identifié, par des caractéristiques temporaires (transcient) lui permettant d'exprimer un état et par sa capacité à être manipulé (handled) par l'utilisateur. Additionnellement, les signes peuvent être utilisés comme des outils, c'est-à-dire manipulés par l'utilisateur afin d'agir sur d'autres signes (par exemple, le vaporisateur du logiciel de dessin utilisé pour colorier une surface ou la clé utilisée pour décrypter un fichier).

Ainsi qu'on le voit, les « signes informatiques » ne sont pas caractérisés par leur forme (son, ensemble de pixels sur l'écran), mais par des caractéristiques abstraites qui rendent compte des aspects dynamiques et interactifs de l'ordinateur.

Cette grille d'analyse permet à l'auteur de fournir une classification des signes d'interface suivant que le signe peut être manipulé ou non, qu'il montre des caractéristiques temporaires ou non et qu'il peut servir d'outil ou non. Chaque catégorie est illustrée par plusieurs exemples et est décrite par un réseau de pétri définissant les états internes du signe (caractéristiques temporaires), les actions qu'il peut subir (aspect manipulable) ainsi que celles qu'il peut faire subir à d'autres signes (aspect outil). Au passage, ceci fait apparaître des « signes fantômes » qui n'ont pas de manifestation permanente dans l'interface mais dont l'interaction avec d'autres signes leur font changer certaines caractéristiques temporaires (ainsi le curseur deviendra-t-il un symbole d'interdiction lorsqu'il est placé au-dessus d'un catalogue dans lequel on ne peut écrire). Ces signes fantômes sont déjà un indice de l'activité de l'ordinateur... L'auteur présente de nombreux exemples d'utilisation de ces signes, illustrés par diverses applications (casse-brique, Word, FullPaint, HyperCard et l'application aux chèques postaux suédois qui sert de fil rouge à l'ensemble de l'ouvrage). Au passage, l'auteur a fait une trouvaille astucieuse pour réaliser le test de commutation : il le fait non en changeant le programme mais en le remplaçant par une autre version du même programme et en observant les caractéristiques variables et les invariants.

Une fois inventoriés, les différents types de signes sont mis en rapport dans la création de signes composites. Cette dernière partie est moins fouillée que la précédente et mériterait certainement un ouvrage complet à son sujet. Elle est particulièrement importante dans l'interaction avec l'utilisateur et dans l'application de signes outils. Une des difficultés de cette tâche est que les signes révèlent leur signification dans l'interaction (entre eux et avec l'utilisateur) et aucun formalisme n'est disponible pour exprimer cette composition.

Cela révèle une difficulté rencontrée par ailleurs (étudiée par la perspective langage/action et, plus récemment, les systèmes multi-agents) : l'interprétation des signes isolés ne permet pas d'accéder à la totalité du sens. Il est nécessaire d'identifier la discussion dont ils font partie (ce qui, en informatique, est souvent rendu par un protocole).

Alors que l'on pourrait s'en tenir à l'aspect interface, l'une des plus originales contributions de l'ouvrage concerne l'analyse de l'approche de programmation objet. Emboîtant le pas de la métaphore « objet », l'au-

teur note que la définition d'un objet dans ce cadre est assez proche de celle d'un signe informatique : une identité, un état, une interface permettant à l'environnement d'interagir avec l'objet et la possibilité d'interagir avec d'autres objets. L'auteur s'est particulièrement intéressé au langage HyperCard dans lequel nombre des objets de base sont des éléments d'interface graphique ce qui en fait le langage de programmation idéal pour la « programmation orientée signe ». Le programme qui contrôle le comportement de l'interface agit comme sa sémiologie<sup>1</sup> (on a ici une sémantique opérationnelle de l'interface). Il interprète les différentes actions de l'utilisateur (dans les deux sens du terme). Le langage de programmation est alors sa méta-sémiologie : un langage de description de sémiologie.

Cependant, l'auteur ne va pas beaucoup plus loin dans cette voie. Il considère que « l'interface est un système de signes manifesté dans les processus [cognitifs] que les personnes créent lorsqu'elles utilisent et interprètent le système ». C'est sans doute oublier qu'un ordinateur ne se contente pas uniquement de manipuler les signes visibles à son interface. L'auteur reconnaît que la fonction « Rechercher » d'un gestionnaire de bases de données est complexe et monopolise une large part de l'activité de l'ordinateur. Mais elle n'est que la réponse à une sollicitation de celui-ci. Ce n'est pas le cas de la tâche de nettoyage ou de profilage de l'utilisation qu'un ordinateur peut accomplir et qui modifient son comportement. Pour aller plus loin, certains ont ébauché ces dernières années une « sémiologie computationnelle » qui appréhende l'ordinateur non seulement comme communiquant par le biais d'un système de signes mais comme manipulateur de signes voire comme l'instigateur d'un véritable processus de création de signes (sémiosis).

Si l'on attribue peu de capacité à l'ordinateur dans l'interprétation des signes, l'auteur ne laisse pas beaucoup plus de place au concepteur : « Un développeur de systèmes informatiques est dans la même position que d'autres producteurs de matériaux symboliques, comme des dramaturges : ils peuvent espérer que leurs intentions vont être réalisées lors de la première. ». En fait, cette attitude permet de ne pas prendre pour argent comptant les affirmations des manuels d'utilisateurs,

1. Ici le terme sémiologie est utilisé pour la description d'un système de signes, en l'occurrence celui révélé par l'interaction entre l'utilisateur et l'interface.

mais de justifier les interfaces en contexte comme ce sera fait dans la partie suivante.

Cette partie est celle qui a le plus vieilli, non pas tant dans son fonds, que dans son corpus : rappelez-vous de l'informatique au début des années 1990. En revanche, l'analyse des signes présentés par l'interface a parfaitement résisté au passage du temps.

### III. Travail

La dernière partie replace les théories présentées dans les parties précédentes au cœur du contexte de travail. Ceci est analysé dans une application de gestion de la compensation des chèques postaux suédois dans laquelle l'auteur est intervenu pour informatiser l'ensemble du processus (c'est-à-dire remplacer le plus possible des caisses et des chèques physiques circulants de bureau en bureau par une application informatique).

L'idée est d'y montrer comment les principes développés précédemment s'appliquent dans un contexte réel où il est possible de les utiliser pour faire évoluer le système (et il s'agit ici de l'application informatique mais aussi de l'organisation du travail) à la satisfaction de ses utilisateurs. Pour cela l'auteur revient au langage utilisé sur le lieu de travail en analysant conjointement le contexte de travail et le langage utilisé par l'acteur. L'analyse utilise deux techniques : les champs sémantiques permettent d'inventorier les unités syntaxiques du langage utilisé par les intervenants, les jeux de langages permettent d'identifier les conversations permettant d'accomplir un certain travail (affecter une tâche, se coordonner...).

Dans la théorie des champs sémantiques, l'analyse des conversations permet de caractériser les points de vue et l'intérêt des locuteurs. Ils sont classés suivant la tâche à accomplir et la position hiérarchique du locuteur. Une fois le sous-langage (ou le sous-lexique) établi, on montre comment les langages s'accordent ou non au logiciel réalisé, à l'instar de ce que l'on appellerait l'alignement d'ontologies, et comment les utiliser pour concevoir le système (par exemple, pour présenter la « possession » d'une formule de chèque ce que rendait précédemment inutile leur possession physique). Un des aspects intéressants de ce travail est le fait que l'on ne cherche pas à reproduire coûte que coûte les langages disponibles avant l'arrivée de l'informatique, mais que l'on tienne compte du changement qu'introduit l'irruption de l'informatique dans le service.

L'analyse des conversations permet d'établir des «

jeux de langage » dont le but est de prolonger la théorie des actes de langage par la prise en compte des conversations permettant d'accomplir une certaine tâche et non plus d'actes isolés. Ces jeux de langage s'apparentent donc aux protocoles requis pour interpréter la composition statique et dynamique de signes dont nous déplorions l'absence précédemment. Au passage, l'auteur propose une superbe justification pour une approche externe (structuraliste) des actes de langage : les intentions et croyances des intervenants ne sont pas accessibles aux méthodes d'investigation linguistiques. On cherche donc à isoler les dialogues en terme de « jeux de langage » et ainsi à construire une typologie des mouvements autorisés (sinon une grammaire des jeux). Bien que ceci aurait pu être très intéressant, l'auteur ne montre pas de jeux de langages remplacés par des interactions avec l'ordinateur. Il évoque en revanche les modifications dans ces jeux de langages après l'informatisation et s'attache longuement à appliquer ces jeux de langages à la conception du système informatique.

Une dernière partie considère la question de savoir qui de l'ordinateur ou de son utilisateur doit avoir le contrôle du processus et, en particulier, du jeu de langage. Andersen considère que les utilisateurs doivent pouvoir contrôler quelle variante de la procédure de travail ils veulent utiliser (que l'interface ne doit pas tourner une variante en norme). Ceci permet à l'organisation du travail d'évoluer. Un argument linguistique pour cela est que la plasticité du matériel est essentielle pour la vie, et donc l'évolution, d'un système de signes.

Cette partie est pour moi la moins aboutie car, si elle applique effectivement les techniques structuralistes à l'analyse de l'organisation du travail, elle ne retourne que très peu au système de signes manipulé par l'ordinateur. Plus souvent les conclusions sont tirées sur des bases sociologiques. C'est plus l'impact de l'informatisation que le système de signes qui est mis en avant. C'est cependant dans la lignée annoncée dès le début par l'auteur : « les signes et leur vie en société ».

### Conclusions

Ce compte-rendu est la lecture d'un informaticien, pas forcément informé mais conscient de ses besoins. Je serais redevable envers quiconque me signalera mes erreurs d'interprétation.

Cet ouvrage est point de départ un stimulant et solide. S'il y en avait d'autres, je serais heureux d'en entendre parler.

Pour ma part, je dépasserais la vision naïve de l'ordinateur comme simple médium interactif. Le livre n'indique pas que cette vision introduit un déséquilibre troublant : alors que la plupart des systèmes de signes sont utilisés pour faire communiquer deux entités, la sémiologie informatique présentée ici dote les énoncés d'une certaine autonomie à cause du caractère interactif du médium. Il n'est pas clair que le système de signes serve à la communication entre deux interlocuteurs : l'interactivité semble permettre à l'utilisateur de jouer avec le système de signes plus que de communiquer.

La perspective d'un Web sémantique, où les ordinateurs peuvent manipuler la connaissance exprimée, unie les perspectives de l'ordinateur comme médium et l'ordinateur comme « processeur ». Il est alors d'autant plus important de disposer de techniques telles que décrites dans cet ouvrage afin de pouvoir contrôler l'expression de l'ordinateur en fonction de celle de l'humain.

En somme, ce livre est clair et montre une application raisonnée de techniques sémiotiques dans un contexte précis d'utilisation des ordinateurs. C'est un cadre d'analyse et de conception très utile. Il est illustré d'exemples variés et pertinents. Il est parfois un peu répétitif, mais cela permet lors d'une lecture séquentielle de disposer des éléments nécessaires rapidement.

Un aspect rafraîchissant de ce texte est son enracinement nordique. Ainsi, peut-on y rencontrer le cercle de Copenhague en linguistique, l'école Simula de programmation orientée-objet, l'approche participative de la conception, mais aussi les legos, Ibsen et Bergman.

Je le recommanderais à tout chercheur qui se demande quelle sorte de langage est utilisé lors de la communication entre ordinateurs et humains (et entre humains lors de communications médiatisées par ordinateurs) et comment en rendre compte de manière systématique.

Cela dit, le livre est plutôt une présentation de recherche en mouvement qu'un manuel. Il faut le lire pour y retrouver les positions que j'ai soulignées. Ceci est visible à l'absence de réelle conclusion ou perspective.

Il existe d'autres recensions de cet ouvrage disponibles sur le Web (demandez-les à votre moteur de recherche favori). Je les ai lues après avoir rédigé celle-ci, ce qui m'a conduit à faire une correction (ne pas utiliser le terme de « registre » utilisé par l'auteur mais celui de sous-langage).

## Report from the XI International Conference Knowledge Acquisition and Management - KAM'2003

The conference took place 16-18 of May in Turawa (Poland) in surroundings of the lovely lake. The main challenge taking into account was that peculiar crossing of technology and wisdom becomes the source of many problems. The same many subjects of multifaceted researchers are recognized. This year we have widened the conference's formula emphasizing the meaning of knowledge management.

Conferences of the Department of Artificial Intelligence Systems (Wroclaw University of Economics) have fairly long tradition; current was 11'th. Professor Andrzej Baborski was the organizer of the conference. Professor created conference as an informal forum, place of meeting for people interested in various aspects of Artificial Intelligence. Most papers and speakers represented various regions of Poland (for

example: prof. Zdzislaw Hippe and his group from Rzeszow University of Informatics and Management, prof. Barbara Debska from Rzeszow Technical University, prof. Antoni Ligeza from Krakow University of Mining and Metallurgy, prof. Witold Chmielarz from Warsaw University and prof. Kazimierz Perekuda from Wroclaw University of Economics). Their participation as well as contribution of foreign guests from Australia (prof. Leszek Maciaszek – Macquarie University Sydney), Belarus (prof. Romuald Rutkowski – Minsk State University of Economics), Belgium (a team of prof. Jeanne Schreurs – Limburg University Center Diepenbeek), France (prof. Eunice Mercier-Laurent – Lyon 3 University) and Yemen (dr Saeed Al-Dobai – Sana University) showed that the conference's importance is increasing.



Sessions covered the following topics: Knowledge Management; Intelligent Systems Technologies; Learning, Knowledge and Intelligent Systems; Agents, Knowledge Representation and Acquisition; Education and Knowledge Management; Chosen Problems of Knowledge Processing; Evolution Algorithms and Expert Systems; Knowledge Discovery; Knowledge

Acquisition and Modeling. Our invited speaker prof. Leszek Maciaszek (Macquarie University Sydney) gave a talk: « Enterprise Information Systems in the Internet Era: Architecture, Technologies and Standards ».

*Mieczyslaw L. Owoc and  
Maciej Pondel*

# CONFÉRENCES

## Calendrier des conférences AFIA

NOVEMBRE 2003			
26 - 27	<b>LFA</b>	Tours	Rencontres Francophones sur la Logique Floue et ses Applications <a href="http://www.rfai.li.univ-tours.fr/lfa03/">http://www.rfai.li.univ-tours.fr/lfa03/</a>
27 - 29	<b>JFSMA</b>	Hammamet	Journées Francophones sur les Systèmes Multi-Agents <a href="http://www.cck.rnu.tn/JFSMA/">http://www.cck.rnu.tn/JFSMA/</a>
DECEMBRE 2003			
3 - 4	<b>CITE</b>	Troyes	Coopération, Innovation, Technologie <a href="http://tech-cico.utt.fr/cite2003/">http://tech-cico.utt.fr/cite2003/</a>
14 - 16	<b>COMETIC</b>	Marseille	Colloque Médiation et Ingénierie des Connaissances <a href="http://www.mediterranee.univ-mrs.fr/cometic/index.htm">http://www.mediterranee.univ-mrs.fr/cometic/index.htm</a>
JANVIER 2004			
28 - 30	<b>RFIA</b>	Toulouse	Reconnaissance des Formes et Intelligence Artificielle <a href="http://www.laas.fr/rfia2004">http://www.laas.fr/rfia2004</a>

## Présentation de la conférence

# COMETIC Colloque Médiation et Ingénierie des Connaissances

Marseille, les 14-16 décembre 2003

<http://www.ejcm.univ-mrs.fr/cometic/>

La société de l'information et ses réseaux produisent et diffusent sans limite de l'information souvent non hiérarchisée, non vérifiée, non structurée... information qui est pourtant la base de la connaissance, considérée comme la principale valeur ajoutée de nos sociétés !

Appréhender l'essor de cette société de l'information et de la connaissance et les mécanismes de production et de diffusion des connaissances en croisant les logiques interdisciplinaires en termes de Sciences de l'Information et de la Communication (SIC) et de Sciences et Technologies de l'Information et de la Communication (STIC) est l'objet central de ce *Colloque Médiation ET Ingénierie des Connaissances*.

L'organisation de cette manifestation est assurée par trois Centres de Recherche :

- Ⓜ Médias, Information et Connaissance ;
- Ⓜ Atelier de Modélisation ;
- Ⓜ Laboratoire de Valorisation de l'Information.

Et par les Universités d'Aix-Marseille II et d'Aix-Marseille III regroupés au sein de l'Équipe d'Accueil 3240, « Connaissances, Systèmes d'Informations et Décision ».

## PRINCIPAUX THEMES

### 1. Nouvelles Médiations, Nouvelles Ingénieries

- Veille et mémoire des organisations
- Diffusion, transmission des connaissances; réseaux et innovation
- Méthodes de recueil, d'acquisition des connaissances et de gestion du cycle de vie des connaissances, y compris visualisation, mesure de la qualité...
- Enjeux des nouvelles médiations (commerce électronique, télé-enseignement, knowledge management, analyse cognitive).

### 2. Ingénierie des connaissances

- Modèles de représentation des données et des connaissances
- Traitement de données volumineuses, entrepôts de données
- Méthodes et algorithmes d'extraction de connaissances à partir de données hétérogènes (bases de données, textes, images, son...), algorithmes d'apprentissage
- Systèmes d'information distribués, déploiement des connaissances, réseaux de diffusion et de partage

### 3. Données, information et organisation

- L'organisation des informations pour la mémoire organisationnelle : catégorisation par la création de classification.
- Connaissances et encodage : concepts, termes, terminologies, ontologies.
- Application des classifications et des encodages en Intelligence Économique.

### 4. Médias, information et connaissances

- Le rôle des médias et de l'information, nouveaux médias et nouveaux enjeux (les coûts, la TNT, Internet et les fractures, enjeux de la diffusion scientifique, audiovisuel et web vidéo...)
- Éthique, société de l'information et connaissance.

Un ouvrage co-édité P-Y. Badillo, P. HASSANALY & J-P. MARCIANO devrait publier les meilleurs exposés des Actes.

## DATES IMPORTANTES

- 30 juin 2003 repoussée au 15 Juillet 2003 : Date limite de soumission des envois des articles
- Septembre 2003 : Notification aux auteurs de l'acceptation de leur[s] article[s]
- 14 Décembre 2003 : Accueil des participants au congrès
- 15 & 16 Décembre 2003 : Date du congrès

## COMITE SCIENTIFIQUE

J.-P. BARTHES :  
Université Technologie de Compiègne  
D. BENSLIMANE : Université de Lyon I  
C. CAUVET : Université Aix - Marseille  
S. A. CERRI : Université de Montpellier  
M. EGEA : Université de Lyon I  
J.-L. ERMINE : Université de Technologie de Troyes  
C. FRASSON : Université de Montréal  
J.-G. GANASCIA : Université Paris VI  
G. GOUARDERES : Université de Pau  
D. HERIN : Université de Montpellier  
C. LEBOEUF : Université de Montpellier I  
B. MIEGE : Université de Grenoble III  
J. PERRIAULT : Université Paris X  
F. SEDES : Université de Toulouse  
F. THIBAUD : M E N

## COMITE d'ORGANISATION

D. AUGEY :  
Laboratoire « Médias, Information et Connaissance »  
D. BOURGEOIS :  
Laboratoire « Médias, Information et Connaissance »  
P.-Y. BADILLO :  
Laboratoire « Médias, Information et Connaissance »  
V. CAMPILLO :  
« Laboratoire de Valorisation de l'Information »  
J. CAUSSANEL : "Atelier de Modélisation"  
P. HAMMAD :  
Laboratoire de Mathématiques Appliquées  
P. HASSANALY :  
« Laboratoire de Valorisation de l'Information »  
S. HILALA : « Atelier de Modélisation »  
J.-P. MARCIANO : « Atelier de Modélisation »  
P.-Y. ROLLAND : « Atelier de Modélisation »  
P. THIBAUD : « Atelier de Modélisation »

## Appel à communications

### LMO 2004 : Langages et Modèles à Objets

Lille, 15-17 mars 2004

<http://www.lifl.fr/lmo2004/>

Depuis 10 ans, les conférences « Langages et Modèles à Objets » constituent un forum d'échange privilégié entre chercheurs s'intéressant aux diverses facettes du concept d'objet, de composant et de modèle (en programmation, en représentation de connaissances, en bases de données, en génie logiciel, en programmation répartie, en intergiciel et en système), afin d'étudier leurs spécificités, leurs points communs et leurs divergences, ainsi que les tendances futures.

Cette 10e édition de LMO se tiendra à Lille dans la semaine du 15 mars 2004. Elle sera adossée à une journée « composants » et sans doute d'autres ateliers connexes.

Nous souhaitons recevoir des propositions de communications originales concernant les thèmes cités ci-dessous. En supplément de ces thèmes traditionnels, l'édition 2004 souhaite mettre l'accent sur le thème transversal de l'« objet de modélisation » qui signe la convergence entre les approches UML (méta-modèles) et représentation de connaissance (ontologies). Dans la mesure où ces modèles sont maintenant destinés à être communiqués (par exemple dans la problématique du web sémantique), il devrait être possible de les compa-

rer et de les faire dialoguer. La comparaison est-elle superficielle ou cache-t-elle des convergences plus profondes? Ce mouvement cache-t-il la convergence entre programmation et représentation par objets, souvent annoncée, toujours attendue? Nous examinerons avec attention les contributions cherchant à éclairer ces interrogations.

La liaison entre les journées composants et LMO sera concrétisée par une session commune présentant les meilleures soumissions aux journées composants qui seront publiées dans les actes de LMO.

## THÈMES

- Programmation par objets : Langages, interprétation, compilation ; modèles d'objets pour la programmation ; objets et types (héritage, sous-typage, inférence, etc.) ; points de vues, rôles, aspects, subjectivité ; réflexivité, métaobjets ; environnements de programmation.
- Composants et objets distribués : Modèles de composants à objets ; interactions de composants ; raisonnement compositionnel ; développement à base de com-

posants, composants réutilisables ; objets et composants distribués, répartis ; acteurs, parallélisme ; objets et Internet ; inter-opérabilité ; intergiciel (middleware) à objets.

- Représentation et traitement des connaissances par objets : Modèles d'objets en représentation des connaissances (logiques de descriptions, graphes conceptuels, objets composites, frames) ; sémantique et mécanismes d'inférence de ces modèles (classification, héritage, filtrage, etc.) ; fouille de données ; les objets dans les ontologies ; le Web sémantique et la galaxie XML.
- Génie des objets : Cycle de vie des objets ; rétro-conception, évolution des programmes, versions ; sûreté des programmes, spécifications formelles ; méthodes d'analyse et de conception objet (UML) ; ingénierie des modèles et des métamodèles ; réutilisation, architectures logicielles réutilisables et à base de composants ; hiérarchies, frameworks, patterns.
- Bases de données à objets : Systèmes de gestion de bases de données objet ; indexation par classification ; fouille de bases d'objets et entrepôts de données objet ; "data-grid" ; objets semi-structurés.
- Applications : Objets et algorithmique ; objets métier ; objets pour les IHM, les télécommunications, le travail coopératif, le multimedia, la chimie, la musique, les sciences de la vie et la médecine, la CAO, l'EAO, etc.

## CALENDRIER

- 7 octobre 2003 : Réception des soumissions
- 25 novembre 2003 : Notification aux auteurs
- 15 janvier 2004 : Réception des copies définitives
- 15 mars 2004 : LMO 2004 (Lille)

## MODALITÉS DE SOUMISSION

Les langues officielles de la conférence sont le français et l'anglais : les articles et les présentations d'auteurs francophones doivent être en français. La première page comportera le titre de l'article, les noms des auteurs, leurs affiliations et adresses (en particulier électronique), ainsi qu'un résumé d'une dizaine de lignes et une liste de mots clés. Pour les articles écrits en français, le résumé et la liste de mots clés seront faits en anglais et en français. Chaque article soumis sera évalué par 3 lecteurs, membres du comité de programme.

Les actes seront publiés par les Editions Hermès. Les soumissions doivent suivre le format Hermès/Lavoisier disponible sur : [http://www.editions-hermes.fr/fr/cons\\_actes.html](http://www.editions-hermes.fr/fr/cons_actes.html)

Elles n'excéderont pas 13 pages.

Les propositions de communications doivent être déposées sur le site de la conférence (<http://www.lifl.fr/lmo2004/>) au plus tard le 7 octobre sous format électronique (PDF ou PostScript).

Afin d'optimiser la relecture, une déclaration préalable d'intention de soumission devra être déposée à la même adresse avant le 3 octobre.

## POSITIONS, PROBLÈMES ET DERNIÈRE MINUTE

On cherchera à ajouter une dose spontanité à la rigueur habituelle en introduisant des sessions de prise de positions, de présentation de résultats de dernière minute et d'exposition de problèmes.

Les contributions courtes seront évaluées légèrement sur la base d'un résumé. Elles ne seront pas publiées dans les actes.

## COMITÉ DE PROGRAMME

Président: Jérôme Euzenat  
(Jerome.Euzenat @ inrialpes.fr), INRIA Rhône-Alpes

- \* Patrick Albert (ILOG, Gentilly)
- \* Françoise André (IRISA/Université de Rennes 1)
- \* Jean-Pierre Briot (LIP6, Paris)
- \* Mireille Blay-Fornarino (I3S, Sophia-Antipolis)
- \* Jean Bézivin (INRIA/IRIN/Université de Nantes)
- \* Isabelle Borne (VALORIA, Vannes)
- \* Cécile Capponi (LIF, Marseille)
- \* Denis Caromel (Université de Nice Sophia-Antipolis/CNRS/I3S/INRIA/IUF)
- \* Bernard Carré (LIFL, Lille)
- \* Philippe Collet (I3S, Sophia-Antipolis)
- \* Noël Conruyt (IREMIA, Université de la Réunion)
- \* Théo D'Hondt (VUB, Bruxelles)
- \* Stéphane Ducasse (IAM, Berne)
- \* Marianne Huchard (LIRMM, Montpellier)
- \* Thérèse Libourel (LIRMM, Montpellier)
- \* Jacques Malenfant (LIP6, Paris)
- \* Philippe Merle (LIFL, Lille)
- \* Amedeo Napoli (LORIA, Nancy)
- \* Jacques Noyé (INRIA/École des Mines de Nantes)
- \* Houari Sahraoui (Université de Montréal)

# 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 V146 N°2 JUNE 2003

- Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries, *G. WIDMER*
- Similarity of personal preferences: Theoretical foundations and empirical analysis, *V. HA, P. HADDAWY*
- The complexity of achievement and maintenance problems in agent-based systems, *I.A. STEWART*
- Dynamic belief revision operators, *A.C. NAYAK, M. PAGNUCCO, P. PEPPAS*

### ARTIFICIAL INTELLIGENCE V147 N°1-2 JULY 2003

- On the undecidability of probabilistic planning and related stochastic optimization problems, *O. MADANI, S. HANKS, A. CONDON*
- Weak, strong, and strong cyclic planning via symbolic model checking, *A. CIMATTI, M. PISTORE, M. ROVERI, P. TRAVERSO*
- SAT-based planning in complex domains: Concurrency, constraints and nondeterminism, *C. CASTELLINI, E. GIUNCHIGLIA, A. TACCHELLA*
- Contingent planning under uncertainty via stochastic satisfiability, *S.M. MAJERCIK, M.L. LITTMAN*
- Equivalence notions and model minimization in Markov decision processes, *R. GIVAN, T. DEAN, M. GREIG*
- Solving factored MDPs using non-homogeneous partitions, *K.-E. KIM, T. DEAN*
- Performance bounds for planning in unknown terrain, *S. KOENIG, C. TOVEY, Y. SMIRNOV*

### ARTIFICIAL INTELLIGENCE V148 N°1-2 AUGUST 2003

- Fuzzy Set and Possibility Theory-Based Methods in Artificial Intelligence, Edited by *D. DUBOIS AND H. PRADE*
- Fuzzy set and possibility theory-based methods in artificial intelligence, *D. DUBOIS, H. PRADE*
- Fuzzy rDFCSP and planning, *I. MIGUEL, O. SHEN*
- A fuzzy constraint based model for bilateral, multi-issue negotiations in semi-competitive environments, *X. LUO, N. R. JENNINGS, N. SHADBOLT, H. LEUNG, J. HO-MAN LEE*
- Fuzzy constraint networks for signal pattern recognition, *P. FELIX, S. BARRO, R. MARIN*
- Representation and fusion of heterogeneous fuzzy information in the 3D space for model-based structural recognition--Application to 3D brain imaging, *I. BLOCH, T. GERAUD, H. MAITRE*
- Agent-oriented epistemic reasoning: Subjective conditions of knowledge and belief, *D. G. SCHWARTZ*
- Testing the descriptive validity of possibility theory in human judgments of uncertainty, *E. RAUFASTE, R. DA SILVA NEVES, C. MARINE*
- Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach, *D. DUBOIS, H. FARGIER, P. PERNY*
- Temporal scenario modelling and recognition based on possibilistic logic, *M. GRABISCH*
- Logical representation and fusion of prioritized information based on guaranteed possibility measures: Application to the distance-based merging of classical bases, *S. BENFERHAT, S. KACI*
- Possibilistic instance-based learning, *E. HULLERMEIER*
- Operations and evaluation measures for learning possibilistic graphical models, *C. BORGELT, R. KRUSE*

### ARTIFICIAL INTELLIGENCE V149 N°1 SEPTEMBER 2003

- Preferences and explanations, *R. PINO-PEREZ, C. UZCATEGUI*
- Belief, information acquisition, and trust in multi-agent systems--A modal logic formulation, *C.-J. LIAU*
- Metaqueries: Semantics, complexity, and efficient algorithms, *R. BEN-ELIYAHU-ZOHARY, E. GUDES, G. IANNI*
- Embodied Cognition: A field guide, *M.L. ANDERSON*
- Embodied artificial intelligence, *R. CHRISLEY*
- Representations, symbols, and embodiment, *M.L. ANDERSON*

### ARTIFICIAL INTELLIGENCE V149 N°2 OCTOBER 2003

- Nonmonotonic inconsistency, *C.B. CROSS*
- Point algebras for temporal reasoning: Algorithms and complexity, *M. BROXVALL, P. JONSSON*
- Two party immediate response disputes: Properties and efficiency, *P.E. DUNNE, T.J.M. BENCH-CAPON*

## SOMMAIRES DES REVUES

- Sound and complete qualitative simulation is impossible, A.C.C. SAY, H.L. AKIN

### AI MAGAZINE

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

#### AI MAGAZINE V24 N°2 SUMMER 2003

- Advances in Artificial Intelligence Research and Applications at IJCAI-03, S. HEDBURG
- In Memoriam: Robert Englemore, B. G. BUCHANAN, T. C. RINDFLEISCH, E. A FEIGENBAUM
- An Overview of RoboCup-2002 Fukuoka/Busan, M. ASADA, O. OBST, D. POLANI, B. BROWNING, A. BONARINI, M. FUJITA, T. CRISTALLER, T. TAKAHASHI, S. TADOKORO, E. SKLAR, G. A. KAMINKA
- SPADES: A System for Parallel-Agent, Discrete-Event Simulation, P. RILEY
- RoboCupJunior: Learning with Educational Robotics, E. SKLAR, A. EGUCHI, J. JOHNSON
- Toward RoboCup without Color Labeling, R. HANEK, T. SCHMITT, S. BUCK, M. BEETZ
- GRACE: An Autonomous Robot for the AAAI Robot Challenge, R. SIMMONS, D. GOLDBERG, A. GOODE, M. MONTEMERLO, N. ROY, B. SELLNER, C. URMSOON, A. SCHULTZ, M. ABRAMSON, W. ADAMS, A. ATRASH, M. BUGAJSKA, M. COBLENZ, M. MACMAHON, D. PERZANOWSKI, I. HORSWILL, R. ZUBEK, D. KORTENKAMP, B. WOLFE, T. MILAM, B. MAXWELL
- Learning-Assisted Automated Planning: Looking Back, Taking Stock, Going Forward, T. ZIMMERMAN, S. KAMBHAMPATI
- The Twenty-Fifth Annual German Conference on Artificial Intelligence (KI-2002), J. KOEHLER, G. LAKEMEYER
- The 2002 Starting Artificial Intelligence Researchers Symposium, T. VIDAL
- Ray Reiter's Knowledge in Action: A Review, D. MCDERMOTT
- An Introduction to Support Vector Machines: A Review, Y. CHEN, I. G. COUNCILL

#### AI MAGAZINE V24 N°3 FALL 2003

- Ontology Research, C. WELTY
- Sweetening WORDNET with DOLCE, A. GANGEMI, N. GUARINO, C. MASOLO, A. OLTRAMARI
- Where Are the Semantics in the Semantic Web?, M. USCHOLD
- WEBODE in a Nutshell, J. C. ARPÍREZ, O. CORCHO, M. FERNÁNDEZ-LÓPEZ, A. GÓMEZ-PÉREZ
- Ontologies for Corporate Web Applications, L. OBRST, H. LIU, R. WRAY
- The Process Specification Language (PSL) Theory and Applications, M. GRÜNINGER, C. MENZEL

- The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata, M. DOERR
- A Framework for the Development of Personalized, Distributed Web-Based Configuration Systems, L. ARDISSONO, A. FELFERNIG, G. FRIEDRICH, A. GOY, D. JANNACH, G. PETRONE, R. SCHÄFER, M. ZANKER
- In Search of the Horowitz Factor, G. WIDMER, S. DIXON, W. GOEBL, E. PAMPALK, A. TOBUDIC

### COGNITIVE SCIENCE

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

#### COGNITIVE SCIENCE V27 N°4 JULY 2003

- A real-world rational agent: unifying old and new AI, P.F.M.J. VERSCHURE, P. ALTHAUS
- Instance-based learning in dynamic decision making, C. GONZALEZ, J.F. LERCH, C. LEBIERE
- Visualization, pattern recognition, and forward search: effects of playing speed and sight of the position on grandmaster chess errors, C.F. CHABRIS, E.S. HEARST
- Intra-sentential context effects on the interpretation of logical metonymy, M. LAPATA, F. KELLER, C. SCHEEPERS
- The effects of self-explaining when learning with text or diagrams, S. AINSWORTH, A. TH LOIZOU

### APPLIED ARTIFICIAL INTELLIGENCE

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

#### APPLIED ARTIFICIAL INTELLIGENCE V17 N°5-6 JULY 2003

- Data preparation for data mining, S. ZHANG, C. ZHANG, Q. YANG
- An agent-based hybrid framework for database mining, Z. ZHANG, C. ZHANG, S. ZHANG
- Gasrule for knowledge discovery, N. N. BINTI ABDULLAH, M. LIQUIÈRE, S. A. CERRI
- Pre-processing of high-dimensional categorical predictors in classification settings, E. TUV, G. RUNGER
- Web-log cleaning for constructing sequential classifiers, Q. YANG, T. LI, K. WANG
- Cooperative strategy for web data mining and cleaning, Y. LI, C. ZHANG, S. ZHANG
- Summarization and categorization of text data in high-level data cleaning for information retrieval, M. SARAVANAN, P. C. REGHU RAJ, S. RAMAN
- Feature selection for the naive bayesian classifier using decision trees, C. A. RATANAMAHATANA, D. GUNOPULOS
- A feature selection bayesian approach for extracting classification rules with a clustering genetic algorithm, E. R. HRUSCHKA, JR., E. R. HRUSCHKA, N. F. F. EBECKEN

## SOMMAIRES DES REVUES

- Numeric mapping and learnability of naive bayes, *H. ZHANG, C. X. LING*
- An analysis of four missing data treatment methods for supervised learning, *G. E. A. P. A. BATISTA, M. C. MONARD*
- A pre-processing method to deal with missing values by integrating clustering and regression techniques, *S.-M. TSENG, K.-H. WANG, C.-I. LEE*
- Toward databases mining: pre-processing collected data, *X. YAN, C. ZHANG, S. ZHANG*

### APPLIED ARTIFICIAL INTELLIGENCE V17 N°7 AUGUST 2003

- Intelligent automatic landing system using time delay neural network controller, *J.-G. JUANG, H.-H. CHANG, W.-B. CHANG*
- The s-ethos system: a methodology for systematic flight analysis centered on human factors, *E. CHOURAQUI, C. DONIAT*
- Optimum tolerance design using constraint networks and relative sensitivity ratio algorithm, *C. C. YANG, V. N. ACHUTHA NAIKAN*
- Application of neural networks to stock prediction in « pool » companies, *O. SAPENA, V. BOTTI, E. ARGENTE*

### COMPUTATIONAL INTELLIGENCE

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

### COMPUTATIONAL INTELLIGENCE V19 N°2 MAY 2003

- Semantic-Based Information Retrieval for Content Management and Security, *B.-H. YUN, C.-H. SEO*
- Blasting Open a Choice Space: Learning Inflectional Morphology for NLP, *M. MCSHANE, S. NIRENBURG*
- Web Discovery and Filtering Based on Textual Relevance Feedback Learning, *W. LAM, W. WANG, C.-W. YUE*
- Inductive Inference by Using Information Compression, *B. CHOI*
- A Study of AdaBoost with Naive Bayesian Classifiers: Weakness and Improvement, *K. MING TING, Z. ZHENG*
- Belief, Logic, and Partial Truth, *P. SNOW*

### MINDS AND MACHINES

JOURNAL FOR ARTIFICIAL INTELLIGENCE PHILOSOPHY AND COGNITIVE SCIENCE

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

### MINDS AND MACHINES V13 N°3 AUGUST 2003

- The Epistemological Foundations of Artificial Agents, *N. J. LACEY, M. H. LEE*

- The Influence of Epistemology on the Design of Artificial Agents, *M.H. LEE, N.J. LACEY*
- What Did You Mean by That? Misunderstanding, Negotiation, and Syntactic Semantics, *W. J. RAPAPORT*

### Book review

- M. J. Wooldridge, Reasoning about Rational Agents, Intelligent Robots and Autonomous Agents Series, Cambridge, MA: The MIT Press, 2000, by *L. SPALAZZI*
- James H. Fetzer, Computers and Cognition: Why Minds Are Not Machines, Dordrecht: Kluwer, 2001, by *R. WYATT*
- John Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole, 2000, by *J. GELLER*
- Gerald Edelman and Giulio Tononi, A Universe of Consciousness: How Matter Becomes Imagination, New York: Basic Books, 2000, by *D. J. COLE*
- Michael Tye, Consciousness, Color, and Content, Representation and Mind Series, Cambridge, MA/London: A Bradford Book, MIT Press, 2000, by *N. J.T. THOMAS*
- Kapa Korta, Ernest Sosa, and Xabier Arrazola, eds., Cognition, Agency and Rationality: Proceedings of the Fifth International Colloquium on Cognitive Science, Philosophical Studies Series 79, Dordrecht/Boston: Kluwer Academic Publishers, 1999, by *H. D. MULLER*

### ARTIFICIAL INTELLIGENCE REVIEW

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

### INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

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

### IJHCS V58 N°4 APRIL 2003

- A method for team intention inference, *T. KANNO, K. NAKATA, K. FURUTA*
- Fuzzy query interface for a business database, *R.A. RIBEIRO, A.M. MOREIRA*
- Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design, *K.M. STANNEY, M. MOLLAGHASEMI, L. REEVES, R. BREAU, D.A. GRAEBER*
- Human performance modeling in temporary segmentation Chinese character handwriting recognizers, *C. WU, K. ZHANG, Y. HU*
- Effects of "gender" of the computer on informational social influence: the moderating role of task type, *E.-J. LEE*
- A unifying framework for intelligent DNS management, *C.-S. CHEN, S.-S. TSENG, C.-L. LIU*

## IJHCS V58 N°6 JUNE 2003

- Human performance and embedded intelligent technology in safety-critical systems, *M. GRABOWSKI, S.D. SANBORN*
- The importance of trust and community in developing and maintaining a community electronic network, *A. OXENDINE, E. BORGIDA, J.L. SULLIVAN, M.S. JACKSON*
- The role of trust in automation reliance, *M.T. DZINDOLET, S.A. PETERSON, R.A. POMRANKY, L.G. PIERCE, H.P. BECK*
- The effects of errors on system trust, self-confidence, and the allocation of control in route planning, *P. DE VRIES, C. MIDDEN, D. BOUWHUIS*
- On-line trust: concepts, evolving themes, a model, *C.L. CORRITORE, B. KRACHER, S. WIEDENBECK*
- The researcher's dilemma: evaluating trust in computer-mediated communication, *J. RIEGELSBERGER, M.A. SASSE, J.D. MCCARTHY*
- Empirical research in on-line trust: a review and critical assessment, *S. GRABNER-KRAUTER, E.A. KALUSCHA*

## IJHCS V59 N°1-2 JULY 2003

- To feel or not to feel: The role of affect in human-computer interaction, *E. HUDLICKA*
- New visions of human-computer interaction: making affect compute, *M.D. MCNEESE*
- Affective computing: challenges, *R.W. PICARD*
- Is affective computing an oxymoron?, *E. HOLLNAGEL*
- Response: Is affective computing an oxymoron?, *E. HUDLICKA*
- Response: is affective computing an oxymoron?, *M.D. MCNEESE*
- From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent, *F.D. ROSIS, C. PELACHAUD, I. POGGI, V. CAROFIGLIO, B.D. CAROLIS*
- Emotion and sociable humanoid robots, *C. BREAZEAL*
- The production and recognition of emotions in speech: features and algorithms, *O. PIERRE-YVES*
- Pupil size variation as an indication of affective processing, *T. PARTALA, V. SURAKKA*
- Physiological responses to different WEB page designs, *R.D. WARD, P.H. MARSDEN*
- Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques, *A. CAMURRI, I. LAGERLOF, G. VOLPE*
- SenToy: an affective sympathetic interface, *A. PAIVA, M. COSTA, R. CHAVES, M. PIEDADE, D. MOURAO, D. SOBRAL, K. HOOK, G. ANDERSSON, A. BULLOCK*
- Nonverbal indicators of malicious intent: affective components for interrogative virtual reality training, *F. MCKENZIE, M. SCERBO, J. CATANZARO, M. PHILLIPS*
- Developing multimodal intelligent affective interfaces for tele-home health care, *C. LISETTI, F. NASOZ, C. LEROUGE, O. OZYER, K. ALVARE*

## IJHCS V59 N°3 SEPTEMBER 2003

- Context-based free-form annotation in XML documents, *W.-S. SOHN, J.-K. KIM, S.-K. KO, S.-B. LIM, Y.-C. CHOY*
- Off to new shores: conceptual knowledge discovery and processing, *G. STUMME*
- The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task, *R. PEARSON, P. VAN SCHAİK*
- Internet attitudes and Internet use: some surprising findings from the HomeNetToo project, *L.A. JACKSON, A.V. EYE, G. BARBATSIS, F. BIOCCA, Y. ZHAO, H.E. FITZGERALD*
- Perceived usefulness, ease of use and electronic supermarket use, *R. HENDERSON, M.J. DIVETT*

## IJHCS V59 N°4 OCTOBER 2003

- Zhang and Dillon Special Issue on HCI and MIS, Edited by *D. PING ZHANG AND A. DILLON*
- HCI and MIS: shared concerns, *P. ZHANG, A. DILLON*
- The evolution of US state government home, *T. RYAN, R.H.G. FIELD, L. OLFMAN*
- Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model, *M.Y. YI, Y. HWANG*
- Predicting e-services adoption: a perceived risk facets perspective, *M.S. FEATHERMAN, P.A. PAVLOU*
- A person-artefact-task (PAT) model of flow antecedents in computer-mediated environments, *C.M. FINNERAN, P. ZHANG*
- Issues and strategies for integrating HCI in masters level MIS and e-commerce programs, *S.S. CHAN, R.J. WOLFE, X. FANG*

## COMPUTATIONAL LINGUISTICS

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

## COMPUTATIONAL LINGUISTICS V59 N°2 JUNE 2003

- A Model for Matching Semantic Maps between Languages (French/English, English/French), *S. PLOUX, H. JI*
- Implementing the Binding and Accommodation Theory for Anaphora Resolution and Presupposition Projection, *J. BOS*
- Document Structure, Richard Power, *D. SCOTT, N. BOUAYAD-AGHA*
- A Probabilistic Account of Logical Metonymy, *M. LAPATA, A. LASCARIDES*

### Book review

- Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins edited by Marie-Hélène Corréard, by *W. HAYNES, M. EVENS*
- Multimodality in Language and Speech Systems edited by Björn Granström, David House, and Inger Karlsson, by *M. JOHNSTON*

- The Semantics of Relationships: An Interdisciplinary Perspective edited by Rebecca Green, Carol A. Bean, and Sung Hyon Myaeng, by *M. LAPATA*
- Recent Advances in Computational Terminology edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, by *R. GAIZAUSKAS*

## NATURAL LANGUAGE ENGINEERING

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

### NLE V9 N°1 MARCH 2003

#### Special issue

- Finite state methods in natural language processing, *L. KARTTUNEN, K. KOSKENNIEMI, G. VAN NOORD*
- Finite state methods for hyphenation, *G. BOUMA*
- Recursion by optimization: on the complexity of bidirectional optimality theory, *G. JÄGER*
- Application of finite-state transducers to the acquisition of verb subcategorization information, *I. ALDEZABAL, M. ARANZABE, K. GOJENOLA, M. OROÑOZ, K. SARASOLA, A. ATUTXA*
- An efficient incremental DFA minimization algorithm, *B. W. WATSON, J. DACIUK*
- Logical specification of regular relations for NLP, *N. VAILLETTE*
- Lenient morphological analysis, *K. OFLAZER*

### NLE V9 N°2 JUNE 2003

- High-level authoring of illustrated documents, *K. VAN DEEMTER, R. POWER*
- Mostly-supervised statistical segmentation of Japanese kanji sequences, *R. KUBOTA ANDO, L. LEE*
- A portable method for acquiring information extraction patterns without annotated corpora, *N. CATALÀ, N. CASTELL, M. MARTÍN*
- A statistical information extraction system for Turkish, *G. TÜR, D. HAKKANI-TÜR, K. OFLAZER*

### NLE V9 N°3 SEPTEMBER 2003

- An approach to the formal specification of lingware, *B. GARGOURI, M. JMAIEL, A. BEN HAMADOU*
- Open-domain textual question answering techniques, *S. M. HARABAGIU, S. J. MAIORANO, M. A. PASCA*
- Analysis of the grammatical functions between adnoun and noun phrases in Korean using Support Vector Machines, *S. LEE, J. SEO, T-Y. JANG*
- Evaluation-driven design of a robust coreference resolution system, *A. POPESCU-BELIS*

## USER MODELING AND USER-ADAPTED INTERACTION

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

### USER MODELING AND USER-ADAPTED INTERACTION V13 N°1-2 FEB-MAY 2003

- A Movie Recommendation System – An Application of Voting Theory in User Modeling, *R. MUKHERJEE, N. SAJJA, S. SEN*
- A System for Building Intelligent Agents that Learn to Retrieve and Extract Information, *T. ELIASSI-RAD, J. SHAVLIK*
- Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents, *J. CASSELL, T. BICKMORE*
- Interactive Improvisational Music Companionship: A User-Modeling Approach, *B. THOM*
- Multi-Agent Multi-User Modeling in I-Help, *J. VASSILEVA, G. MCCALLA, J. GREER*

### USER MODELING AND USER-ADAPTED INTERACTION V13 N°3 AUGUST 2003

- Personalizing the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE, *K.A. PAPANIKOLAOU, M. GRIGORIADOU, H. KORNILAKIS, G. D. MAGOULAS*
- Probabilistic Student Modelling to Improve Exploratory Behaviour, *A. BUNT, C. CONATI*

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

### COMPUTER SPEECH AND LANGUAGE V17 N°4 OCTOBER 2003

- Acoustic model clustering based on syllable structure, *I. SHAFRAN, M. OSTENDORF*
- The effect of pruning and compression on graphical representations of the output of a speech recognizer, *Y. LIU, M. P. HARPER, M. T. JOHNSON, L. H. JAMIESON*
- Modeling partial pronunciation variations for spontaneous Mandarin speech recognition, *Y. LIU, P. FUNG*
- Non-linear feature extraction for robust speech recognition in stationary and non-stationary noise, *O. ZHU, A. ALWAN*
- Preliminary experiments in speaker verification using time-dependent largest Lyapunov exponents, *A. PETRY, D. A. COUTO BARONE*
- Language modelling for Russian and English using words and classes, *E. W. D. WHITTAKER, P. C. WOODLAND*

## MACHINE LEARNING

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

### MACHINE LEARNING V52 N°1-2 JULY-AUGUST 2003

- Editorial: Methods in Functional Genomics, *P. SEBASTIANI, I. S. KOHANE, M. F. RAMONI*
- Relation Between Permutation-Test P Values and Classifier Error Estimates, *T. HSING, S. ATTOOR, E. DOUGHERTY*
- Boosting and Microarray Data, *P. M. LONG, V. BERLIAN VEGA*
- Analysis and Visualization of Gene Expression Microarray Data in Human Cancer Using Self-Organizing Maps, *S. HAUTANIEMI, O. YLI-HARJA, J. ASTOLA, P. KAURANIEMI, A. KALLIONIEMI, M. WOLF, J. RUIZ, S. MOUSSES, O-P. KALLIONIEMI*
- Self-Organizing Latent Lattice Models for Temporal Gene Expression Profiling, *B-T. ZHANG, J. YANG, S. W. CHI*
- Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data, *S. MONTI, P. TAMAYO, J. MESIROV, T. GOLUB*
- Inclusion of Textual Documentation in the Analysis of Multidimensional Data Sets: Application to Gene Expression Data, *S. RAYCHAUDHURI, H. SCHÜTZE, R. B. ALTMAN*
- On Learning Gene Regulatory Networks Under the Boolean Network Model, *H. LÄHDESMÄKI, I. SHMULEVICH, O. YLI-HARJA*
- External Control in Markovian Genetic Regulatory Networks, *A. DATTA, A. CHOUDHARY, M. L. BITTNER, E. R. DOUGHERTY*

### MACHINE LEARNING V52 N°3 SEPTEMBER 2003

- Tree Induction for Probability-Based Ranking, *F. PROVOST, P. DOMINGOS*
- Feature Weighting in k-Means Clustering, *D. S. MODHA, W. SCOTT SPANGLER*
- Inference for the Generalization Error, *C. NADEAU, Y. BENGIO*

### MACHINE LEARNING V53 N°1-2 OCTOBER 2003

- A Microchoice Bound for Continuous-Space Classification Algorithms, *Y. GAT*
- Theoretical and Empirical Analysis of ReliefF and RreliefF, *M. ROBNIK-ŽIKONJA, I. KONONENKO*
- Online Ensemble Learning: An Empirical Study, *A. FERN, R. GIVAN*
- Programming by Demonstration Using Version Space Algebra, *T. LAU, S. A. WOLFFMAN, P. DOMINGOS, D. S. WELD*
- Improved Rooftop Detection in Aerial Images with Machine Learning, *M.A. MALOOF, P. LANGLEY, T.O. BINFORD, R. NEVATIA, S. SAGE*

## NEURAL NETWORKS

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

### NEURAL NETWORKS V16 N°5-6 JUNE 2003

- Welcome to the special issue: the best of the best, *D.C. WUNSCH, M. HASSELMO, D. WANG, G.K. VENAYAGA-MOORTHY*
- Adaptive force generation for precision-grip lifting by a spectral timing model of the cerebellum, *A. ULLOA, B.J. RHODES*
- Radial basis function neural networks for nonlinear Fisher discrimination and Neyman-Pearson classification, *D. CASASSENT, X.-W. CHEN*
- Intrinsic generalization analysis of low dimensional representations, *X. LIU, A. SRIVASTAVA, D. WANG*
- Application of four-layer neural network on information extraction, *M. HAN, L. CHENG, H. MENG*
- Subject independent facial expression recognition with robust face detection using a convolutional neural network, *M. MATSUGU, K. MORI, Y. MITARI, Y. KANEDA*
- A generalized feedforward neural network architecture for classification and regression, *G. ARULAMPALAM, A. BOUZERDOUM*
- Hierarchical cognitive maps, *H. VOICU*
- Modeling goal-directed spatial navigation in the rat based on physiological data from the hippocampal formation, *R.A. KOENE, A. GORCHETCHNIKOV, R.C. CANNON, M.E. HASSELMO*
- An efficient training algorithm for dynamic synapse neural networks using trust region methods, *H.H. NAMARVAR, T.W. BERGER*
- Temporal binding as an inducer for connectionist recruitment learning over delayed lines, *C. GUNAY, A.S. MAIDA*
- Developments in understanding neuronal spike trains and functional specializations in brain regions, *R.A. SANTIAGO, J. MCNAMES, K. BURCHIEL, G.G. LENDARIS*
- Shaping up simple cell's receptive field of animal vision by ICA and its application in navigation system, *L. ZHANG, J. MEI*
- eLoom and Flatland: specification, simulation and visualization engines for the study of arbitrary hierarchical neural architectures, *T.P. CAUDELL, Y. XIAO, M.J. HEALY*
- Associative morphological memories based on variations of the kernel and dual kernel methods, *P. SUSSNER*
- Adaptive double self-organizing maps for clustering gene expression profiles, *H. RESSOM, D. WANG, P. NATARAJAN*
- An accelerated procedure for recursive feature ranking on microarray data, *C. FURLANELLO, M. SERAFINI, S. MERLER, G. JURMAN*
- Pattern completion through phase coding in population neurodynamics, *A. GUTIERREZ-GALVEZ, R. GUTIERREZ-OSUNA*
- Passive dendritic integration heavily affects spiking dynamics of recurrent networks, *G.A. ASCOLI*

- Abductive reasoning with recurrent neural networks, *A.M. ABDELBAR, E.A.M. ANDREWS, D.C. WUNSCH*
  - Neural networks with chaotic recursive nodes: techniques for the design of associative memories, contrast with Hopfield architectures, and extensions for time-dependent inputs, *E. DEL-MORAL-HERNANDEZ*
  - Simple and conditioned adaptive behavior from Kalman filter trained recurrent networks<sup>\*</sup>, *L.A. FELDKAMP, D.V. PROKHOROV, T.M. FELDKAMP*
  - Learning robot actions based on self-organising language memory, *S. WERMTER, M. ELSHAW*
  - Autonomous mental development in high dimensional context and action spaces, *A. JOSHI, J. WENG*
  - Chaos control and synchronization, with input saturation, via recurrent neural networks, *E.N. SANCHEZ, L.J. RICALDE*
  - Proper orthogonal decomposition based optimal neurocontrol synthesis of a chemical reactor process using approximate dynamic programming, *R. PADHI, S.N. BALAKRISHNAN*
  - Numerical solution of elliptic partial differential equation using radial basis function neural networks, *L. JIANYU, L. SIWEI, Q. YINGJIAN, H. YAPING*
  - Statistical efficiency of adaptive algorithms, *B. WIDROW, M. KAMENETSKY*
  - On structure-exploiting trust-region regularized nonlinear least squares algorithms for neural-network learning, *E. MIZUTANI, J.W. DEMMEL*
  - Stochastic resonance in noisy threshold neurons, *B. KOSKO, S. MITAIM*
  - Quantum optimization for training support vector machines, *D. ANGUITA, S. RIDELLA, F. RIVIECCIO, R. ZUNINO*
  - On the quality of ART1 text clustering, *L. MASSEY*
  - Extension neural network and its applications, *M.H. WANG, C.P. HUNG*
  - Fuzzy least squares support vector machines for multiclass problems, *D. TSUJINISHI, S. ABE*
  - Evolving efficient learning algorithms for binary mappings, *J.A. BULLINARIA*
  - A network for recursive extraction of canonical coordinates, *A. PEZESHKI, M.R. AZIMI-SADJADI, L.L. SCHARF*
  - Automatic basis selection techniques for RBF networks, *A. GHODSI, D. SCHUURMANS*
  - Data smoothing regularization, multi-sets-learning, and problem solving strategies, *L. XU*
  - Million city traveling salesman problem solution by divide and conquer clustering with adaptive resonance neural networks, *S.A. MULDER, D.C. WUNSCH*
  - A practical sub-space adaptive filter, *A. ZAKNICH*
  - Pharmacodynamic population analysis in chronic renal failure using artificial neural networks-a comparative study, *A.E. GAWEDA, A.A. JACOBS, M.E. BRIER, J.M. ZURADA*
  - Electronic nose based tea quality standardization, *R. DUTTA, K.R. KASHWAN, M. BHUYAN, E.L. HINES, J.W. GARDNER*
  - A novel neural network-based survival analysis model, *A. ELEUTERI, R. TAGLIAFERRI, L. MILANO, S. DE PLACIDO, M. DE LAURENTIIS*
  - Divide-and-conquer approach for brain machine interfaces: nonlinear mixture of competitive linear models, *S.-P. KIM, J.C. SANCHEZ, D. ERDOGMUS, Y.N. RAO, J. WESSBERG, J.C. PRINCIPE, M. NICOLELIS*
  - Stochastic error whitening algorithm for linear filter estimation with noisy data, *Y.N. RAO, D. ERDOGMUS, G.Y. RAO, J.C. PRINCIPE*
  - New internal optimal neurocontrol for a series FACTS device in a power transmission line, *J.-W. PARK, R.G. HARLEY, G.K. VENAYAGAMOORTHY*
  - Design of an adaptive neural network based power system stabilizer, *W. LIU, G.K. VENAYAGAMOORTHY, D.C. WUNSCH*
  - On neural network techniques in the secure management of communication systems through improving and quality assessing pseudorandom stream generators, *D.A. KARRAS, V. ZORKADIS*
  - Multimedia authenticity protection with ICA watermarking and digital bacteria vaccination, *H. SZU, S. NOEL, S.-B. YIM, J. WILLEY, J. LANDA*
  - Interpolation processes in the visual perception of objects, *P.J. KELLMAN*
  - Laminar cortical dynamics of visual form perception, *S. GROSSBERG*
  - Moving objects appear to slow down at low contrasts, *S. ANSTIS*
  - Neural models of motion integration and segmentation, *E. MINGOLLA*
- NEURAL NETWORKS V16 N°8 OCTOBER 2003**
- Recurrent neural networks with trainable amplitude of activation functions, *S.L. GOH, D.P. MANDIC*
  - Solving the XOR problem and the detection of symmetry using a single complex-valued neuron [rapid communication], *T. NITTA*
  - A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes [review article], *S. GROSSBERG, D.V. REPIN*
  - A neural network simulating human reach-grasp coordination by continuous updating of vector positioning commands, *A. ULLOA, D. BULLOCK*
  - A model synapse that incorporates the properties of short- and long-term synaptic plasticity, *A.R. SARGSYAN, A.A. MELKONYAN, C. PAPTAEODOROPOULOS, H.H. MKRTCHIAN, G.K. KOSTOPOULOS*
  - ∑ Back-propagation learning of infinite-dimensional dynamical systems, *I. TOKUDA, R. TOKUNAGA, K. AIHARA*
  - ∑ Controlling chaos in a chaotic neural network, *G. HE, Z. CAO, P. ZHU, H. OGURA*

## SOMMAIRES DES REVUES

- Neural independent component analysis by 'maximum-mismatch' learning principle, *S. FIORI*
- Necessary and sufficient condition for absolute stability of normal neural networks, *T. CHU, C. ZHANG, Z. ZHANG*

### Book review

- Artificial Immune Systems: A New Computational Intelligence Approach, - L.N. de Castro and J. Timmis (Eds); Springer, 2002, by *F. AZUAJE*

### ARTIFICIAL INTELLIGENCE IN MEDICINE

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

#### AI IN MEDICINE V28 N°1 MAY 2003

- A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer, *P.J.G. LISBOA, H. WONG, P. HARRIS, R. SWINDELL*
- Active subgroup mining: a case study in coronary heart disease risk group detection, *D. GAMBERGER, N. LAVRAC, G. KRSTACIC*
- Analyzing tumor gene expression profiles, *C. Peterson, M. RINGNER*
- An Epicurean learning approach to gene-expression data classification, *A. ALBRECHT, S.A. VINTERBO, L. OHNO-MACHADO*
- Extracting multisource brain activity from a single electro-magnetic channel, *C.J. JAMES, D. LOWE*
- Linking clinical data using XML topic maps, *R. SCHWEIGER, S. HOELZER, D. RUDOLF, J. RIEGER, J. DUDECK*

#### AI IN MEDICINE V28 N°2 JUNE 2003

- Knowledge-based neurocomputing in medicine, *I. CLOETE, K. ROHR*
- Inductive bias strength in knowledge-based neural networks: application to magnetic resonance spectroscopy of breast tissues, *C.W. OMLIN, S. SNYDERS*
- A model for single and multiple knowledge based networks, *G. BOLOGNA*
- Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue, *M.E. FUTSCHIK, A. REEVE, N. KASABOV*
- Explaining the output of ensembles in medical decision support on a case by case basis, *R. WALL, P. CUNNINGHAM, P. WALSH, S. BYRNE*
- Knowledge-based approach to septic shock patient data using a neural network with trapezoidal activation functions, *J. PAETZ*

#### AI IN MEDICINE V28 N°3 JULY 2003

- Temporal abstraction and inductive logic programming for arrhythmia recognition from electrocardiograms, *G. CAR-RAULT, M.-O. CORDIER, R. QUINIOU, F. WANG*
- Urinary nucleosides as potential tumor markers evaluated

by learning vector quantization, *F. DIETERLE, S. MULLER-HAGEDORN, H.M. LIEBICH, G. GAUGLITZ*

- Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines, *C. LU, T. VAN GESTEL, J.A.K. SUYKENS, S. VAN HUFFEL, I. VERGOTE, D. TIMMERMAN*
- Brain volumes characterisation using hierarchical neural networks, *S. DI BONA, H. NIEMANN, G. PIERI, O. SALVETTI*
- Comparison of tree-based methods for prognostic stratification of survival data, *M. RADESPIEL-TROGER, T. RABENSTEIN, H.T. SCHNEIDER, B. LAUS*

#### AI IN MEDICINE V29 N°1-2 SEPTEMBER 2003

- Artificial Intelligence in Medicine Europe AIME '01, *S. QUAGLINI*
- Integrating classification trees with local logistic regression in Intensive Care prognosis, *A. ABU-HANNA, N. DE KEIZER*
- VComprehensible evaluation of prognostic factors and prediction of wound healing, *M. ROBNIK-SIKONJA, D. CUKJATI, I. KONONENKO*
- Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection, *P. ANTAL, G. FANNES, D. TIMMERMAN, Y. MOREAU, B. DE MOOR*
- Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data, *M. ZAFFALON, K. WESNES, O. PETRINI*
- Transductive reliability estimation for medical diagnosis, *M. KUKAR*
- GenePath: a system for inference of genetic networks and proposal of genetic experiments, *B. ZUPAN, I. BRATKO, J. DEMSAR, P. JUVAN, T. CURK, U. BORSTNIK, J.R. BECK, J. HALTER, A. KUSPA, G. SHAULSKY*
- Integrating model-based decision support in a multi-modal reasoning system for managing type 1 diabetic patients, *S. MONTANI, P. MAGNI, R. BELLAZZI, C. LARIZZA, A.V. ROUDSARI, E.R. CARSON*
- Using OncoDoc as a computer-based eligibility screening system to improve accrual onto breast cancer clinical trials, *B. SEROUSSI, J. BOUAUD*
- Using lexical disambiguation and named-entity recognition to improve spelling correction in the electronic patient record, *P. RUCH, R. BAUD, A. GEISSBUHLER*

### INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

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

#### IJAR V33 N°2 JUNE 2003

- Evaluation of a verbal-numerical probability scale, *C. WITTEMAN, S. RENOUIJ*

## SOMMAIRES DES REVUES

- Searching for the dimension of valued preference relations, *J. GONZALEZ-PACHON, D. GOMEZ, J. MONTERO, J. YANEZ*
- On implicative closure operators in approximate reasoning, *R.O. RODRIGUEZ, F. ESTEVA, P. GARCIA, L. GODO*
- A survey on universal approximation and its limits in soft computing techniques, *D. TIKK, L.T. KOCZY, T.D. GEDEON*
- On the measurability of knowledge acquisition and query processing, *W. RODDER*

### IJAR V33 N°3 AUGUST 2003

- An alternative characterization of a Bayesian network, *S.K.M. WONG, T. LIN*
- Comparison of multiagent inference methods in multiply sectioned Bayesian networks, *Y. XIANG*

- Fusion of possibilistic knowledge bases from a postulate point of view, *S. BENFERHAT, S. KACI*
- The reasonableness of necessity, *P. SNOW*

### IJAR V34 N°1 SEPTEMBER 2003

- Constructing a logic of plausible inference: a guide to Cox's theorem, *K.S. VAN HORN*
- Level sets and minimum volume sets of probability density functions, *J. NUNEZ GARCIA, Z. KUTALIK, K.-H. CHO, O. WOLKENHAUER*
- Definition and classification of semi-fuzzy quantifiers for the evaluation of fuzzy quantified sentences, *F. DIAZ-HERMIDA, A. BUGARIN, S. BARRO*

## RÉSUMÉS HABILITATIONS ET THÈSES

### Méthodologie et Architectures pour la Conception de Systèmes Multi-Agents

Michel Occello

Habilitation à Diriger les Recherches de l'Université Joseph Fourier de Grenoble, Spécialité : INFORMATIQUE.

#### Résumé

L'ensemble des travaux présentés dans cet HDR se concentre dans le domaine de l'Intelligence Artificielle Distribuée (IAD) et des Systèmes Multi-Agents (SMA). Notre préoccupation a toujours été la décentralisation de l'intelligence. Focalisée d'abord sur les systèmes tableaux noirs (« blackboards »), elle a ensuite progressé vers une réelle notion de collectif à travers les systèmes multi-agents.

Nous avons toujours considéré les SMA comme permettant de proposer une nouvelle approche méthodologique pour la construction d'applications informatiques complexes, le postulat étant que les systèmes sont aujourd'hui constitués d'entités distribuées, hétéro-

gènes, travaillant à des niveaux d'abstraction différents, qui coopèrent entre elles et souvent avec des humains dans un environnement très dynamique.

Notre méthode de travail est fondée sur une recherche sur des modèles de base, un souci constant de généricité et de réutilisabilité, une abstraction de processus reproductibles à partir de la mise en oeuvre des modèles sur des applications réelles, conduisant à l'émergence d'un savoir-faire méthodologique par une approche à la fois théorique et expérimentaliste.

Nous présentons de nombreux résultats au niveau des modèles, de la démarche et de la constitution d'outils génériques d'aide à la conception, ainsi que des applications.

Michel Occello  
Université Pierre Mendès France – IUT de Valence  
51 Rue Barthélémy de Laffemas – BP 29  
26901 Valence Cedex 9

Mél : Michel.Occello@iut-valence.fr  
Tél : 04 75 41 88 40  
Fax : 04 75 41 88 44

## Dossier « Planification et heuristiques »

A paraître dans le bulletin numéro 56

### Coordonnateurs :

Patrick Fabiani (*ONERA/CERT, patrick.fabiani@cert.fr*), Éric Jacopin (*CREC Saint-Cyr, eric.jacopin@lip6.fr*),  
Vincent Vidal (*CRIL, vidal@cril.univ-artois.fr*)

### Objectifs

La construction de plans, i.e. la planification, est par nature une activité heuristique : un plan est un guide pour l'activité qui évite les alternatives impossibles ou trop complexes. Et de fait GPS, le premier système d'IA à être aussi un planificateur utilisait des heuristiques pour choisir les actions les plus prometteuses.

Que ce soit pour la planification en ligne ou hors ligne, réactive ou délibérative, d'actions ou de déplacements avec ou sans incertitudes, une heuristique permet d'orienter d'autant mieux les décisions qu'elle utilise, ou qu'elle résume, les propriétés les plus fines du problème, jusqu'à permettre parfois de prouver que le plan obtenu est la solution optimale du problème. Dans certains domaines ces propriétés sont très efficacement utilisées en faisant appel aux représentations adéquates : la géométrie en planification de mouvement donne des heuristiques indispensables. La planification de mouvement, pour spécifique qu'elle soit, a toujours été une source d'inspiration pour la recherche dans les graphes (plus courts chemins par exemple). Calculer automatiquement une heuristique s'apparente en toute généralité à un problème d'apprentissage du domaine.

Pour autant, l'utilisation d'heuristiques en planification a parfois été assimilée à une résolution ad hoc, ne fonctionnant que sur des domaines spécifiques.

L'aspiration à la généralité a attiré la planification vers l'indépendance d'un domaine particulier et l'utilisation d'heuristiques a longtemps été synonyme de peu d'intérêt de la part de la communauté ; jusqu'à la fin des années 90 où des heuristiques additives très simples ont abouti à des records de performance qui n'ont guère été égalés depuis.

Ce dossier, a pour but de faire le point sur les travaux en cours de la communauté francophone, portant sur l'utilisation d'heuristiques pour la planification. Si vous utilisez des heuristiques et que vous construisez des plans (et peu importe la représentation de ce plan), ce dossier est pour vous !

### Qui peut soumettre ?

Toute équipe de recherche de la communauté francopho-

ne appartenant à un établissement public ou industriel, impliquée dans des actions de recherche et de développement portant sur l'utilisation d'heuristiques pour la planification.

### Comment soumettre ?

Les contributions devront présenter succinctement l'équipe et les principaux travaux concernant le thème « Planification et heuristiques » : concepts, algorithmes, propriétés et critères d'optimalités, outils, expériences, applications et mises en œuvre. Elles seront organisées selon le schéma suivant :

Identification de l'équipe: noms, adresse, site Web, personne à contacter (téléphone, mél) ;

Membres de l'équipe concernés par le thème ;

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, merci d'envoyer votre contribution :

\*\*\*\*\* *avant le 1er décembre 2003* \*\*\*\*\*

à l'une des trois adresses électroniques suivantes :

**patrick.fabiani@cert.fr, eric.jacopin@lip6.fr**

**vidal@cril.univ-artois.fr.**

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

Adhésion individuelle et abonnement		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
<b>Nom :</b> Prénom : Affiliation : Adresse postale : N° de téléphone : N° de télécopie : Adresse électronique : Activités (à titre professionnel / à titre privé (rayez la mention inutile) ) :			
	Consultation du bulletin sur WEB (pour une personne)	Envoi du bulletin papier + un accès pour consultation du bulletin sur WEB	
<input type="checkbox"/> Adhésion simple :	30 Euros	60 Euros	
<input type="checkbox"/> Adhésion étudiant (sur justificatif) :	15 Euros	30 Euros	
<input type="checkbox"/> Adhésion de soutien	Sans objet	90 Euros	
<input type="checkbox"/> Abonnement au bulletin sans adhésion	Sans objet	55 Euros	
<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
<b>Organisme :</b> <b>Adresse postale commune aux bénéficiaires couverts par cette adhésion :</b> Nom et prénom du représentant : Fonction : Mél : Tél : Fax : Adresse postale :			
Le tarif d'adhésion comprend une partie fixe et une partie par bénéficiaire <b>Coordonnées des bénéficiaires (10 maximum) :</b>			
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'ECCA pour constituer un fichier européen.			
<b>Veillez trouver un règlement (à l'ordre de l'AFIA) de ..... Euros</b>			
<b>Trésorier AFIA :</b> Marc AYEL, LIA-Université de Savoie, 73376 Le Bourget du Lac cedex. <b>Mode d'adhésion :</b> De préférence, en ligne via le site Internet de l'AFIA : <a href="http://www.afia-france.fr">http://www.afia-france.fr</a> A défaut, cette page doit être envoyée au trésorier. <b>Modes de paiement :</b> 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. <b>TVA non applicable, article 293B du CGI</b>			

## Les dossiers du *Bulletin de l'AFIA*

Systèmes d'Information, connaissances et Intelligence Artificielle.....	Bulletin n°55
Web sémantique.....	Bulletin n°54
L'IA dans le RNTL.....	Bulletin n°53
IA et diagnostic.....	Bulletin n°52
Temps, espace et évolutif .....	Bulletin n°51
Equipes d'IA en France .....	Bulletin n°49/50
IA et Médecine .....	Bulletin n°48
Fouille de données .....	Bulletin n°46/47
IA et document .....	Bulletin n°44
IA et connexionisme .....	Bulletin n°43
IA et Vie Artificielle.....	Bulletin n°42
IA et CHM .....	Bulletin n°41
IA et EIAH.....	Bulletin n°40
Plates-formes multi-agents .....	Bulletin n°39
IA et WEB.....	Bulletin n°38
Mémoires d'entreprises.....	Bulletin n°36
IA et logique .....	Bulletin n°35
Ingénierie des connaissances.....	Bulletin n°34
IA et Télécommunications.....	Bulletin n°33
IA et Terminologie .....	Bulletin n°32
Décision et IA .....	Bulletin n°31
Raisonnement IA et Image.....	Bulletin n°30
Raisonnement temporel et spatial .....	Bulletin n°29
Systèmes Multi-agents .....	Bulletin n°28
IA et robotique.....	Bulletin n°27
I.A . et biologie moléculaire .....	Bulletin n°26
A. et droit .....	Bulletin n°25
I.A. et fusion de données .....	Bulletin n°24
I.A. et musique .....	Bulletin n°23
Apprentissage .....	Bulletin n°22
Les explications dans les SBC.....	Bulletin n°20
Pétrole-Chimie.....	Bulletin n°19
Le raisonnement à partir de cas .....	Bulletin n°18
I.A. et temps-réel.....	Bulletin n°17
Planification et action.....	Bulletin n°16
Traitement automatique des langues .....	Bulletin n°15
I.A. et médecine.....	Bulletin n°14
Diagnostic à base de modèles .....	Bulletin n°13
Validation des SBC .....	Bulletin n°12
Le connexionisme .....	Bulletin n°11
I.A. et jeux .....	Bulletin n°10
E.I.A.O. ....	Bulletin n°9
I.A. et gestion .....	Bulletin n°8
Conception et I.A. ....	Bulletin n°7
Intelligence artificielle distribuée.....	Bulletin n°6
Acquisition des Connaissances.....	Bulletin n°5
IA et ordonnancement.....	Bulletin n°4

## SOMMAIRE DU BULLETIN N° 55

Editorial. ....	3
La Vie de l'AFIA .....	4
Présentation de Laboratoires .....	8
Dossier : Systèmes d'Information, Connaissance et Intelligence Artificielle .....	23
Revue bibliographique .....	55
Compte rendu de conférence .....	59
Conférences .....	60
Sommaires des Revues .....	72
Résumés Habilitations et Thèses .....	72
Appels à dossiers .....	73

### CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Eté</i>
Réception des contributions: <b>15 décembre</b>	Réception des contributions: <b>15 juin</b>
Sortie le <b>31 janvier</b>	Sortie <b>31 juillet</b>

<i>Printemps</i>	<i>Automne</i>
Réception des contributions: <b>15 mars</b>	Réception des contributions: <b>15 septembre</b>
Sortie le <b>30 avril</b>	Sortie le <b>31 octobre</b>