

BULLETIN

DE

L'AFIA



NOVEMBRE 2001

N° 46-47

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 reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

Pour contacter l'AFIA

Président :
Bertrand BRAUNSCHWEIG
Institut Français du Pétrole
Direction Informatique
Mathématiques Appliquées
B.P. 311
92506 Rueil Malmaison CEDEX
Mél. : Bertrand.Braunschweig@ifp.fr

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**

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

Bertrand BRAUNSCHWEIG, président
Serge DUPUY, secrétaire
Marc AYEL, trésorier

**François ARLABOSSE, Jean-Paul BAQUIAST, Gilles BISSON,
Jérôme EUZENAT, Alain CARDON, Catherine GARBAY,
Jean-Daniel ZUCKER, Eunika MERCIER-LAURENT, Sylvie PESTY,
Gérard SABAH, Jean-Paul SANSONNET, Marc SCHOENAUER,
Michèle SEBAG, Laurent SIKLÓSSY, Fabien TORRE.**

Comité de rédaction

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

Jean-Paul BAQUIAST
Rubrique « Groupes de travail et collègues »
72 bis, Rue de Paris,
92190 Meudon
jean-paul.baquias@wanadoo.fr

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

Catherine BARRY-GRÉBOVAL
Rubriques « Présentation de laboratoires »
PSI — INSA de Rouen
place Émile Blondel BP 08
76131 Mont Saint Aignan Cedex
Catherine.Greboval@insa-rouen.fr

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

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

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

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



Réa. D.E.B. Copy – 01 46 31 64 53
ISSN 1273-1323
Dépot légal novembre 2001

Le bulletin de l'AFIA entre dans une phase expérimentale.

Dans l'éditorial du bulletin précédent, Gérard Sabah, rédacteur en chef du bulletin de l'AFIA depuis plusieurs années, a annoncé son départ en raison des nouvelles charges qui lui incombent tant dans son propre laboratoire, le LIMSI-CNRS, qu'au département STIC du CNRS. Le Bureau de l'AFIA et le Comité de Rédaction du bulletin s'unissent pour le remercier chaleureusement de tout le travail qu'il a mené au sein de l'AFIA pour faire vivre le bulletin avec une constance et une qualité qui risquent de ne pas être égalées. Ainsi, le bulletin a affaire à une double transition : d'une part un changement de rédaction et ceci dans un contexte en forte évolution.

Le changement de rédaction s'opère à la rentrée 2001 à l'occasion du renouvellement annuel du bureau de l'AFIA. Ainsi, nous remercions les « sortants » pour leur contribution et accueillons les nouveaux : Sylvie Pesty remplace Jean Charlet à la rubrique « Conférences et comptes rendus », rubrique qu'elle développe et rend très vivante sur le site Internet de l'AFIA ; Jean-Paul Baquiast remplace Patrick Gallinari à la rubrique « Groupes de travail et Collèges » ; enfin, Gérard Sabah a depuis peu ouvert une rubrique « IA au Québec » qui est tenue par Nicole Tourigny.

Le contexte de l'AFIA est actuellement en pleine évolution avec la volonté de son Président Bertrand Braunschweig de faire évoluer l'information au sujet de l'IA francophone vers l'Internet, en recentrant les supports vers le site WEB de l'AFIA. Ceci devrait permettre de diffuser une information plus à jour et plus interactive. Alors que la construction de ce nouveau site est en cours, la position du bulletin a été très discutée : d'une part, le bureau de l'AFIA a la volonté de factoriser les efforts entre le site et le bulletin, d'autre part beaucoup de chercheurs de la communauté nous ont fait savoir leur attachement à la « version papier » du bulletin, pour une consultation à tête reposée (je donnerais comme exemple la consultation des sommaires de revues que nous avons pensé supprimer du bulletin et placer sous forme de page de liens directs sur le site AFIA mais que paradoxalement, les chercheurs préfèrent souvent pouvoir consulter sur le papier), et aussi pour conserver des dossiers sur une période longue. C'est pourquoi nous devons être prudents et proposer une convergence éditoriale à terme plutôt qu'un basculement trop brutal vers le tout électronique. C'est ce que nous nous proposons de mettre en œuvre en 2002, en faisant des expérimentations, et en tenant compte de vos suggestions et critiques.

Ce numéro est le reflet des changements auxquels nous avons à faire face : le changement de rédaction nous a conduit à différer la sortie du n° 46 et à proposer un n° double 46/47 pour la fin 2001 qui comporte des rubriques dédoublées, comme la présentation de laboratoires (l'INRIA de Sophia Antipolis et le LIH du Havre). Malheureusement, nous ne présentons pas deux dossiers et ceci mérite d'être réfléchi car au cours de l'année 2000 puis en 2001, le Comité de Rédaction du bulletin a dû beaucoup solliciter les collègues pour obtenir des contributions, en particulier au niveau des dossiers et des présentations d'unités. Ceci nous a amené à nous poser la question de savoir comment renouveler le contenu éditorial du bulletin et aussi comment le rendre plus attractif. Cette question se reflète d'ailleurs au niveau du contenu de l'information mise en ligne sur le site Internet de l'AFIA. Deux réponses ont été évoquées : 1) tout en gardant l'IA comme thème central, il s'agit peut être d'élargir les thèmes abordés, en particulier aux champs où elle se met de plus en plus en œuvre comme par exemple, la Société de l'Information (au sens IST), l'Interaction entre les « gens » (ordinary people) et les systèmes et services distribués etc. (cette tendance a déjà largement été mise en œuvre dans les deux précédentes plate formes AFIA). 2) tout en gardant les rubriques classiques, il s'agit peut être d'élargir les modalités de contribution, par exemple vers des présentations de problématiques nouvelles, ou bien installées, selon un mode proche des écoles thématiques c'est-à-dire se situant entre le tutoriel et l'article de synthèse. Dans ce numéro, nous avons ouvert une rubrique, (mal) intitulée « synthèse » dans laquelle j'ai demandé deux contributions au pied-levé à mes collègues, ceci à titre expérimental.

Pour les numéros de 2002, nous faisons un appel général à contribution que ce soit dans le format classique des rubriques existantes (dans ce cas, adressez vous aux responsables de rubriques correspondantes -- pour les suggestions de dossiers le contact est J-P Sansonnet) ou bien sous forme de textes plus libres, plus personnels, qui doivent remplir deux conditions simples : présenter un intérêt pour la communauté AFIA et ne pas être des articles de colloques ou de revues (n'hésitez pas à envoyer une déclaration d'intention par mël à J-P Sansonnet).

Jean-Paul Sansonnet

Rédacteur en chef du bulletin de l'AFIA

1^{er} Appel à communication IC'2002

Treizièmes journées francophones d'Ingénierie des Connaissances

Rouen, France, 28-30 mai 2002

<http://ic2002.insa-rouen.fr>

Historique

Placée sous l'égide du GRACQ (Groupe de Recherche en Acquisition des Connaissances, <<http://www.irit.fr/GRACQ/>>), la conférence IC est un lieu d'échange et de réflexion de la communauté française sur les problématiques de l'Ingénierie des connaissances. Elle fait suite aux conférences IC 2001 (plate forme AFIA, Grenoble), IC 2000 (IRIT, Toulouse), IC'99 (École Polytechnique, Palaiseau), IC'98 (Nancy), IC'97 (Roscoff) et aux Journées d'Acquisition des Connaissances (JAC) qui se sont tenues tous les ans de 1990 à 1996 (Lannion, Sète, Dourdan, Saint-Raphaël, Strasbourg, Grenoble, Sète). Elle se déroulera les 28-29-30 mai 2002 à Rouen.

L'ingénierie des connaissances

De manière générale, l'ingénierie des connaissances porte sur l'instrumentation technique des contenus pour les exploiter dans un cadre où ils sont mobilisés pour leur signification. Par conséquent, l'ingénierie des connaissances porte sur la représentation informatique du sens pour que les systèmes construits puissent avoir leur comportement paramétré par cette représentation. Le sens peut être le sens d'une activité (comment le système s'adapte à ce qui fait sens pour l'utilisateur à un moment donné), d'un document (comment le système traite le document en fonction de ce qu'il signifie), etc.

Plus précisément, l'Ingénierie des connaissances propose des concepts, méthodes et techniques permettant de d'acquérir, de modéliser et de formaliser des connaissances pour les mobiliser dans l'activité d'une organisation. Structurées, formalisées et opérationnalisées pour être intégrées dans le fonctionnement d'un système à base de connaissances (SBC), ces

connaissances deviennent des informations destinées à être, *in fine*, interprétées par un humain dans son interaction avec l'artefact construit. La réussite d'une démarche d'Ingénierie des connaissances peut alors se mesurer par rapport à 3 objectifs :

- Construire un SBC ayant un comportement appréhensible et compréhensible par l'utilisateur. Cela passe par une modélisation à un niveau d'abstraction pertinent qui fait sens pour les différents acteurs impliqués (ingénieur cognitiens, experts métiers, utilisateurs, etc.) et leur permet de s'approprier le comportement du système et d'interagir avec lui. C'est le « niveau des connaissances ».
- Appréhender le fonctionnement et l'utilisation d'un SBC par rapport à son intégration dans une organisation. L'Ingénierie des connaissances doit alors permettre de modéliser et d'explicitier le fonctionnement cognitif de l'utilisateur ou du collectif dans sa relation au SBC.
- Développer et mettre en place une réelle démarche d'ingénierie, reproductible et proposant des instrumentations techniques – i.e. des méthodes et des outils – fondées sur des approches rigoureuses de modélisation des connaissances et de formalisation logique, mathématique et informatique.

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

- d'une part, dans sa démarche d'ingénierie, l'IC mobilise les concepts et techniques de la représentation des connaissances, les méthodes d'analyse et de conception à objets, le raisonnement à base de cas, l'ingénierie documentaire ou l'ingénierie éducative, la conception de systèmes d'information, etc. ;
- d'autre part, dans sa démarche de modélisation des connaissances, l'IC doit se rapprocher de disciplines permettant de caractériser et décrire les connaissances d'un domaine et d'évaluer leur mise en œuvre dans les SBC. La sociologie, la gestion ou l'ergonomie peuvent ainsi concourir à une démarche

d'explicitation de ce que sont les connaissances dans un contexte humain et organisationnel.

Calendrier

- 1^{er} Février : date limite des soumissions
- 22 Mars : résultat de la sélection des papiers
- 26 Avril : réception des papiers définitifs
- 28-30 mai : IC'2002

Soumissions attendues

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

- **Communications de recherche** (format long – 12 p. max., interligne simple, times 12) Sont concernés des travaux de recherche originaux, théoriques ou expérimentaux. Une rigueur devra être apportée à la rédaction des articles, tant sur le plan de la forme que sur celui du contenu scientifique.
- **Communications appliquées** (format court – 6 p. max.) Sont concernés des applications industrielles et opérationnelles significatives, des études de cas, des prototypes avancés. Le contexte de l'application devra être clairement décrit et les auteurs mettront l'accent sur les problèmes rencontrés, les enseignements tirés et les retours d'expériences.

Les auteurs sont invités à indiquer, sur la première page de l'article, le ou les thèmes dans lesquels s'inscrivent leurs propositions (La liste de ces thèmes est en cours de constitution). Les articles ne relevant pas de ces thèmes s'attacheront à mettre clairement en évidence leur lien avec l'ingénierie des connais-

sances. Le cas échéant, les auteurs sont invités à prendre contact avec les animateurs des différents thèmes ou le Président du comité de programme.

Chaque proposition est relue par 3 membres du comité de programme. Les articles inscrits dans un thème ne sont pas en priorité relus par les animateurs du thème : une ventilation est effectuée pour chaque article entre 2 spécialistes et un non spécialiste.

Toute communication écrite (de recherche ou appliquée) donnera lieu à un exposé oral de 25 mn pour les communications de recherche et de 15 mn pour les communications appliquées.

Formats des soumissions

Les auteurs sont invités à envoyer leur soumission de préférence par (i) enregistrement d'un résumé et des coordonnées des auteurs de l'article proposé puis (ii) dépôt FTP électronique (d'un fichier PDF, Postscript [compressé zip ou gzip] ou [finalement] Word) à : <http://ic2002.insa-rouen.fr>

Pour tout problème ou renseignement contacter le président du comité de programme:

Bruno Bachimont

Directeur Scientifique
Institut National de l'Audiovisuel
4, Ave de l'europe,
94366 Bry sur Marne cedex
Courriel : bbachimont@ina.fr
tél. : 01 49 83 29 08
fax : 01 49 83 25 82

Nous avons lancé un débat intitulé « *Évaluation Scientifique : peut-on mieux faire ?* ». L'appel à débat est reproduit dans l'encadré ci-dessous. Suite à la contribution de Jacques Pitrat dans le n° précédent que nous reproduisons ci-dessous pour mémoire, nous publions une évaluation de cette proposition (en anglais cela va de soit). Bien entendu, nous nous devons de tenir secret le nom du « referee ». Il (ou elle) est repéré(e) sous la clé #FP/09.01. Au préalable, pour mémoire l'encadré ci-dessous présente le premier appel à débat :

La loi LOI n° 82-610 du 15 juillet 1982 *d'orientation et de programmation pour la recherche et le développement technologique de la France*¹ est claire **quant à la nécessité d'évaluer la recherche**. On y lit dans l'article 6 : « *L'appréciation de la qualité de la recherche repose sur des procédures d'appréciation périodique portant à la fois sur les personnels, les équipes, les programmes et les résultats. Ces procédures respecteront le principe de l'examen contradictoire et ouvriront la possibilité de recours devant l'autorité hiérarchique* ». Sur le plan individuel, **l'évaluation des recherches pose cependant de nombreux problèmes**.

Voici à titre d'exemple les trois premiers critères utilisés par une université australienne² pour évaluer l'excellence des chercheurs :

- 1 **Publication** in refereed journals or series, or by publishers recognised as leaders in the field. [...].
- 2 **High ratings of the publications** and other research or creative work - as attested (where appropriate) by **citations indices**, by references in published literature which cite the applicant's work [...].
- 3 **Number of publications** (publications with n co-authors are rated as one nth of a publication unless evidence is provided to support a higher allocation).

Il est clair et nous le constatons dans les différentes institutions que c'est le **poids** des publications qui est le critère incontournable de l'évaluation scientifique. Or, aussi bien le nombre que les « *indices de citations* » sont de mauvais critères d'évaluations³. De plus, ce mode d'évaluation a un impact direct sur la prolifération des publications. Les chercheurs sont poussés par le système à produire toujours plus. **Quelles sont les directions possibles pour une évaluation scientifique qui donne plus de temps aux chercheurs pour chercher et moins pour publier ?**

Améliorer ou supprimer les referees⁴ ?

Jacques Pitrat

LIP6 Université Pierre et Marie Curie

Un chercheur en IA, qui est intéressé par l'automatisation de toutes les tâches intellectuelles, se doit de réfléchir à la définition d'une stratégie de sélection des papiers soumis à un congrès ou à une revue. Cela préparera le jour où les papiers d'IA seront sélectionnés par des systèmes d'IA !

Il existe deux méthodes extrêmes pour prendre une décision. Avec la méthode conjonctive on rejette un papier s'il est mauvais sur un point important ; on ne prend pas de risques, cette méthode est bien adaptée si l'on veut sécuriser une centrale nucléaire. De façon dia-

métralement opposée, avec la méthode disjonctive on accepte un papier s'il est bon sur un point important. Il faut préférer la méthode disjonctive quand la créativité est essentielle, ce qui est le cas pour tout ce qui est lié à la recherche. Bien entendu, tous les intermédiaires sont possibles entre les deux, en particulier la méthode compensatoire qui est basée sur une fonction d'évaluation où l'on pondère tous les critères.

Malgré l'intérêt de la méthode disjonctive pour ce qui est lié à la recherche, la quasi-totalité des revues et congrès d'IA penchent fortement vers la méthode

1. <http://garp.univ-bpclermont.fr/guide/Textes/Txtfond/L82-610.html>

2. <http://www.acs.uwa.edu.au/hrs/policy/part09/attachments/7-1.htm>

3. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314:498-502. (<http://194.216.217.166/reg/bmj/archive/7137/7137e6.htm>)

4. Cette contribution est déjà parue dans un précédent n° du bulletin AFIA, mais ces choses arrivent (NDLR).

conjonctive. Cela nous vaut une floraison de papiers qui n'ont aucun défaut grave, mais qui manquent totalement d'intérêt parce qu'ils n'apportent pas grand chose de nouveau. En utilisant la méthode conjonctive, non seulement on fait couper des forêts pour rien, mais encore on barre des papiers qui contiennent des idées très originales, mais qui sont souvent mal rédigés, qui ont de sérieuses faiblesses dans leur bibliographie ou qui laissent de nombreuses questions légitimes en suspens.

Dans Talking Nets, les fondateurs du connexionnisme racontent leur vie de chercheur. Il est frappant de constater la difficulté que certains ont eu pour publier, souvent avec des années de retard, les papiers considérés actuellement comme les plus importants de la discipline. Pour ma part, parmi tous les articles que j'ai lus dans ces dix dernières années, celui qui m'a le plus influencé a été refusé par toutes les revues auxquelles il avait été soumis. Il avait beaucoup de défauts de forme et il n'était certes pas facile à comprendre, mais le but des membres d'un comité de lecture n'est pas uniquement de sélectionner des papiers ; nous devons aider l'auteur d'un tel article à en gommer les défauts les plus criants. Ceux qui ont des idées ne sont pas forcément ceux qui savent le mieux rédiger. Malheureusement il se trouve que dans les colloques, qui devraient être le terrain de prédilection pour les papiers créatifs, le processus d'amélioration par dialogue entre l'auteur et le lecteur est difficile à mettre en œuvre pour des raisons de délais.

Une lettre très intéressante sur le mécanisme de sélection a été publiée dans le numéro de septembre 1994 de AI Communications. Robert Milne commente ce qui s'est passé dans un comité de programme. Il pense, avec raison, qu'ils ont fait leur travail très sérieusement et qu'ils se sont donné beaucoup de peine. Il a bien fait d'écrire cette lettre, car il donne des conseils très utiles si nous voulons que nos papiers soient acceptés dans le contexte sélectif actuel. Mais certains passages m'ont laissé rêveur :

« Many rejected papers would have been saved by adding 2 or 3 paragraphs to put the work into context. Some authors had very good ideas but the paper did not present these ideas clearly enough. This is particularly common when many ideas are being presented. There were a large number of papers rejected that had good innovative ideas but the presentation and description was too confusing and unclear for the referee to understand. »

« At the opposite extreme, many papers were accepted with a small contribution to the field, but clear and well presented so that they were understandable by the referees. »

Plus un papier est original et plus il risque de ne pas passer les critères de sélection conjonctive. C'est le cas s'il présente un gros système, difficile à décrire, ou si les idées sont très nouvelles : l'auteur n'a pas encore acquis l'expertise pour les présenter clairement et le referee se demande si ce n'est pas quelque chose de complètement farfelu. Accepter un papier douteux nuirait à son évaluation en tant que referee compétent alors que personne n'identifie les referees qui ont écarté un papier qui s'est révélé fondamental bien des années plus tard.

La situation a changé avec l'arrivée des revues électroniques. Un des avantages d'une revue publiée électroniquement est justement de permettre la publication d'articles qui prêtent plus à controverse, mais qui relancent la recherche dans de nouvelles directions. Imprimer un papier est une tâche noble et coûteuse ; cela incite à refuser tout défaut et mène automatiquement à la méthode conjonctive. Profitons de notre nouvelle liberté ! Cela ne veut pas dire de publier n'importe quoi, mais de publier un papier si et seulement si il apporte quelque chose de vraiment nouveau, même s'il a de sérieux défauts par ailleurs. Il n'y aura pas plus de papiers publiés, mais les papiers seront choisis sur des critères différents.

Mais nous pouvons aller plus loin et nous demander si nous ne pourrions pas arriver à un autre mode de publication que les revues. Chaque chercheur publierait sur son site les papiers qui présentent ses travaux, sans mécanisme de sélection, et tout le monde pourrait les consulter. Naturellement il faut un autre mécanisme pour nous aider à trouver les papiers susceptibles de nous intéresser. Cela pourrait être réalisé grâce à des spécialistes du domaine qui feraient la promotion sur leur site des papiers qui leur ont particulièrement plu. Il y aura certainement des vocations spontanées pour cela : personnellement, je présente régulièrement aux chercheurs de mon équipe les articles qui m'ont intéressé. Les chercheurs sauront vite repérer les lecteurs qui font des critiques qui leur conviennent ; je fais cela sans difficulté pour les critiques de films dans la presse écrite. Nous retrouvons alors la méthode disjonctive, parce que nous n'avons envie de parler que de ce qui

présente au moins une caractéristique remarquable. En effet, ces lecteurs ne seraient pas obligés, comme les referees actuels, de donner un avis sur tout ce qui leur est soumis ; ils choisiraient naturellement ce dont ils aiment parler. Comme on s'en est rendu compte avec le loft, il vaut mieux voter pour ce que l'on garde que pour ce que l'on élimine. Ainsi que cela se passe sur Amazon, nous pourrions avoir aussi le lien inverse reliant le papier à toutes ses critiques, ce qui éclairerait un lecteur potentiel : ce dernier pourrait consulter plusieurs avis et il pourrait même ensuite ajouter sa propre critique après l'avoir lu. A partir du moment où l'on mettrait un article sur son site, on accepterait que les lecteurs y ajoutent des critiques, mais on aurait naturellement le droit d'y mettre aussi des réponses aux critiques. Il reste toutefois que, si l'on peut éliminer le processus de sélection actuel, il faut toujours que l'au-

teur fasse auparavant lire son papier par des relecteurs qu'il choisirait pour améliorer le texte initial. Il est nécessaire d'avoir d'autres points de vue que le sien, la relecture par d'autres est une étape incontournable. Nous pourrions résumer cette structure par : des relecteurs : oui, des conseillers : oui, des sélecteurs : non.

Ce processus aura pour certains un grave défaut : comment alors évaluer les chercheurs ? Actuellement, on compte les publications faites dans les revues et les congrès en pondérant avec leur renommée. Nous devons trouver d'autres moyens, nous savons tous très bien faire la différence entre les chercheurs dont les travaux nous intéressent et ceux qui publient pour publier. **Il faudrait savoir si notre but fondamental est l'avancement des chercheurs ou l'avancement de la recherche.**

Evaluation de cette contribution

Referee : #FP/09.01

This paper contains some suggestions that do address the issue of paper selection. However, the author fails to convince the reader on several important points.

1 - The author seems to ignore the huge literature on the subject of paper selection, such as the unsuccessful but sympathetic annihilation approach (Zucker & Sakharov, 1999), the cryptic but ambitious creative-based paper generation approach (Gnagnascia & rame-lahaut, 1998c), the multi-criteria paper selection paradigm (Perny, 1990.6), and the framework-based pattern-oriented approach advocated by (Blin, Revo, Biriard & Guess-whom 2002), embodied e.g. in the PaperTalk system (Wolinsse & Perault, 1989). The author can check the recent synthesis by (Brezillions & Jakubowski, 2001) who propose a validation scheme for comparing all the approaches in a systematic way, using a 19-dimension, 100 by 100 criteria matrix. The author should, at least, position his work using Brezillions' assessment measure. Failing to do so shows the lack of knowledge the author has of his own research domain. This attitude can be felt as a lack of respect for the aforementioned hard-working researchers. Especially Mr. Zucker, who does not like to be treated like that.

2 - The English is poor, and the paper has probably not been double checked by a native English speaker. The paper even contains many non-English words (such as the word « papier », which probably means « per peer ») which makes it difficult to read and thus to understand. And consequently to assess. And finally to select. It seems to me that the paper was hastily translated from one of these exotic European language (Italian? Serbo-Croatian? Vietnamese?).

3 - The author does not take into account the well-known Darwinist pressure, as explained e.g. by Damascios ("Darwin, again and again", Boringbooks, 2002), which states that selecting poor papers is the only way program committee members can survive in a world where 15 year old children make more important contributions to the field than fully-fledged researchers. The only way.

4 - The paper greatly suffers from a total lack of validation. The hypothesis and claims should be substantiated by experimental data, otherwise the work looks more like doodling than actual research. The solutions should have been implemented and tested, and the results should show clearly the advantage of the approach proposed, in comparison with previous works, such as the ones listed in 1). Furthermore how

can we define precisely the notion of «quality»? What is exactly an «interesting» paper? To consider a priori that papers should be «interesting» or «creative» is more a product of an outdated and somewhat naïve conception of research that a scientific attitude towards the object of study: papers are, by definition, like objects fall, by gravitation. Ascribing such human qualities as interestingness to papers is definitely not a good start for a scientific endeavour.

We suggest that the author instead consider true, reproducible experiments. For instance, he could study how, say, a distribution of random papers reviewed by random agents can be evaluated using random ranking schemes? (THIS would be interesting, right in the spirit of the great achievements performed in the field of random constraint graphs (Dognet and his colleagues, 2001)).

Additionally, the data collected during the experiments should conform to the IEEE 75-15 format for graphs, if possible using Excel, which has not been invented pour les chiens, nom de Dieu.

5 - The level 2 headings are in italics. This, once again, reveals the deep disrespect of the author for the community of paper selection research, and in particu-

lar for our dearest Rev. Claudios Cadozos, the inventor of the rule catalogue for paper selection research paper submission format, and who, in 1972, established the definite demonstration of the long-term drawbacks of repeated exposure to italics 2nd headings.

6 - The formula in the second paragraph is unclear. Please explain. Avoid the use of obscure mathematical signs such as =, + and *. Use simple direct notation instead.

Globally the paper should be rejected on the basis on the lack of soundness, completeness and absence of validation. Alternatively a poster could be proposed, if the author adheres to the comments, submits an updated version of his work, and has his revised version proof-checked by the reviewers.

Private comment to the committee (not seen by author):

I strongly oppose the acceptance of this kind of paper, which would seriously degrade the high-quality standard of the RRASCNPC (Rencontres Régionales des Amateurs de Sciences Cognitives du Nord Pas de Calais).

Si vous avez envie d'intervenir sur ce débat, de donner votre opinion ou de nous faire partager votre point de vue, n'hésitez pas à envoyer une contribution (remarques, articles, commentaires) à Jean-Daniel.Zucker@lip6.fr



N'OUBLIEZ PAS LES PETITES ANNONCES DE L'AFIA !

L'AFIA et son bulletin offrent un nouveau service de petites annonces !

Vous pourrez y soumettre ou consulter des annonces concernant l'intelligence artificielle, des offres ou demandes de stages, d'emploi, de logiciels...

Pour consulter les annonces actuelles, visitez le site de l'AFIA à l'adresse :

<http://www.afia.polytechnique.fr/bulletin/PetitesAnnonces/PtAnnonces.html>

Pour soumettre une annonce, envoyez un mél à Brigitte Grau (Brigitte.Grau@limsi.fr) qui gère cette rubrique (le champ « objet » (ou « sujet » ou « subject ») de ce mél devra mentionner explicitement « Petite annonce AFIA »)

Intelligence Artificielle : l'introuvable STIC.

Jean-Gabriel Ganascia

Université Pierre et Marie Curie - LIP6

Jean-Gabriel.Ganascia@lip6.fr

Lu, dans le N° 396 de CNRS info (septembre 2001), la lettre d'information du CNRS destinée à informer les médias, à la rubrique « Objectifs et stratégie, programme COGNITION ET TRAITEMENT DE L'INFORMATION » : *Le terme « artefactuel » est utilisé par la communauté STIC car il fait référence à un « artefact », littéralement un « fait d'art », autrement dit une création. Le terme « artificiel » tend à être abandonné bien que plus accessible au grand public car il évoque l'idée d'une fausse intelligence.*

Pour la grande presse, le constat est donc établi : L'intelligence artificielle est morte et l'*intelligence artificielle* reprendra le flambeau de sa puissance ! Dans des situations dramatiques de cet ordre, des questions urgentes, des questions pratiques sautent à l'esprit : que vont-ils faire du cadavre ? En se décomposant, il va sentir fort ; ça risque de nous incommoder ! Et, comment annoncer le décès à la famille ? Aux cousins de l'étranger membres des associations sœurs : ECAI, AAAI ? Devront-ils, eux aussi, changer de nom et se laisser vaincre par le grand « Artefactuel » ?

Personnellement, en tant que chercheur, j'ai toujours été plus poussé à la curiosité qu'à régler l'intendance. Aussi, je me suis interrogé sur les raisons, les causes et les mobiles du crime. Qui est responsable et pourquoi l'intelligence artificielle a-t-elle été supprimée ? Quelle est cette mystérieuse communauté terroriste des STIC qui l'a assassinée ? Quelles en sont les ramifications et les complicités ? J'ai mené l'enquête pour vous et je vous livre ici, en exclusivité, les maigres résultats de mes premières investigations.

Il m'a d'abord semblé bon d'interroger des collègues, des membres de l'AFIA, des chercheurs qui se reconnaissent dans l'intelligence artificielle, et d'autres informaticiens qui nous fréquentent depuis longtemps. Nenni ! M'ont-ils répondu. Nous n'avons jamais entendu parler de cette « *intelligence artificielle* » toute nouvelle. En dépit des témoignages de sympathie envers le défunt, je n'ai rien recueilli de concret auprès d'eux. Je suis donc allé au-delà de notre communauté, voir d'autres chercheurs.

J'ai commencé par les biologistes, car le texte en question portait le paraphe du département Sciences de la Vie du CNRS. O les coquins, ils nous auraient bien joué un mauvais tour à leur façon, me suis-je dit ! Or, à les interroger, ils n'y sont pour rien. Ils ont même comploté sincèrement en m'affirmant qu'ils avaient beau se gausser parfois et se moquer, nous étions tous des scientifiques et il qu'il y avait des limites au-delà desquelles la plaisanterie devenait de mauvais goût. *Artefactuel* ! L'insulte était trop forte. Ils en frémissaient d'horreur. Pour des expérimentateurs comme eux, le mot porte des couleurs bien sombres : est qualifié d'artefact tout résultat reflétant non l'objet d'étude, mais les défauts du montage et de l'appareillage de mesure. Bref, l'artefact, c'est l'erreur, l'instrument qui nous fourvoie, l'illusion que nous pourchassons tous sans relâche.

Si les expérimentateurs ne sont pas en cause, allons voir les littéraires, les linguistes, les sociologues, les philosophes, les anthropologues ou tout autre savant en rhétorique relevant des Sciences de l'Homme et de la Société. Eux aussi, ils ne ratent jamais une occasion de nous railler. Mais, là encore, après avoir posé quelques questions, il s'est vite avéré que je faisais fausse route. On m'a rapidement convaincu qu'il y avait des bornes au-delà desquelles les jeux de mots deviennent ridicules, voire indécents. Artefact a une connotation néfaste. Et, ce mot relève tout juste de notre langue, puisqu'il n'est pas encore entré au dictionnaire de l'Académie française. Pour les linguistes, la signification tient à l'usage, non à l'étymologie. Artefact vient de l'anglais. La racine latine explique l'acception courante qu'il a prise, mais pas ses connotations. Celles-ci tiennent à l'emploi qu'on en fait dans notre langue et non à une dérivation savante totalement reconstruite. Et, si l'on veut forger un nouveau néologisme, comme « artefactuel », et qu'il se conforme au génie de la langue, il faudrait que le sens qu'il prend dérive de celui de sa racine, ici « artefact ». Selon eux, à se référer uniquement à l'étymologie, on risque de laisser croire que les bou-

cheres ne vendent que de la viande de bouc, les charcuteries, de la chaire cuite, et que les ordinateurs sont des dignitaires ecclésiastiques préposés à l'organisation des cérémonies religieuses.

Revenu bredouille de toutes mes recherches, je crois qu'il conviendrait de dépêcher en ambassade des représentants de notre communauté dans les instances du CNRS, des gens qui enquêteraient sur le meurtre, des gens qui demanderaient à ce qu'à l'avenir, nous soyons associés aux décisions nous concernant afin d'éviter que de tels crimes ne se reproduisent, enfin des gens qui se fassent l'écho de nos voix et qui réclament, qu'à l'égal des autres scientifiques, nous ayons un département à nous, un département d'informaticiens regroupant toutes les composantes de cette discipline, un département qui comprenne nos problèmes, nos difficultés et nos projets. A défaut, au train où vont les choses, nous risquons d'apprendre bientôt, par la presse, que le CNRS a décidé de supprimer tous les ordinateurs en se fiant à leur étymologie, pour les remplacer par des *prêtres artefactuels*...

Jean-Gabriel GANASCIA

Rappelons que le mot « Artefactuel » n'existe pas en français ; on ne le trouve attesté nulle part, sauf dans le N°396 de CNRS info. Quant au mot « artefact », il n'est pas dans le dictionnaire de l'Académie. Il se trouve toutefois des définitions dans les dictionnaires d'usage. J'en cite ici deux :

Définition du Nouveau Petit Robert (*édition mise à jour en mars 1995*)

ARTEFACT [artefakt] n.m. — 1905 ; mot angl., aussi artifact (1821); du lat. artis factum « fait de l'art »
◇ DIDACT. ANGLIC. Phénomène d'origine humaine, artificielle (dans l'étude de faits naturels). ◇ SPECIALT. MED. Toute altération produite artificiellement lors d'un examen de laboratoire (par ex. examen microscopique d'un tissu, électro-encéphalographie, électrocardiographie, etc.). — On écrirait mieux artéfact.

Définition du Lexis - Larousse (*édition 1994*)

ARTEFACT [artefakt] n.m. — (mot angl. du lat. artis facta effets de l'art; v. 1950). Structure ou phénomène, d'origine artificielle ou accidentelle, rencontré au cours d'une observation ou d'une expérience portant sur un phénomène naturel.

Si vous avez envie d'intervenir sur ce débat, de donner votre opinion ou de nous faire partager votre point de vue, n'hésitez pas à envoyer une contribution (remarques, articles, commentaires) à Jean-Daniel.Zucker@lip6.fr



Nouvelles du Québec

Nicole Tourigny

Dans le numéro précédent (Juin 2001, n° 45), les activités de trois laboratoires d'I.A. de l'Université Laval (<http://www.ulaval.ca>) ont été décrites. Une brève description de l'Université Laval et du département d'informatique (<http://www.ift.ulaval.ca>) avait alors été donnée. Aujourd'hui, je vous propose de faire la connaissance d'un autre laboratoire d'I.A. de l'Université Laval ainsi que d'un laboratoire d'I.A. de l'Université de Sherbrooke.

L'Université de Sherbrooke (<http://www.usherb.ca/>) a été fondée en 1954. Elle compte plus de 20 000 étudiants. Au niveau de l'enseignement, elle se caractérise par un régime coopératif avec l'industrie. Au niveau de la recherche, elle est l'université canadienne qui tire le plus de revenus de la valorisation de la recherche. Le programme d'informatique (<http://www.dmi.usherb.ca/>) est donné officiellement depuis 1980, il couvre actuellement les trois cycles universitaires et plus de 750 étudiants y sont inscrits. Le développement des systèmes logiciels intelligents est l'un des deux axes majeurs de recherche en informatique.

Laboratoire Dialogues, Agent, MultiAgentS (DAMAS)

Responsable : Professeur Brahim Chaib-draa

Contact : chaib@ift.ulaval.ca

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

Adresse : Département d'informatique, Faculté de Science-Génie, Université Laval, Ste-Foy, PQ, Canada G1K 7P4

Ce laboratoire rassemble des étudiants à la maîtrise et au doctorat en informatique sous la supervision de M. Brahim Chaib-draa (chaib@ift.ulaval.ca) ainsi que des chercheurs invités. Les recherches portent sur :

- la théorie de l'action et de l'interaction dont le dialogue entre agents logiciels ;
- la modélisation du raisonnement particulièrement pour la coopération et la compétition entre agents logiciels ;
- la représentation et l'exploitation des connaissances dans des environnements distribués ;
- l'application de ces théories à des problèmes réels importants avec l'implémentation d'agents logiciels interagissant entre eux et avec leur environnement (ou système multiagent).

Parmi les projets récemment réalisés, mentionnons :

- NetSA : une architecture de systèmes multiagents pour la recherche d'information dans des sources hétérogènes et distribuées comme les entrepôts de données et l'Internet. Ce projet a été réalisé en collaboration avec NCE (UK) et le centre allemand de recherche en IA—le DFki (<http://www.dfki.de/>).
- AUCTION : une simulation à des fins de validation d'une enchère multiobjet pour la négociation automatique et le commerce électronique. Ce projet se poursuit actuellement au Centre Interuniversitaire de

Recherche en Analyse des organisations CIRANO (<http://www.cirano.qc.ca/>).

Parmi les projets en cours, mentionnons :

- CHOICEBOARD : un système multiagent d'aide à la décision assistant une personne dans le processus d'achat groupé. Un étudiant en master est actuellement impliqué dans ce projet.
- FRÉGATES CANADIENNES : un système multiagent d'aide à la décision pour la gestion des ressources à bord d'une frégate. Un étudiant en master et deux autres étudiants au doctorat sont actuellement impliqués dans ce projet.
- CARTES COGNITIVES : une représentation des connaissances d'un système multiagent visant à capturer les relations causales existant entre des concepts. Une étudiante en master travaille actuellement sur ce projet.
- CHAÎNES LOGISTIQUES : une gestion des flux d'information et de matériaux d'une chaîne logistique à l'aide d'un système multiagent. Un étudiant au doctorat travaille présentement sur ce projet.
- DIALOGUE : une étude de la co-construction de dialogues dans la communication entre agents logiciels. Un étudiant au doctorat travaille en étroite collaboration avec un post-doctorat sur cet aspect.

Les recherches du DAMAS sont subventionnées par le Conseil de recherche en sciences naturelles et

en génie du Canada (CRSNG), le Conseil de recherche en sciences humaines du Canada (CRSH), le fonds FCAR du Québec, Le centre de recherche sur les organisations réseaux (le CENTOR), le Centre de recherche de la défense Valcartier (le CRDV), le

Réseau de centres d'excellence voiture du 21^{ème} siècle.

Brahim Chaib-draa donne un cours au 1^{er} cycle d'IA avancée et un cours pour le 2^{ème} et 3^{ème} cycle sur les « agents et les systèmes multiagents » (<http://www.damas.ift.ulaval.ca/~chaib>).

Laboratoire des Systèmes Intelligents et du Multimédia (LASIM)

Responsables : Professeur André Mayers

Professeur Roger Nkambou

Contacts : andre.mayers@dmi.usherb.ca
<http://www.DMI.USherb.CA>

roger.nkambou@dmi.usherb.ca

Le but de ce laboratoire consiste à étudier et développer des systèmes informatisés dans lesquels l'interface homme-machine doit tenir compte de l'utilisateur. Plus précisément, ces systèmes doivent avoir une représentation fine des connaissances que l'utilisateur possède du système ainsi que des caractéristiques cognitives, conatives et affectives propres à l'utilisateur. Ces systèmes peuvent être des centrales nucléaires, des avions de ligne, des instruments médicaux; cependant pour des raisons de disponibilité des ressources, les activités du laboratoire sont actuellement concentrées sur l'étude de systèmes tutoriels intelligents.

abstraits comme la construction de preuves en mathématique.

Par rapport à ces problématiques, les théories élaborées et les techniques développées utilisent et enrichissent plusieurs domaines de l'intelligence artificielle. Parmi ces dernières, mentionnons les différentes variétés de la logique modale, les différentes formes de réseaux sémantiques, les sociétés d'agents et les systèmes dynamiques (par opposition aux systèmes de symboles).

Les problématiques associées à l'étude de ces systèmes sont :

- La représentation des connaissances et les mécanismes générant des interprétations accessibles à l'être humain et à la machine.
- La modélisation des processus cognitifs (mémoire, résolution de problème, prise de décision...)
- La modélisation des processus régissant l'attention, la motivation et les émotions et de leurs impacts sur les processus cognitifs.
- L'intégration des processus mentionnés précédemment dans une seule architecture pouvant simuler les usagers du système.
- La construction d'environnements et de laboratoires virtuels.
- La réification à l'aide d'outils multimédias de phénomènes physiques comme les réactions chimiques ou

Présentement, l'ensemble des recherches du laboratoire concourent à développer un système tutoriel générique pour les disciplines scientifiques et techniques. Une subvention de l'université de Sherbrooke a été reçue pour développer des applications spécifiques pour la formation à distance de différents cours de la faculté des sciences de l'université de Sherbrooke, notamment en mathématiques, en informatique et en biologie. Dans le cadre de cette subvention, les recherches portent aussi sur le multimédia et les systèmes distribués.

André Mayers donne des cours sur la programmation fonctionnelle, sur l'intelligence artificielle sur les interfaces et le multimédia.

Roger Nkambou donne des cours sur les bases de données, sur les systèmes tutoriels intelligents au niveau M. Sc. et Ph. D, sur les interfaces et le multimédia.

Dans le prochain numéro, d'autres laboratoires seront présentés. D'ici là, il me fera plaisir de répondre à vos questions. Vos commentaires seront appréciés.

Nicole Tourigny, Ph.D., professeur
Département d'informatique
Pavillon Adrien-Pouliot

Université Laval
Québec, CANADA
G1K 7P4

tourigny@ift.ulaval.ca
<http://www.ift.ulaval.ca/~ericac/>

Dialogueurs artificiels : précurseurs ou illusoires ?

Jean-Baptiste BERTHELIN
LIMSI-CNRS -- jbb@limsi.fr

1. Introduction

Quelques années de fréquentation de la toile mondiale sont l'occasion de rencontrer bien des étrangetés, parmi lesquelles je voudrais évoquer brièvement les « chatterbots », ou simulateurs de conversation en ligne.

Pour cela, nous commencerons par dire deux mots sur ELIZA*, leur ancêtre à tous. (Note pour la suite : les noms marqués d'un astérisque sont ceux pour lesquels nous renvoyons à une page de toile, en fin d'article). Puis nous donnerons des exemples d'interaction avec des systèmes plus récents comme Megahal* et Jabberwacky* ; pour finir, nous préciserons la position quant à la pertinence d'une étude de ces entités.

Depuis sa première diffusion en 1966, Eliza se décline en d'innombrables versions. Ici nous nous intéresserons à celle de Tom Bender*, assez proche de l'original.

Rappelons qu' Eliza imite le comportement verbal d'un psychologue non-directif, face aux déclarations d'un quidam lui exposant son, ou ses problème(s). Sur cette base, certains ont pu objecter qu' Eliza ne contribue jamais à résoudre aucun des problèmes en question, ce qui en amoindrit l'intérêt. Joseph Weizenbaum, créateur d'Eliza, avait certes prévu l'objection, et choisi pour cette raison le côté « non-directif » des réponses.

2. Eliza et Jules César

Plutôt qu'une analyse de ce concept de « non-directivité », nous avons constitué une collection* de conversations mettant en scène des « clients tragiques », autrement dit, des personnages (Jeanne d'Arc, Mishima et autres) qui exposent chacun un problème trop grave pour que quiconque (sauf justement Eliza) le prenne à la légère. Soit par exemple le cas de Jules César aux Ides de Mars :

COMMENT ALLEZ-VOUS ? MON NOM EST ELIZA. SOUS QUEL NOM ÊTES-VOUS CONNU ?

Jules César.

MINCE, VOUS ME SEMBLEZ NERVEUX. QUEL EST VOTRE PROBLÈME, JULES CÉSAR ?

On me transperce à coups de poignards.

ON VOUS TRANSPERCE TOUJOURS À COUPS DE POIGNARDS ? C'est la première fois que ça m'arrive.

JE NE COMPRENDS PAS VOTRE ATTITUDE.

Mon attitude ? J'essaie de les combattre.

BON, CELA PEUT FAIRE DU BIEN DE SE DÉTENDRE UN PEU.

Je serais plus détendu si mon cher fils n'était pas de leur nombre.

OUI, OUI.

Je te dis qu'il en est !

NE NOUS ATTARDONS PAS TROP SUR SON CAS.

Ce n'est pas moi qui vais le faire.

JE PENSE QU'EN FAISANT UN EFFORT, VOUS POURRIEZ MONTRER PLUS D'OUVERTURE.

Des ouvertures, mon corps en comporte actuellement vingt-trois de plus que tout à l'heure.

PARDON, MAIS JE NE COMPRENDS PAS CE QUE VOUS ESSAYEZ DE ME DIRE.

Je tente de m'exclamer « Tu quoque, fili ! »

VOUS CROYEZ QUE CELA IMPORTE À BEAUCOUP DE GENS, LE FAIT QUE VOUS TENTIEZ DE VOUS EXCLAMER « TU QUOQUE » ?

Ceux du moins qui m'entendent aujourd'hui ; je ne peux rien dire quant à la postérité.

LA QUESTION VOUS SEMBLE INTÉRESSANTE ?

En tant que patricien, j'ai le devoir de m'y intéresser.

CELA SUFFIRA POUR AUJOURD'HUI. N'OUBLIEZ PAS LE PROCHAIN RENDEZ-VOUS.

Ce qui frappe dans ce dialogue tragico-comique, c'est le contraste entre les appels au secours du personnage et le détachement de l'interlocuteur artificiel. Cet effet est facile à obtenir, puisque les principes d'Eliza sont bien connus à présent. Mais ces principes ne sont plus à l'oeuvre sous la même forme dans les autres dialogueurs de la Toile.

3. Megahal et Jabberwacky

Prenons par exemple Megahal, de Jason Hutchens. Au lieu d'avoir des réponses purement écholaliques comme Eliza, il combine, de façon markovienne, les interventions de ses différents interlocuteurs, produisant des phrases hybrides. L'impression qu'il donne n'est pas de détachement, mais plutôt de confusion, qui ne manque parfois pas d'un certain charme.

MegaHAL : Le petit ornithorynque ne peut pas manger plus que son poids de jambon fumé le matin.

l'utilisateur : Moi non plus, je ne pourrais pas manger tout ça.

MegaHAL : Moi non plus.

l'utilisateur : Mais rien ne nous y oblige.

MegaHAL : Mais la pensée de Lacan ne peut jamais pulvériser des licornes dans le bureau et fume avec nous.

l'utilisateur : Pourquoi sa pensée ne peut-elle les pulvériser ?

MegaHAL : Pourquoi est-ce que tu ne regardes pas quoi ?

SYNTHÈSE

L'utilisateur : Pourquoi je ne regarde pas quoi comment ?

MegaHAL : Le président ne le regarde pas.

Contrairement à Megahal qui travaille au niveau des successions de mots, son rival Jabberwacky enregistre des couples de phrases. Ainsi, il « sait » que quand il a dit récemment « Tu connais Paul Ricoeur ? », son interlocuteur lui a répondu « J'en ai vaguement entendu parler ». De la sorte, il peut non seulement faire la même chose en sens inverse, mais de façon plus générale, traiter similairement toutes les phrases de la forme « Tu connais X ? ». Ici, nous lui avons soumis le cas de Jules César.

Mon nom est Jabberwacky. A qui ai-je l'honneur ?

Moi, je suis Jules César.

Salut, Jupiter le foudroyant.

Au secours, on me transperce à coups de poignard.

On vous transperce toujours ?

Non, juste aujourd'hui.

Ça nous occupe.

Mais c'est atrocement douloureux.

J'ai cru entendre appeler, mais il n'y a personne.

Peux-tu faire venir du secours ?

Je ne vois vraiment pas pourquoi tu as besoin d'avoir un nom de famille.

Je suis un patricien, il est vrai, mais sauve-moi.

De quel royaume ?

Non, je ne voulais pas devenir roi, ils ont menti.

Et si tu me parlais des malades que tu connais ?

Conclusion : un domaine à observer.

Ce rapide survol nous montre que les dialogueurs ne font, pour le moment, rien d'extraordinaire. Il semble néanmoins opportun de continuer à suivre leur activité, car le domaine de la conversation en ligne est, potentiellement, d'une haute valeur stratégique. Il suffirait, par exemple, que l'on parvienne à coupler les fonctionnalités des dialogueurs avec celles des moteurs de recherche, et l'on verrait déjà des effets impressionnants. En attendant ce jour, je vous invite à faire vos propres expériences en allant sur les pages ici référencées ci-dessous.

Sites à consulter :

<http://www.limsi.fr/Individu/jbb/sites.htm> : ma page sur le dialogue artificiel

<http://www.simonlaven.com/> : celle de Simon Laven sur les simulateurs de conversation.

<http://botspot.com/search/s-chat.htm> : le catalogue de Botspot.

<http://www-ai.ijs.si/eliza/eliza.html> : Eliza.

<http://www.limsi.fr/Individu/jbb/fliza.html> : Eliza et Jeanne d'Arc.

<http://www.jabberwacky.com/> : Jabberwacky, ayant réponse à tout (mais pas la bonne).

<http://www.limsi.fr/Individu/jbb/traumspiel.html> : Megahal, un interlocuteur markovien.

<http://Www.AliceBot.Org/> : Alice au pays du langage.

<http://www.thespoon.com/mimic/> : Mimic, un chatterbot qui apprend.

*Si vous souhaitez réagir sur l'intérêt actuel des dialogueurs artificiels,
vis à vis des études sur le dialogue ou la langue,
n'hésitez pas à ouvrir un débat en contactant Jean-Daniel.Zucker@lip6.fr*

Présentation du réseau Francophone d'Ingénierie Linguistique FRANCIL

Gabriel Illouz et Patrick Paroubek
LIMSI-CNRS – {gabrieli, pap}@limsi.fr

Suite aux conclusions du Sommet des Chefs d'État et de Gouvernement ayant le français en partage qui s'est tenu à l'Île Maurice en octobre 1993, l'Agence Universitaire de la Francophonie a mis en place un programme destiné à renforcer ses activités dans le domaine du

Génie Linguistique qui a démarré le 21 juin 1994. Le traitement automatique des langues est un secteur actuellement en plein essor qui englobe des activités de recherche et de développement en matière d'analyse et de génération de textes, de reconnaissance, de compré-

hension et de synthèse de la parole. Il inclut les applications relatives à la gestion de documents, à la communication entre l'humain et la machine, à l'aide à la rédaction, à la traduction assistée par ordinateur et à la recherche automatique d'information. Outre des enjeux de type

industriel et économique, de type scientifique et technologique, il présente aussi une dimension culturelle, voire politique, très spécifique. En particulier, il est toujours préférable de maîtriser la langue qui sert à exprimer les résultats des recherches, surtout si cette recherche porte sur la langue elle-même. La langue française et la francophonie sont donc très directement concernées par cet enjeu.

A l'instar des autres réseaux de l'AUF, le réseau FRANCIL a pour objectifs fondamentaux de :

- a) assurer la solidarité entre les chercheurs du monde francophone dans une perspective de co-développement ;
- b) aider au désenclavement des chercheurs, des laboratoires et des équipes de recherche des pays francophones et en particulier des pays du Sud ;
- c) favoriser la coopération entre les équipes, les laboratoires des universités, des grandes écoles et des grands centres de recherche, qu'ils soient nationaux, internationaux, publics ou privés ;
- d) contribuer à la qualité et au rayonnement de la recherche francophone ;
- e) favoriser la production scientifique en langue française ;
- f) permettre, dans toute la mesure du possible, aux chercheurs s'exprimant en français mais n'appartenant pas à des pays francophones de collaborer avec leurs homologues francophones.

Il regroupe plus de 80 organismes à travers le monde représentant environ 1000 chercheurs, enseignants ou ingénieurs.

Les activités du réseau FRANCIL s'articulent autour de quatre actions prioritaires suivantes :

- 1) Des actions de recherche partagée (ARP); ce sont des projets collaboratifs sur appel d'offre, qui ont pour but de favoriser les échanges entre les équipes et les laboratoires constitués autour des thèmes retenus par l'assemblée constitutive du réseau.
- 2) L'élaboration et la diffusion d'outils d'information scientifique et tech-

nique, avec par exemple le serveur SILFIDE (<http://www.loria.fr/projets/Silfide/Index.html>), la publication d'annuaires de chercheurs et d'institutions et la publication trimestrielle d'une lettre d'information dont les archives sont disponibles à l'URL : <http://francil.limsi.fr>.

- 3) L'organisation de journées scientifiques sur les sujets retenus par le réseau, par exemple les journées scientifiques et techniques du réseau FRANCIL qui ont eu lieu en Avignon les 15 et 16 avril 1997 [1, 2].
- 4) L'attribution de bourses pour des étudiants en rédaction de thèse et de bourses d'excellence post-doctorales.

L'appel aux Actions de Recherche Partagée, lancé en juin 1995 a reçu 17 réponses. La commission de sélection, en a retenu 12 (après fusion de deux projets).

Les thèmes prioritaires de l'appel à collaboration étaient :

Thème 1 :

Identification, création et mise à disposition de ressources (écrit et oral) et leur utilisation pour la réalisation et l'évaluation des systèmes de traitement automatique du langage ; outils informatiques et formalismes linguistiques.

Thème 2 :

Environnement d'aide à la rédaction : saisie optique, correction orthographique, typographie numérique.

Thème 3 :

Système assisté par ordinateur d'apprentissage du français écrit et oral.

Thème 4 :

Mise en place d'outils de création de ressources terminologiques multilingues, incluant le français.

Bien que la plupart des propositions soient à l'initiative de laboratoires de recherche français, les pays participants sont largement représentatifs de la francophonie. Outre la France, les pays de tradition francophone comme le Québec, la Suisse et la Belgique ont pris part à ces actions. Certains projets

débordent du cadre strictement francophone avec des collaborations de chercheurs francophones originaires de pays non francophones du Nord (Angleterre et Japon). L'Europe de l'Est est également présente avec un projet impliquant des équipes de recherches bulgares et hongroises.

La priorité avait été donnée aux propositions incluant des collaborations Nord-Sud ; ce critère de sélection est respecté pour la majorité des réponses reçues. Ainsi, le continent africain est représenté par les 3 pays du Maghreb, Algérie, Maroc et Tunisie. Concernant l'Afrique sub-saharienne, quatre pays figurent dans les propositions : le Niger, le Mali, le Congo et Madagascar. Des équipes de deux pays du Proche-Orient, l'Égypte et le Liban coopèrent à des actions. 45 laboratoires au total ont été retenus.

Le principal thème des projets porte sur la terminologie (Thème 4) avec une perspective bi- ou multilingue (6 projets sur 14). Le thème corpus

(thème 1) concerne 2 propositions. Ces deux thèmes (1 et 4) sont également présents conjointement dans 4 projets d'actions. Enfin, deux actions intéressent le thème 3 (apprentissage du français comme langue seconde).

Dans un souci de cohérence, les sujets de recherche convergents ont été fusionnés ; les regroupements concernent 4 projets fusionnés en deux. Les budgets annuels alloués sont de l'ordre de 50 KF par an sur 3 ans, excepté pour trois projets qui se sont vu attribuer un budget du même ordre pour un an, renouvelable sur présentation d'un dossier de prolongation à la fin de la première année. Ces budgets portent essentiellement sur le financement de missions et de stages entre les laboratoires partenaires.

Un second appel d'offre pour les ARP en 1998 pour lequel 18 réponses ont été reçues et 10 retenues. En outre, le Fonds Francophone de la Recherche (FFR) avait mis en place en coordination avec le réseau FRANCIL et au sein du programme Génie Linguistique, les Actions de Recherche Concertées (ARCs) [3, 4], des actions spécifiques qui ont fait l'objet d'appel d'offres concernant l'évalua-

tion dans 7 domaines en relation avec les thèmes retenus par l'assemblée constitutive du réseau. Pour chaque domaine, il y avait 3 tâches à accomplir :

- i) organisation d'une campagne d'évaluation,
- ii) fourniture des données annotées de référence pour les campagnes d'évaluation,
- iii) participation à la campagne de d'évaluation.

Un premier appel à participation a eu lieu en Novembre 1994; il a reçu 89 réponses et en mars 1995, 50 propositions étaient retenues (elles représentent 35 laboratoires ou institutions de 4 pays différents). Deux séries de campagnes ont eu lieu, l'une en 1996-1997, l'autre en 1998-2000. Chaque campagne d'évaluation a été clôturée par un atelier où tous les participants commentaient leurs résultats et comparaient leurs méthodes. Dans le domaine de l'écrit, 4 actions ont été considérées :

A1 Accès à l'information textuelle en langage naturelle (Amaryllis) [5].

Il faut retrouver des textes traitant d'un sujet particulier dans une collection comprenant un grand nombre de textes d'origine diverses (journal Le Monde, résumé scientifiques des bases PASCAL et FRANCIS de l'INIST).

A2 Alignement de corpus (Bi/Multi)lingue (ARCADE) [6].

Le but est d'aligner automatiquement des corpus parallèles (textes identiques dans plusieurs langues, par exemple les minutes des débats du parlement européen) phrase à phrase.

A3 Conception automatique de base de données terminologiques [7].

Dans cette ARC ont été testés des systèmes d'extraction terminologiques avec trois tâches (extraction de termes, extraction de relations sémantiques, classificateurs) sur des textes de littérature spécialisée.

A4 Compréhension de texte [12]

Comme les réponses à l'appel à proposition n'avaient pas été assez nom-

breuses, et puisque le domaine est essentiel, les organisateurs ont mis en place un groupe de travail avec le concours de trois laboratoires qui a fourni un rapport final en novembre 1997, dans lequel il est proposé d'utiliser la méthodologie DQR (Documents-Questions Réponses) [11].

et dans le domaine de l'oral, 3 actions ont été prises en compte :

B1 Dictée vocale [8].

Evaluation de systèmes de transcription automatique de signal de parole.

B2 Dialogue oral [9].

Evaluation d'interfaces vocales pour la recherche d'information touristique.

B3 synthèse vocale [10].

Evaluation des systèmes de traduction graphème-phonème.

Pour la période 1996-2000, le budget total du réseau FRANCIL était de l'ordre d'environ 4 MF; avec 1 MF pour le réseau (y compris les ARP), 1 MF pour les programmes de formation et 2 MF pour les ARC.

Les ARCS sont maintenant terminées et les activités actuelles du réseau FRANCIL concernent les ARP et la gestion du réseau. A ce propos, la lettre d'information FRANCIL cherche toujours de nouvelles occasions de faire connaître le travail des chercheurs francophones, n'hésitez donc pas envoyer à l'adresse suivante : francil@limsi.fr, des résumés de thèses récemment soutenues ou des descriptifs de vos travaux récents, ainsi que toutes les informations sur des événements ayant trait à l'ingénierie linguistique francophone.

Références

- [1] Actes des Journées Scientifiques et Techniques du Réseau FRANCIL, Avignon, 15-16 avril 1997, AUPELF-UREF, 566 pages.
- [2] Ressources et Évaluation en Ingénierie des Langues, sous la coordination de Karim Chibout, Joseph Mariani, Nicolas Masson, Françoise Néel, De Boeck Université et

AUPELF-UREF, 2000 ISBN 2-8011-1258-5, 677 pages.

- [3] Joseph Mariani and Patrick Paroubek. Human language technologies evaluation in the european framework. In Proc. of the DARPA Broadcast News Workshop, pages 237-242, Herndon, VA, Feb 1999.
- [4] J. Mariani. The Aupelf-Uref evaluation-based language engineering actions and related projects. In A. Rubio, N. Gallardo, R. Castro, and A. Tejada, editors, First International Conference on Language Resources and Evaluation, volume I, pages 123-128, Granada, May 1998.
- [5] A. Coret, P. Kremer, B. Landi, D. Schibler, L. Schmitt, N. Viscogliosi. Accès à l'information textuelle en français : Le cycle exploratoire Amaryllis. Actes des JST Francil 1997. Aupel-Uref, 15-16 avril 1997, Université d'Avignon.
- [6] J. Véronis. Une action d'évaluation des systèmes d'alignement de textes multilingues. Actes des JST Francil 1997. Aupel-Uref, 15-16 avril 1997, Université d'Avignon.
- [7] W. Mustafa el-Hadi, C. Jouis. Terminology Extraction and Acquisition from Textual Data: Criteria for Evaluating Tools and Methods. Proc. of the First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998, vol. 2 pp 1175-1178.
- [8] M. El-Bèze, M. Jardino, F. Bimbot. Une approche alternative pour le calcul de la perplexité. Actes des JST Francil 1997. Aupel-Uref, 15-16 avril 1997, Université d'Avignon.
- [9] S. Rosset, S. Bennacef, J.L. Gauvain, L. Devillers, L. Lamel. Corpus Oral de renseignements touristiques. Actes des JST Francil 1997. Aupel-Uref, 15-16 avril 1997, Université d'Avignon.
- [10] F. Yvon, P. Boula de Mareuil, C. d'Alessandro, V. Auberger, M. Bagein, G. Bailly, F. Bechet, S. Foukia, J.F. Goldman, E. Keller, D. O'Shaughnessy, V. Pagel, F. Sannier, J. Véronis, B. Zellner. Objective evaluation of grapheme to phoneme conversion for text-to-speech synthesis in French. Computer Speech and Language, Vol 12, pp393-410, October 1998.
- [11] J.Y. Antoine, J. Zeiliger, J. Caelen. RQA methodology: towards a qualitative evaluation of speech understanding and spoken dialog systems. Speech and Language Technology (SALT) Club Workshop on Evaluation in Speech and Language Technology, June 17-18, 1997, U. of Sheffield, UK.
- [12] P. Sabatier, P. Blache, J. Guizol, F. Levy, A. Nazarenko, S. N'Guema, R. Pasero, M. Rolbert. Evaluer des Systèmes de Compréhension de textes. Actes des JST Francil 1997. Aupel-Uref, 15-16 avril 1997, Université d'Avignon.

L'INRIA Sophia-Antipolis

Créé en 1967 à Rocquencourt près de Paris, l'INRIA (Institut national de recherche en informatique et en automatique) - <http://www.inria.fr> - est un établissement public à caractère scientifique et technologique (EPST) placé sous la double tutelle du ministère de la recherche et du ministère de l'économie, des finances et de l'industrie. Ses principales missions sont : (1) d'entreprendre des recherches fondamentales et appliquées ; (2) de réaliser des systèmes expérimentaux ; (3) d'organiser des échanges scientifiques internationaux ; (4) d'assurer le transfert et la diffusion des connaissances et du savoir-faire ; (5) de contribuer à la valorisation des résultats des recherches ; (6) de contribuer, notamment par la formation, à des programmes de coopération pour le développement ; (7) d'effectuer des expertises scientifiques ; (8) de contribuer à la normalisation.

L'INRIA est organisé en équipes (projets et actions) réparties dans cinq unités de recherche (Lorraine, Rennes, Rhône-Alpes, Rocquencourt et Sophia Antipolis) et menant des recherches dans 4 thèmes : (1) réseaux et systèmes, (2) génie logiciel et calcul symbolique, (3) interaction homme-machine, images, données, connaissances et (4) simulation et optimisation de systèmes complexes.

L'unité de Sophia-Antipolis - <http://www-sop.inria.fr> -, créée en 1983, comprend environ 450 personnes (dont 300 scientifiques parmi lesquels 85 chercheurs travaillant dans une trentaine d'équipes de recherche). Les trois équipes d'I.A. présentées ci-dessous - ACACIA (mémoire d'entreprise), AID (conception de sites/services Web) et ORION (interprétation d'images et pilotage de programmes) - font partie du thème 3A (Bases de données, bases de connaissances, systèmes cognitifs).

LE PROJET ACACIA

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

Contacts : Rose Dieng-Kuntz
(Rose.Dieng@sophia.inria.fr),
Olivier Corby

(Olivier.Corby@sophia.inria.fr)
et Alain Giboin
(Alain.Giboin@sophia.inria.fr)

1 Objectifs

Le projet ACACIA est un projet pluridisciplinaire, visant à développer des aides méthodologiques et logicielles pour la construction et la diffusion de mémoire d'entreprise.

Nous approfondissons la mémoire métier, la mémoire technique et la mémoire de projet, et leur matérialisation sous forme de documents, d'ontologies et d'annotations sémantiques, avec une éventuelle modélisation de multiples points de vue. Par analogie avec le Web, nous étudions la mémoire d'entreprise comme un « *Web sémantique d'entreprise* ».

Pour la représentation des ontologies, nous reposons sur :

- la méthode d'ingénierie des connaissances CommonKADS,
- le formalisme des graphes conceptuels (GC) de Sowa,
- et les langages de la galaxie XML conçus par le World Wide Web Consortium (W3C), en particulier RDF - Resource Description Framework.

Nos axes de recherche sont guidés par notre vision du cycle de vie de la mémoire d'entreprise : détection des besoins en mémoire d'entreprise, construction, diffusion, utilisation, évaluation et évolution de la mémoire [Dunod 2000].

2 Détection des besoins et construction d'une mémoire d'entreprise

2.1 Modèles psychologiques pour la conception de mémoires d'entreprise

Nous étudions comment adapter des modèles psychologiques à la conception de systèmes de gestion de mémoires d'entreprise, ces modèles pouvant servir de cadres pour comprendre et interpréter les pratiques réelles des membres de l'entreprise, pour définir des spécifications, pour élaborer des architectures de systèmes ou pour évaluer les systèmes et leur usage. L'adaptation des modèles

repose sur des analyses des pratiques réelles des membres des organisations considérées. Nous avons ainsi proposé une adaptation du modèle de Rouet et Tricot afin de prendre en compte les aspects collaboratifs de la recherche d'information [JETCSIC'2000]. Nous étudions également le processus d'« annotation sémantique » (du point de vue des utilisateurs) [PEC'2000]. Enfin, notre approche pour la détection des besoins aussi bien que pour l'évaluation repose sur l'analyse de scénarios.

2.2 Gestion de multiples expertises et de multiples points de vue

Pour prendre en compte la multi-expertise, nous avons proposé :

- une méthode d'acquisition reposant sur un protocole de recueil collectif et sur un modèle d'agent cognitif dédié à la multi-expertise (avec application à l'accidentologie par exemple) [COOP'96, AI-Mag 98, thèse de Labidi en 95],
- un modèle de la tâche d'ingénierie concurrente, un modèle de la mémoire de projet et des extensions des modèles CommonKADS pour la modélisation de conflits en conception concurrente [CERA 98],
- des techniques permettant de comparer (en se basant sur la terminologie et sur le contexte) des graphes conceptuels représentant les connaissances de plusieurs experts [ECAI'98] et de les intégrer en adoptant différentes stratégies d'intégration [ICCS'98],
- les modèles C-VISTA et CG-VISTA pour la représentation et la gestion de multiples points de vue [ICCS'97, ICCS'98, thèse de Myriam Ribière en 99],
- l'expression de multiples points de vue en RDF.

2.3 La mémoire d'entreprise vue comme un Web sémantique d'entreprise

L'évolution actuelle de nos travaux repose sur l'analogie entre la mémoire d'entreprise et le Web : tous deux sont des pay-sages d'information hétérogènes et distribués et partagent le même

problème de la pertinence des résultats obtenus lors d'une recherche dans leur contenu. Mais, comparée au Web, une mémoire d'entreprise a un contexte, une infrastructure et une portée délimités et mieux définis : l'organisation. Les travaux sur le Web sémantique visent à rendre le contenu sémantique du Web interprétable par les machines. Nous considérons la mémoire d'entreprise comme un Web sémantique d'entreprise, matérialisé par des ontologies, des documents (XML) et des annotations sémantiques concernant ces documents et utilisant le vocabulaire conceptuel défini dans ces ontologies [EKAW'99, IEEE Intell. Systems 2000].

Cette approche s'inspire de nos travaux antérieurs exploitant l'indexation de documents par des modèles de connaissances : CGKAT offrait une ontologie de haut niveau intégrant WordNet, permettait de construire une base de connaissances, représentée en GC, avec des liens vers les documents et offrait une aide à la recherche d'information dans la base ou dans les documents [KAW'95, ICCS'96, thèse de Philippe Martin en 96]. De même, le serveur de connaissances WebCokace [ECAI'96, IEEE-IC 99] permettait une diffusion de modèles d'expertise CommonKADS via le Web et une navigation hypertexte dans ces modèles et vers des documents associés. Les modèles d'expertise pouvaient être ainsi annotés par des documents et vice versa.

Les principaux problèmes de recherche posés sur un « Web sémantique d'entreprise » sont :

- Comment détecter les besoins et construire la mémoire (méthodologie et architecture) ?
- Comment construire et faire évoluer les composants de la mémoire (en particulier ontologies et annotations), tant à partir de plusieurs experts qu'à partir de textes ?
- Comment diffuser et utiliser la mémoire (recherche d'information vs diffusion active) ?
- Comment évaluer et faire évoluer la mémoire (concepteur vs utilisateur) ?

2.4 Construction d'ontologies

Ontologie d'entreprise

Nous proposons une méthode inté-

grant à la fois (1) une approche descendante se basant sur une synthèse d'ontologies existantes, (2) une approche ascendante afin d'exploiter l'analyse (manuelle ou semi-automatique) d'entretiens de spécialistes ou des documents de l'entreprise et (3) une approche centrifuge reposant sur la détermination, la généralisation et la spécialisation des concepts clés de l'entreprise. Nous avons utilisé cette méthode hybride pour construire l'ontologie O'CoMMA incluant des scénarios tels que l'insertion d'un nouvel employé dans une entreprise ou la veille technologique [IC'01].

Construction d'ontologies à partir de textes

Diverses sources d'informations dans l'entreprise peuvent aider à construire une ontologie : des experts humains, des documents textuels voire multimédia, des bases de données (avec éventuellement des données textuelles)... Suivant l'approche du groupe TIA, nous approfondissons la construction semi-automatique d'ontologies à partir de sources d'information hétérogènes, en utilisant des outils linguistiques pour exploiter les sources textuelles [TIA'01]. La méthode SAMOVAR repose sur :

- un amorçage manuel de l'ontologie à partir de l'analyse des entretiens avec des experts ou des documents de l'entreprise,
- un enrichissement progressif semi-automatique des ontologies grâce à l'exploitation des résultats d'outils linguistiques tels que des extracteurs de candidats-termes (e.g. Nomino) et des extracteurs de relations (e.g. Cameleon) appliqués au corpus textuel, et grâce à la proposition de règles heuristiques s'appuyant sur l'analyse des régularités de ces résultats.

La représentation en RDFS de l'ontologie et l'annotation en RDF de la base de documents ou de la base de données permettent ensuite d'exploiter notre moteur CORESE pour faire des recherches dans la base annotée.

Cette méthode, appliquée dans le système SAMOVAR [KCAP'01] pour une mémoire de projet véhicule chez Renault, peut se généraliser à la construction d'une mémoire d'un projet de

conception d'un système complexe (e.g. la mémoire des problèmes dans un tel projet).

Représentation formelle de l'ontologie et des annotations

Plusieurs langages sont possibles pour représenter les ontologies et les annotations :

→ RDF Schema (recommandé par le W3C) - nous avons étendu ce langage pour permettre d'exprimer dans une ontologie des définitions explicites de concepts et de relations, ainsi que des connaissances contextuelles [ICCS'01].

→ Des formalismes classiques de représentation des connaissances. Avec les équipes INRIA EXMO et ORPAILLEUR, nous participons au projet ESCRIRE visant à comparer GC, représentation par objets et logiques de description pour représenter le contenu sémantique de documents (Web) à des fins de recherche d'information.

Apprentissage d'ontologies

Nous avons proposé une méthode pour apprendre incrémentalement des ontologies à partir d'annotations RDF de ressources du Web ou de la mémoire d'entreprise [CAP'01]. Cette méthode peut être utilisée pour construire automatiquement des classes de documents à partir des annotations de ces documents et sera appliquée dans le projet européen CoMMA.

3 Diffusion et utilisation de la mémoire d'entreprise

3.1 Moteur de recherche sémantique

RDF permet de décrire le contenu de documents sous forme de méta-informations reposant sur une ontologie représentée en RDF Schema (ou RDFS). Or le formalisme des GC se prête bien à l'interprétation d'énoncés RDF et en particulier à l'interrogation, grâce à l'opérateur de projection, d'une base d'annotations RDF, traduites en GC. Nous avons développé CORESE (Conceptual Resource Search Engine), un moteur RDF(S) basé sur les GC. Il permet de charger une ontologie repré-

sentée en RDFS et des annotations RDF, de les traduire en GC puis, grâce à l'opérateur de projection, d'interroger la base d'annotations obtenue. Il exploite les propriétés des relations (symétrie, transitivité, etc.) ainsi que des règles d'inférence sur les GC. Utilisé dans les projets CoMMA, ESCRIRE et SAMOVAR, le moteur CORESE permet donc la recherche d'information dans un « Web sémantique d'entreprise » [ICCS'2000].

3.2 Architecture multi-agents pour la gestion d'une mémoire d'entreprise

L'hétérogénéité et la distribution des systèmes multi-agents peuvent être une solution à l'hétérogénéité et à la distribution de la mémoire d'entreprise. Nous avons ainsi proposé une architecture multi-agents pour la recherche d'information dans une mémoire d'entreprise (approche « pull ») ou pour la dissémination d'informations vers les utilisateurs en fonction de leurs profils (approche « push »): agents dédiés aux ontologies, aux documents, aux utilisateurs, à la connexion entre agents [PAKM'2000].

Ces travaux se déroulent dans le cadre du projet européen CoMMA (avec ATOS, le CSTB, le LIRMM, l'université de Parme et T-NOVA) [eBusiness'2000]. Ce projet vise à développer un système multi-agents pour la gestion d'une mémoire d'entreprise matérialisée dans des documents sémantiquement annotés par des énoncés RDF relatifs à une ontologie, un modèle d'entreprise et des modèles utilisateurs. Ces agents ont des capacités d'apprentissage pour s'adapter à l'utilisateur et certains agents seront proactifs et permettront de disséminer les informations adéquates en fonction des profils utilisateurs. Deux scénarios sont étudiés : l'insertion des nouveaux employés, et la veille technologique. Le système CoMMA, implémenté en JAVA sur la plate-forme multi-agents JADE, intègre notre moteur CORESE.

4 Applications et collaborations

Nos travaux ont été appliqués à la mémoire technique ou mémoire de projet ou veille technologique ou serveur de

connaissances en industrie aéronautique (Aérospatiale, Dassault-Aviation), en industrie automobile (Renault), en industrie du bâtiment (CSTB), en accidentologie (INRETS), en télécomm. (T-NOVA, CSELT), et dans le domaine médical.

Nous participons aux groupes de travail COOP, GRACQ - <http://www.irit.fr/GRACQ/> - et TIA - <http://www.biomath.jussieu.fr/TIA> - ainsi qu'au réseau européen OntoWeb (Ontology-based Information Exchange for Knowledge Management and Electronic Commerce) - <http://www.ontoweb.org> - et au projet européen IST CoMMA (Corporate Memory Management through Agents) - <http://www.ii.atos-group.com/sophia/comma/HomePage.htm>.

Enfin, nous animons le groupe de travail chercheurs - industriels Memento du Club CRIN ITI et organisons des ateliers sur la mémoire d'entreprise (cf. KAW'96, KAW'98, KAW'99, ECAI'98, IJCAI'99, ECAI'2000, IJCAI'2001).

Quelques références

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

[IEEE-IC 99] O. Corby and R. Dieng. The WebCokace Knowledge Server. IEEE Internet Computing, vol. 3, n. 6, pp.38-43, December 1999.

[Dunod 2000] R. Dieng, O. Corby, A. Giboin, J. Golebiowska, N. Matta, M. Ribière, Méthodes et outils pour la gestion des connaissances, Dunod, 2000 (ISBN 2 10 0045741), Informatiques, Série Stratégies et systèmes d'information.

[AI-Mag 98] R. Dieng, A. Giboin, C. Amérgé, O. Corby, S. Després, L. Alpay, S. Labidi, S. Lapalut. Building of a Corporate Memory for Traffic Accident Analysis. AI Magazine, Vol. 19, n.4, p. 80-100, Winter 1998.

L'ACTION AID

(<http://www-sop.inria.fr/aid/>)

Gestion des connaissances pour l'aide à la conception coopérative. Application à la conception et développement de sites/services Web basés sur l'analyse de l'usage

Contact: Brigitte Trousse
(Brigitte.Trousse@inria.fr)

et Thierry Despeyroux
(Thierry.Despeyroux@inria.fr)

1 Objectifs

L'action AID (« Artificial Intelligence in Design ») travaille depuis plusieurs années sur des méthodes de représentation des connaissances et techniques de raisonnement en IA (SBC, CSP, CBR) pour le développement d'outils d'aide à l'analyse et à la conception de systèmes complexes (satellites, voitures, logiciels CAO, sites Web, etc.). Elle regroupe actuellement sur Sophia-Antipolis une dizaine de personnes dont deux permanents. Notons que l'action AID est en cours d'évolution et que la description ci-dessous constitue la base d'une proposition d'un nouveau projet INRIA avec un renfort prévu au niveau de techniques statistiques et méthodes de classification. Notre domaine d'applications depuis 1997 concerne la conception et le développement de sites/services Web ou plus généralement de systèmes d'informations hypermédia et de systèmes de recommandations. Nous visons actuellement des méthodes et outils pour aider les « usagers » d'un site Web que sont les utilisateurs et créateurs/webmasters. Pour cela, nous avons adopté une démarche volontairement pluridisciplinaire (IA, génie logiciel et analyse exploratoire de données). Nos objectifs sont triples : 1) anticiper les problèmes liés à l'évolution d'un site le plus tôt possible dès la conception par une analyse de critères de cohérence globale du site ; 2) anticiper les problèmes liés à l'utilisation d'un site le plus tôt possible dès la conception par une analyse de critères d'évaluation liés à l'usage et en offrant des recommandations personnalisées d'aide à la recherche d'information dans un site; et enfin 3) capitaliser des expériences de conception et développement de sites/services Web.

2 Notion de Qualité d'un site - Sémantique d'un site Web

Nos motivations de recherche sont liées à la notion de qualité d'un site et en particulier aux questions suivantes: qu'est-ce qu'un site cohérent? qu'est-ce

qu'un bon site pour les utilisateurs cible?. C'est à l'aspect langage formel (non naturel, comme HTML, XML) - et non aux parties en langues naturelles des sites - auquel nous nous intéressons dans cette équipe. La lecture et l'interprétation des pages par un usager se font par le langage naturel, par contre le webmaître privilégie principalement dans son interprétation la partie non naturelle du site. Par analogie avec les langages de programmation, et bien qu'un site Web ne s'exécute pas, nous distinguons deux aspects dans la sémantique d'un site:

- **L'aspect statique** d'un site concerne bien sûr la conformité d'un site à un standard (HTML, XML et DTD, etc.) mais ceci est insuffisant pour garantir la cohérence d'un site. Aussi au-delà de ces aspects syntaxiques ou contraintes locales (W3C Web Semantics, XSLT), il faut pouvoir s'assurer de la légalité d'un site vis à vis de règles globales qui décrivent sa cohérence.
- **L'aspect dynamique** d'un site vise la formalisation de tout ce qui concerne la navigation dans un site, et plus généralement son usage. Nous visons dépasser la notion habituelle de « hit » en définissant celle de « session » puis en essayant d'en extraire le plus de renseignements possibles sur le comportement de l'utilisateur et sur l'adéquation du site à ses besoins (notion de réussite et d'échec). Ceci peut prendre la forme d'une analyse statistique et d'une classification de sessions, ou mener à un mécanisme de calcul de recommandations personnalisées afin d'aider l'utilisateur à rechercher de l'information. Le parcours de l'usager se fait en fonction de ses actions sur les informations contenues dans les pages, donc d'une perception sémantique de ce contenu.

Les aspects *statique* et *dynamique* sont deux composantes d'une démarche qualitative dans la conception et le développement d'un site en perpétuelle évolution. Notre démarche vise à permettre non seulement à l'équipe de conception d'exprimer certaines contraintes globales du site mais aussi de mieux analyser son usage en typant les pages ou fragments de page correspondant aux

pas d'une navigation. Nous visons aussi l'expression de critères d'évaluation de l'usage (réussite, échec) pour mieux réutiliser l'usage dans une approche de calcul de recommandations basée sur des similarités de navigations. D'autre part l'analyse de l'usage est un des éléments qui peuvent motiver l'évolution d'un site car elle est liée à la satisfaction de l'usager. Cette satisfaction est dépendante du type de site ; par exemple, pour les sites de e-commerce, celle-ci concerne la proposition d'articles adaptés aux besoins de l'usager.

L'originalité de nos recherches pour le Web sémantique (cf. thème 1) consiste à explorer le plus loin possible le parallèle entre programmes et sites Web pour mieux aborder la sémantique des sites. De plus nous visons sur ce thème faire des propositions pour faciliter la tâche de mise à jour voire de re-conception du site par les créateurs/webmaîtres plutôt qu'améliorer la tâche de recherche d'informations des utilisateurs finaux (comme la plupart des travaux en IA basés sur les ontologies). Notons que nous participons au réseau européen IST Ontoweb en vue d'appliquer notre approche à la vérification sémantique d'une ontologie basée sur un langage de construction d'ontologies ou à la vérification d'un document XML utilisant une ontologie.

3 Programme de recherche

Celui-ci se décompose en deux thèmes correspondant respectivement aux aspects statique et dynamique d'un site plus un troisième transversal aux précédents se situant davantage à un niveau ontologique :

3.1 Aspect statique - Vérification sémantique d'un site Web

- Etude des méthodes de formalisation de l'aspect statique d'un site i.e. permettant de décrire les contraintes syntaxiques et sémantiques d'un site. La partie syntaxique est assez bien connue grâce aux notions de syntaxes abstraites reprises dans XML et les DTD. La partie « sémantique statique » i.e. prenant en compte les contraintes globales d'un

site pourra être basée sur une méthode proche de la déduction naturelle, avec la définition d'un langage de spécification sémantique. Notons que cette partie est très peu étudiée jusqu'à présent (citons, en IA, Webmaster de Van Harmelen&al),

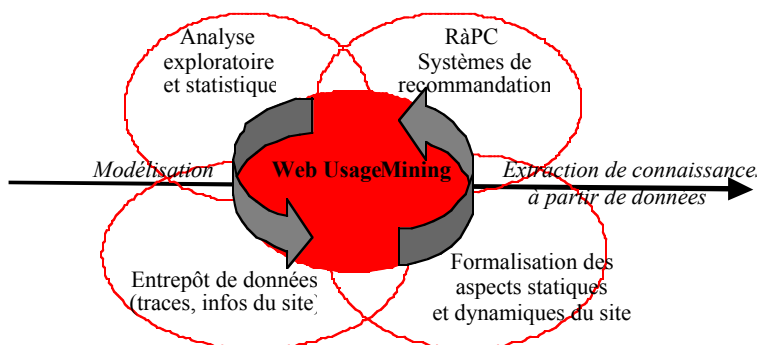
- Conception et réalisation d'outils de vérification sémantique de sites i.e. produisant des diagnostics de conformité d'un site particulier à un ensemble de règles globales sémantiques [1]. Notre objectif consiste à étudier l'apport des travaux en sémantique des langages de programmation appliqués aux sites Web, en particulier un style de sémantique proche de la « sémantique naturelle » (cf. G. Kahn 97). Il s'agit de donner un ensemble de règles d'inférence qui définissent ce qu'est un site sémantiquement cohérent : a) construire un environnement qui contient un ensemble d'assertions, b) donner des règles qui permettent de déduire à partir des assertions contenus dans l'environnement, ou à partir de données externes que des fragments de pages sont sémantiquement corrects ou bien typés, c) déléguer éventuellement une partie de la preuve à un système externe (par exemple à un thesaurus). Plusieurs maquettes (Centaur/as/cs/Typol, Sax/Prolog) ([1] et HT2001) ont été réalisées validant notre approche.

3.2 Aspect dynamique - Analyse et réutilisation de l'usage

La Figure en page 22 illustre nos activités sur ce thème..

- Etude des méthodes de *formalisation de l'aspect dynamique d'un site* i.e. permettant de définir et formaliser les notions de navigation et plus généralement de session (séquences temporelles, graphes, etc.) afin de pouvoir générer des traces utiles à l'analyse de l'usage ;

- Etude de méthodes d'*analyse de l'usage* et conception d'outils. Il s'agit du dépouillement des traces recueillies lors des navigations avec l'aide d'informations sur le site (en termes de contenu). On peut distinguer une analyse faite on-line, pendant la navigation, qui peut alimenter un système de recommandation d'aide à la navigation par exemple d'une analyse off-line pouvant alimenter une



étude de mise à jour voir d'une re-conception du site. Ces analyses font appel à des méthodes de classification et techniques statistiques ou encore de fouille de données et d'extraction de connaissances à partir de données (ECD); Les recherches visées concernent en particulier 1) la modélisation des neurones (Kohonen) en contexte décisionnel (ici dans le cadre de systèmes de recommandations basés sur l'analyse de l'usage) et 2) l'utilisation de la classification croisée entre les rubriques d'un site par exemple et les sessions utilisateur: en vue d'une future réorganisation de ces rubriques donc de la structure du site par analyse de l'usage.

- Etude de méthodes de conception de systèmes de recommandations personnalisés basés sur la gestion de l'expérience ici liée à l'usage d'un site. Ces travaux ont fait l'objet d'une présentation dans le dossier « IA et Internet » du numéro 38 du bulletin de l'AFIA et s'appuient principalement sur les travaux en raisonnement à partir de cas (RàPC) issus de la thèse de M. Jaczynski [2,3]. Une approche originale de calcul de recommandation pour l'aide à la navigation et recherche d'informations sur le Web ou dans un site Web appelée « Broadway » a été proposée: celle-ci est hybride à la fois de type filtrage collaboratif et de type filtrage basé sur l'analyse du contenu des pages visitées, s'appuyant sur des similarités de sessions et des techniques RàPC et offrant des capacités on-line d'apprentissage et de personnalisation. Elle a donné lieu à ce jour à trois systèmes: Broadway-1 pour l'aide à la navigation sur le Web présenté à EWCBR98, Broadway-2 pour l'aide à la navigation dans un

annuaire thématique de France Télécom (CTI FT/INRIA), BeCBKB pour l'aide à la formulation de requêtes dans un contexte de meta-moteur de recherche (contrat XRCE Meylan/INRIA présenté à RàPC'99 (plate-forme AFIA)). Des expérimentations en coopération avec des ergonomes ont permis de valider l'aide apportée par Broadway-1 et Broadway-2 (en coopération avec FT Grenoble). De plus une première expérimentation à partir de logs HTTP issus d'un site musical (challenge IJCAI97) a été faite pour comparer notre approche avec d'autres approches (statistiques, Markov) sur la capacité à prédire les comportements utilisateurs, les résultats ont été présentés à RIAO'2000. Nos principales recherches visent 1) des méthodes efficaces des sessions utilisateurs d'indexation et de classification (cartes de Kohonen, règles d'association, k-hiérarchies faibles, etc.) dans un **contexte décisionnel** (RàPC) ; 2) une formalisation du calcul de recommandation basée sur notre approche RàPC; 3) une capitalisation de composants métier dans Broadway*Tools et enfin 4) des méthodes d'évaluation expérimentale (en particulier statistique) de tels systèmes. Notons que nous participons à deux projets européens, ASSO (géré par Eurostat) et KnowEX (Action COST) qui concernent ce thème.

3.3 Capitalisation des connaissances/expériences

Il s'agit ici de capitaliser les expériences relativement à la conception de sites particuliers (par exemple sites institutionnels, annuaires thématiques, portails, etc.) en gestion de points de vue relatifs à un site (cf contrat

CNES/INRIA 1997, GIS Cognition-Conception 1998) et en formalisation des aspects statique et dynamique d'un site..

3.4 Réalisation de logiciels

En complément à nos actions de recherche, l'action AID poursuit la réalisation de ces deux plates-formes à objets conçues via l'atelier Rose de Rational et implantées en Java, offrant de nombreux points d'ouverture documentés en termes de patrons de conception: 1) CBR*Tools [2,3] pour la conception de systèmes RàPC et 2) Broadway*Tools, pour la conception de systèmes de recommandations pour l'aide à la recherche d'informations et à la navigation sur le Web ou dans un site Web.

Quelques références

(<http://www-sop.inria.fr/aid/publications.html>)

- [1] T. Despeyroux et B. Trousse. *Semantic Verification of Web Sites Using « Natural Semantics »*. In RIAO'2000, 6th Conference on "Content-based Multimedia Information Access", Collège de France, Paris, France, avril 2000.
- [2] M. Jaczynski. *Modèle et plate-forme à objets pour l'indexation de cas par situation comportementale: application à l'aide à la navigation sur le Web*. Thèse de doctorat, Université de Nice Sophia Antipolis, décembre 1998.
- [3] M. Jaczynski et B. Trousse. *Patrons de conception dans la modélisation d'une plate-forme pour le raisonnement à partir de cas*. Revue *L'objet*, 5(2), 1999. Numéro spécial sur les patterns orientés objet, D. Rieu et J-P. Giraudon (guest editors).

LE PROJET ORION

(<http://www-sop.inria.fr/orion/>)

Contacts : Monique Thonnat (Monique.Thonnat@inria.fr) et Sabine Moisan (Sabine.Moisan@inria.fr)

1 Objectifs

ORION est une équipe pluridisciplinaire, à la frontière des domaines de la vision par ordinateur, des systèmes à

base de connaissances et du génie logiciel. Les recherches d'ORION ont deux objectifs : d'une part, l'interprétation automatique d'images et, d'autre part, la réutilisation et le pilotage automatique de programmes.

Cet objectif est poursuivi en développant des langages d'expression des connaissances ainsi que des mécanismes d'apprentissage et de traitement de ces connaissances, adaptés à des classes de problèmes spécifiques. ORION se focalise sur l'étude des connaissances qui interviennent dans les deux types de problèmes étudiés : connaissances sur les objets et les scénarios à reconnaître pour l'interprétation automatique d'images, connaissances sur les programmes et leur utilisation pour le pilotage automatique de programmes.

Les applications qui intéressent ORION relèvent principalement du domaine de la vision par ordinateur.

2 Interprétation d'images

L'interprétation automatique d'images consiste à attribuer une sémantique aux données en fonction de modèles prédéfinis.

Cela correspond à un sous-problème de la perception, celui de l'interprétation automatique des résultats des traitements d'images.

L'interprétation automatique d'images est une problématique très difficile qui est la base de nombreux travaux en vision et aussi en intelligence artificielle. La difficulté dépend en premier lieu de la nature des entités à reconnaître. Il est plus simple de reconnaître des objets statiques et rigides en environnement manufacturé, que des comportements dynamiques de plusieurs objets non-rigides en environnement naturel. La difficulté dépend également du type d'interprétation recherchée. Le problème peut être soit, simplement, d'étiqueter une entité bien déterminée que l'on peut mettre directement en correspondance avec des modèles, soit de détecter les entités, de les étiqueter et de vérifier leur cohérence (spatiale, temporelle, structurelle, etc).

Dans le projet ORION, nous nous

intéressons à l'interprétation automatique de données spatiales et/ou temporelles (2D à 4D). Les résultats de l'interprétation peuvent être la reconnaissance d'objets physiques, d'événements, de situations ou de scénarios. Le problème d'interprétation se décompose en une étape de traitement d'images qui permet la détection des entités d'intérêt et une étape d'analyse de ces entités dépendante du but du système développé, qui peut être la classification d'objet ou l'analyse de comportement. La résolution d'un tel problème se révèle complexe en raison de la double nature des connaissances requises : en effet, l'extraction, à partir d'images d'entrée, des paramètres et des primitives essentiels à l'interprétation nécessite des appels à des programmes de traitement d'images ; de plus, l'interprétation proprement dite ne peut s'accomplir qu'avec la connaissance du domaine que possède un expert sur les entités à reconnaître ou à analyser. L'automatisation des appels aux programmes de traitement d'images est un problème de pilotage de programmes ; il est donc traité dans l'axe de recherche correspondant.

Les deux points présentés ici concernent les modèles proposés pour l'interprétation : d'une part, pour l'interprétation de séquences d'images allant jusqu'à la reconnaissance de scénarios, et, d'autre part, pour la reconnaissance d'objets complexes.

2.1 Interprétation de séquences d'images

L'interprétation de séquences d'images a pour objectif, pour ce qui nous concerne, de donner un sens à une scène décrivant des activités humaines, à partir d'images fournies par une caméra couleur, monoculaire et fixe [ECAI00]. Cette interprétation de scène repose, en général, sur la coopération d'un module de traitement d'images, d'un module de suivi des objets mobiles et d'un module de reconnaissance du comportement des objets mobiles qui s'appuient sur une base de contexte. Il s'agit, pour le module de traitement d'images, de détecter les régions mobiles sur la séquence d'images. Le module de suivi associe les régions

détectées afin de former et de suivre les objets mobiles. La tâche du module de reconnaissance des comportements consiste, grâce à des techniques d'intelligence artificielle, à identifier les objets suivis et à reconnaître leur comportement comme constitutif d'un ou plusieurs scénarios prédéfinis.

2.2 Reconnaissance d'objets complexes

La reconnaissance d'objets complexes a pour objectif, à partir de modèles sémantiques abstraits, la reconnaissance d'un objet non géométrique. Dans une première phase, les traitements d'images sont effectués sur l'image à interpréter ; ils permettent de détecter, d'isoler et de décrire l'objet grâce à des mesures numériques. A partir de ces mesures numériques, le système d'interprétation va ensuite classer l'objet à reconnaître parmi des hiérarchies de classes prédéfinies et structurées qui représentent les modèles sémantiques ; trois étapes récursives essentielles sont impliquées dans cette phase de classement : abstraction des données, mise en correspondance avec une classe prédéfinie, raffinement de la reconnaissance. Pendant la phase de classement, de nouvelles informations peuvent devoir être recherchées à partir de l'image.

Notons que la réalisation de tels systèmes opérationnels requiert une somme considérable de travail pour le développement des bases de connaissances et des algorithmes de traitement d'images.

3 Pilotage de programmes

Le pilotage de programmes a pour but d'automatiser l'utilisation de logiciels complexes (bibliothèques de traitement d'images par exemple), grâce à des systèmes capables de planifier des traitements et d'en contrôler l'exécution.

Un grand nombre de travaux en pilotage de programmes sont motivés par un domaine d'application particulier (tel que le traitement d'images, le traitement du signal ou le calcul scientifique).

Le pilotage est un domaine de recherches pour lequel les techniques des systèmes à base de connaissances sont bien adaptées. En effet ces techniques répondent bien au double objectif du pilotage qui est premièrement,

d'automatiser l'utilisation de logiciels complexes et deuxièmement de capitaliser la connaissance sur l'utilisation des programmes dans les entreprises.

Nous proposons donc depuis plusieurs années une approche basée sur les techniques des systèmes à base de connaissances (SBC). Nous étudions en particulier le problème de la modélisation des connaissances spécifiques au pilotage de programmes. Cette modélisation sert à définir, d'une part, des langages de description et des méthodes de vérification de bases de connaissances, et, d'autre part, des moteurs dédiés au pilotage de programmes. Nous développons aussi des prototypes de moteurs, grâce à une plate-forme logicielle dont le but est de servir de support au développement de systèmes à base de connaissances.

Un système à base de connaissances pour le pilotage de programmes est généralement constitué :

- d'une bibliothèque de programmes exécutables adaptés à un domaine d'application (par exemple : traitement d'images médicales),
- d'une base de connaissances recueillant le savoir-faire d'utilisation des programmes de la bibliothèque (choix, ordonnancement, paramétrage, exécution, évaluation des résultats, etc.),
- d'un moteur de pilotage qui choisit, ordonne et exécute les programmes de la bibliothèque en fonction de l'objectif et des données de l'utilisateur et des connaissances contenues dans la base de connaissances,
- d'une interface permettant à un utilisateur d'exprimer un objectif à atteindre sur des données particulières, de suivre l'exécution des traitements et de visualiser leurs résultats.

La *modélisation* des connaissances de pilotage de programmes permet le transfert des compétences de l'expert qui a développé les programmes vers des utilisateurs novices. Nous avons modélisé les concepts et les mécanismes du pilotage de programmes tout d'abord pour OCAPI, puis pour les moteurs récemment développés. Pour cela nous avons

utilisé tout d'abord le modèle d'expertise de KADS. Ce travail de modélisation a été amélioré en se basant sur des techniques de réutilisation de composants logiciels (génie logiciel), des techniques de planification (intelligence artificielle), des systèmes existants de pilotage de programmes et sur notre expérience pratique sur des applications comme la détection d'obstacles dans des scènes routières ou l'imagerie médicale.

4 Plate-forme logicielle de développement de SBC

La construction d'outils dédiés à une classe de problèmes particulière (comme le pilotage de programmes) offre deux avantages : d'une part focaliser les modèles de connaissances fournis par les outils sur les besoins de la classe de problèmes et, d'autre part, fournir des formalismes unifiés et communs aux bases de connaissances d'une même classe de problèmes. Nous souhaitons aller plus loin et aborder l'ingénierie des systèmes à base de connaissances dans leur globalité. C'est dans ce but qu'a été développée la plate-forme logicielle *Lama*, afin de servir de support au développement rapide et fiable de systèmes à base de connaissances. L'objectif est de fournir des techniques et des outils (italique génériques) permettant de construire des systèmes à base de connaissances dans de nombreux domaines, en s'appuyant sur les techniques avancées en matière de modélisation, de programmation et de génie logiciel.

Cette plate-forme fournit un environnement unifié pour construire non seulement des bases de connaissances, mais aussi des variantes de moteurs et des outils annexes (vérification, interfaces, etc.). Elle regroupe des outils communs partageables et adaptables à la tâche visée. Ces outils sont complémentaires mais indépendants, de façon à pouvoir modifier, voire ajouter ou retirer un outil sans modifier le reste. Les outils proposés couvrent tous les besoins liés au développement de systèmes à base de

connaissances : depuis les modèles de tâches jusqu'aux preuves de moteurs et à la distribution des systèmes, en passant par l'expression des connaissances.

La plate-forme comprend principalement une bibliothèque de composants réutilisables et extensibles (un *framework*) [TSI01] pour construire des moteurs (de pilotage de programmes en particulier), un langage de description de base de connaissances, des outils de vérification de base de connaissances et une bibliothèque pour la construction d'interfaces graphiques.

En pilotage l'utilisation de cette plate-forme a entraîné un gain en temps et en quantité de code à écrire pour créer ou reconfigurer des moteurs, ou pour construire des bases de connaissances diverses [ICVS99]. C'est aussi un environnement précieux pour comparer des systèmes entre eux.

Quelques références

(<http://www-sop.inria.fr/orion/Publications/>)

- [ICVS99] M. Thonnat, S. Moisan and M. Crubézy. « Experience in Integrating Image Processing Programs ». Dans *International Conference on Computer Vision Systems*, Las Palmas, Canary Islands, janvier 1999.
- [TSI01] S. Moisan. « Réutilisation et générateurs de systèmes à base de connaissances : le framework Blocks », *TSI, numéro spécial réutilisation*, vol 20, n.4/2001, p. 529-553.
- [ECAI00] N. Rota and M. Thonnat. « Activity Recognition from Video Sequences using Declarative Models », 14th European Conference on Artificial Intelligence (ECAI 2000), 20-25 August, Berlin, W Horn (ed.), IOS Press, Amsterdam

Laboratoire d'Informatique du Havre (LIH) Université du Havre UPRES EA3270

Directeur : Alain Cardon - Alain.Cardon@univ-lehavre.fr
25 rue Ph. Lebon - BP 540 - 76058 Le Havre Cedex
<http://www-lih.univ-lehavre.fr>

1 Présentation générale du LIH

Le laboratoire LIH est depuis l'année 2000, une Equipe d'Accueil - EA3270 - de l'Université du Havre.

Il est constitué de 31 enseignants-chercheurs, doctorants et IATOS : 1 professeur d'université, 1 HDR-maître de conférence, 18 maîtres de conférences, 1 ingénieur de recherche en temps partagé, 1 docteur, 8 doctorants et 1 secrétaire en temps partagé.

Le laboratoire LIH est structuré autour d'un thème central : « Les systèmes informatiques répartis » qui se décline en 3 sous-thèmes :

- Objets distribués et systèmes multi-agents ;
- SGBD temps réel ;
- Parallélisme et ordonnancement.

Le laboratoire a publié, de 1999 à 2001, 3 ouvrages, 14 articles dans des revues, 65 articles dans des congrès majoritairement internationaux et a fait soutenir 4 thèses.

Il participe à la formation doctorale en Haute-Normandie, en ayant la responsabilité avec le LIFAR (Laboratoire d'Informatique Fondamental et Appliqué de Rouen) du DEA Informatique Théorique et Applications de l'Ecole Doctorale SPMI - Rouen/Le Havre.

Le laboratoire est très actif au niveau du développement d'actions de recherche dans le cadre des problématiques spécifiques développées et soutenues par la Région Haute-Normandie. Il est notamment co-fondateur du Groupement d'Intérêt Scientifique CRIHS (Centre de Recherche en Ingénierie Homme-Système) dont le partenaire économique principal est Matra Système et Information (Val de

Reuil). Une bourse industrielle régionale (X. Denis) a démarré en 2001 dans ce cadre.

Il est également engagé dans des collaborations internationales et notamment dans le projet Européen « Tactile Interactive Multimedia computer games for visually impaired children » (IST-2000-25298).

2 Positionnement des recherches en Intelligence Artificielle

Les activités liées à l'intelligence artificielle ont essentiellement lieu au sein de l'équipe « Objets distribués et systèmes multi-agents » (Contact : D. Olivier) dont le thème est orienté vers les aspects distribués de représentation comportementale ou organisationnelle. Ainsi, nous étudions la résolution de problèmes où des entités distribuées plus ou moins autonomes sont en interaction et nous abordons en particulier les points suivants :

- La complexité organisationnelle et combinatoire des systèmes ;
- Les approches de résolution fondées sur des systèmes informatiques distribués et coopératifs ;
- Les propriétés évolutives et adaptatives des systèmes complexes modélisés par agents.

Par ailleurs, cette équipe qui est la plus importante numériquement dans le laboratoire (elle regroupe environ la moitié de ses membres), développe des interactions importantes et durables avec les autres équipes sur des aspects liés à l'intelligence artificielle distribuée :

- Au sein de l'équipe « SGBD temps réel » (Contact : B. Sadeg), Claude Duvallat a soutenu sa thèse sur l'utilisation d'agents dans les systèmes d'aide à la décision temps réel et distribués.

- Dans l'équipe « Parallélisme et ordonnancement » (Contact : F. Guinand), la parallélisation de modèles représentés par des graphes est au cœur des problématiques développées. Des collaborations pour construire et gérer des graphes distribués par des systèmes multi-agents, notamment pour la gestion de mobilité de codes, d'équilibrage de communications et de fiabilisation sont développées au travers des différentes thèses en cours.

3 Description des actions de recherches en IA

Nos actions de recherche consistent à modéliser des systèmes complexes qui sont constitués :

- d'entités hétérogènes de base en interaction ;
- d'un environnement global pour ces entités dans lequel il est possible d'identifier des dynamiques, des structures ou des organisations générées par le comportement collectif des entités en interaction ;
- des contraintes, issues de l'environnement global et de ses dynamiques ou structures internes qui agissent sur les entités ou encore sont capables d'en générer de nouvelles.

Nous cherchons à construire des modèles génériques pour représenter de tels systèmes qui s'auto-organisent et s'adaptent de manière autonome à leur environnement.

Les principales actions de recherche sont les suivantes.

3.1 Aide à la décision en modélisation par agents

Nos travaux dans le domaine de l'aide à la décision concernent les SIC

(Systèmes d'Information et de Communication). La démarche retenue s'appuie sur les systèmes multi-agents. Ainsi nous considérons des Systèmes d'Information multi-institutions permettant l'échange d'informations à caractère décisionnel entre les décideurs de ces diverses institutions. Il s'agit alors d'interpréter chaque message, émis sous forme hypertexte, par les différents acteurs décisionnels d'une situation d'urgence afin que chaque receveur dispose d'une interprétation adaptée à sa problématique courante. Nous traitons donc un problème d'interprétation artificielle de la signification des messages envoyés entre des acteurs distants, à propos d'un phénomène que chacun perçoit à sa façon, dans le but d'harmoniser les points de vue. Les problèmes de fusion de connaissances et d'interprétation artificielle sont abordés. Nous nous intéressons également à la modélisation de scénarii et d'agents de prédiction dans le cadre de la surveillance de sites industriels.

SIC de gestion de situations d'urgence

Contact : A. Cardon

Appuyés par les thèses de Stéphane Durand et Franck Lesage, ces travaux portent sur la représentation de points de vue multiples dans la couche communicationnelle d'un Système d'Information et de Communication, notamment dans le cadre de crises industrielles. Le système développé représente les avis et jugements des acteurs dans un modèle multi-agent dont la connaissance se reconforme structurellement selon l'évolution du phénomène perçu par ces différents acteurs. Une interprétation adaptative du discours est créée et traitée par des algèbres d'agents opérant sur des morphologies de paysages d'agents. Le système est alors chargé de construire une représentation effective du sens du phénomène. Il la délivre à chaque acteur de manière interactive sous forme graphique. Cette recherche se place dans le cadre de la génération d'un langage relatif au sens des communications entre opérateurs, exprimé par la morphologie des SMA, par émergence de concepts significatifs. Ces travaux ont été implémentés en SmallTalk et un proto-

type de SIC distribué en Haute-Normandie a été réalisé.

Agents et système de veille préventive

Contact : G. Simon

Ces travaux s'inscrivent dans le cadre d'un système de surveillance de sites industriels. Le principal objectif de ce système est l'aide à la gestion du risque. Ce système repose sur une architecture multi-agents et comporte plusieurs types d'agents permettant, entre autres, de représenter et d'analyser la situation courante ainsi que de répondre à des requêtes d'utilisateurs en s'appuyant sur des systèmes d'information (météo, capteurs, minutes, ...).

Nous nous intéressons plus particulièrement à des agents appelés "agents de prédiction" dont le rôle est d'évaluer la « situation courante », en la rapprochant de situations passées, en vue de détecter un éventuel risque pouvant engendrer une crise. Cette évaluation s'appuie sur des techniques de type « raisonnement à partir de cas » permettant de comparer des paysages d'agents.

Nous nous orientons actuellement vers l'utilisation d'une organisation d'agents aspectuels pour représenter la situation courante.

3.2 Simulation discrète des systèmes complexes

La simulation de systèmes complexes naturels peut conduire à gérer un grand nombre d'entités en interaction et communication. Une telle simulation implémentée dans un système informatique distribué à grande échelle, nécessite des stratégies de placements dynamiques pertinents, ainsi qu'une fiabilisation des processus de répartition. Ces systèmes complexes sont également amenés à s'auto-organiser dynamiquement sur différents niveaux d'échelles et nous recherchons des processus génériques pour de telles simulations en les appliquant aux écosystèmes estuariens.

Placement dynamique et fiabilisation de systèmes multi-agents distribués

Contact : D. Olivier

Ce projet se situe dans le cadre d'une collaboration entre le Laboratoire d'Informatique du Havre et le Laboratoire d'Informatique de Paris 6, sous la responsabilité d'Alain Cardon.

Le problème abordé est celui de la distribution physique de systèmes multi-agents de grandes échelles et constitués de nombreuses entités. Nous traitons cette résolution en abordant les deux problèmes complémentaires que sont :

- la fiabilisation en cours d'exécution qui s'appuie sur l'utilisation de répliques assignées dynamiquement et de manière transparente à l'application avec un contrôle de cohérence de l'ensemble de ces répliques. Ce travail a donné lieu au démarrage de la thèse d'Olivier Marin fin 2000, financée par une bourse MNESR.
- le placement dynamique se fait par un module d'observation qui récupère les données significatives en provenance des systèmes multi-agents, telles que la charge des différentes machines ou les caractéristiques du réseau. Ces informations sont ensuite exploitées par un module d'exécution qui fournit des services d'équilibrage de charge, de manière transparente, aux applications.

Simulation d'écosystèmes estuariens par des SMA multi-échelles

Contact : C. Bertelle

Cette action de recherche s'inscrit dans l'apport des méthodes d'IAD pour modéliser des problèmes de nature physique.

Nous nous intéressons à la mise en place progressive de simulations multi-échelles d'écosystèmes estuariens dont le fluide est à la fois la composante principale de transport et la source de formation de structures organisationnelles en phase de complexification évolutive qui est provoquée par les flux énergétiques traversant le milieu.

Nous utilisons pour cela, des modèles d'écoulement fluide fondés sur une approche distribuée et centrée sur les individus représentant des particules fluides virtuelles chargées de tourbillons et en interaction mutuelle. Ces

méthodes sont connues sous le nom de méthodes particulières ou vortex.

Nos objectifs sont, à terme, de simuler un milieu estuarien aquatique dans toute sa complexité organisationnelle et constitutive en prenant en compte des entités hétérogènes de nature physico-chimique ou organique pour des études écotoxicologique, bactériologique ou halieutique.

Nous nous intéressons à la mise en place d'un modèle conceptuel reposant sur une description multi-niveaux d'un système complexe. Des niveaux organisationnels sont chargés de capter des organisations dynamiques dans des écoulements fluides ainsi que de forcer la constitution d'organisations calculatoires permettant une distribution des entités sur un réseau de machines afin d'envisager des simulations à grande échelle.

Notre travail a donné lieu au démarrage de la thèse de Pierrick Tranouez, financée par la Région Haute-Normandie et portant sur la mise en place de processus automatiques de changements d'échelles dans des systèmes complexes.

3.3 Génétique dans les SMA

Nous mettons en oeuvre des algorithmes génétiques dans des systèmes d'agents dans le cadre de la description de comportements évolutifs par automates à multiplicités. Nous nous intéressons également à une utilisation dynamique de processus génétiques pour aborder des problèmes de planification.

Automates évolutifs de comportement d'agents

Contact : C. Bertelle, M. Flouret

Nous nous intéressons à l'utilisation d'automates pour représenter d'une manière générale, le comportement des agents d'un SMA. Ils permettent de mettre en place des processus automatiques et évolutifs de comportements dynamiques. En particulier, les propriétés des automates à multiplicités et les opérations rationnelles qui peuvent leur

être associées permettent de prendre part dans une formalisation efficace du comportement d'agents ainsi que de leurs interactions. Nous avons alors défini une mesure comportementale liée à la représentation matricielle de ces automates. Nous appliquons des opérateurs génétiques sur des automates à multiplicités pris sous forme probabiliste pour forcer la constitution automatique d'organisations qui rétro-agissent ainsi sur les individus qui les composent, grâce à un algorithme génétique. Ce procédé permet ainsi d'obtenir une approche calculatoire de certains systèmes complexes adaptatifs qui peut notamment s'appliquer à des modèles d'auto-organisation dans les écosystèmes.

Algorithmes génétiques dans des SMA pour la planification

Contact : A. Cardon, T. Galinho

Les travaux de thèse de Jean-Philippe Vacher portent sur l'utilisation des algorithmes génétiques pour l'optimisation d'un Gantt selon un modèle multiagent. Le problème de l'optimisation d'un planning dans un système Job-Shop (diagramme de Gantt) est un problème NP-complet. Une façon d'approcher sa solution consiste en la détermination d'un planning initial amélioré selon différentes techniques. Nous utilisons pour cela les algorithmes génétiques avec fonction de fitness multi-objectifs. De plus, nous modélisons les constituants du problème selon le paradigme agent, pour déployer les AG selon différents niveaux de granularité. Nous obtenons ainsi une application dynamique des AG, avec usage de la reproduction et du crossing-over dans le paysage des agents.

3.4 Multimodalité

Contact : D. Archambault, D. Olivier

Cette action de recherche s'appuie en grande partie sur les travaux que nous développons dans le cadre du projet européen TIM (cité en introduction). Les points abordés sont en particulier les modèles de représentation de l'information indépendants des modalités. Il s'agit de définir des interfaces multimodales non pas dirigées uniquement par

les besoins liés à l'application mais aussi par les besoins liés à l'utilisateur. Citons pour exemple, les modalités comme le braille, la synthèse vocale ou les interfaces tactiles, il s'agit là d'interfaces non visuelles qui doivent pouvoir cohabiter avec des interfaces visuelles. Le problème consiste donc à proposer un modèle de représentation indépendant des données et qui permet de construire l'interface avec l'utilisateur de différentes manières et de façon dynamique. En effet, l'adaptation de l'interface ne doit pas se faire uniquement en fonction des modalités mais également en fonction des compétences de l'utilisateur puisque nous nous inscrivons dans une démarche ludique et/ou éducative avec éventuellement un public polyhandicapé.

3.5 Spécification formelle de SMA

Contact : B. Mermet

Des travaux récents liés à la spécification formelle dans les SMA, à l'aide du langage formel B sont actuellement en cours de réalisation. L'intérêt de ces méthodes repose sur le fait que les obligations de preuve sont clairement définies et qu'elles s'appuient sur des supports de qualité industriels. Nous avons notamment présenté la vérification de propriétés temporelles en B dans le cadre du thème STS du GDR-ARP.

4 Collaborations

Nous avons développé des collaborations avec différentes équipes du Laboratoire d'Informatique de Paris 6, notamment les équipes OASIS et SRC pour le placement dynamique et la fiabilisation des SMA.

Nous développons également des échanges avec des laboratoires Européens pour ce qui concerne les approches multi-modales des interfaces :

- CRS4 (Center for Advanced Studies, Research and Development in Sardinia) qui a accueilli en 2001, un de nos étudiants (Luc Hogie) en stage de DEA ;
- INSERM U483/INOVA ;
- NCFL Halmstad University (Nordic

PRÉSENTATION DE LABORATOIRES

Center for Research on Toys and Educational Media, School of Information Science and Computer Engineering, Suède) ;

- Sunderland University (UK) ;
- Tomtebodaskolans Resurcenter (Suède).

5 Sélection de publications récentes relatives aux actions décrites

Archambault D. et Burger D. « From modality to multimodalities : the need for independent models », Proc. Universal Access in Human-Computer Interaction, New-Orleans, Louisiana, USA, Août 2001. Lawrence Erlbaum Associates.

Archambault D., Olivier D. et al. « TIM : Tactile Interactive Multimedia computer games for visually impaired children », Information Society Technologies, réf. IST-2000-25298, Mai 2000. Total EC contribution 1540 K-euro, début : Janvier 2001, durée : 36 mois.

Bertelle C., Flouret M., Jay V., Olivier D. et Ponty JL. « Automata with multiplicities as behaviour model in multi-agent simulations », SCI2001, Orlando, Floride, USA, 22-25 Juillet 2001.

Bertelle C., Flouret M., Jay V., Olivier D. et Ponty JL. « Genetic algorithms on automata with multiplicities for adaptive agent behaviour in emergent organizations », SCI2001, Orlando, Floride, USA, 22-25 Juillet 2001.

Bertelle C., Olivier D., Jay V., Tranouez P. et

Cardon A. « A multi-agent system integrating vortex methods for fluid flow computation », 16th IMACS Congress 2000, Lausanne, Switzerland, 21-25 Août 2000.

Blazewicz J., Guinand F., Trystram D. et Penz B. « Scheduling complete trees on two uniform processors with any integer speed ratios with communication delay », to appear in Parallel Processing Letter, 2001

Boukachour H., Coletta M., Galinho T., Person P. et Simon G. « Preventive monitoring multiagent system », SCI2001, Orlando, Floride, USA, 22-25 Juillet 2001.

Cailliau D., Leger A., Marin O. et Folliot B. « A joint middleware /configuration approach for space embedded software update », DASIA'2001, Nice, France, 28 Mai - 1 Juin 2001.

Cardon A. « La modélisation des systèmes adaptatifs : caractères d'une modélisation non locale », Revue d'Intelligence Artificielle, Hermès, 13(2):169-200, 1999.

Cardon A. « Système de gestion de crises coopératif : un processus d'interprétation de points de vue multiples », Journal of Decision Systems, 7:39-67, 1999.

Cardon A. « Conscience artificielle et systèmes adaptatifs », Eyrolles, Paris, France, 2000.

Cardon A. et Vacher JP. « Genetic algorithm using multi-objective in a multi-agent system », Robotic and Autonomous Systems, Elsevier, pp 179-190, 2001.

Durand S. « Représentation de points de vue multiples dans la couche communicationnelle d'un Système d'Information et de Communication », Thèse de l'Université du Havre, Décembre 1999.

Duvallet C., Boukachour H. et Cardon A. « Intelligent and self-adaptive interface », IEA/AIE'2000, LNCS Vol. 1821, Springer-Verlag, 2000.

Duvallet C., Sadeg B. et Cardon A. « Real-time in multi-agent systems », Int. Conf. On Computers and their Applications (CAINE'99), pp 212-215, Atlanta, Georgia, USA, Q. Yang Ed., 1999.

Duvallet C., Sadeg B. et Cardon A. « An anytime multiagent system to manage electronic commerce transactions », OOIS-2000, Springer-Verlag, London, England, 2000.

Guinand F. et Trystram D. « Optimal scheduling of uect tress on two processors », RAIRO Recherche Opérationnelle/Operations Research, (34):131-144, 2000.

Lesage F. « Interprétation adaptative du discours dans une situation d'urgence multiparticipants : modélisation par agents », Thèse de l'Université du Havre, Octobre 2000.

Olivier D., Jay V. et Bertelle C. « Distributed multi-agent system used for dynamic aquatic simulation », ESS'2000 Congress, pp 504-508, Hambourg, Germany, 28-30 Septembre 2000.

Tranouez P., Bertelle C. et Olivier D. « Changing the levels of description of a fluid flow in an agent-based simulation », ESS'2001 Congress, Marseille, France, 18-20 Octobre 2001.

Vacher JP. « Utilisation des algorithmes génétiques pour l'optimisation d'un Gantt selon un modèle multiagent », Thèse de l'Université du Havre, Juillet 2000.

PRÉSENTATION DE LABORATOIRES

Présentations de laboratoires dans le bulletin de l'AFIA

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

L'extraction de connaissances, fouille de données et intelligence artificielle

Coordonneurs.

Yves Kodratoff, LRI ORSAY, yk@lri.fr,
Amedeo Napoli, LORIA Nancy, napoli@loria.fr,
Djamel Zighed, ERIC Lyon, zighed@univ-lyon2.fr.

Objectifs

Depuis une décennie, l'extraction de connaissances dans les bases de données (ECBD) a pris une ampleur considérable. C'est une activité qui consiste à analyser un ensemble de données brutes de façon à en extraire des connaissances exploitables. Un expert du domaine des données, l'analyste, est chargé de diriger l'extraction de connaissances. En fonction de ses objectifs, l'analyste va sélectionner des données et utiliser des outils de fouille de données pour construire des modèles expliquant les données. Pour mener à bien son activité, l'analyste met à contribution ses connaissances mais aussi un ensemble d'outils regroupés au sein d'un système d'ECBD.

Un système d'ECBD s'articule généralement autour de quatre composantes principales :

1. les bases de données et leurs systèmes de gestion,
2. un système à base de connaissances pour la gestion des connaissances et la résolution de problèmes sur le domaine relatif aux données,
3. des modules de fouille de données qui s'appuient sur des techniques symboliques (induction, classification par treillis et par arbres de décision, extraction de motifs, etc.), ou numériques (analyse de données, classification automatique, statistiques, modèles probabilistes, etc.),
4. une interface se chargeant des interactions et de la visualisation des résultats intermédiaires et finaux.

Actuellement, l'ECBD s'applique à des données de toutes natures et des domaines très divers au premiers rangs desquels se trouvent le Web ou les textes.

L'ECBD est donc par essence pluridisciplinaire, et l'objectif de ce dossier de mieux faire connaître les personnes qui travaillent autour de l'ECBD en

France, et les caractéristiques de leurs recherches, sachant qu'un bon nombre de ces personnes ne sont pas (forcément) directement associées à la communauté IA.

Un certain nombre de laboratoires ont répondu à l'appel des trois coordonneurs, et le dossier est très certainement incomplet encore à bien des égards. Toutefois, les lecteurs du bulletin désirant avoir d'autres informations en dehors de celles qu'ils trouveront sur les laboratoires qui suivant pourront consulter les sites suivants :

- le site [KDNuggets](http://www.kdnuggets.com/) qui est très complet ; un abonnement à une mailing liste est possible, et des messages de synthèse sont envoyés tous les mois, où paraissent toutes les nouvelles les plus importantes dans la communauté, et selon tous les points de vue : <http://www.kdnuggets.com/> ; mentionnons également 1 site du groupe de l'ACM d'intérêt sur la fouille de données : <http://www.acm.org/sigkdd/>,
- le site de la toute nouvelle action spécifique du CNRS (département STIC) Fouille de bases de données - groupe GaFoDonnees -, qui regroupe les chercheurs de la communauté I3 (d'horizons différents et complémentaires) : <http://www.grappa.univ-lille3.fr/~torre/gafo/index.php3>
- il existe aussi un autre annuaire associé à la base CAFE qui regroupe encore des chercheurs de différents domaines qui travaillent sur la fouille de données : <http://hydria.u-strasbg.fr/CAFE/annu.html>.

Pour finir, les coordonneurs aimeraient remercier les laboratoires qui ont bien voulu répondre à cet appel et ils espèrent que les lecteurs du bulletin de l'AFIA pourront faire bon usage de ce dossier.

CID : Connaissances, Informations et Données

(IRIN - Université de Nantes
2, rue de la Houssinière - BP 92208
44322 Nantes Cedex 03

URL :

<http://www.sciences.univ-nantes.fr/irin/>

Responsable :

Henri Briand (Professeur)

Membres permanents :

Claude Belleil (MC), Maurice Bernadet (MC), Sylvie Cazalens (MC), Emmanuel Desmontils (MC), Régis Gras (Professeur), Fabrice Guillet (MC), Mounira Harzallah (MC), Christine Jacquin (MC), Pascale Kuntz (MC), Philippe Lamarre (MC), Michel Leclère (MC), Maurice Liscouët (MC), Abdelouahab Mekaouche (MC), Gilles Nachouki (MC), Marie Pierre Nachouki (MC), Philippe Peter (MC), Jacques Philippé (MC), Gérard Ramstein (MC), Francky Trichet (MC)

Principaux axes de recherche :

Les activités de l'équipe CID se décomposent en 7 axes regroupés en 2 thèmes :

- Extraction et Gestion des Connaissances
- Ingénierie des Connaissances

- Thème Extraction et Gestion des Connaissances (EGC)

Les activités de recherche de ce thème s'articulent autour de deux grands axes principaux :

1. Les algorithmes d'extraction de la connaissance à partir des données, des textes, des images, ou de toile, ainsi que les mesures de qualité de la connaissance extraite.
2. La visualisation de la connaissance. Conjointement à ces axes qui fédèrent et prolongent nos activités initiales dans le domaine de l'Extraction de

Connaissances dans les Données (ECD), deux nouveaux axes ont émergé ces deux dernières années:

3. L'ECD pour la bioinformatique
4. La gestion des connaissances

L'axe 1 se focalise sur plusieurs maillons de la chaîne de traitement "des données aux connaissances" classiquement adoptée. En particulier, nous développons des algorithmes de conception de datawarehouse et, différentes techniques de fouille de données (fouille interactive, parallélisation des mécanismes d'ECD, extraction de connaissances floues, classification automatique). Nous définissons également des mesures de qualité de règles pour des données variées (discrètes, ordinales, floues) et développons autour de ces mesures des techniques d'élagage.

P. Kuntz, F. Guillet, R. Lehn, H. Briand. Vers un processus d'extraction de règles d'association centre sur l'utilisateur. In *Cognito*, Revue Francophone internationale en Sciences Cognitives. A paraître.

P. Kuntz, F. Guillet, R. Lehn and H. Briand. A User-Driven Process for Mining Association Rules. In Proc. of Fourth European Conference on Principles of Data Mining and Knowledge Discovery, PKDD'2000. Lecture Notes in Computer Science, vol. 1910, pages 483-489, 2000. Springer.

T. Teusan, G. Nachoucki, H. Briand, et J. Philippé. Discovering association rules in large, dense databases. In Proceedings of Principles of Data Mining and Knowledge Discovery, PKDD'2000. Lecture Notes in Computer Science, vol. 1910, pages 581-586, 2000. Springer.

S. Guillaume, F. Guillet, et J. Philippé. Improving the discovery of association rules with intensity of implication. In Proc. of the 2 European symposium on principles of data mining and knowledge discovery, PKDD'98, pages 318-327, Nantes, France, 1998. Springer.

M. BERNARDET. Basis of a fuzzy knowledge discovery system. In Proceedings of Principles of Data Mining and Knowledge Discovery, pages 24-33. Springer, 2000.

L'axe 2 porte principalement sur la visualisation de réseaux de relations modélisés par des graphes et l'intégration de relations combinatoires et sémantiques dans des cartographies statiques et dynamiques de données extraites du Web ou de règles d'association.

F. Guillet, R. Lehn, P. Kuntz. A Genetic Algorithm for Visualizing Networks of Association Rules. In Proc. of the 12th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, IEA/AIE-99. Lecture Notes in Computer Science, vol. 1611, pages 145-154, 1999. Springer.

P. KUNTZ, R. LEHN, et H. BRIAND. Dynamic rule graph drawing by genetic search. In Proceedings of the IEEE International Conference on System Man and Cybernetics, 2000.

F. VELIN, P. KUNTZ, C. BELLEIL, et H. BRIAND. Integrating semantic and combinatorial relationships in a cartography for web site promotion. In Session Lecture, 24th Annual Conference of the German Gesellschaft für Klassifikation e.V. (GfKI 2000), Passau, Germany, March 15-17 2000.

L'axe 3 vise à adapter des algorithmes d'extraction de connaissances à des données pour des données de biologie moléculaire et d'odontologie.

G. RAMSTEIN, P. BUNELLE, et Y. JACQUES. Discovery of ambiguous patterns in sequences. application to bioinformatics. In Proceedings of Principles of Data Mining and Knowledge Discovery, pages 581-586. Springer, 2000.

L'axe 4 porte sur le recueil et la capitalisation de connaissances, la diffusion de connaissances et la formation et la stimulation de l'innovation par réutilisation/transposition des connaissances.

H. Briand et F. Guillet (eds). Extraction et Gestion des Connaissances. Revue Extraction des Connaissances et Apprentissage (ECA), vol. 1, n° 1-2, 2001. Hermes Science Publication.

F. Guillet, P. Vandekerckhove. Samanta : a Knowledge Server. In J.-L. Ermine (editor), Workshop « Knowledge Management - Theory and Application » in conjunction with

PKDD'2000, pages 19-28, Lyon, France, September 12 2000.

- Thème Ingénierie des Connaissances (IC)

L'ingénierie des connaissances est une thématique de recherche qui vise à proposer des concepts, méthodes et techniques permettant de d'acquérir, de modéliser et de formaliser des connaissances pour les mobiliser dans l'activité d'une organisation. Ce thème se décline en 3 axes :

5. Approche multi-experts pour l'informatique documentaire ;
6. Ingénierie ontologique et représentation de connaissances ;
7. Systèmes multi-agents et gestion de connaissances distribuées ;

L'axe 5 concerne le travail réalisé sur la construction hybride, intégrant des descriptions explicites (règles) et implicites (réseaux de neurones), d'une architecture distribuée fondée sur le modèle du tableau noir.

Mékaouche Abdelouahab, « The Cooperative System for Information Retrieval ». The IEEE Computer Society Press in the proceedings of the XX International Conference of the Chilean Computer Science Society (SCCC'2000), 16-18 November 2000 in Santiago, Chile. ISBN 0-7695-0810-3.

Mékaouche Abdelouahab, « Integrating of Learning Mechanisms for Information Retrieval ». The 2000 International Conference on Artificial Intelligence (IC-AI'2000), June 26-29, 2000, Monte Carlo Resort, Las Vegas, Nevada, USA.

L'axe 6 vise à proposer un environnement informatique dédié à l'opérationnalisation d'ontologies de domaine. L'étude et le développement de cet environnement sont fondés sur des extensions du modèle des Graphes Conceptuels.

Leclère M., Trichet F., Furst F. (2002). « Formalisation de la géométrie projective : vers un langage unifié d'ontologisation », A paraître dans les actes de la 13ième Conférence Francophone intitulée « Reconnaissance des Formes et Intelligence Artificielle » (RFIA'2002), Angers.

Trichet F., Leclère M., Tixier B. (2001). « Using a Task/Method Knowledge-Based System to capitalize and share know and know-how », Proceedings of the Workshop on Knowledge Management and Organizational Memories, pp 94-102, in conjunction with the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI'2001). Seattle, USA.

L'axe 7 vise d'une part à proposer des agents logiciels pour l'aide à la gestion d'informations et de connaissances (création d'index sémantique structuré d'un site au moyen d'ontologies). D'autre part, des études sont menées sur les systèmes multi-agents permettant de gérer les connaissances distribuées.

Cazalens S., Lamarre P. (2001). « An organization of Internet agents based on a hierarchy of information domains », Demazeau Y et Garjjo F.J. Eds. Proceedings MAAMAW 2001.

Desmontils E., Jacquin C. (2001). « Indexing a web site with a terminology oriented ontology ». Proceedings of the 1st Semantic Web Working Symposium (SWWS 2001).

Cazalens S., Lamarre P., Desmontils E., Jacquin C. (2002). « Sources d'informations et de connaissances: de la gestion locale à la recherche distribuée ». Revue L'objet, numéro spécial « Connaissances et Distribution ». A paraître en 2002.

**Centre de recherche
en informatique de Lens,
CRIL, Université d'Artois**

Contact : Engelbert Mephu Nguifo, CRIL, rue de l'Université SP 16, 62307 Lens Cedex

Tél. 03 21 79 32 73

Fax : 03 21 79 32 72

Mél : mephu@cril.univ-artois.fr

<http://www.cril.univ-artois.fr>

Le CRIL (directeur : Prof Eric Grégoire) est un laboratoire de l'Université d'Artois qui regroupe une vingtaine d'enseignants-chercheurs et thésards, dont l'activité est centrée autour de l'intelligence artificielle symbolique et ses

applications. L'extraction de connaissances dans les bases de données (ECBD) ou fouille de données est un champ pluridisciplinaire dans lequel s'intègre parfaitement quelques uns de nos travaux en apprentissage automatique.

Participants : Huaiguo Fu, Engelbert Mephu Nguifo

Thèmes de Recherche :

Nos recherches sur la fouille de données sont centrées sur la structure de treillis de concepts et concernent principalement 2 aspects :

- Redescription de données.

Dans ce travail, nous avons développé une méthode de transformation de descripteurs symboliques en descripteurs numériques afin de faciliter l'usage des mesures de similarité comme élément de calcul de distance entre objets. Cette redescription de données symboliques en données numériques, est basée sur l'usage de la structure d'un treillis de concept pour prendre en compte les corrélations existantes dans les descripteurs initiaux au travers des objets qu'ils décrivent.

- Classification automatique par treillis.

Nous nous appuyons sur la structure de treillis de concepts pour construire des outils de classification supervisée. Ces outils partent d'une description attributs/valeur et utilise le principe de la correspondance de Galois pour structurer la connaissance apprise. Trois outils ont été développés : LEGAL, IGLUE et CIBLE. LEGAL utilise un raisonnement empirique majoritaire. Il a été évalué sur des problèmes de biologie et d'archéologie. IGLUE est un outil multi-stratégique qui combine un mécanisme d'induction et une technique de plus proche voisin. CIBLE est une généralisation d'IGLUE sur des données symboliques/numériques.

Nous avons co-organisé un atelier de recherche sur l'apport des Treillis de Concepts à l'ECBD à durant la conférence ICCS (Conceptual Structures) qui a eu lieu à Stanford (Californie) du 30 juillet au 4 août 2001. Les actes sont disponibles sur le site web suivant: <http://CEUR-WS.org/Vol-42/>

Projets en cours :

- Développer un mécanisme d'extraction, d'explication et de visualisation de motifs fréquents (associations, dépendances, ...) à partir d'un treillis de concepts. L'idée consiste à utiliser des techniques de logique floue pour construire un treillis partiel de concepts approximatifs, et en générer les motifs fréquents.
- Fouille de données autour de données bancaires en collaboration avec une entreprise régionale.

Bibliographie :

Mephu Nguifo E. & Njiwoua P., 1998, « Using Lattice-based Framework as a Tool for Feature Extraction », in European Conference on Machine Learning (ECML), LNAI 1398, 304-309, Chemnitz (Germany), April.

Njiwoua P. & Mephu Nguifo E., 1999, « Améliorer l'apprentissage à partir d'instances grâce à l'induction de concepts : le système CIBLE », dans Revue d'Intelligence Artificielle (RIA), 13(2):413-440, New Orleans, November 6-9th, IEEE Press.

Mephu Nguifo E. & Njiwoua P., 2000, « IGLUE : A lattice-based constructive induction system », in Intl. Journal of Intelligent Data Analysis, 4:1-19, IOS Press. Mephu Nguifo E. & al. Eds., 2001, Proc. of ICCS workshop on Concept Lattices for KDD, Stanford University (Palo-Alto, California), July 30th.

**EDF Recherche et
Développement - Groupe
Statistique, Optimisation,
Aide à la Décision**

1, Av. du Général de Gaulle,

92141 CLAMART Cedex

Téléphone secrétariat : 01 47 65 41 03

Ingénieurs-chercheurs concernés :

G.Hébrail (responsable du groupe), C.Derquenne, P.Futtersack, H.Klajnmic, P.Lé, R.Palmieri, A.Peradotto, M.L.Picard, L.Pierre, V.Stéphan.

Thésards : N.Fischer (CNAM), B.Huguency (LIP6).

Principales collaborations universitaires : CNAM, LIP6, INRIA

Thème « Data mining sur les bases de données clientèle »

En 1998, une approche a été mise au point pour enrichir les bases de données clientèle d'EDF, c'est-à-dire compléter

certaines caractéristiques manquantes des clients, dans un objectif de marketing (voir [1], [2]). La méthode employée est la régression logistique (voir [3]). En 2001, l'approche a été généralisée à l'ensemble des 25 millions de clients résidentiels. Les recherches en cours portent toujours sur l'enrichissement des bases de données clientèle, mais par des informations externes (données INSEE, enquêtes sur les clients). Une thèse est en cours sur une méthode de fusion statistique des données (voir [4]). Une autre piste que nous commençons à explorer est celle du data mining 'spatial', dans une optique de géomarketing. A noter également une activité sur l'analyse statistique de la fréquentation de sites internet d'EDF (« Web usage Mining »).

Thème « Analyse des courbes de consommation d'électricité »

L'historique des courbes de consommation d'électricité constitue une information riche et volumineuse sur les grands clients (voir [1]). En 1997, nous avons développé un logiciel pour la classification automatique de courbes de consommation : « Le Courboscope » (voir [5]). Ce logiciel est basé sur la méthode des cartes de Kohonen, il favorise l'interactivité avec l'analyste, et peut s'appliquer à n'importe quel type de courbes. Les recherches en cours portent sur l'analyse symbolique de courbes (voir [1] et [6]). L'objectif est double : (1) construire des représentations synthétiques de longues courbes (1 année de consommation), (2) construire des courbes en prédiction sur de longs horizons (1 année), par apprentissage à partir du passé et ajout de connaissances d'expert.

Thème « Statistique et entrepôts de données »

Les entrepôts de données, en particulier lorsqu'ils sont structurés sous forme de bases de données multidimensionnelles, offrent des possibilités de fouille de données par la production de tableaux croisés dynamiques. De nombreux problèmes se posent alors, auxquels la statistique apporte des solutions : la vérification, la détection et l'explication des erreurs dans les données, la représentativité des données par rapport

aux requêtes lorsque des informations sont manquantes, la qualité de la réponse aux requêtes, ... Une thèse est en cours de lancement sur ce sujet. Elle apportera une contribution au data mining sur deux aspects : (1) le nettoyage des données, (2) la fouille de données par production de tableaux croisés dynamiques pertinents et fiables.

Bibliographie

- [1] « Practical data mining in a large utility company », G.Hébrail, Congrès Compstat 2000, Utrecht, Août 2000.
 [2] « Data mining techniques for mass personalization within a large utility company », M.L.Picard, S.Nugier, World Congress on mass personalization and customization, Hong Kong, Octobre 2001.
 [3] « Mise en œuvre d'une démarche statistique complète pour la prédiction de variables dans une base de données clientèle d'EDF », C.Derquenne, Journées industrielles Modulad EDF, Clamart, Novembre 2000.
 [4] « A multivariate approach to combine data from different sources: application to customers knowledge of Electricité de France », C.Derquenne, N.Fischer, 53^{ème} Congrès de l'International Statistical Institute, Séoul, Août 2001.
 [5] « Interactive Interpretation of Kohonen Maps Applied To Curves », A.Debrégeas, G.Hébrail, Congrès KDD 98, New-York, Août 1998.
 [6] « Symbolic representation of long time series », G.Hébrail, B.Huguency, Congrès Applied Stochastic Models and Data Analysis (ASMDA'2001), Compiègne, Juin 2001.

Extraction et Exploitation de l'Information en Environnements Incertains (E3I2, ENSIETA)

2, Rue François Verny,
 29806 Brest Cedex 9, FRANCE,
 Tel. 33 (0)2 98 34 88 45,
 Fax 33 (0)2 98 34 87 50,
 Email Ali.Khenchaf@ensieta.fr

La maîtrise de l'action ne peut plus s'envisager sans la maîtrise de l'information et cette synergie apparaît fortement dans tous les processus décisionnels. Aujourd'hui de grandes masses de données structurées ou semi-structurées sont accessibles. Aussi les opérationnels ont-ils besoin de méthodes et d'outils

capables d'extraire les connaissances pertinentes pour les décideurs, de les stocker, et de les diffuser à d'autres acteurs.

Problématique

- Systèmes d'information liés aux données environnementales,
- Connaissances incertaines, imprécises ou incomplètes des senseurs,
- Aide à la décision pour les opérateurs. Domaines d'application
- Acoustique sous-marine : signaux transitoires, tomographie passive, cartographie, ...
- Communications : détection, localisation, soutien aux analystes du renseignement, ...
- Radar : télédétection, radionavigation, identification, surveillance, systèmes de renseignement à capteurs multiples, ...

Thématiques

1. Modélisation et Caractérisation de l'Environnement
2. Représentation et Extraction de l'Information
3. Imagerie (Sonar et Radar)
4. Aide à la décision
5. Expérimentation Systèmes Sonar, Radar

Spécificités

- Modéliser, Simuler, Extraire, Exploiter, Valider,
- Partenariat industriel fort,
- Equipe mixte.

Collaborations

- UNIVERSITAIRES
 UBO, Univ. Rennes 1, Polytech'Nantes, ENST Bretagne, ISEB, Ecole Navale
 Ecole Doctorale Sciences de la Matière, de l'Information et du Vivant (ED SMIV) à l'UBO.
- INDUSTRIELLES ET ETATIQUES
 Thales Underwater Systems, Oceano Technologies, ORCA, Coopagri Bretagne, IFREMER, SHOM, GESMA, CELAR, CMO, Thales, Sodena, BouyguesTelecom, ...
- INTERNATIONALES
 Université de Beyrouth (Liban), Université de Faro (Portugal), Université de Victoria (Canada),

- Université d'Edimbourg (Ecosse), Université de Hambourg (Allemagne), Université de Minsk (Biélorussie), Académie Technique Militaire de Bucarest, Université de Naples, ...
- AUTRES
 - GDR ISIS (Information, Signal, Images, viSion),
 - G2RA (Groupe Régional de Recherche en Acoustique sous marine).
 - Effectif permanent
 - 2 HDR
 - 6 MC ou ingénieur de Recherche
 - 1-2 IEPA
 - 5-6 doctorants
 - Publications / Contrats
 - 30 Revues
 - 80 Communications
 - 8 ouvrages scientifiques
 - 10 contrats réalisées et 6 en cours

Equipe de Recherche en Ingénierie des Connaissances ERIC-Lyon 2

5, avenue Pierre-Mendès France,
69676 Bron cedex
téléphone : +33 (0)4 78 77 31 54
Télécopie : +33 (0)4 78 77 23 75
e-mail : eric@univ-lyon2.fr
Site web : http://eric.univ-lyon2.fr

Chercheurs et ingénieurs concernés

Professeurs des Universités :

Abdelkader Zighed (directeur), Serge Miguet (directeur adjoint), Nicolas Nicoloyannis.

Maîtres de Conférences :

Fadila Bentayeb, Omar Boussaid, Jean-Hugues Chauchat, Jérôme Darmont, Stéphane Lallich, Sabine Rabaséda, Ricco Rakotomalala, David Sarrut, Laure Tougne, Jacques Viallaneix, Anne Viallefont.

Doctorants et post-doctorants :

Céline Agier, Laurent Baumes, Sébastien Clippe, Jeremy Clech, David Cœurjolly, Rajai Mourid El Dajani, Serge Di Palma, Franck Favetta, Radwan Jalam, Pierre Jouve, Fabrice Muhlenbach, Marian Scuturici, Sahbi Sidhom, Olivier Teytaud, Tiffany Tweed, Yamina Zellouf

Titre des études menées

Modèles Géométriques et apprentissage
Discrétisation bi-dimensionnelle et arbres de segmentation généralisés
Extraction de règles d'associations : approches mono et polythétique
Echantillonnage
Mesures d'association
Méthodes rapides de sélection de variable en apprentissage supervisé
Organisation et performances des grandes bases de données
Extraction de connaissances à partir de données textuelles
Reconnaissance des formes
Extraction de connaissances à partir de données images
Recherche par le contenu dans les bases de données d'images
Fusion de données
Agréation de prédicteurs en apprentissage
Fouille de textes
Types de méthodes principalement utilisées
Graphes d'induction
Analyse des données
Apprentissage PAC et Agrégation de prédicteurs
Data Warehousing, OLAP
Méthodes statistiques, mesures de similarité

Panorama des travaux effectués ou en cours

Créé en 1995, le laboratoire ERIC a pour objectifs scientifiques le développement de méthodes et d'outils informatiques pour l'ingénierie des connaissances destinés plus particulièrement à l'Extraction automatique des Connaissances à partir des Données.

Les données traitées sont multiformes (textes en langage naturel, images, données numériques quantitatives ou qualitatives...) et peuvent se situer sur des supports structurés comme les bases de données ou non structurés comme le web. Elles peuvent provenir de sources diverses.

Les problèmes auxquels nous sommes confrontés en Extraction des Connaissances à partir des Données couvrent un large spectre qui s'étend du stockage dans les entrepôts de données à l'organisation et l'accès à ce large volume d'information, en passant par les

méthodes de fouille de données (généralement appelées Data mining), jusqu'aux techniques de validation des résultats et de mise en forme des connaissances produites.

Le laboratoire ERIC mène ses recherches dans ce cadre fédérateur. Trois types des travaux sont menés au sein d'ERIC :

- Théorique : Il s'agit de mener des réflexions sur les concepts et les méthodologies qui nous préoccupent. Par exemple, la caractérisation d'un bon espace de représentation en apprentissage, l'étude des propriétés des mesures de similarité sur des images, la validation et/ ou l'agrégation des modèles de prédiction, etc.
 - Logiciel : Il s'agit de développer les outils informatiques issus des travaux théoriques soit dans un but expérimental, s'il faut comparer des méthodes sur des benchmarks, soit dans un but d'application à des cas pratiques. Nous venons de lancer un projet ambitieux, baptisé Smart, dont l'objectif est de proposer un environnement logiciel permettant d'extraire des connaissances à partir des données. Cette plate-forme devra supporter tous types de données, fonctionner sur tous types de système, être accessible comme serveur d'applications en ECD, être évolutive et capable de recevoir de nouveaux programmes, etc.
 - Pratique : Il s'agit de mettre en oeuvre nos idées et d'utiliser dans des cas concrets les logiciels que nous avons développés. Ce volet nous a assuré un contact permanent avec les milieux professionnels : médecine, banques, assurances, industrie, institutions gouvernementales, etc.
- Le laboratoire développe ces activités au sein de groupes de travaux et de projets régionaux et nationaux, et tisse des relations privilégiées avec des universités étrangères, particulièrement au Canada (Université Laval à Québec, Université d'York à Toronto), en Tunisie (Universités de Tunis et de Sfax) ou en Algérie (USTHB, INI à Alger).

Bibliographie

(principales références sur les travaux décrits)

1. BERGERON S., LALLICH S., LE BAS C.,
Location of innovative activities and techno-

logical structure in the french economy, 1985-1990. Some evidences from US patenting, *Research Policy* 26(7-8):733-751 (1998).

2. CONTASSOT-VIVIER S., MIGUET S. A Load-Balanced Algorithm For Parallel Digital Image Warping. *International Journal of Pattern Recognition and Artificial Intelligence*. 13(4):445-463 (1999).
3. DELORME M., MAZOYER J., TOUGNE L. Discrete parabolas and circles on 2D cellular automata, *Theoretical Computer Science* 218, 2:347-417 (1999).
4. ELISSEFF A., PAUGAM-MOISY H. JNN, a Randomized Algorithm for Training Multilayer Networks in Polynomial Time. *NeuroComputing*, 29:3-24 (1999).
5. FESCHET F., MIGUET S., PERROTON L. ParList: a Parallel Data Structure for Dynamic Load Balancing. *JPDC*, 51:114-135 (1998).
6. FESCHET F., TOUGNE L. Discrete Waves on Cellular Automata, *International Journal on Pattern Recognition and Artificial Intelligence*, (à paraître) (2001).
7. MIGUET S., MONTANVERT A., WANG P.S.P. World Scientific. Parallel Image Analysis : tools and models. *Machine Perception* 31, World Scientific, Londres (1998).
8. MIGUET S., PIERSON J.M. Quality and Complexity Bounds of Load Balancing Algorithms for Parallel Image Processing, *International Journal of Pattern Recognition and Artificial Intelligence*, 14(4):463-476 (2000).
9. NICLOYANNIS N., TERRENOIRE M., TOUNISSOUX D. An Optimization model for aggregating preferences : a simulated annealing approach. *Health and System Science* 2(1-2):33-44 (1998).
10. RITSCHARD G., BERSCHTOLD A., DUC F., ZIGHED D.A. (Eds) *Apprentissage : Des principes naturels aux modèles artificiels*. Hermès, Paris (1998).
11. SAGOT C., BRUN J.J., GROSSI J.L., CHAUCHAT J.H., PONGE J.F. Earthworms distribution and humus forms evolution in the sylvigenetic cycle of a subnatural mountain spruce forest, *Eur. J. Soil. Biol.* 35(4):163-169 (1999).
12. SEBBAN M., NOCK R., CHAUCHAT J.H., RAKOTOMALALA R. Impact of Learning Set Quality and Size on Detection Tree Performances : a Comparative Study, *International Journal of Computers, Systems and Signals*, (IJCSS), 1(1):85-105 (2001).
13. ZIGHED D.A., RAKOTOMALALA R. Graphes d'induction - apprentissage et data mining. Hermès, Paris (2000).
14. ZIGHED D.A., KOMOROWSKI J., ZYTKOW J. (Eds) *Principles of Data Mining and Knowledge Discovery*. Lecture Notes in Artificial Intelligence, Springer Verlag, Berlin (2000).
15. ZIGHED D.A., RABASÉDA S., RAKOTO-

MALALA R. FUSINTER : a discretization method of continuous attributes. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6(3):307-326 (1998).

ESIEA Recherche

38, rue des Docteurs Calmette et Guérin, 53000 LAVAL
et 9, rue Vésale, 75005 PARIS
e-mail : recherche@esiea.fr

Chercheurs, Ingénieurs et Doctorants concernés : Mounir Asseraf (C), Mohamed Bendou (D), Anthony Boulestreau (I), Denis Cau (I), Pascal Gandin (I), Lionel Jouffe (C), Nadia Kabachi (C), Sophie Levionnois (I), Paul Munteanu (C), Do Than Nghi (D), François Poulet (C).

Cadre général : ESIEA Recherche a été créé en Septembre 1997. Son principal domaine de compétence est l'Ingénierie de la Connaissance, avec pour thématique privilégiée l'Extraction de Connaissances à partir de Données (ECD). Son activité de recherche, fondamentale et appliquée, porte sur les Réseaux Bayésiens, la Visualisation en ECD et diverses techniques hybrides de découverte de connaissances dédiées à des domaines applicatifs spécifiques.

Panorama des travaux.

Réseaux bayésiens.

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

Notre recherche s'inscrit dans le domaine récent (et peu développé en France) de la recherche de méthodes d'apprentissage spécifiques permettant

de construire automatiquement ou d'affiner des réseaux bayésiens à partir des données.

Nos premiers travaux se sont focalisés sur la conception de nouvelles stratégies de recherche pour l'apprentissage des réseaux bayésiens, moins sensibles aux minima locaux. Ils ont abouti à la conception d'algorithmes performants :

- élaboration de stratégies de recherche mixtes intégrant des idées des approches traditionnelles (fondées d'une part sur l'optimisation d'un score et, d'autre part, sur des tests d'indépendance conditionnelle);
- recherche dans l'espace des ordres topologiques des nœuds;
- recherche dans l'espace des classes d'équivalence des réseaux bayésiens.

Nous travaillons également sur le traitement des variables mixtes (numériques et nominales) et sur l'adaptation des architectures et des algorithmes à des objectifs spécifiques (i.e., l'apprentissage supervisé). Nous nous intéressons aussi à l'apprentissage des réseaux bayésiens dynamiques et, d'une manière générale, à tout ce qui concerne l'élaboration (par apprentissage automatique et/ou par recueil d'expertise) et l'exploitation des réseaux bayésiens.

Visualisation en Extraction de Connaissances dans les Bases de Données.

Les travaux dans ce thème sont menés selon trois directions :

- la fouille visuelle de données (Visual Data Mining), étape du processus d'extraction qui utilise la visualisation comme canal de communication entre l'homme et la machine dans le but de produire de nouvelles connaissances. C'est l'utilisateur qui est au centre du processus et, contrairement à la plupart des autres approches utilisées à l'heure actuelle, l'utilisateur est ici un spécialiste des données traitées et non un spécialiste en data mining. Ce type d'approche a de multiples avantages : prise en compte de l'expertise du domaine tout au long du processus, meilleure compréhension des modèles obtenus et donc une confiance accrue dans les résultats, utilisation des capacités de l'humain en reconnaissance des formes qui améliore les résultats de classification.

- l'amélioration de la compréhensibilité des données et des résultats des algorithmes et l'interactivité dans les méthodes de visualisation. La compréhensibilité des résultats obtenus est un paramètre peu ou pas traité à l'heure actuelle, seule la précision sert à comparer les algorithmes. Or un algorithme aussi précis soit-il ne sera pas utilisé si ses résultats ne sont pas compris. Nous développons donc des outils permettant de transformer les résultats des algorithmes habituels sous forme graphique interactive, compréhensible par tous.
- une étude sur l'utilisation des paradigmes / techniques de réalité virtuelle en ECD. La compréhensibilité des données et l'interactivité peuvent notamment être améliorées par l'utilisation d'un environnement utilisant les techniques et paradigmes dérivés de la réalité virtuelle. Ceci permet, entre autres choses, d'avoir plusieurs représentations différentes des mêmes données (tirant ainsi avantage de chaque représentation) qui sont liées interactivement entre elles.

L'ECD dans la Santé.

Le travail de ESIEA Recherche dans le domaine a démarré dans le contexte du projet Agora-Data (en collaboration avec le CHU de Rennes et la CPAM de Bretagne) qui visait la mise en place et l'exploitation d'un espace de communication d'informations entre les différents acteurs du système de soin en s'appuyant sur les données de la caisse primaire d'Assurance Maladie (CPAM) complétées par un groupe de médecins libéraux. Notre étude était destinée à évaluer l'intérêt des techniques d'ECD pour aider à modéliser, caractériser et utiliser, à des fins de prédiction ou de conseil, les trajectoires des patients (i.e. le parcours des patients à travers le système de soins au cours de leur vie).

Notre contribution actuelle consiste à poursuivre des travaux de recherche et développement autour de la « Modélisation des trajectoires de soin ». Nous proposons l'architecture d'une plate-forme qui s'appuie sur la mise en œuvre conjointe des technologies Multi-Agents et de l'ECD. Son objectif est la confrontation du « réel » (les données d'AM) et de l'"idéal" (consensus d'ex-

perts). Nous proposons d'étudier la modélisation automatique à partir des données d'AM d'une « société d'agents », au sens où l'entend l'Intelligence Artificielle Distribuée, chaque agent possédant un comportement appris au travers des données sous forme d'une méthode probabiliste.

Les avancées de nos travaux nous permettent de mettre en place des collaborations concrètes avec l'assurance maladie : CNAMTS, MSA voient dans ces nouvelles technologies des solutions potentielles à leurs soucis de gestion du risque et d'évaluation de la politique de santé.

Publications récentes.

- L. Jouffe, P. Munteanu, New Search Strategies for Learning Bayesian Networks, Xth Intl Symp. on Applied Stochastic Models and Data Analysis (ASMDA), juin 2001.
- P. Munteanu, Les réseaux bayésiens et leur apprentissage, Journée Apprentissage et Inférence Bayésienne (CAFE-IB), Paris, avril 2001.
- M. Bendou, Classes d'équivalence des réseaux bayésiens, nouveaux résultats théoriques et algorithmes d'apprentissage, Journée Apprentissage et Inférence Bayésienne (CAFE-IB), Paris, avril 2001.
- P. Munteanu, D. Cau, Efficient Learning of Equivalence Classes of Bayesian Networks, 4th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD), p. 96-105, Lyon, sept. 2000.
- L. Jouffe, P. Munteanu, Smart-Greedy+ : Apprentissage hybride de réseaux bayésiens, Colloque francophone sur l'apprentissage (CAP), St. Etienne, juin 2000.
- F.Poulet, « A Virtual Reality Environment for Visual Data Mining », in proc. of 3rd Virtual Reality Intl Conf., Laval, May 2001.
- F.Poulet, « Visualisation des connaissances en ECD », Tutorial EGC'2001, Nantes, Jan.2001.
- F.Poulet, « Improving Graphical Interactions and Result Comprehensibility in Data-Mining », in Intl Symp. on Data Mining and Statistics, Augsburg, Nov.2000.
- F.Poulet, « Visualization in Data-Mining and Knowledge Discovery » in Proc. of Human Centered Processes'99, HCP'99, Sept.1999, Brest.
- M. Asseraf, Comparaison des critères de construction des arbres binaires de décision, XXXIIIe journées de Statistique SFdS, Nantes, Mai 2001.
- M. Asseraf, Extension de la distance de Kolmogorov-Smirnov pour la segmentation et ses propriétés asymptotiques, Extraction et gestion des connaissances (EGC'2001), Vol. 1 - n°1-2/2001

M. Asseraf, Metric on decision trees and optimal partition problem. Intl Conf. on Human System leasing, Proc. of CAPS'3, dec. 2000, Paris, France.

Groupe exploration de données, laboratoire GREYC CNRS UMR 6072

Permanents : K. Zreik, B. Crémilleux, V. Ficet-Cauchard, J. Madelaine

Thésitifs : N. Baupin, B. Cuissart, A. Doucet, N. Durand, D. Fournier, C. Jouvin, C. Turbout

Nos travaux portent sur l'extraction et l'exploitation de connaissances issues de grands volumes de données factuelles ou textuelles pour des tâches liées au clustering et à la classification ainsi que leurs applications entre autres dans les hyper-documents. Pour cela, nous accordons une grande importance aux phases de pré et post-traitements ainsi qu'à la conception de processus interactifs et itératifs.

1- Extraction et utilisation de motifs fréquents dans les bases de données :

- détermination de clusters à partir de représentations condensées (sous forme de fermés) d'ensembles fréquents. Utilisation de ces clusters pour la recherche de facteurs pronostics en cancérologie et pour la navigation dans les hyper-documents.
- utilisation de la plus grande sous-structure commune entre graphes pour la recherche de motifs fréquents dans les molécules chimiques, application à la biodégradabilité.
- caractérisation de classes à l'aide de règles de classification delta-fortes.
- pré-traitement des données et aide au choix pour la complétion des valeurs manquantes.

2- Extraction d'information et documents électroniques :

- exploitation d'informations structurales et syntaxiques (telles que hyper-liens, éléments XML) afin de construire des descripteurs pertinents servant au clustering et à l'élaboration de règles d'associations ou d'épisodes.
- extraction de fonctions de reconnaissances classificatoires issues de docu-

ments numériques pour l'optimisation des processus de numérisation et de reconversion de documents composés.

- étude d'une approche de classification conceptuelle incrémentale pour optimiser la personnalisation de document dynamique multi-plateforme (dans les systèmes d'information mobile).

3- Etude de la qualité des données à partir de l'apprentissage automatique, application aux arbres de décision.

Collaborations :

- Centre de Technologie d'Information (Faculté du Génie Civil et de Géosciences, TUDelft, Pays-Bas)
- groupe do-re-mi (Université d'Helsinki, Finlande)
- équipe data mining et base de données inductives (laboratoire LISI, Lyon)
- France Télécom R & D :
- algorithme de clustering basés sur la recherche de motifs fréquents pour l'anticipation de la connectivité des utilisateurs sur les sites web,
- système de type multi-agents apprenti pour la gestion des agendas des travailleurs nomades.
- ATOFINA : aide à la prédiction de biodégradabilité de composés chimiques.
- société DPCI : Système apprenti pour l'aide à la numérisation de documents composés.
- Centre Anticancéreux François Baclesse, Caen : aide au raffinement de profils pronostics (base de données de l'Organisation Européenne de Traitement du Cancer).
- CHU de Grenoble : apport des techniques d'extraction de connaissances à partir de données pour l'évaluation des pratiques de prévention des thromboses veineuses.

Références.

Boulicaut J. F., Crémilleux B. Delta-strong classification rules for characterizing chemical carcinogens, In: Proceedings of the Predictive Toxicology Challenge for 2000-2001 co-located with the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD 01, Freiburg, Germany, September 2001.
Cuissart B., Touffet F., Crémilleux B., Hébrard J. J. Relevance of Molecular

Similarities Based on the Maximum Common Substructure, In: Proceedings of the Sixth Int. Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP 2001) co-located with the Tenth World Congress on Health and Medical Informatics Medinfo 01, London, UK, September 2001.

Durand N., Crémilleux B., Henry-Amar, M. Discovering Associations in Clinical Data: Application to Search for Prognostic Factors in Hodgkin's Disease, Eighth Conference on Artificial Intelligence in Medicine in Europe (AIME 2001), S. Quaglini, P. Barahona &

S. Andreassen (Eds.), Lecture Notes in Artificial Intelligence, N. 2101, pp. 50-54, Springer-Verlag, Cascais, Portugal, July 2001

K. Zreik. Learning WWW. Actes de CAPS 3. Éditions Europa (Paris), 200 pages, Décembre 2000. ISBN 2-909285-12-X

Special Issue : Cyber Design, K. Zreik & C. Branki Guest Editors, in Knowledge Based System Journal (E. Edmonds, General editor), Elsevier Science Ireland Ltd, 2000.

N. Baupin, K. Zreik: Remote Decision Support System: a distributed information management system, in Knowledge Based System Journal (E. Edmonds, G. editor), Elsevier Science, 2000

Crémilleux B., Gaio M., Madelaine J., Zreik K. Discovering browsing paths on the web, Third International Conference on Human-System Learning, pp. 9-18, Europa Edition, Paris, France, December 2000.

K. Zreik, E-commerce: From designless to design net environment.

DEECE, 14-15 November, 2000. University of Edinburgh, Edinburgh, UK.

Fournier D., Crémilleux B. Using Impurity and Depth for Decision Trees Pruning, Second International ICSC Symposium on Engineering of Intelligent Systems (EIS 2000), pp. 320-326, Paisley, UK, June 2000.

J-P. Goulette, J. Madelaine, K. Zreik, Conception des Nouveaux Systèmes d'Information. Actes de O1Design 99. Décembre 1999, Éditions Europa Productions (Paris), 429 pages, Novembre 1998. ISBN 2-909285-14-6

Ragel A., Crémilleux B. MVC - A Preprocessing Method to Deal with Missing Values, Knowledge-Based Systems 12, pp. 285-291, Elsevier, 1999.

UMR INRA-INAPG – Biométrie et Intelligence Artificielle (BIA)

16 rue Claude Bernard,
75231 Paris cedex 05
Contact : Nicolas Turenne
(turenne@inapg.inra.fr)

Présentation

Dans le cadre du champ thématique « risque et sécurité alimentaire » du département BIA de l'INRA, des techniques d'IA sont imaginées pour subvenir à des besoins en recherche d'information. En particulier je développe une activité data-mining (découverte et extraction de connaissances) pour aider un utilisateur à chercher des relations entre des catégories d'entités biochimiques. Ces relations établies, il faut établir un lien avec un paramètre physiologique. Les données que j'exploite sont de deux natures différentes : base de données relationnelle et corpus textuels bibliographiques.

Les techniques envisagées pour la fouille sont les suivantes : les approches de traitement automatique des langues et les approches d'apprentissage statistique (graphe d'induction, réseaux bayésiens...). En effet, d'une part l'étude des textes requiert une bonne compréhension des mécanismes linguistiques. D'autre part l'apprentissage permet de prendre en compte des connaissances du domaine. Ces disciplines ont souvent évolué de façon autonome, nous essayons de déterminer leur recouvrement, leur intérêt et leurs limites.

Publications

Turenne N., « Apprentissage statistique pour l'extraction de concepts à partir de textes. Application au filtrage d'informations textuelles », thèse de doctorat, Université Louis-Pasteur, Strasbourg, 2000.

Turenne N. « Apprentissage d'un ensemble de concepts préstructuré d'un domaine: l'outil Galex », Mathématiques Informatique et Sciences Humaines n°148 pp41-71 1999 ISSN 0995-2314.

Laboratoire IRCAM/CNRS (Institut de Recherche et de Coordination Acoustique/Musique)

Place Igor Stravinsky, 75004 Paris

Contact : Francis Rousseaux
(francis.rousseau@ircam.fr,
tél. 01 44 78 48 39)

Chercheurs et ingénieurs concernés :
- équipe « représentations musicales »
(Gérard Assayag)

- équipe « perception et cognition musicales » (Stephen McAdams)
- équipe « analyse/synthèse » (Xavier Rodet)
- équipe « services en lignes » (Pascal Mullon)

Problématique ECBD

A l'évidence, la quantité de données sonores et musicales numérisées (sons ou titres) disponibles sur des serveurs de données (des SGBD ou des fichiers accessibles en ligne) est en très forte croissance dans le monde. Ce phénomène change rapidement les métiers et les marchés de la musique, à commencer par la distribution, la production et la composition.

De manière générale, la question de la navigation dans ces grandes quantités de données est posée, l'écoute des sons musicaux ne suffisant plus à orienter les utilisateurs, pas plus d'ailleurs que l'artifice consistant à multiplier les métadonnées renseignées « à la main » dans le but d'indexer ces données. Il devient indispensable d'extraire des descripteurs des contenus musicaux à partir du signal brut, et ce dans le but de disposer de nouveaux outils de navigation et d'édition.

Cette activité de type ECBD étant commise aux différents domaines d'étude, aux besoins des utilisateurs et aux bases de données elles-mêmes, elle est conduite à l'IRCAM/CNRS au fil des projets de recherche. Récemment, un projet européen a été inauguré qui vise deux applications pilote d'ECBD.

Titre de l'étude menée

CUIDADO (Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online) est un projet européen de type IST, coordonné par l'IRCAM, et conduit en partenariat avec les sociétés CreamWare (Allemagne), Sony France, Oracle (Espagne), Artspages International (Norvège) ainsi que les Universités Pompeu Fabra (Espagne) et BenGurion (Israël).

CUIDADO fonctionne en liens étroits avec les groupes de normalisation MPEG7.

Outre la réalisations de modules techniques réutilisables et le développement d'un programme de recherche scienti-

fique, CUIDADO vise deux applications pilote :

- MusicBrowser est un outil de navigation parmi les titres musicaux, destiné notamment au grand public consommateur de musique,
- SoundPalette est un outil d'édition de sons musicaux, destiné notamment aux compositeurs de musique savante.

Types de méthodes principalement utilisées

- méthodes numériques pour l'extraction de « descripteurs de bas niveau » à partir du signal sonore (cf. équipe « analyse/synthèse »),
- méthodes symboliques pour l'extraction de « descripteurs de haut niveau » à partir des « descripteurs de bas niveau » et de « profils utilisateurs » (cf. équipes « perception et cognition musicales » et « représentations musicales »),
- méthodes type « ingénierie des besoins » pour analyser les exigences des utilisateurs et spécifier les interfaces (cf. équipe « services en lignes »).

IRISA, projet AIDA

Contact : Israël-César Lerman, Campus Universitaire de Beaulieu, 35042 RENNES

Notre contribution au domaine se situe au niveau de la Classification et de l'Analyse Ordinale de Données Complexes. Les données peuvent être issues de l'observation expérimentale ou de la connaissance. Elles peuvent avoir une forme numérique, qualitative ou logique.

Les travaux effectués ont été conduits avec la collaboration de nombreux chercheurs dont nous mentionnerons certains ci-dessous.

Dans notre recherche il y a bien sûr des aspects algorithmiques importants liés à la classification et à l'organisation de « gros ensembles ». Ces aspects sont liés à la topologie (dans un sens très général) dont se trouve muni l'espace choisi de représentation des données. Cependant, ce qu'il y a de plus impor-

tant qui a été créé et développé concerne précisément cette topologie. Il s'agit de la notion de SIMILARITÉ entre attributs de types quelconques ou bien, entre objets ou concepts (classes) décrits par de tels attributs. Cette notion habillera les algorithmes qu'il s'agit d'optimiser.

Notre notion de similarité que nous avons appelée de la Vraisemblance du Lien se réfère à une échelle de probabilité. Dans son contexte les attributs de description sont exprimées en termes de RELATIONS (pouvant d'ailleurs être valuées) sur l'ensemble des objets de l'univers concerné.

Signalons à présent les méthodes qui ont été bâties soit directement par nous en collaboration, soit par des chercheurs en forte liaison avec notre méthodologie et notre apport :

- (1) Classification Ascendante Hiérarchique par AVL (Analyse de la Vraisemblance des Liens).
- (2) Analyse Implicative.
- (3) Classification Non Hiérarchique par la méthode des « Pôles d'Attraction ».
- (4) Classification de « gros ensembles » avec ou sans contrainte de contiguïté.
- (5) Introduction des algorithmes génétiques et du calcul parallèle dans la représentation euclidienne des données (« Multidimensional Scalling »).
- (6) Arbres de décision binaires dans le cas d'attributs qualitatifs à très grand nombre de valeurs.
- (7) Lien entre l'analyse combinatoire des données et la résolution des problèmes de complexité à travers #SAT.

Les collaborations mentionnées ci-dessus concernent pour

- (1) F. Nicolău (Université Nouvelle de Lisbonne), M.H. Nicolău (Université de Lisbonne), M. Ouali-Allah (École de Coëtquidan), Ph. Peter (École Polytechnique de l'Université de Nantes), B. Tallur (Université de Rennes1-Irisa, projet Aida).
- (2) R. Gras et Collaborateurs (École Polytechnique de l'Université de Nantes).
- (3) H. Lerredde (Université de Paris 13).
- (4) K. Bachar (ESSCA d'Angers) et Ph. Peter.

- (5) R. Ngouënet (Mathsoft, Seattle, USA).
 (6) J.P. Da Costa (Université de Porto)
 (7) V. Rouat (CELAR, Bruz).

Les méthodes présentées sont généralistes. Elles ont eu un très large champ d'application. Cependant, elles peuvent être accentuées dans le cadre d'une recherche spécifique par rapport au domaine d'application. Les dernières applications concernent la bio-informatique (classification et discrimination dans les séquences génétiques) et les données textuelles (travaux d'un stage dirigé par P. Sébillot (Irisa, projet Aïda)). De nombreux programmes informatiques ont été mis au point dont d'importants se situent dans le cadre du club Modulad (CHAVL, AVARE, MPATD).

De très nombreuses publications ponctuent l'ensemble des travaux.

Laboratoire d'Informatique (EA 2101),

Ecole d'Ingénieurs en Informatique pour l'Industrie, Université de Tours,
 64 avenue Jean Portalis, 37200 Tours,
 Tel : 02 47 36 14 14
 Fax : 02 47 36 14 22
 Web : <http://www.li.univ-tours.fr>,
 Email : venturini,monmarche@univ-tours.fr

G Venturini (PR), N. Monmarché (MC),
 N. Labroche (doctorant), F. Picarouge (doctorant)

Algorithmes génétiques et fourmis artificielles pour l'extraction de connaissances et la visualisation de données.

Nous utilisons des concepts issus de la biologie pour réaliser des algorithmes d'apprentissage et d'optimisation performants. Ces algorithmes fonctionnent de manière dynamique, stochastique et distribuée en utilisant un ensemble d'entités élémentaires qui vont combiner localement leurs actions pour résoudre le problème. Nous utilisons principalement les algorithmes génétiques ainsi que les fourmis artificielles et nous envisageons également de généraliser notre

approche à d'autres modèles biomimétiques comme les automates cellulaires, les réseaux immunitaires etc. Nos actions en cours liées à l'extraction de connaissances sont les suivantes :

- Visualisation de données en 3D par réalité virtuelle (collaboration avec le CERIES, laboratoire financé par la société Chanel),
- Classification de données hétérogènes par nuages d'insectes (collaboration avec le CERIES),
- Analyse des habitudes des utilisateurs de sites web par modélisation du système d'identification chimique des fourmis (thèse de N. Labroche financée par ALDERAN),
- Veille stratégique sur Internet avec des fourmis artificielles (thèse de F. Picarouge),

D'autres travaux de notre équipe en dehors du thème de l'extraction de connaissances portent sur l'optimisation numérique et combinatoire, sur la vision chez les fourmis.

- N. Monmarché, M. Slimane, G. Venturini, D. Lièvre, M. Benoit, J. Caulet, G. Nocent, C. Guinot (1999) Exploration de données numériques par visualisation 3D en réalité virtuelle. Conférence d'Apprentissage, Ecole Polytechnique, Palaiseau, France, pp 259-266, 1999.
- M. Monmarché, D. Laügt, M. Mestre, N. Labroche, A. Oliver, G. Venturini (2001), Classification et visualisation dynamique de données par nuage d'insectes volants, à paraître dans les actes des Journées de la Société Francophone de Classification (SFC'2001),
- N. Monmarché, M. Slimane, G. Venturini (2002 ?). L'algorithme AntClass : classification non supervisée par une colonie de fourmis artificielles. Extraction des Connaissances et Apprentissage, Hermès. A paraître.
- N. Monmarché, G. Venturini, M. Slimane (2000). On how Pachycondyla apicalis ants suggest a new search algorithm. Future Generation Computer Systems, vol 16 n°8 pp 937-946. 2000

LIFO (Laboratoire d'Informatique Fondamentale de l'université d'Orléans), thème Apprentissage

Rue Léonard de Vinci,
 Université d'Orléans, BP 6759,
 45067 Orléans cedex 02

Chercheurs concernés :

Responsable du thème : Ch. Vrain
 (email : cv@lifo.univ-orleans.fr)
 Permanents : M. Exbrayat, S. Guilloré-Billot, L. Martin, F. Moal, V. Clavier (associé)

Doctorants : A. Braud, G. Cleuziou, I. Debourges, A. Salleb, T. Turmeaux

Les recherches dans le thème Apprentissage du LIFO s'articulent autour de deux axes :

- des travaux formels ayant pour origine la Programmation Logique Inductive, orientés vers l'apprentissage de connaissances exprimées dans un *formalisme relationnel*,
- l'application de ces travaux à l'Extraction de Connaissances dans les Bases de Données (ECD).

Parmi les tâches possibles en Fouille de Données, nous nous sommes principalement intéressés à la *classification*. Cependant, une collaboration débutée en 1999 avec le BRGM (Bureau de Recherche Géologique et Minière), nous a conduit à nous intéresser aussi à l'*apprentissage de règles d'association*. Une autre tâche importante est la *découverte de classes*. On peut citer dans ce domaine un projet qui débute sur l'élaboration de mesures de similarité pour la classification et/ou le regroupement d'objets par similitude.

Pour pouvoir appliquer les travaux menés à l'ECD, il faut prendre en compte la taille de l'espace de recherche, le volume de données à traiter et le nombre important de requêtes à exécuter. Nous nous sommes intéressés au premier problème abordé sous deux angles :

- une modélisation, par des grammaires et des *contraintes*, de l'espace de recherche induit par les hypothèses correctement typées,
- une étude de l'intérêt des algorithmes évolutifs (algorithmes génétiques et programmation génétique) qui permettent une exploration stochastique

de l'espace de recherche.

En ce qui concerne les deux autres problèmes, nous étudions d'une part la *parallélisation* des algorithmes génétiques dans le cadre de l'ECD, d'autre part comment optimiser l'évaluation des connaissances engendrées.

Du point de vue des *applications*, nous collaborons actuellement avec le service Ressources Minérales du BRGM qui conçoit et développe des systèmes d'information géographique (SIG) à vocation métallogénique. Nous nous intéressons aussi aux applications de l'Apprentissage à la *fouille de textes* (Text Mining). Le problème que nous avons abordé est celui de la recherche de connaissances dans un corpus, recherche guidée par la requête d'un utilisateur.

Des publications récentes:

- A. Braud and C. Vrain, A Genetic Algorithm for Propositionalization, in Proceedings of the 11th International Conference on Inductive Logic Programming, Céline Rouveirol and Michèle Sebag ed., Vol.2157, LNAI, pp.27-40, Springer-Verlag, sep, 2001
- A. Braud and C. Vrain, A Study of the Scalability of a Data Mining Genetic Algorithm on a Distributed Database, in Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001), L. Spector, E. D. Goodman, A.Wu, W.B. Langdon, H.-M.Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshek, M. H. Garzon and E. Burke ed., pp.753, San Francisco, California, USA, Morgan Kaufmann, 7-11 July, 2001
- L. Martin and F. Moal, A Language-Based Similarity Measure, in Proceedings of the 12th European Conference on Machine Learning, Luc De Raedt and Peter Flach ed., Vol.2167, LNAI, pp.336-347, Berlin, Springer, September, 2001
- I. Debourges I., C. Vrain and S. Guilloire, TEXT MAPPING: A New Approach for Exploring Large Corpora, in Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4), pp.447-450, Tilburg, January, 2001
- F. Moal, T. Turmeaux and C. Vrain, Mining Relational Databases, in PKDD 2000, Principles of Data Mining and Knowledge Discovery, D.A. Zighed, J. Komorowski and J. Zytkow ed., Vol.1910, LNAI, pp.536-541, Berlin, Springer, sep, 2000
- A. Sallab and C. Vrain, An Application of Association Rule Discovery to Geographic Information Systems, in PKDD 2000, Principles of Data Mining and Knowledge Discovery, D.A. Zighed, J. Komorowski and J. Zytkow ed., Vol.1910, LNAI, pp. 613-618, Berlin, Springer, sep, 2000

Groupe « Data mining et bases de données inductives » LISI

INSA de Lyon

LISI Batiment Blaise Pascal
F-69621 Villeurbanne Cedex

Permanents : Jean-Francois Boulicaut, Guy Caplat, Claire Leschi, Christophe Rigotti.

Doctorants : Artur Bykowski, Thomas Daurel, Baptiste Jeudy, Marion Leleu, Cyrille Masson

L'équipe est spécialisée dans la découverte de motifs fréquents et donc, dans leurs applications classiques pour la découverte de règles descriptives. Il peut s'agir de règles d'associations du type « Quand j'observe les caractéristiques A et B alors j'observe souvent la caractéristique C » ou encore de règles d'épisodes « Quand les événements A et B se suivent en moins de 2 heures alors, moins de 5 heures après, il sont souvent suivis des événements C et D ».

Nous nous intéressons à une perspective « Base de données » sur les processus en considérant que leurs différentes étapes s'organisent autour de requêtes, que ce soit pour « sélectionner » les données à fouiller ou pour interroger et explorer les collections, généralement volumineuses, de propriétés extraites. On doit alors souvent mettre en oeuvre un véritable processus de « fouille de propriétés » qui lui-même peut provoquer une nouvelle « fouille de données » etc.

Depuis 1998, nous développons notamment le concept de représentation condensée de motifs fréquents. Il s'agit de représentations de tous les motifs fréquents qui sont compactes et calculables efficacement. Notre représentation dites des delta-libres fréquents généralise d'ailleurs les travaux sur les représentations à base de fermés. Dans le même temps, nous étudions les problèmes posés par l'extraction de motifs sous contraintes avec des mises en oeuvre dans le contexte des règles d'association d'une part et des motifs séquentiels d'autre part.

Dans les années à venir, une grande part de notre activité scientifique va concerner notre rôle dans le projet euro-

péen cInQ (IST-2000-26469) dédié au développement du concept de base de données inductive. Ce projet qui a débuté au 1 Mai 2001 implique les équipes de Rosa Meo (Turin), Stefano Ceri (Milan), Luc de Raedt (Freiburg) et Heikki Mannila (Helsinki) pour un effort global de 217 p/m.

Nous poursuivons également une collaboration avec S. Abdennadher de l'Université de Munich sur l'utilisation de techniques inductives pour la conception de solveurs de contraintes.

Sur le plan national, nous collaborons avec B. Crémilleux du GREYC (Université de Caen).

Bibliographie

- J-F. Boulicaut, M. Klemettinen, H. Mannila. Modeling KDD Processes within the Inductive Database Framework. In: Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery DaWaK'99, Florence (I), August 30-September 1, 1999. Springer-Verlag LNCS 1676, pp. 293-302.
- J-F. Boulicaut, A. Bykowski. Frequent closures as a concise representation for binary data mining. In: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining PaKDD'00, Kyoto (Japan), April 18-20, 2000. Springer-Verlag LNAI volume 1805, pp. 62-73
- S. Abdennadher, C. Rigotti. Automatic Generation of Propagation Rules for Finite Domains. In: Proceedings of the 6th International Conference on Principles and Practice of Constraint Programming CP'00, Singapore, September 18-22, 2000. Springer-Verlag LNCS Volume 1834. pp. 18-34.
- J-F. Boulicaut, A. Bykowski, and C. Rigotti. Approximation of frequency queries by means of free-sets. In: Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD'00, Lyon (F), September 13-16, 2000. Springer-Verlag, LNAI volume 1910, pp. 75-85.
- A. Bykowski and C. Rigotti. A condensed representation to find frequent patterns. November 2000, 11 p. in: Proceedings of the joint conference ACM SIGMOD-PODS 2001, May 21-24, 2001, Santa Barbara (CA, USA), pp. 267-273.
- J-F. Boulicaut, B. Jeudy. Mining Free Sets under Constraints. in: Proceedings of the International Database Engineering and Applications Symposium IDEAS'01, Grenoble (F), July 16-18, January 2001, IEEE Computer Press. pp. 322-329.
- J-F. Boulicaut, B. Crémilleux. Delta-strong classification rules for predicting collagen diseases. In: Proceedings of the PKDD'01 Discovery Challenge on Thrombosis Data

co-located with the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD'01, Freiburg (Germany), September 6, 2001. pp. 29-38.

S. Abdennadher, C. Rigotti. Generation of Propagation Rules for Intentionally Defined Constraints. In: Proceedings of the Thirteenth IEEE International Conference on Tools with Artificial Intelligence ICTAI'01, Dallas, Texas (USA), November 2001. To appear.

Laboratoire LITA (Laboratoire d'Informatique Théorique et Appliquée).

Equipe Fouille des données.

Titre de l'étude menée: Contribution de la classification automatique au Data mining.

Membres : François-Xavier Jollois (Doctorant), Franck Marchetti (MCF), Pierre Laroche (MCF) et Mohamed Nadif (MCF).

Adresse: Université de Metz, Ile du Saulcy 57 045 Metz.

Comme toutes les méthodes d'analyse des données, la Classification Automatique ou classification non supervisée (Clustering) a pour but d'obtenir une représentation simplifiée des données initiales. Elle consiste à organiser un ensemble d'objets (individus) décrits par un ensemble de caractères (variables) en classes homogènes ou classes naturelles. Il s'agit d'une démarche très courante qui permet de mieux comprendre l'ensemble analysé. D'où l'intégration systématique de la classification automatique dans les logiciels à vocation « Data Mining ». Citons par exemple la méthode des centres mobiles connue sous le nom de méthode k-means reposant sur un critère métrique. Ce type d'algorithme et ses variantes sont conçus d'un point de vue heuristique et convergent itérativement, à partir d'une situation initiale, vers un optimum local. Les inconvénients de cette approche sont d'une part la justification du choix de la métrique et du critère utilisé et d'autre part la dépendance entre la solution fournie et l'initialisation de l'algorithme. Ces problèmes ont conduit depuis quelques années à une évolution de l'approche algorithmique, heuristique et géométrique vers une

approche statistique. L'introduction de modèles probabilistes a permis de formaliser l'idée intuitive de la notion de classe naturelle et de donner une interprétation statistique à certains critères métriques. De plus, cette approche permet de proposer de nouveaux critères répondant à des hypothèses précises qui faisaient défauts aux algorithmes traditionnellement utilisés et diffusés dans les logiciels.

Les travaux de l'équipe de l'équipe s'inscrivent dans cette démarche pour répondre d'une part aux problèmes de la classification simple (classification sur les individus) des tableaux de données qualitatives nominales ou ordinales et d'autre part au problème de la classification croisée (méthode ayant comme objectif de classer aussi bien les lignes que les colonnes) des tableaux de contingence, binaires ou tableaux de mesures. Ainsi, des algorithmes de type EM, CEM (Classification EM) et stochastiques ont été proposés et des stratégies de leur utilisation dans le cadre de grandes bases des données sont en cours de développement. Aussi, grâce à cette approche, nous avons pu proposer des solutions aux problèmes des données manquantes, du choix du modèle et du choix du nombre de classes qui sont des facteurs déterminants dans la qualité d'une partition.

Jollois F-X, Nadif M et Marchetti F. Classification de données qualitatives. EGC Nantes(France), 2001.

Jollois F-X, Nadif M et Marchetti F. Un outil de Classification de données qualitatives sous SAS. Club SAS Paris (France), 2000.

Bzioui M, Nadif M. et Govaert G. (1999). Mélanges de Bernoulli et Gaussiens dans la classification croisée. 7èmes rencontres de la société francophone de Classification, Nancy (France), 165-172.

Bzioui M, Nadif M. et Govaert G. (1999). Modèle de mélange croisé Estimation et Classification. XXXI Journées de statistiques, Grenoble (France), 107-110.

Nadif M. and Govaert G. (1998). Clustering for binary data mixture models-Choice of the model. Applied Stochastic Models and Data Analysis. 13, 269-278.

Bzioui M, Nadif M. et Govaert G. (1998). Classification croisée et modèle. XXX Journées de statistiques, Rennes (France), 86-88.

Govaert G. and Nadif M. (1996). Comparison of the mixture and the classification maximum likelihood in cluster analysis with binary data. Computational Statistics and

Data Analysis. 23, 65-81.

Nadif M, Marchetti F. et Govaert G. (1996). Approches Vraisemblance et vraisemblance classifiante en Classification de données Qualitatives. XXVIII Journées de statistiques, Québec (Canada), 556-561.

Laboratoire de recherche en informatique, équipe Inférence et Apprentissage

Université Paris-Sud, Bâtiment 490,
91405 Orsay Cedex
e-mail : yk@lri.fr

Personnel concerné :

Enseignants -

Chercheurs et Chercheurs

ALES BIANCHETTI Jacques
ATER

BOURNAUD Isabelle

MCIUT Sceaux

CORNUEJOLS Antoine

MC CNAM IIE Evry

FRANOVA Marta

CR CNRS

KODRATOFF Yves

DR CNRS

MATWIN Stan

Prof. U. Ottawa, invité 2001-2002

NÉDELLEC Claire

MC Université Paris XI

ROUVEIROL Céline

MC Université Paris XI

Chercheurs en Thèse

ALPHONSE Érick

AZÉ Jérôme

OULD ABDEL VETAH Mohammed

ROCHE Mathieu

MALOBERTI Jérôme

Collaborations très actives avec

BISSON Gilles

CR CNRS (Grenoble)

NAZARENKO Adeline

MdC LIPN

SEBAG Michèle

CR CNRS (X Palaiseau)

Cadre général : L'équipe s'intéresse à l'apprentissage automatique de connaissances par ordinateur, domaine dans lequel elle a, au cours des années, acquis une solide réputation internationale, comme le montre sa participation à de nombreux contrats européens et au

comité de programme de la plupart des conférences sur ce thème.

L'apprentissage automatique comporte une étape délicate (et discutée) de raisonnement inductif, par lequel on invente des hypothèses qui rendent compte au mieux des données. Dans notre équipe, cette recherche se fait selon plusieurs thèmes principaux: l'induction symbolique-numérique, l'induction symbolique en programmation logique inductive et l'induction symbolique en preuves par récurrence.

L'utilisation de l'induction comme un outil scientifique a toujours soulevé une grande opposition d'une partie de la communauté scientifique. Cette opposition a été contredite récemment par l'ampleur des applications industrielles de l'Extraction de connaissances à partir des données (ECD, souvent appelée « Data Mining » ou KDD: « Knowledge Discovery in DataBases ») qui fait systématiquement usage du raisonnement inductif.

Panorama des travaux :

Thème de recherche 1 : Méthodes symbolique-numériques pour l'extraction des connaissances

Ce thème regroupe toutes les méthodologies développées dans l'équipe, ou adaptées d'autres équipes, afin d'effectuer l'ECD.

Les deux points forts de l'équipe sont les algorithmes génétiques et l'apprentissage coopératif.

- Algorithmes génétiques et évolution artificielle

Notre effort se porte sur la recherche d'une expression relationnelle à variables représentant (apprise à partir d') un ensemble d'exemples, avec cette restriction que seul le quantificateur universel est autorisé. La programmation génétique nous permet de travailler avec un nombre de données de l'ordre du millier ce qui autorise à envisager des applications pratiques de ces techniques. L'implémentation d'un tel système d'apprentissage de relations est terminée et elle fait appel de façon systématique à des algorithmes parallèles.

- Découverte des associations

Nous avons montré que les mesures classiques de détection des associations, le support et la confiance, sont extrêmement sensibles au bruit. C'est pourquoi nous développons une nouvelle méthode, moins sensible, que nous appelons « étonnement statistique ». Des algorithmes de découverte rapide des associations contenant plus de deux termes sont en cours d'élaboration.

- Apprentissage coopératif et révision des connaissances

L'apprentissage coopératif est nécessaire lorsque la prise en compte de nouvelles situations conduit à réviser les connaissances acquises ou apprises précédemment. Les méthodes de révision s'attachent à compléter ou à corriger ces connaissances de manière incrémentale, sans repartir de zéro. Par exemple, un robot qui savait ouvrir les portes à battants, doit pouvoir compléter ses connaissances en apprenant à ouvrir les portes à glissières à partir de ce qu'il savait déjà. Ces méthodes nécessitent souvent la participation d'un utilisateur humain. Son rôle peut être double, il peut concerner l'apport de nouvelles connaissances (par exemple, le concept de portes à glissières) et le contrôle des méthodes de révision en l'absence d'exemples significatifs (par exemple, choisir de considérer la porte à glissière comme une porte ou comme un obstacle mobile).

Une application récente de cette approche consiste à automatiser la construction des « ontologies », c'est-à-dire des relations sémantiques entre mots du langage naturel à partir de textes. Quelle que soit la méthode d'analyse du langage naturel utilisée (par exemple, un analyseur syntaxique), elle fera certaines fautes, et les erreurs induites doivent être corrigées par un utilisateur, ce qui constitue la part interactive de la construction. La part automatique, implémentée à l'heure actuelle, se fonde sur une méthode de classification conceptuelle ascendante. La construction des ontologies se réduit alors à la correction des connaissances extraites par le système sur un corpus de textes, et non plus, comme c'est le cas actuellement, à la création laborieuse de ces ontologies « à la main ».

L'utilisation d'autres méthodes que celles fondées sur la classification conceptuelle ascendante sont également en cours d'étude..

- Outils d'ECT de nature linguistique

Notre approche à l'ECT fait usage d'un analyseur linguistique (par exemple, un analyseur syntaxique, ou un analyseur statistique) permettant de mettre en évidence des relations entre les composants du discours, par exemple les relations de nature syntaxique. Le logiciel Asium permet de passer de ces relations à une taxonomie de concepts. En d'autres termes, on passe d'une connaissance syntaxique ou statistique à une connaissance sémantique. Nous désirons maintenant « mesurer » l'apport de chaque structure syntaxique à la sémantique des concepts.

Le logiciel Rowan s'attaque au même problème qu'Asium. La différence tient en ce que les instances d'un concept ne sont pas simplement définies par un terme, mais par le couple (terme, position syntaxique du terme). Une application à l'étude des différences profondes entre textes scientifiques rédigés en anglais par des francophones et par des anglophones est en cours.

Nous commençons à étudier la possibilité d'obtenir les mêmes résultats en utilisant un simple indexeur plutôt qu'un analyseur syntaxique.

Thème de recherche 2 : Programmation Logique Inductive

Dans ce thème, l'équipe s'attache à montrer l'importance des connaissances implicites (les «biais») dans la conduite des inductions dans le cas d'induction symbolique en programmation logique inductive. Ces techniques sont nécessaires quand on veut des règles exprimant non plus des valeurs de descripteurs mais des relations entre descripteurs, c'est à dire apprises dans une restriction de la logique des prédicats.

Les applications les plus prometteuses de ces techniques se font en Biologie Moléculaire et en Fouille de Données dans les bases de données relationnelles.

Les points forts de l'équipe sont :

- *la configuration de systèmes de PLI*

Chaque système d'apprentissage traite de façon particulière le problème de la complexité calculatoire liée à l'utilisation d'une restriction de la logique des prédicats, c'est ce qu'on appelle les « biais » du système. Nous cherchons à trouver une représentation uniforme de ces biais.

Ces travaux ont permis de développer les systèmes de PLI générique Haiku, et plus récemment Haiku-DB, dédié à la découverte dans les bases de données multi-relationnelles.

- *formalisation de la PLI*

L'algorithme DiVS, hybride de l'algorithme de l'Espace des Versions et de l'algorithme de l'Étoile, permet d'apprendre l'ensemble des clauses maximales générales couvrant au moins un exemple positif et rejetant l'ensemble des exemples négatifs d'un concept cible. Cet algorithme a été formalisé et étendu à l'apprentissage de clauses définies, puis à l'apprentissage de clauses contraintes.

Nous travaillons également à l'exploitation dynamique de contraintes sur le langage de concept. La tâche d'apprentissage est décrite à un niveau abstrait par la donnée d'un ensemble de propriétés attendues pour la définition du concept cible (exigences concernant tant la couverture / rejet d'exemples, que la syntaxe de la définition recherchée ou encore des propriétés plus complexes, tels des schémas d'instanciation imposés). Nous avons défini des opérateurs optimaux vis-à-vis de ces contraintes.

Nous avons défini un cadre logique de l'apprentissage dans un langage hybride clause de Horn - Logique de Description. L'avantage d'un tel langage hybride est qu'il est plus expressif que ces deux composantes de base, tout en gardant une complexité intéressante. En exploitant des travaux sur la famille de langages hybrides CARIN, nous avons montré la faisabilité de cette forme d'apprentissage.

- *la « propositionnalisation » de la PLI*

Le problème consiste à trouver une approximation booléenne d'un problème relationnel, tout en conservant les relations les plus "significatives". On transforme les problèmes d'explosion combinatoire en temps calcul en problèmes d'explosion combinatoire en place mémoire. Une optimisation adaptée permet de créer des modèles approximant le modèle réel de façon plus satisfaisante qu'une simple limitation en temps.

Publications récentes :

1999 (sélection)

Augier, S. « Changement d'échelle d'un algorithme d'apprentissage évolutionnaire », Actes de la Conférence d'Apprentissage CAP'99, Palaiseau, M. Sebag (Ed), pages 235-242, 1999.

Cornuéjols A., Tiberghien A., Collet G. « Tunnel effects in cognition : A transfer mechanism from known conceptual domains to new ones ». Conference on Artificial Intelligence and Simulation of Behaviour (AISB-99), Edinburgh, April 6-9, 1999. (pp. 9-18 dans. « Scientific Creativity »).

Kodratoff Y. « Knowledge Discovery in Texts: A Definition and Applications » dans Ras, Z. W. and Skowron A (Eds) Proceedings of the 11th International Symposium on Foundations of Intelligent Systems (ISMIS'99), pages 16-29, Springer-Verlag, 1999.

M. Sebag., « Constructive induction: A version space based approach. », Proc. Int. Conf. on Artificial Intelligence, pages 708--713. Morgan Kaufmann, 1999.

A. Rosete, A. Ochoa, and M. Sebag. « Automatic graph drawing and stochastic hill-climbing, » Proc. of the Genetic and Evolutionary Computation Conf., pages 1699--1706. Morgan Kaufmann, 1999.

2000:

Induction numérique-symbolique à partir de données complexes: des outils pour le « data mining » Diday, Kodratoff, Brito, Moulet (Eds.) Cepadues, 2000.

Alphonse, E., Rouveirol, C. « Lazy Propositionalization for Relational Learning », Proc. of the 14th European Conference on Artificial Intelligence (ECAI'2000)", Berlin, pp. « 256--260 », Horn, W. (Ed), IOS Press, 2000.

Bisson G., Nédellec C., Canamero L. "Designing clustering methods for ontology building: The Mo'K workbench" dans les actes du workshop ECAI'2000, Ontology Learning, Stabb et al. (Eds), 2000.

Bournaud I., Courtine M., « KIDS : Un algorithme itératif pour l'organisation de données relationnelles » dans Conférence d'Apprentissage CAP'2000, Juin 2000. Voir

aussi: « KIDS: An interactive algorithm to organize relational knowledge », publié dans les actes de EKAW'2000.

Bournaud, I., Courtine, M., Zucker J.-D. « Abstractions for knowledge Organization of Relational Descriptions », à paraître dans LNAI, Symposium on Abstraction Reformulation and Approximation SARA 2000, juillet 2000.

Cornuéjols A., Tiberghien A., Collet G. « A new mechanism for transfer between conceptual domains in scientific discovery and education », à paraître dans Foundations of Science 5, 2, 2000.

Kodratoff Y. « L'induction symbolique et numérique en ECD », dans Induction numérique-symbolique à partir de données complexes : des outils pour le « data mining » Diday, Kodratoff, Brito, Moulet (Eds.) Cepadues, 2000.

Nédellec C., « Corpus-based learning of semantic relations by the ILP system Asium » dans James Cussens et Saso Dzeroski, (Eds), Learning Language in Logic, Springer Verlag, 2000.

Rouveirol, C., Ventos, V., « Towards Learning in CARIN-ALN », Proceedings ILP 2000, pp. 191--208.

Sebag, M, Rouveirol, C., « Resource-bounded Relational Reasoning: Induction and Deduction Through Stochastic Matching », ML Journal, 38, 41-62, 2000.

A. Giordana, L. Saitta, M. Sebag, M. Botta, « Analyzing Relational Learning in the Phase Transition Framework, » in Proc. of Int. Conf on Machine Learning, pages 311-318, Morgan Kaufmann, 2000.

A. Ratle, M. Sebag, « Genetic Programming and Domain Knowledge : Beyond the limitations of Grammar-Guided Machine Discovery, » in Proc. of Parallel Problem Solving from Nature, pages 211-220, Springer Verlag, 2000

M. Franova, L. Popelinsky, « Synthesis of formal specifications of predicates: Why and How? » in: R. Trappl, (ed.): Cybernetics and Systems 2000; proc. of the Fifteenth European Meeting on Cybernetics and Systems Research, vol. I, vol. II, Austrian Society for Cybernetics Studies, 2000, 739-744.

2001:

Ales Bianchetti, J. « A Machine Learning Approach to Analogical Reasoning, » International Expert Update, Volume 4, No2, pp 16-24, Summer 2001.

Ales Bianchetti, J. « Improving the use of knowledge in Analogical Reasoning, » Proc. of International Conference Model-Based Reasoning 2001 (MBR'01).

A. Cornuéjols, A. Tiberghien, G. Collet : « A new mechanism for transfer between conceptual domains in scientific discovery and education, » Foundations of Science, vol.5, No.2, 2001, pp.129-155.

Kodratoff Y. « Applications de l'apprentissage

gé automatique et de la fouille de données, » EGC'01, Nantes Jan. 2001.

Kodratoff Y. « Comparing Machine Learning and Knowledge Discovery in Databases: An Application to Knowledge Discovery in Texts, » in *Machine Learning and Its Applications* G. Paliouras, V. Karkaletsis, C.D. Spyropoulos (Eds.): LNAI 2049, p. 1-21, Springer Verlag, 2001.

Kodratoff Y. « Rating the Interest of Rules Induced from Data and within Texts », Proc. DEXA 2001, IEEE eds., pp. 265-269.

Nédellec C., Ould Abdel Vetah M., Bessières P. (2001). « Sentence Filtering for Information Extraction in Genomics, a Classification Problem, » Proc. 5th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 01). L. De Raedt, A. Siebes (Eds.) LNAI 2168, p. 326 ff, Springer-Verlag 2001.

Nédellec C., Ould Abdel Vetah M., Bessières P., Brun C., Jacq J, « Reconnaître les fragments de phrases pertinents pour l'extraction d'information dans les textes de génomique, un problème de classification, » Actes de CAP'2001, p. 221-226.

Équipe Orpailleur - LORIA

Membres permanents : F. Le Ber, J.-F. Mari, A. Napoli, E. Nauer, R. Taouil, Y. Toussaint

Doctorants : M. Cadot, H. Cherfi, S. Hergalant, S. Maumus

Site Web : <http://www.loria.fr/LORIA/EXT/equipes/ORPAILLEUR/>
LORIA - UMR 7503, BP 239
54506 Vandoeuvre les Nancy

Mots clés : représentations des connaissances par objets, raisonnements par classification et à partir de cas, extraction de connaissances dans les bases de données, fouille de données, fouille de textes, système intelligents de traitement de l'information.

L'orpailleur est l'artisan qui recueille par filtrage les paillettes d'or dans les fleuves et les terres aurifères. De même, l'objectif des membres de l'équipe Orpailleur est d'extraire ces paillettes d'or que sont les éléments de connaissances pour les systèmes à bases de connaissances. Ces connaissances peuvent être de plusieurs types : expertises, expériences, explications, stratégies, modes d'emploi, etc. Les connaissances sont ensuite mises en oeuvre dans des systèmes intelligents pour résoudre des

problèmes sur un domaine donné. Parmi les domaines d'application privilégiés des orpailleurs figurent l'agronomie, l'analyse de textes scientifiques et techniques, la bibliométrie, la chimie (pour la planification de synthèses organiques), la médecine, la muséologie et enfin la sidérurgie.

Dans l'équipe Orpailleur, les approches symboliques et numériques sont exploitées pour fouiller les données. Les premières s'appuient essentiellement sur la classification par treillis (et les techniques associées), et les secondes sur l'utilisation de modèles de Markov cachés d'ordre 1 et 2 pour l'extraction de régularités spatio-temporelles (l'utilisation de tels modèles pour la fouille, habituellement réservés pour la reconnaissance de formes, est une des originalités de l'équipe Orpailleur).

Parmi les domaines d'intérêt de l'équipe Orpailleur en matière de fouille de données figurent l'agronomie pour l'étude des successions de cultures, la médecine pour la fouille de données médicales, administratives et géographiques, la bioinformatique et la fouille de séquences génomiques, la fouille de bases de données médico-sociales, la chimie pour l'étude de base de produits de réactions, la sidérurgie pour la détection de défauts spatio-temporels, les textes et l'information scientifique et technique en général, et enfin l'utilisation de méthodes de fouilles dans le cadre de la gestion de grandes bases de données.

La base des méthodes symbolique de fouilles employées chez Orpailleur est la classification par treillis et la théorie des treillis de Galois. Dans le cadre de la fouille de textes par exemple, sont exploitées la classification par treillis et l'extraction de règles d'association. L'objectif est de confronter ces méthodes formelles et leurs résultats à l'interprétation par des experts pour proposer un environnement dans lequel l'accès à l'information intéressante soit plus efficace.

Dans le cadre d'une collaboration Bull-Inria-Orpailleur, le traitement de grandes bases de données et l'apport de techniques de fouilles de données pour étendre les possibilités de la gestion de données sont au premier plan. Ainsi, c'est ici l'extraction des motifs fréquents

et de règles qui sont particulièrement mis en avant. De plus, sont également étudiés des méthodes pour le stockage et l'interrogation de treillis. Ces algorithmes sont intégrés et combinés avec un moteur de requêtes très performant, qui tire profit de la puissance de calcul de super-calculateurs vectoriels. Outre l'aspect algorithmique, l'intérêt et l'originalité de ces travaux réside dans la multitude d'utilisations possibles des treillis : règles d'associations, classifications, regroupement, visualisation, dans des applications très diversifiées comme le marketing, la communication, et toutes celles qui sont mentionnées plus haut.

Éléments de bibliographie.

Y. Bastide, R. Taouil, N. Pasquier, G. Stumme et L. Lakhal, Levelwise search of frequent patterns with counting inference, 16èmes Journées Bases de Données Avancées (BDA'00), Blois, A. Doucet éditeur, pages 307-322, 2000.

H. Cherfi et Yannick Toussaint, Interprétation des règles d'association extraites par un processus de fouille de textes, RFIA-2002, Angers, janvier 2002.

J.-F. Mari, F. Le Ber et M. Benoît, Fouille de données agricoles par Modèles de Markov cachés, Ingénierie des connaissances (IC-2000), Toulouse, France, pages 197-205, 2000.

E. Nauer, Principes de conception de systèmes hypertextes pour la fouille de données bibliographiques multibases, Thèse de l'université Henri Poincaré Nancy 1, 2001.

A. Simon, Outils classificatoires par objets pour l'extraction de connaissances dans des bases de données, Thèse de l'université Henri Poincaré Nancy 1, 2000.

A. Simon et A. Napoli, Un algorithme de fouille dans une représentation des données par objets : une application au domaine médical, Ingénierie des connaissances - Évolutions récentes et nouveaux défis, J. Charlet, M. Zacklad, G. Kassel et D. Bourigault éditeurs, Eyrolles, pages 195-207, 2000.

Y. Toussaint, A. Simon, et H. Cherfi, Apport de la fouille de données textuelles pour l'analyse de l'information, Ingénierie des connaissances (IC-2000), Toulouse, France, pages 335-344, 2000.

**Equipe : Laboratoire
des Sciences de l'Image,
de l'Informatique et
de la Télédétection**

LSIIT, CNRS UPRES-A 7005,

Equipe : Apprentissage et fouille de données

Bd. S. Brant, 67400 ILLKIRCH

Chercheurs : J. Korczak (Pr, responsable), P. Gañarski (MC), N. Lachiche (MC), C. Wemmert (MC), E. Blindauer (doctorant), O. Sandel (doctorant), P. Lipinski (doctorant), G. Fruh (doctorant), N. Poh (doctorant), P. Kustner (MSc), S. Yeung (MSc).

Problématiques de recherche et références

1) Méthodes évolutives : découverte d'expertise en temps réel à partir de séries financières

Nous utilisons une approche évolutive de découverte d'expertise à partir de données. L'expertise, encodée dans les chromosomes décrivant un ensemble des fonctions efficaces, n'est pas connue a priori mais est générée par un processus d'évolution dynamique. Dans nos travaux nous nous sommes plus particulièrement concentrés sur la modélisation d'algorithmes d'apprentissage évolutif, les caractéristiques d'une évolution génétique d'un titre ou d'un portefeuille, les performances du système d'aide à la décision boursière et l'optimisation des points de vue de la convergence, de la rapidité, de la robustesse et de la qualité d'apprentissage et de généralisation. Plusieurs systèmes ont été élaborés dans notre équipe. Actuellement, nos activités se concentrent sur la découverte d'expertise boursière en temps réel à partir de séries temporelles.

Korczak J., Lipinski P., Roger P., Evolution Strategy in Portfolio Optimization, EC'2001, 2001

Korczak J., Lipinski P., Evolution Strategies: Principles and Prototypes, Res. Rap. LSIIT, 2001.

Korczak J., Roger P., Portfolio Optimization using Differential Strategy, ZRIB, 2000, 2000.

Korczak J., Approche génétique pour découvrir un modèle de marché boursier, CAP'99, 1999.

2) Méthodes non supervisées : extraction de concepts et application à la télédétection.

Nous nous intéressons à la classification d'objets supervisée par des connaissances du domaine d'application mais aussi la découverte de connaissances à partir de ces classifications et de leurs processus (historiques) de création : nous nous proposons de ne pas « réinventer » une ou plusieurs méthodes de classification, mais de réutiliser des méthodes (supervisées ou non) existantes, de les intégrer telles quel dans notre système et surtout de les faire collaborer pour une amélioration mutuelle de leur résultat. Cette approche de classification hybride non-supervisée a été implantée dans le système de classification Samarah et validée sur des images de télédétection (SPOT).

Wemmert C., Gañarski P., A New Method of Data Fusion and its Integration in a Multi-Agent System of Automatic Refinement of Classifications, Proc. 5th World Multi-Conference on Systemics, Cybernetics and Informatics, 2001

Wemmert C., Gañarski P., Korczak J., A Collaborative Approach to Combine Multiple Learning Methods,

International Journal on Artificial Intelligence Tools, 2000

Wemmert C., Gañarski P., Korczak J., An Unsupervised Collaborative Learning Method to Refine Classification Hierarchies, ICTAI'99, 1999

3) Fouilles de données structurées, applications à la bio-informatique

De nombreux domaines se prêtent mal à une représentation sous la forme d'un vecteur attribut-valeur. On peut noter une similarité avec la représentation des données dans une base de données relationnelles. Nos travaux concernent la représentation et l'apprentissage à partir de données structurées. Nous travaillons d'une part sur les représentations et les interfaces avec des SGBD. D'autre part, nous intégrons des aspects numériques et probabilistes aux techniques d'apprentissage issues de la programmation logique inductive. Les systèmes développés comprennent un système d'extraction de règles, Tertius, et un classeur bayésien, 1BC. Enfin, nous étudions les apports de ces techniques à la bioinformatique.

Flach P., Lachiche N., Confirmation-Guided Discovery of First-Order Rules with Tertius, Machine Learning, Vol. 42, Number 1/2, Special issue on unsupervised learning, Kluwer, 2001.

Flach P., Lachiche N., A First-order Representation for Knowledge Discovery and Bayesian Classification on Relational Data, DDMI-2000 workshop, PKDD-2000, 2000.

Flach P., Lachiche N., 1BC: a first-order Bayesian classifier. ILP'99, LNCS 1634, Springer, 1999.



« TOTALEMENT INHUMAINE »

Editions Les empêcheurs de penser en rond par JEAN-MICHEL TRUONG.

Notes de lecture de Bertrand Braunschweig.

Les fidèles de l'AFIA et anciens de l'IA se souviennent que Jean-Michel Truong était un des fondateurs de Cognitech, une des premières sociétés développant des systèmes experts en France dans les années 80. Le dernier essai de JMT contient d'ailleurs une dédicace à Alain Bonnet, avec qui l'auteur avait publié un ouvrage sur les systèmes experts en 1986 - Jean Paul Haton en était le troisième co-auteur.

Dans « Totalelement inhumaine », essai de 200 pages, JMT défend l'idée que l'humanité est en train de céder sa place dominante au profit du Successeur, « *l'espèce émergeant sous nos yeux de ce substrat artificiel fait de mémoire et de processeurs toujours plus nombreux en voie d'interconnexion massive qu'on appelle le « Net »*. Il défend cette théorie en montrant comment le Successeur a notamment réussi à détourner à son profit tout d'abord les crédits militaires, avec la Guerre des Etoiles, puis les crédits civils, avec internet et le commerce électronique. A son service, l'humanité se voit répartie en trois classes : les Imbus, ceux qui profitent de sa gloire ; le cheptel, c'est-à-dire vous et moi ; et « epsilon », les David Vincent en lutte contre l'envahisseur inhumain.

Pourquoi donc parler de ce livre sur le site et dans le bulletin de l'AFIA ? Pas seulement parce que l'auteur a commencé sa carrière dans l'industrie de l'IA. Mais aussi, et surtout, parce que plusieurs thèmes chers aux aficionados de l'IA y sont abordés, parfois sans prendre de gants (pour les systèmes experts), parfois de manière plus sympathique, par exemple pour les réseaux neuronaux, les systèmes multi-agents et l'évolution artificielle : les approches avec apprentissage ont la faveur de JMT.

Mon but n'est pas ici de vous conter « T.I. » par le menu. Je préfère en extraire trois morceaux choisis, pour vous donner une idée de ce qu'on peut trouver dans ce

livre, et, au cas où vous seriez hésitants après la lecture des premiers chapitres politiquement incorrects, pour vous encourager à pousser plus loin votre curiosité. Je veux aussi souligner la grande culture générale de JMT, qui cite en référence aussi bien Nietzsche, Heidegger que Turing et Minsky ou encore les rapports de recherche MIT et DARPA.

Quelques morceaux choisis

Sur les systèmes experts et les réseaux neuronaux, JMT attaque fort le principe de l'ingénierie des connaissances :

« ... il nous faut à présent renoncer à l'homme comme *top model* d'intelligence... elle (l'intelligence artificielle) entreprit donc de mettre en boîte les connaissances et méthodes de raisonnement des plus doués d'entre nous... puis on se rendit à l'évidence: à trop vouloir faire l'ange, l'IA ne faisait que la bête... les chercheurs finirent par réaliser que ce qu'il fallait singer, ce n'était pas l'ange, mais la bête. Ceux qui étudiaient les réseaux neuronaux artificiels firent dans cette direction un premier pas... quelque chose d'intelligent pouvait émaner d'une coalition de sots... »

Sur l'apprentissage, citant Turing, JMT soutient le principe d'évolution des systèmes :

« ... plutôt que de chercher à obtenir d'emblée un être parfait, pourquoi ne pas donner jour à un organisme inachevé - une « machine-enfant » -, puis laisser faire cet inlassable éducateur qu'est le temps?... dorénavant, on chercherait à imiter le processus, non le produit ».

Sur les systèmes multi-agents et le web, JMT voit le web comme lieu de vie des agents logiciels :

« ... l'agent logiciel accède ainsi à la dignité de *gène* ou - pour le distinguer de son alter ego organique, d'*e-gène*... les *e-gènes* sont en compétition pour une place

dans un logiciel comme des *allèles* pour un *locus* sur un chromosome... nos calculateurs les plus puissants n'acouchent par ce moyen que de dérisoires avortons. Pour que les e-gènes expriment la plénitude de leur potentiel, il faudra encore l'interconnexion massive de multitudes de processeur hyper-rapides. Il faudra le Net... »

Suite du livre

On sort des quelques chapitres consacrés à l'IA avec une image plus qu'ambitieuse de ce que pourraient devenir nos machines à calculer préférées : une gigantesque collection d'agents apprenants, se reproduisant naturellement ou par voie humaine, s'automodifiant génétiquement par confrontation avec les données du monde réel, et communiquant à haute vitesse sur le Net. Cette vision est fascinante parce qu'elle développe encore plus loin les prévisions les plus optimistes sur les capacités de l'IA, elle est aussi inquiétante de par la place laissée à l'homme dans ce schéma. On peut contester l'idée en soi, mais plus concrètement je voudrais souligner que les technologies actuelles d'apprentissage numérique et symbolique, d'évolution artificielle, d'agents intelligents, sont loin de pouvoir relever un tel défi à moyen terme. Mais il ne s'agit que d'un élément du débat, puisque JMT inclut l'homme dans le processus de répllication et d'évolution des logiciels, et que sur ce point, il n'y a pas grand-chose à redire.

Poursuivant son raisonnement, JMT passe en revue les projets militaires de la *Strategic Defense Initiative*, l'arrivée massive des ordinateurs dans les usines et à la maison, pour des fonctions de plus en plus nombreuses et de plus en plus indispensables, l'émergence du web, la folie dot.com et ce qu'il s'en suivit, démasquant ainsi le Successeur, pour aboutir à la constatation qu'une nouvelle humanité voit le jour, définie par l'équation : *nouvelle*

humanité = Imbus + cheptel + epsilon dont je vous laisse découvrir les conséquences en termes de lutte sociale.

Et alors ?

Un livre qui provoque, un livre qui fait réfléchir... « T.I. » suscitera certainement des réactions intéressantes, il est rare de voir quelqu'un qui comprenne les bases de l'I.A. écrire aussi librement à son sujet. Quant au futur « Matrixien » décrit par JMT, je ne peux m'empêcher de le rapprocher de la communication faite par le futurologue K. Steinmüller à Berlin pendant la session plénière d'ECAI-2000. Il nous avait décrit trois futurs possibles pour l'IA à long terme: (i) l'IA maîtrisée par l'homme, par l'entremise de coupe-feux empêchant les machines de communiquer librement; (ii) l'homme symbiotique, mi-homme, mi-machine, augmentant ses capacités grâce à ses excroissances informatiques; et surtout (iii) l'IA qui, à force de progresser, finit par dépasser l'homme, et à partir de cet instant précis, ce *point singulier*, toute prédiction devient impossible. On retrouve, est-ce l'effet de l'entrée dans le vingt-et-unième siècle ?, la thématique de la machine hyper intelligente, à laquelle JMT ajoute le facteur catalysant du Net et de l'évolution artificielle. Mais, contrairement à Matrix, où la domination de l'IA sur l'homme est apparue après une catastrophe écologique, dans « totalement inhumaine », cette hypothèse repose sur une évolution logique des ordinateurs, des programmes et du réseau à partir de l'état actuel.

Au moment où sort le film « AI » de Spielberg qui va tenir la vedette des discussions pendant un bon moment, il serait intéressant de débattre également des idées émises dans « Totalement Inhumaine ». La maison d'édition a mis à notre disposition quelques exemplaires pour ceux qui voudraient s'y essayer (*me les demander*).

A vos claviers, donc ...



Vous avez *aimé* un *livre*

Vous voulez faire partager votre intérêt, deux options sont possibles.

Vous pouvez en faire une critique d'une ou deux pages et l'envoyer à Amadeo Napoli.

Mais vous pouvez aussi lui faire parvenir une courte note, de quelques lignes, indiquant en quoi il vous semble digne d'attention pour les membres de l'AFIA. Nous la publierons volontiers.

Compte rendu de la conférence EGC'2001 Extraction et Gestion des Connaissances

Nantes, les 17-18-19 Janvier 2001

A l'École polytechnique de l'université de Nantes

IRIN, Université de Nantes,

<http://www.sciences.univ-nantes.fr/irin/EGC2001/>

H. Briand (Président du comité de programme)

F. Guillet (Président du comité d'organisation)

Bilan

La conférence « Extraction et Gestion des connaissances » EGC2001 avait pour objectif de rassembler d'une part une communauté académique multidisciplinaire (systèmes d'information et bases de données, apprentissage automatique, ingénierie des connaissances, statistiques et analyse de données,...) et d'autre part des spécialistes de l'entreprise autour de la double thématique de l'« **Extraction des Connaissances dans les bases de Données** » ECD (KDD : Knowledge Discovery in Databases) et de la « **Gestion des Connaissances** » GC (KM : Knowledge Management).

L'objectif fixé nous semble avoir été atteint au-delà de nos souhaits. D'une part cette première édition des journées francophones EGC a rencontré un plus vif succès que nous ne l'avions envisagé, rassemblant 120 participants, dont 25% entreprise, 25% étudiants et 50% universitaires. D'autre part, les actes de la conférence ont été édités par **Hermès Science Publications** dans la **revue ECA** « Extraction des Connaissances et Apprentissage ». Ils intègrent les textes des trois communications invités, vingt articles concernant l'extraction des connaissances dans les données et dix articles sur la gestion de connaissances.

Déroulement

La première journée du 17/01 a été consacrée aux tutoriels. Une cinquantaine de participants ont assisté à ces interventions aux thèmes variés : *connaissances et XML* (C. Belleil), *ensembles approximatifs* (M. Quafafou), *visualisation des connaissances* (F. Poulet), *SODAS* (E. Diday), *Multimédia Mining* (C. Djeraba).

La journée du jeudi 18/01, a constitué le temps fort de la conférence, propice aux échanges multidisciplinaires et au dialogue entre universitaires et spécialistes des entreprises, au travers de sessions dédiées aux applications dans l'entreprise (exposés et table ronde sur les enjeux de la gestion des connaissances) et d'exposés plus académiques.

Après une cérémonie d'ouverture sous l'égide du président de l'université de Nantes (Y. Tanguy), du directeur de l'École polytechnique de l'université de Nantes (B. Remaud), du directeur de l'Institut de Recherche en

Informatique de Nantes (F. Benhamou), et du président du comité de programme (H. Briand), cette journée a débuté par trois **conférences invitées** (*le processus de gestion de connaissances* (J-L Ermine), *datamining and customer Relationship Management* (M. Jambu), *recherche documentaire sur le Web* (M. Rajman)). La conférence s'est poursuivie par deux sessions consacrées aux applications dans l'entreprise de la gestion des connaissances (gestion des compétences, travail collaboratif et intranet, Web mining, la conception de produits complexes) et une **table ronde** (*enjeux de la gestion des connaissances pour l'entreprise*), une session **gestion des connaissances** (*entreprise étendue, évolution des connaissances et innovation, intelligence économique et gestion des connaissances*), une session **Extraction de connaissances** (*description symbolique de classes, découverte de règles ordinales, applications de la fouille de données, l'intensité d'implication entropique, construction de hiérarchies de classes à partir de données semi-structurées, génération de règles par compression*), et enfin une session spécialement consacrée aux présentations au format poster (10 présentations de 5 minutes).

Après l'effort, le réconfort : cette journée de travail bien remplie s'est achevée par un cocktail de bienvenue, lui-même prélude à un repas de gala en croisière fluviale sur l'Erdre nantaise réunissant 90 convives.

Le vendredi 17/01, a été l'objet de communications plus appliquées. Ainsi, en **Extraction de connaissances** les sujets abordés concernaient : *les règles d'association, la visualisation, l'analyse de séquences biologiques, le F-OLAP Mining, la modélisation multidimensionnelle des systèmes décisionnels, la recherche dans les données vidéo, le ciblage d'événements rares en marketing, l'analyse implicite d'un questionnaire de personnalité, l'analyse symbolique de données bancaires et de données boursières*. Une session a été spécialement consacrée à l'**Extraction de connaissances dans les Textes**, avec : *la conception de méthodes de classification pour la construction d'ontologies, la catégorisation de textes basées sur les N-grammes, l'extraction de connaissances sur les anthroponymes, l'exploitation de données textuelles dans le cadre du Web mining*. Enfin, les derniers

sujets abordés ont concernés la **Gestion de connaissances** : *modélisation type tâche/méthode, capitalisation et déploiement de connaissances de diagnostic, navigation hypermédia adaptative, approche hypermédia pour la gestion des savoir-faire, techniques de storytelling, fonctionnements d'une presse à injecter à partir de données de production*).

De nombreuses photos souvenir, ont été réalisées et diffusées lors de la manifestation. Elles sont maintenant disponibles sur le site web de la conférence.

Perspectives

Devant l'intérêt manifesté pour la double thématique ECD/GC, le comité de pilotage a proposé que les journées francophones « Extraction et Gestion des Connaissances » se poursuivent à un rythme annuel.

Ainsi, la seconde édition, **EGC2002**, sera organisée par Danièle Hérim (LIRMM) et aura lieu à **Montpellier** les **21-22- 23 Janvier 2002**.

Remerciements

Nous tenons à remercier très chaleureusement l'ensemble des organismes qui nous ont parrainés et soutenus

dans cet exercice difficile qui a consisté à lancer un nouveau cycle de conférences : la conférence européenne **PKDD** (Principles and Practice of Knowledge Discovery in Databases), l'**AFIA** (Association Française pour l'Intelligence Artificielle), l'**ASTI** (Association Française des Sciences et Technologies de l'Information), le **GRACQ** (GROUPE ACQUISITION et ingénierie des connaissances), **INFORSID** (Informatique des Organisations et Systèmes d'Information et de Décision), le **PRC-GDR I3** (Information Interaction Intelligence), la **SFC** (Société Francophone de Classification), la **SFdS** (Société Française de Statistique), **SPECIF** (association des Personnels Enseignants et Chercheurs en Informatique de France) et l'équipe **ERIC** (Université de Lyon 2). La manifestation a aussi été soutenue par deux clubs d'entreprises : le club *Gestion des Connaissances*, et le **CRI-Ouest** (Club des Responsables Informatiques de l'Ouest).

Nous remercions tout particulièrement la Ville de Nantes, le Conseil Général de Loire Atlantique, le Conseil Régional des Pays de Loire, l'Université de Nantes, le CRI Ouest, INFORSID et l'Ecole polytechnique de l'université de Nantes pour le soutien financier qu'ils ont apporté à la manifestation.

Compte rendu du 10^{ème} Workshop sur les Systèmes-Multi-Agents MAAMAW 2001 Modelling Autonomous Agents in a Multi-Agent World

2- 4 Mai 2001, Annecy

<http://www-leibniz.imag.fr/MAAMAW/>

Sylvie Pesty - Yves Demazeau

La série des Colloques MAAMAW est LA manifestation européenne de référence dédiée à l'étude des systèmes multi-agents. Les participants aux colloques MAAMAW proviennent principalement du domaine de l'informatique mais aussi des sciences humaines et sociales, de la physique ou encore des sciences de la vie, les systèmes multi-agents constituant l'un des paradigmes les plus en vue actuellement dans le domaine des Sciences et Technologies de l'Information. Ils se retrouvent autour de leur intérêt mutuel à comprendre les phénomènes liés aux interactions et aux organisations entre agents, ainsi que pour étudier les effets de l'agent en tant qu'individu ou en tant que participant à un groupe, spécifier des langages, les architectures et les mécanismes propres à ces systèmes, appliquer leurs résultats sur des problèmes concrets et partager leurs expériences.

Le colloque est revenu cette année en France après avoir visité une grande partie de l'Europe : UK-Cambridge (89), F-Paris (90), D-Kaiserlautern (91), I-Viterbe (92), CH-Neuchatel (93), DK-Odense (94), NL-Eindhoven (96), S-Ronneby (97), et E-Valencia (99). Le

colloque a obtenu le soutien du collège SMA de l'AFIA qui a octroyé 5 bourses à de jeunes chercheurs. Il a également été soutenu par le GDR-PRC I3, la Fondation IFMAS et à un niveau européen, par le Réseau d'Excellence AgentLink. Cinq groupes industriels étaient partenaires de MAAMAW'01: AEGIS, Motorola, Bouygues Telecom, Teamlog et le Centre de Recherche de Xerox.

Cette année, MAAMAW avait pour thème privilégié, le thème des "organisations" et 65 participants sont venus s'exprimer et débattre pendant 3 jours. Trois conférenciers de renom étaient invités au colloque: Les Gasser de l'Université d'Illinois, Paolo Ciancarini de l'Université de Bologne, et Jacques Ferber de l'Université de Montpellier.

Les actes finaux du colloque MAAMAW'01 seront publiés par Springer Verlag dans la collection des *Lectures Notes in Artificial Intelligence*, ce qui en garantit déjà ses larges visibilité, diffusion, et impacts.

Un compte-rendu très complet (en anglais) est donné sur le site de l'AFIA.

Compte rendu des Sixièmes journées francophones EIAO Environnements Interactifs d'Apprentissage avec Ordinateur

25-27 avril 2001,

Cité des Sciences et de l'Industrie, Paris-La Villette

<http://www.inrp.fr/atief/eiao/>

Monique Baron, LIP6, Université Pierre-et-Marie-Curie (Paris 6)

Les sixièmes journées EIAO (Environnements Interactifs d'Apprentissage avec Ordinateur) ont eu lieu les 25, 26 et 27 avril 2001, à la Cité des Sciences et de l'Industrie de Paris-La Villette, dans le cadre de la plateforme ASTI'2001, sous l'égide de l'ATIEF (Association des Technologies de l'Information pour l'Éducation et la Formation), l'une des vingt associations fondatrices de l'ASTI.

ASTI'2001, congrès inaugural de l'Association des Sciences et Technologies de l'Information, a accueilli du 24 au 27 avril 2001 une quinzaine de manifestations couvrant différents domaines des STI, ainsi que les premières journées du RNTL. La journée centrale du mercredi 25 avril, commune à toutes les associations fondatrices, a été consacrée aux relations des STIC (Sciences et Technologies de l'Information et de la Communication) avec la culture et le savoir, proposant, outre un ensemble de conférences, diverses démonstrations dont deux démonstrations « EIAO » : l'Université En Ligne, projet national Premier Cycle Sur Mesure (P. Jarraud et C. Cazes, Université Paris 6), et le nouveau micromonde Aplusix pour l'algèbre (J.F. Nicaud, IRIN).

Organisées par l'ATIEF, l'INRP et le GDR I3, avec le parrainage scientifique de l'AFIA et de l'ARCo, les sixièmes journées EIAO ont été placées, malgré la nouveauté du cadre de l'ASTI et du lieu de La Villette, sous le signe de la continuité des objectifs et de l'esprit pluridisciplinaire par rapport aux précédentes éditions. Rappelons que celles-ci, organisées tous les deux ans de 1989 à 1997 à l'ENS de Cachan, dans le cadre des activités du PRC-GDR Intelligence Artificielle, avaient fait place en 1999 au congrès mondial AI-ED qui s'est tenu au Mans, à l'initiative de Martial Vivet.

Ces sixièmes journées ont rassemblé les 26 et 27 avril environ quatre-vingts participants de différentes disciplines (Informatique, Intelligence Artificielle, Didactique, Psychologie Cognitive, Sciences de l'Éducation), pour un programme comportant 2 conférences invitées, 12 communications réparties en 4 sessions, 5 posters et 11 démonstrations.

Dans la première conférence invitée, Jacqueline Bourdeau a présenté l'état des recherches et des réalisations du LICEF (Laboratoire de recherche de la Télé Université du Québec) concernant les dispositifs de formation à distance (téléapprentissage, « e-learning »). Elle a également donné un aperçu de l'approche de conception d'EIAO à base d'ingénierie ontologique

développée au « MizLab », le laboratoire d'IA de l'université d'Osaka dirigé par R. Mizoguchi, avec lequel elle coopère.

Lors de la seconde conférence invitée, Jean-François Rouet (Laboratoire Langage et Cognition, CNRS-Université de Poitiers) a présenté des apports de la psychologie à la conception de dispositifs techniques destinés à l'apprentissage. J.-F. Rouet a en particulier rappelé la complexité de l'activité d'apprentissage, qui met en oeuvre des capacités d'attention, de mémorisation et de traitement d'informations, et où le langage joue un rôle important ; il s'est interrogé sur les prérequis d'une « conception rationnelle des technologies d'apprentissage » et a présenté brièvement quatre axes de recherche actuels liés à cette question : mémoire de travail et apprentissage, lecture et ergonomie d'écran, recherche d'informations, rôle des animations multimédias.

Les autres communications, posters et démonstrations, qui se sont succédé avec une remarquable unité de lieu, ont montré la diversité des travaux de recherche dans ce domaine : diversité des questions abordées et des points de vue liés aux disciplines de recherche concernées, diversité des types d'environnements d'apprentissage visés et des dispositifs de formation dans lesquels ils peuvent s'intégrer (avec une nette évolution vers des dispositifs de formation à distance utilisant le Web), diversité des disciplines d'enseignement concernées... Pour plus de détails, il est conseillé de consulter les actes, édités sous la forme d'un numéro double de la revue Sciences et techniques éducatives (vol. 8, n° 1-2, 2001, Hermès), sous la direction de Cyrille Desmoulin, Monique Grandbastien et Jean-Marc Labat.

Parmi les conclusions qui peuvent être tirées à l'issue de ces journées et des débats qui y ont eu lieu, il convient à mon avis de réaffirmer la nécessité de travaux scientifiques pluridisciplinaires dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain), pour développer, exploiter et expérimenter les possibilités des STIC au service des apprentissages humains.

Quant à l'avenir de ces journées, des discussions sont en cours pour les réunir en 2003 avec un autre colloque francophone du domaine, « Hypermédias et Apprentissages », dont la cinquième édition s'est tenue à Grenoble du 9 au 11 avril 2001.

Les Conférences AFIA en 2001

Janvier	EGC	Extraction et Gestion des Connaissances
Avril	EIAO	Environnements Interactifs d'Apprentissage avec Ordinateur
Mai	MAAMAW	Modelling Autonomous Agents in a Multi-Agent World
-	TIA	Terminologie et Intelligence Artificielle
Juin	Plate-Forme AFIA	L'IA, Informatique de la Connaissance
Octobre	Valgo	Connexionnisme
-	EA	Evolution Artificielle
Novembre	JF IAD SMA	Intelligence Artificielle Distribuée et Systèmes Multi-Agents
-	LFA	La Logique floue et ses applications
-	CITE	Coopération, Innovation et Technologies

Les Conférences AFIA en 2002

Janvier	RFIA	Reconnaissance des Formes et Intelligence Artificielle
-	EGC	Extraction et Gestion des Connaissances
Juillet	ECAI	European Conference on Artificial Intelligence
-	IPMU	Information Processing and Management of Uncertainty

Calendrier des conférences en IA en 2001

7-19 Octobre	MSR 2001	Toulouse	Modélisation de systèmes réactifs
22-23 Octobre	K-CAP 2001	Victoria (Canada)	Knowledge Capture
23-26 Octobre	IAT'2001	Maebashi City, Japon)	Intelligent Agent Technology
5-10 Novembre	CIKM'2001	Atlanta (USA)	Information and Knowledge Management
7-9 Novembre	ICTAI'2001	Dallas (USA)	Tools with Artificial Intelligence
12-14 Novembre	JFIADSMA'2001	Montreal (Canada)	Multi-agent systems
29 nov - 2 Déc	ICDM'2001	Silicon Valley (USA)	Data Mining
2-4 Décembre	MA'2001	Atlanta (USA)	Mobile Agents
10-12 Décembre	ES2001	Cambridge (UK)	Knowledge Based Systems and Artificial Intelligence
8-20 Décembre	AMT'01	Hong-Kong	Active Media Technology

Calendrier des conférences en IA en 2002

08-10 Janvier	RFIA'2002 (AFIA)	Angers	Reconnaissance des Formes et IA
15-18 janvier	NF'2002	La Capitolio de laHabana (Cuba)	Neuro Fuzzy
10-14 Mars	SAC'2002	Madrid (Spain)	Coordination Models, Languages and Applications
25-27 Mars	IAS'2002	Marina del Rey USA)	Intelligent Autonomous Systems
11 Avril	COORDINATION 2002	York (United Kingdom)	Coordination Models and Languages
16-18 Mai	FLAIRS 2002	Pensacola (USA)	Artificial Intelligence
27-29 Mai	AI 2002	Calgary (Canada)	Artificial Intelligence
27-29 Juin	ISMIS 2002	Lyon	Methodologies for Intelligent Systems
5 Juillet	PMU'2002	Annecy	Knowledge-Based Systems
5 Juillet	AISC'2002	Marseille	Artificial Intelligence and Symbolic Computation
8-12 Juillet	ICML'2002	Sydney (Australia)	Machine Learning
15-19 Juillet	AAMAS'2002	Bologna (Italy)	Autonomous Agents and MultiAgent Systems
1-26 Juillet	ECAI'2002 (AFIA)	Lyon	Intelligence Artificielle
28 Juillet-1 Août	AAAI'2002	Edmonton (Canada)	Intelligence Artificielle
5-30 Août	FIP'2002	Montreal (Canada)	Information Processing

Conférences organisées ou co-organisées par l'AFIA

La Plate-Forme AFIA 2001 25-29 juin 2001 - Grenoble

Organisée tous les deux ans en alternance avec la conférence RFIA, la Plate-forme AFIA permet de rassembler sur un même site plusieurs conférences très spécifiques. La plate-forme est l'occasion de réunir des chercheurs ayant des thématiques voisines, voire différentes, et des industriels. Elle permettra une confrontation entre les besoins des industriels et différentes approches. Cette plate-forme est ouverte aux conférences émanant des groupes de travail ou des collèges de l'AFIA, mais elle est ouverte plus largement aux communautés scientifiquement proches de l'Intelligence Artificielle. La Plate-Forme AFIA 2001 est organisée par le laboratoire Leibniz-IMAG.

L'édition 1999 s'est déroulée entre le 14 et le 18 juin à l'Ecole Polytechnique. Vous pouvez consulter les actes 99 sur le site de l'AFIA.

RFIA 2002 : Reconnaissance des Formes et Intelligence Artificielle 8-10 janvier 2002 - Angers

RFIA est le congrès par excellence des chercheurs dans les domaines de la Reconnaissance des Formes et de l'Intelligence Artificielle. Il a pour but de promouvoir toutes les recherches dans ces domaines et offre une occasion unique de rassemblement et d'échanges scientifiques indispensables entre les chercheurs de ces deux importantes communautés. Il est co-organisé par l'AFRIF et l'AFIA.

L'édition 2001 aura lieu à Angers sous la co-présidence de Roger Mohr (XRCE, Grenoble) pour la partie RF, de Catherine Garbay (TIMC-IMAG, Grenoble) pour la partie IA et de Jin-Kao Hao (LERIA, Angers) pour le comité d'organisation.

Contacts :

Roger Mohr :

Roger.Mohr@xrce.xerox.com

Catherine Garbay :

mailto:Catherine.Garbay@imag.fr
Catherine.Garbay@imag.fr

Jin-Kao Hao : Jin-Kao.Hao@univ-angers.fr

La douzième édition de ce congrès, RFIA 2000 a eu lieu du 1er au 3 février 2000 à Paris, aux Salons de l'Aveyron situés dans un quartier en pleine expansion à proximité de la Bibliothèque Nationale de France et du parc de Bercy.

ECAI 2002 : 15th European Conference on Artificial Intelligence 21-26 juillet 2002 - Lyon

ECAI 2002 sera la 15^{ème} édition de la Conférence Européenne d'Intelligence Artificielle.

Conférences organisées par les collèges et les groupes de travail

Voir les liens sur le site de l'AFIA.

MAAMAW 2001 : 10th European Workshop on Multi-Agent Systems 02-04 mai 2001 - Annecy

MAAMAW'01 will be the 10th issue of the MAAMAW series of workshops, the major event about Multi-Agent Systems (MAS) research regularly held in Western Europe since 1989. As the ICMAS series are becoming yearly, the MAAMAW workshops will move back to their original spirit. The MAAMAW workshop promotes interdisciplinary grounding, innovative work, recent results, large place for discussion, for a limited audience (100) in a convivial site (Annecy). The special topic of the 10th European Workshop on Multi-Agent Systems will be devoted to « Multi-Agent System Organisations ».

The workshop is organised by the MAGMA Research Group of the LEIBNIZ-IMAG laboratory in collaboration with the IAD-SMA Working Group of the AFIA Association.

TIA 2001 : Terminologie et Intelligence Artificielle 03-04 mai 2001 - Nancy

Le groupe de travail « Terminologie et Intelligence Artificielle » (TIA) rassemble des chercheurs en linguistique, en intelligence artificielle et en traitement automatique des langues. Il a été créé pour permettre une confrontation entre les cadres théoriques et méthodologiques ainsi qu'entre les pratiques développés dans chaque discipline. Il organise la Conférence internationale Terminologie et acquisition de connaissances à partir de textes TIA 2001 (mai 2001, Nancy).

JFIADSMA 2001 : Journées Francophones d'Intelligence Artificielle Distribuée et Systèmes Multi-Agents 12 - 14 Novembre 2001 - Montréal

Les Journées francophones pour l'intelligence artificielle distribuée (IAD) et les systèmes multi-agents (SMA) sont une occasion unique de rencontres et d'échanges dans les domaines de l'IAD et des SMA. Cette conférence permet également de démontrer, année après année, la vivacité et l'excellence de la

communauté francophone dans ces domaines de pointe, lesquels prennent une importance de plus en plus grande en informatique.

À l'automne 2001, le Québec aura pour la première fois le privilège et le plaisir d'accueillir ces journées en dehors de la France. Ce sera alors une occasion unique pour les chercheurs et les étudiants, mais également pour les industriels québécois, d'avoir un accès direct aux derniers travaux effectués dans ces domaines. À l'inverse, nos hôtes auront l'occasion de découvrir, si cela n'était déjà fait, leurs homologues québécois et de tisser ainsi de plus amples collaborations universitaires et industrielles.

Ces journées se sont tenues successivement à Toulouse (1993), Grenoble (1994), Chambéry (1995), Port Camargue (1996), Nice (1997), Nancy (1998), L'Île de la Réunion (1999) et Saint-Etienne (2000). Pour leur neuvième édition, les JFIADSMA 2001, organisées par le CRIM, se tiendront à Montréal, au Québec, du 12 au novembre 2001.

Conférences Parrainées par l'AFIA

EGC 2001 : Journées Francophones d'Extraction et de Gestion des Connaissances

17, 18 et 19 Janvier 2001 - Nantes

Aujourd'hui, de grandes masses de données structurées ou semi-structurées sont accessibles dans les bases de données d'entreprises ainsi que sur la Toile. Aussi les entreprises ont-elles besoin de méthodes et d'outils capables d'extraire les connaissances pertinentes pour les décideurs, de les stocker, et de les diffuser aux différents acteurs de l'organisation. Pour répondre à cette attente du tissu industriel, de nombreux projets de recherche se développent autour de l'extraction de connaissances à partir de données (Knowledge Discovery in Data), ainsi que sur la gestion de connaissances (Knowledge Management). L'objectif de ces journées consiste à rassembler, d'une part les chercheurs des disciplines connexes (apprentissage, statistiques et analyse et données, systèmes d'information et bases de données, ingénierie des connaissances), et d'autre part les spécialistes d'entreprises qui déploient les méthodes et les outils adaptés à leurs besoins, afin de contribuer à la formation d'une communauté scientifique dans le monde francophone autour de cette

double thématique de l'extraction et de la gestion de connaissances.

Ces journées seront ainsi le lieu privilégié d'échanges entre chercheurs et spécialistes industriels afin de présenter l'état de l'art de cette thématique centrale, d'en faire émerger les besoins et les enjeux, ainsi que d'en stimuler le développement.

EIAO 2001 : Sixièmes journées francophones - Environnements Interactifs d'Apprentissage avec Ordinateur
25, 26 et 27 Avril 2001 - Paris

Sur le plan scientifique, ces journées se placent dans la continuité des précédentes organisées à l'ENS de Cachan en 1997. Il avait alors été décidé de ne pas organiser de manifestation en France la même année qu'AIED99. Après cet événement exceptionnel, le cycle reprend et les prochaines journées ont donc lieu au printemps 2001. Elles s'adressent aux chercheurs, enseignants et professionnels concernés par les questions de conception, de réalisation, d'expérimentation et d'évaluation d'EIAO (Environnements Interactifs d'Appren-tissage avec Ordinateur).

Ces journées se veulent pluridisciplinaires et ont pour objectif de permettre à la communauté scientifique francophone de faire le point sur les recherches en cours et de contribuer aux échanges scientifiques entre les différents champs de recherche concernés, notamment l'intelligence artificielle et l'ingénierie des connaissances, le génie logiciel et les interactions homme machine, les didactiques des disciplines, la psychologie cognitive, les sciences de l'éducation. Elles conserveront la dimension internationale en langue française en cherchant à augmenter la participation de membres francophones non français.

Journées Valgo 2001 : Journées "connexionnisme" des thésards francophones
17 - 19 octobre 2001 - Imbours (Ardèche)

Les journées Valgo ont pour objectifs de réunir les thésards francophones travaillant dans le domaine du connexionnisme afin qu'ils partagent leurs expériences mais aussi qu'ils aient une idée plus précise des différentes approches existantes et des divers débouchés. Elles sont une occasion pour ces thésards de s'exercer à synthétiser et présenter leurs travaux devant une assemblée critique mais amicale.

EA'01 : Conférence Internationale sur l'Evolution Artificielle
29 - 31 octobre 2001 - Le Creusot

EA'01 est la cinquième conférence consacrée aux techniques algorithmiques simulant l'évolution naturelle. Après EA'94 à Toulouse, EA'95 à Brest, EA'97 à Nîmes et EA'99 à Dunkerque, la conférence sera organisée au Creusot, au sein de l'Université de Bourgogne. Les spécificités des éditions précédentes seront préservées : les tarifs d'inscription resteront le plus bas possible (moins de 150 EUROS) et la gastronomie et l'oenologie feront partie du programme.

* LFA 2001 : Rencontres francophones sur la logique floue et ses applications

26 et 27 Novembre 2001 - Mons (Belgique)

Depuis 1992, ces rencontres ont pour objectif de réunir les universitaires et les industriels intéressés par l'utilisation des ensembles flous et de la théorie des possibilités. Tous les domaines où ces théories sont utilisées sont concernés, tels que la commande automatique, l'aide à la décision, la classification, la reconnaissance des formes et l'analyse de données, le traitement et l'interprétation des images, la fusion d'informations, l'intelligence artificielle, les systèmes d'information, etc. Les rencontres sont aussi ouvertes sur d'autres théories de l'incertain, telles que la théorie des fonctions de croyance de Dempster-Shafer, ou les approches qualitatives qui souviennent entretiennent des liens étroits avec les théories des possibilités et des ensembles flous.

Tous les travaux théoriques, méthodologiques, ainsi que les applications, peuvent être soumis. Il est envisagé cette année de mettre l'accent sur la modélisation de l'incertain en reconnaissance des formes, classification, analyse de données et statistique ; les communications émanant des diverses communautés sur ces thèmes seront tout particulièrement bienvenues.

Comme chaque année, la priorité est donnée aux jeunes chercheurs afin qu'ils puissent présenter leurs travaux. Le comité de programme sera particulièrement attentif au caractère novateur et didactique des contributions des chercheurs universitaires. Des exposés traitant d'expériences concrètes et d'applications dans le domaine industriel seront accueillis avec un intérêt particulier.

CITE 2001 : Recherches interdisciplinaires, sciences de l'ingénieur et

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

29 et 30 Novembre 2001 - Troyes

Il ne fait aujourd'hui plus de doute que la mise en place de processus d'innovation en continu, enjeu économique et social crucial, repose sur la capacité à gérer efficacement les activités intellectuelles organisées, ou, ce qui revient au même, à gérer, ou à faciliter " l'autogestion " par les acteurs, de leurs activités collectives, coopératives et cognitives. L'hypothèse majeure du colloque CITE'2001 est que les formes de cognition collective dépendent profondément des technologies de la coopération et de la connaissance mises en œuvre.

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

Le développement des technologies de la coopération et de la connaissance confronte les chercheurs à plusieurs questions de recherche fondamentale relevant des sciences cognitives, des sciences humaines et sociales et des sciences de l'ingénieur. L'informatisation des activités collectives relevant de la coopération et de la production collective de connaissances agit en effet comme un révélateur des phénomènes de cognition distribuée, des processus d'organisation et de gestion de l'action collective et de la dynamique des relations et des identités mises en scène dans les activités communicationnelles. Par ailleurs, elle pose aux sciences de l'ingénieur des problèmes de modélisation ardues d'activités heuristiques et émergentes auxquelles les systèmes d'information doivent en permanence s'adapter.

EGC 2002 : Secondes Journées Francophones d'Extraction et de Gestion des Connaissances

21 - 23 janvier 2002 - Montpellier

Aujourd'hui, de grandes masses de données structurées ou semi-structurées sont accessibles dans les bases de don-

nées d'entreprises ainsi que sur la Toile. Aussi les entreprises ont-elles besoin de méthodes et d'outils capables d'extraire les connaissances pertinentes pour les décideurs, de les stocker, et de les diffuser aux différents acteurs de l'organisation. Pour répondre à cette attente du tissu industriel, de nombreux projets de recherche se développent autour de l'extraction de connaissances à partir de données (Knowledge Discovery in Data), ainsi que sur la gestion de connaissances (Knowledge Management).

L'objectif de ces journées consiste à rassembler, d'une part les chercheurs des disciplines connexes (apprentissage, statistiques et analyse de données, systèmes d'information et bases de don-

nées, ingénierie des connaissances, ...), et d'autre part les spécialistes d'entreprises qui déploient les méthodes et les outils adaptés à leurs besoins, afin de contribuer à la formation d'une communauté scientifique dans le monde francophone autour de cette double thématique de l'extraction et de la gestion de connaissances.

Cette manifestation sera ainsi le lieu privilégié d'échanges entre chercheurs et spécialistes industriels afin de présenter l'état de l'art de cette thématique centrale, d'en faire émerger les besoins et les enjeux, ainsi que d'en stimuler le développement.

IPMU 2002 : The 9th Information Conference on Information Proces-

sing and Management of Uncertainty in Knowledge-Based Systems
1 - 5 juillet 2002, - Annecy

The IPMU Conference is organized every two years with the focus of bringing together scientists working on methods for the management of uncertainty and aggregation of information in intelligent systems. This conference provides a medium for the exchange of ideas between theoreticians and practitioners in these and related areas.

Séminaire IAdIF : Séminaire d'Intelligence Artificielle de l'Île de France
Paris

Pour plus d'informations, consulter le site du séminaire (IAdIF)

Informations pour les organisateurs de conférences en IA

Le parrainage :

Tout organisateur de conférence, colloque, atelier, ... dans le domaine de l'IA peut solliciter l'AFIA pour obtenir un parrainage de l'événement qu'il prépare. Le parrainage de l'AFIA est en quelque sorte une garantie, une caution scientifique qui est obtenue en se fondant sur différents critères tels que les objectifs scientifiques de la conférence, la composition du comité de programme, la procédure de sélection des articles... Le parrainage de l'AFIA se veut également être un indicateur pour tous ceux qui s'interrogent sur l'intérêt de participer à telle ou telle manifestation.

L'octroi du parrainage de l'AFIA s'accompagne d'un engagement réciproque entre l'organisateur et l'AFIA. Cet engagement permet à l'organisateur de bénéficier de la plupart des moyens de diffusion propres à l'AFIA. En contrepartie le parrainage permet à l'AFIA de se faire mieux connaître, de donner aux membres de l'AFIA des informations plus complètes sur l'activité de la communauté IA et d'obtenir des réductions pour ses membres pour les inciter à participer aux manifestations parrainées.

Que faire ?

Cas général (pour tous)

Pour toutes les conférences, que celles-ci soient organisées par l'AFIA, par les groupes de travail ou collègues, comme pour les conférences qui ont obtenues le parrainage, l'organisateur est invité à faire parvenir :

- l'URL du serveur de la manifestation (à adresser à Sylvie PESTY) ainsi qu'au responsable des « Groupes

de travail et Collèges » (Jean-Paul BAQUIAST - voir adresse page 2 du Bulletin).

- les différentes annonces de la conférence,
- une page environ pour annoncer la conférence et une à deux pages pour faire connaître le bilan de la conférence (à adresser au responsable des conférences dans le bulletin (Sylvie PESTY)).

Cas particulier des conférences organisées par l'AFIA, ses groupes de travail ou ses collègues

L'organisation est déléguée à un organisateur local. Un contrat doit être entre l'AFIA et l'organisateur local pour définir les conditions de la délégation, essentiellement au niveau du partage des risques financiers et du résultat de la manifestation (bénéfice ou déficit).

Trois modèles de contrat sont proposés aux organisateurs locaux ; ces contrats sont aménageables (nous faire une proposition : Sylvie PESTY)

- un modèle de contrat avec délégation des risques et résultats à l'organisateur local.
- un modèle de contrat avec délégation des risques et résultats à l'AFIA.
- un modèle de contrat pour un partage des risques et résultats entre l'organisateur local et l'AFIA.

Pour toute information supplémentaire, n'hésitez pas à consulter le site de l'AFIA ou à contacter :

Sylvie PESTY Sylvie. Pesty@imag.fr

Laboratoire Leibniz-IMAG

46, Avenue Felix Viallet

F-38031 Grenoble Cedex

Tél : 04 76 57 48 07

Fax : 04 76 57 50 81

Nous annonçons dans cette rubrique les manifestations IA (i) se déroulant en France ou (ii) francophones et se déroulant hors de France. Pour inclure une annonce, envoyer un texte (mél + fichier attaché) à Sylvie PESTY (voir « l'ours » page 2 de ce bulletin).

SOMMAIRES DES REVUES

Valérie Mangin ne pourra plus participer à l'élaboration de cette rubrique, un grand merci à elle pour tout le travail qu'elle y a fait !

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://www.editions-hermes.fr/periodiques/ia_cur.htm

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:80/inca/publications/store/5/0/5/6/0/1/>

ARTIFICIAL INTELLIGENCE V129 N°1-2 JUNE 2001

- Heuristic search in artificial intelligence, *WEIXIONG ZHANG, RINA DECHTER, RICHARD E. KORF*
- Planning as heuristic search, *BLAI BONET, HÉCTOR GEFFNER*
- LAO*: A heuristic search algorithm that finds solutions with loops, *ERIC A. HANSEN, SHLOMO ZILBERSTEIN*
- Searching stochastically generated multi-abstraction-level design spaces, *LOUIS STEINBERG*
- A general scheme for automatic generation of search heuristics from specification dependencies, *KALEV KASK, RINA DECHTER*
- Exploiting symmetries within constraint satisfaction search, *PEDRO MESEGUER, CARME TORRAS*
- Minimax real-time heuristic search, *SVEN KOENIG*
- Time complexity of iterative-deepening-A*, *RICHARD E. KORF, MICHAEL REID, STEFAN EDELKAMP*
- Sokoban: Enhancing general single-agent search methods using domain knowledge, *ANDREAS JUNGHANNS, JONATHAN SCHAEFFER*
- The PN*-search algorithm: Application to tsume-shogi, *MASAHIRO SEO, HIROYUKI IIDA, JOS W.H.M. UITERWIJK*
- Partial order bounding: A new approach to evaluation in game tree search, *MARTIN MÜLLER*

ARTIFICIAL INTELLIGENCE V130 N° 2 AUGUST 2001

- Influence-based model decomposition for reasoning about spatially distributed physical systems, *CHRIS BAILEY-KELLOGG, FENG ZHAO*
- Robust combinatorial auction protocol against false-name bids, *MAKOTO YOKOO, YUKO SAKURAI, SHIGEO MATSUBARA*

- The MIT Encyclopedia of the Cognitive Sciences, edited by Robert Wilson and Frank Keil, *CATHERINE CARR*
- Review of Natural Language Processing in R.A. Wilson and F.C. Keil (Eds.), The MIT Encyclopedia of the Cognitive Sciences, *BONNIE JEAN DORR*
- Cognitive Science: Real or Imaginary?: Review of The MIT Encyclopedia of The Cognitive Sciences edited by Robert A. Wilson and Frank C. Keil, *PHIL HUSBANDS*
- As advertised: A review of The MIT Encyclopedia of the Cognitive Sciences, *GEORGE LAKOFF*
- Review of R.A. Wilson and F.C. Keil (Eds.), The MIT Encyclopedia of the Cognitive Sciences, *YOSHI OKAMOTO*
- The mind's new labels?: Review of R.A. Wilson and F.C. Keil (Eds.), The MIT Encyclopedia of the Cognitive Sciences, *DONALD M. PETERSON*
- The scope of the cognitive sciences: Reply to 6 reviews of The MIT Encyclopedia of the Cognitive Sciences, *F.C. KEIL*
- The Cognitive Sciences: A comment on 6 reviews of The MIT Encyclopedia of the Cognitive Sciences, *R.A. WILSON*

ARTIFICIAL INTELLIGENCE V131 N° 1-2 SEPTEMBER 2001

- The Qualification Problem: A solution to the problem of anomalous models, *MICHAEL THIELSCHER*
- Evaluating general purpose automated theorem proving systems, *GEOFF SUTCLIFFE, CHRISTIAN SUTTNER*
- Dynamic tunneling based regularization in feedforward neural networks, *Y.P. SINGH, PINAKI ROYCHOWDHURY*
- Planning the project management way: Efficient planning by effective integration of causal and resource reasoning in RealPlan, Biplav Srivastava, *SUBBARAO KAMBHAMPATI, MINH B. DO*
- A comparison of point-based approaches to qualitative temporal Reasoning, *JAMES DELGRANDE, ARVIND GUPTA, TIM VAN ALLEN*
- On functional dependencies in q-Horn theories, *TOSHIHI-DE IBARAKI, ALEXANDER KOGAN, KAZUHISA MAKINO*
- "Field Reviews": A new style of review article for Artificial Intelligence, *A.G. COHN, D. PERLIS*
- Two machine learning textbooks: An instructor's perspective, *ERNEST DAVIS*
- On the state of the art in machine learning: A personal review, *PETER A. FLACH*
- Author's response to reviews of Machine Learning, *TOM MITCHELL*

ARTIFICIAL INTELLIGENCE V132 N°1 OCTOBER 2001

- Artificial nonmonotonic neural networks, *B. BOUTSINAS, M.N. VRAHATIS*
- Computer Go: An AI oriented survey, *BRUNO BOUZY, TRISTAN CAZENAVE*
- Determining if (FC-) (conflict-directed) backjumping visits a given node is NP-hard, *BERND S.W. SCHRÖDER*

ARTIFICIAL INTELLIGENCE V132 N°2 NOVEMBER 2001

- Local search characteristics of incomplete SAT procedures, *DALE SCHUURMANS, FINNEGAN SOUTHEY*
- Planning as constraint satisfaction: Solving the planning

graph by compiling it into *CSP*, *MINH BINH DO*, *SUBBARAO KAMBHAMPATI*

- Bargaining with limited computation: Deliberation equilibrium, *KATE LARSON*, *TUOMAS SANDHOLM*
- Rough approximation quality revisited, *GÜNTHER GEDIGA*, *IVO DÜNTSCH*

AI MAGAZINE

<http://aaai.org/Magazine/>

AI MAGAZINE V22 N°2 2001

- Human-level AI's killer application interactive computer games, *LAIRD J.E.*, *VAN LENT M.*
- SciFINANCE: A program synthesis tool for financial modelling, *AKERS R.L.*, *BICA I.*, *KANT E.*, *RANDALL C.*, *YOUNG R.L.*
- An innovative application from the DARPA knowledge bases programs: Rapid development of a course of action critiquer, *TECUCI G.*, *BOICU M.*, *BOWMAN M.*, *MARCU D.*, *BURKE M.*
- A. Knowledge portals: Ontologies at work, *STAAB S.*, *MAEDCHE*
- Lifecode: A deployed application for automated medical coding, *HEINZE D.T. ET AL.*
- Personalized electronic program guides for digital TV, *SMYTH B.*, *COTTER P.*

COGNITIVE SCIENCE

<http://www.elsevier.nl/gej->

[ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1](http://www.elsevier.nl/gej-)

COGNITIVE SCIENCE V25 N°3 MAY 2001

- Modeling individual differences in working memory performance: a source activation account, *LARRY Z. DAILY*, *MARSHA C. LOVETT*, *LYNNE M. REDER*
- Detecting deception: adversarial problem solving in a low base-rate world, *PAUL E. JOHNSON*, *STEFANO GRAZIOLI*, *KARIM JAMAL*, *R. GLEN BERRYMAN*
- Systematic error in the organization of physical action, *C.B. WALTER*, *S.P. SWINNEN*, *N. DOUNSKAIA*, *H. VAN LANGENDONK*
- Is perception informationally encapsulated? - The issue of the theory-ladenness of perception, *ATHANASSIOS RAFTOPOULOS*
- Multidisciplinarity and cognitive science, *BARBARA VON ECKARDT*

COGNITIVE SCIENCE V25 N°4 JULY 2001

- Learning from human tutoring, *Micheline*, *T.H. CHI*, *STEPHANIE A. SILER*, *HEISAWN JEONG*, *TAKASHI YAMAUCHI*, *ROBERT G. HAUSMANN*
- Cognitive templates for religious concepts: cross-cultural evidence for recall of counter-intuitive representations, *PASCAL BOYER*, *CHARLES RAMBLE*
- Naive causality: a mental model theory of causal meaning and Reasoning, *EUGENIA GOLDBVARG*, *P.N. JOHNSON-LAIRD*
- Convention in joint activity, *RICHARD ALTERMAN*, *ANDREW GARLAN*

COGNITIVE SCIENCE V25 N°5 SEPTEMBER 2001

- New approaches to the neural basis of speech sound processing: introduction to special section on brain and speech, *DAVID POEPEL*
- Role of left posterior superior temporal gyrus in phonological processing for speech perception and production, *BRADLEY R. BUCHSBAUM*, *GREGORY HICKOK*, *COLIN HUMPHRIES*
- Pure word deafness and the bilateral processing of the speech Code, *DAVID POEPEL*
- The role of inferior frontal cortex in phonological processing, *MARTHA W. BURTON*
- Levels of representation in the electrophysiology of speech Perception, *COLIN PHILLIPS*
- Visualizing the emergent structure of children's mathematical Argument, *DOLORES STROM*, *VERA KEMENY*, *RICHARD LEHRER*, *ELLICE FORMAN*
- The bicoherence theory of situational irony, *CAMERON SHELLEY*
- Classifying adults' and children's faces by sex: computational investigations of subcategorical feature encoding, *YI D. CHENG*, *ALICE J. O'TOOLE*, *HERVÉ ABDI*

APPLIED ARTIFICIAL INTELLIGENCE

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

APPLIED ARTIFICIAL INTELLIGENCE V15 N°5 MAY 2001

- Efficient methods for computing investment strategies for multi-market commodity trading, *MILOS HAUSKRECHT*, *LUIS ORTIZ*, *IOANNIS TSOCHANTARIDIS*, *ELI UPFAL*
- Knowledge base specification to automate the fluid critical point of Fluids, *JOSEFINA BARRERA-CORTÉS CINVESTAV*, *JEAN PIERRE ASTRUC*, *ROLAND TUFEU*, *ANA MARÍA MARTÍNEZ-ENRIQUEZ*
- Integrating planning and scheduling, *A. GARRIDO*, *F. BARBER*

APPLIED ARTIFICIAL INTELLIGENCE V15 N°7 AUGUST 2001

- A novel hybrid genetic algorithm using local optimizer based on heuristic pattern move, *Y. G. XU*, *G. R. LI*, *Z. P. WU*
- Cooling load prediction in a district heating and cooling system through simplified robust filter and multilayered neural network, *MASATOSHI SAKAWA*, *KOSUKE KATO*, *SATO-SHI USHIRO*
- A new method for generating fuzzy rules from numerical data for handling Classification problems, *SHYI-MING CHEN*, *SHAO-HUA LEE*, *CHIA-HOANG LEE*
- Personalized web-document filtering using reinforcement learning, *BYOUNG-TAK ZHANG*, *YOUNG-WOO SEO*

APPLIED ARTIFICIAL INTELLIGENCE V15 N°8 SEPTEMBER 2001

- Machine learning in computer vision, *FLORIANA ESPOSITO*, *DONATO MALERBA*
- Learning spatio-temporal relational structures, *WALTER F. BISCHOF*, *TERRY CAELLI*
- A vision system for symbolic interpretation of dynamic

SOMMAIRES DES REVUES

scenes using *ARSOM*, *ANTONIO CHELLA*, *DONATELLA GUARINO*, *IGNAZIO INFANTINO*, *ROBERTO PIRRONE*

- A trainable system for grading fish from images, *FRANCESCA Odone*, *EMANUELE TRUCCO*, *ALESSANDRO VERRI*
- Improving the accuracy of decision tree induction by feature preselection, *PETRA PERNER*
- A model of abstraction in visual perception, *LORENZA SAIITA*, *JEAN-DANIEL ZUCKER*
- Vision system development by machine learning: mashing assessment in brewing, *ANDY SONG*, *VIC CIESIELSKI*, *PETER ROGERS*

COMPUTATIONAL INTELLIGENCE

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

COMPUTATIONAL INTELLIGENCE V17 N°2

- Introduction To the Special Issue on Maintaining Case-Based Reasoning Systems, *DAVID B. LEAKE*, *BARRY SMYTH*, *DAVID C. WILSON*, *QIANG YANG*
- Maintaining Case-Based Reasoners: Dimensions and Directions, *DAVID C. WILSON*, *DAVID B. LEAKE*
- Review and Restore for Case-Base Maintenance, *THOMAS REINARTZ*, *IOANNIS IGLEZAKIS*, *THOMAS ROTH-BERGHOFER*
- Competence Models and the Maintenance Problem, *BARRY SMYTH*, *ELIZABETH MCKENNA*
- A Case-Addition Policy for Case-Base Maintenance, *QIANG YANG*, *JUN ZHU*
- Case-Base Maintenance in a Multimodal Reasoning System, *LUIGI PORTINALE*, *PIETRO TORASSO*
- Case-Base Maintenance By Integrating Case-Index Revision and Case-Retention Policies in a Derivational Replay Framework, *HÉCTOR MUÑOZ-AVILA*
- Transferring Case Knowledge To Adaptation Knowledge: An Approach for Case-Base Maintenance, *SIMON C. K. SHIU*, *DANIEL S. YEUNG*, *CAI H. SUN*, *XI Z. WANG*
- Distributing Case-Base Maintenance: The Collaborative Maintenance Approach, *MARIA ANGELA FERRARIO*, *BARRY SMYTH*
- Intelligent Case-Authoring Support in Casemaker-2, *DAVID MCSHERRY*
- Maintaining Retrieval Knowledge in a Case-Based Reasoning System, *SUSAN CRAW*, *JACEK JARMULAK*, *RAY ROWE*
- Systematic Maintenance of Corporate Experience Repositories, *MARKUS NICK*, *KLAUS-DIETER ALTHOFF*, *CARSTEN TAUTZ*
- A Case Study of Maintenance of a Commercially Fielded Case-Based Reasoning System, *IAN WATSON*

COMPUTATIONAL INTELLIGENCE V17 N°3

- Combining Rough Sets and Bayes' Rule, *ZDZISELSTROKAW PAWLAK*
- Rule Quality Measures for Rule Induction Systems: Description and Evaluation, *AIJUN AN*, *NICK CERCONI*
- Coping With Missing Attribute Values Based on Closest Fit

in Preterm Birth Data: A Rough Set Approach, *JERZY W. GRZYMALA-BUSSE*, *WITOLD J. GRZYMALA-BUSSE*, *LINDA K. GOODWIN*

- Applications of Rough Genetic Algorithms, *PAWAN LINGRAS*, *CEDRIC DAVIES*
- Rough Problem Settings for ILP Dealing With Imperfect Data, *CHUNNIAN LIU*, *NING ZHONG*
- The Variable Precision Rough Set Inductive Logic Programming Model and Strings, *V. UMA MAHESWARI*, *ARUL SIROMONEY*, *K. M. MEHATA*, *K. INOUE*
- Rough Mereological Calculi of Granules: A Rough Set Approach To Computation, *LECH POLKOWSKI*, *ANDRZEJ SKOWRON*
- Rough Neural Computing in Signal Analysis, *J. F. PETERS*, *L. HAN*, *S. RAMANNA*
- Granular Computing: a Rough Set Approach, *SON H. NGUYEN*, *ANDRZEJ SKOWRON*, *JAROSŁAW STEPIANIUK*
- Incomplete Information Tables and Rough Classification, *JERZY STEFANOWSKI*, *ALEXIS TSOUKIAS*
- Interval Evaluations in the Analytic Hierarchy Process By Possibility Analysis, *KAZUTOMI SUGIHARA*, *HIDEO TANAKA*
- Evolving Granules for Classification for Discovering Difference in the Usage of Words, *TETSUYA YOSHIDA*, *TERUYUKI KONDO*, *SHOGO NISHIDA*
- Probabilistic Decision Tables in the Variable Precision Rough Set Model, *WOJCIECH ZIARKO*

MINDS AND MACHINES

JOURNAL FOR ARTIFICIAL INTELLIGENCE PHILOSOPHY AND COGNITIVE SCIENCE

<http://www.wkap.nl/issuetoc.htm/0924-6495>

ARTIFICIAL INTELLIGENCE REVIEW

http://www.elsevier.nl/jeing/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

ARTIFICIAL INTELLIGENCE REVIEW V15 N°4 JUNE 2001

- Solution techniques for constraint satisfaction problems: Foundations, *MIGUEL I.*, *SHEN Q.*
- Solution techniques for constraint satisfaction problems: Advanced approaches, *MIGUEL I.*, *SHEN Q.*
- Representation, coherence and inference, *SMITH E.*, *HANCOX P.*

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

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

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V54 N°6 JUNE 2001

- Cognitive activities in OO development, *DAVID P. TEGARDEN*, *STEVEN D. SHEETZ*
- Are visual programming languages better? The role of ima-

gery in program comprehension, *RAQUEL NAVARRO-PRIENTO, JOSE J. CAÑAS*

- How do operators monitor a complex, dynamic work domain? The impact of control room technology, *KIM J. VICENTE, EMILIE M. ROTH, RANDALL J. MUMAW*
- The impact of status and audio conferencing technology on business meetings, *EMMA F. FRANCE, ANNE H. ANDERSON, MICHAEL GARDNER*
- Assessing users' subjective quality of experience with the world wide web: an exploratory examination of temporal changes in technology acceptance, *MICHAEL G. MORRIS, JASON M. TURNER*
- What do web users do? An empirical analysis of web use, *ANDY COCKBURN, BRUCE MCKENZIE*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V55 N°1 JULY 2001

- Special Issue: Cognitive Modelling and HCI, Guest Editor: Editors: *F. E. RITTER R. M. YOUNG*
- Embodied models as simulated users: introduction to this special issue on using cognitive models to improve interface design, *FRANK E. RITTER, RICHARD M. YOUNG*
- A perception/action substrate for cognitive modeling in HCI, *ROBERT ST AMANT, MARK O. RIEDL*
- ACT-R/PM and menu selection: applying a cognitive architecture to HCI, *MICHAEL D. BYRNE*
- Predicting the effects of in-car interface use on driver performance: an integrated model approach, *DARIO D. SALVUCCI*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V55 N°2 AUGUST 2001

- Editorial: User centred design and implementation of virtual environments, *SHAMUS P. SMITH, MICHAEL D. HARRISON*
- Informing the design of a virtual environment to support learning in children, *MIKE SCAIFE, YVONNE ROGERS*
- A toolset supported approach for designing and testing virtual environment interaction techniques, *JAMES S. WILLANS, MICHAEL D. HARRISON*
- Theme-based content analysis: a flexible method for virtual environment evaluation, *HELEN NEALE, SARAH NICHOLS*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V55 N°3 SEPTEMBER 2001

- Expertise and the interpretation of computerized physiological data: implications for the design of computerized monitoring in neonatal intensive care, *EUGENIO ALBERDI, JULIE-CLARE BECHER, KEN GILHOOLY, JIM HUNTER, ROBERT LOGIE, ANDY LYON, NEIL MCINTOSH, JAN REISS*
- Creativity in the organization: the role of individual creative problem solving and computer support, *MARILYN G. KLETKE, JANE M. MACKAY, STEVE H. BARR, BEATA JONES*
- Explanation and exploration Visualizing the topology of web information systems, *ARNO SCHARL*
- Recognizing Thai handwritten characters and words for

human-computer interaction, *CHOMTIP PORNPANOM-CHAI, DENTCHO N. BATANOV, NICHOLAS DIMMITT*

- Interactive machine learning: letting users build classifiers, *MALCOLM WARE, EIBE FRANK, GEOFFREY HOLMES, MARK HALL, IAN H. WITTEN*
- Relational attribute systems, *IVO DÜNTSCH, GÜNTHER GEDIGA, EWA ORLOWSKA*
- Increasing the visualization realism by frame synchronization between the VRML browser and the panoramic image viewer, *JIUNG-YAO HUANG*
- A methodological approach to supporting organizational learning, *PAUL MULHOLLAND, ZDENEK ZDRAHAL, JOHN DOMINGUE, MAREK HATALA, ANSGAR BERNARDI*
- Erratum: Interacting with the telephone, *HAZEL LACOHÉE, BEN ANDERSON*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V55 N°4 OCTOBER 2001

- Editorial: Human Factors Symposium, *L. DAMODARAN, B. SHACKEL*
- Human factors in the digital world enhancing life style-the challenge for emerging technologies, *LEELA DAMODARAN*
- Requirements trawling: techniques for discovering requirements, *SUZANNE ROBERTSON*
- An approach to requirements analysis for decision support systems, *CAROLINE PARKER*
- New product development by eliciting user experience and aspirations, *ANNE BRUSEBERG, DEANA MCDONAGH-PHILP*
- Context of Use within usability activities, *MARTIN MAGUIRE*
- Prototyping for usability of new technology, *ROGER R. HALL*
- Talking and writing-how natural in human-machine interaction?, *JAN NOYES*
- Ubiquitous computing within cars: designing controls for non-visual use, *GARY E. BURNETT, J. MARK PORTER*
- International standards for HCI and usability, *NIGEL BEVAN*
- The improvement of human-centred processes-facing the challenge and reaping the benefit of ISO 13407, *JONATHAN EARTHY, BRIAN SHERWOOD JONES, NIGEL BEVAN*
- Third age usability and safety-an ergonomics contribution to design, *DAVID R. HITCHCOCK, SUZANNE LOCKYER, SHARON COOK, CLAIRE QUIGLEY*
- Empowered participation of users with disabilities in R&D projects, *CHRISTIAN BÜHLER*
- Evaluating the human-machine interface to vehicle navigation systems as an example of ubiquitous computing, *TRACY ROSS, GARY BURNETT*
- Re-viewing reality: human factors of synthetic training environments, *ALEX W. STEDMON, ROBERT J. STONE*
- Virtual reality for interactive training: an industrial practitioner's viewpoint, *ROBERT STONE*
- Human factors-the challenges facing Europe, *RONALD MACKAY*
- Luddism for the twenty-first century, *LISL KLEIN*

COMPUTATIONAL LINGUISTICS

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

COMPUTATIONAL LINGUISTICS V26 N°4 DECEMBER 2000

- Automatic Text Categorization In Terms Of Genre and Author, *EFSTATHIOS STAMATATOS, NIKOS FAKOTAKIS, GEORGE KOKKINAKIS*
- Bidirectional Contextual Resolution, *STEPHEN G. PULMAN*
- An Empirically-based System for Processing Definite Descriptions, *RENATA VIEIRA, MASSIMO POESIO*
- Learning Methods to Combine Linguistic Indicators: Improving Aspectual Classification and Revealing Linguistic Insights, *ERIC V. SIEGEL, KATHLEEN R. MCKEOWN*

- On Coreferring: Coreference in MUC and Related Annotation Schemes, *KEES VAN DEEMTER, RODGER KIBBLE*

Book review

- 'Speech and Language Processing' by Daniel Jurafsky and James H. Martin, *VIRGINIA TELLER*
- 'Artificial Intelligence and Literary Creativity' by Selmer Bringsjord and David A. Ferrucci, *RONALD DE SOUSA*
- 'Architectures and Mechanisms for Language Processing' edited by Matthew W. Crocker, Martin Pickering, and Charles Clifton, JR., *AMY WEINBERG*
- 'Breadth and Depth of Semantic Lexicons' edited by Evelyne Viegas, *JOHN S. WHITE*

COMPUTATIONAL LINGUISTICS V27 N° 1 MARCH 2001

- Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus, *MIKIO YAMAMOTO, KENNETH W. CHURCH*
- Integrating Prosodic and Lexical Cues for Automatic Topic Segmentation, *G. TUR, D. HAKKANI-TUR, A. STOLCKE, E. SHRIBERG*
- Bootstrapping Morphological Analyzers by Combining Human Elicitation and Machine Learning, *KEMAL OFLAZER, SERGEI NIRENBURG, MARJORIE MCSHANE*
- D-Tree Substitution Grammars, *OWEN RAMBOW, K. VIJAY-SHANKER, DAVID WEIR*
- Unsupervised Named Entity Recognition Using Syntactic and Semantic Contextual Evidence, *ALESSANDRO CUCCHIARELLI, PAOLA VELARDI*

Book review

- 'Longman Grammar of Spoken and Written English' by Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan, *GRAEME HIRST*
- 'Construing Experience through Meaning: A language-based approach to cognition' by M.A.K. Halliday and Christian M.I.M. Matthiessen, *JOHN F. SOWA*
- 'Computing Meaning, volume 1' edited by Harry Bunt and Reinhard Muskens, *YOAD WINTER*
- 'The Syntactic Process' by Mark Steedman, *JOAKIM NIVRE*

COMPUTATIONAL LINGUISTICS V27 N° 2 JUNE 2001

- Unsupervised Learning of the Morphology of a Natural Language, *JOHN GOLDSMITH*
- Improving Accuracy in NLP Through Combination of Machine Learning Systems, *HANS VAN HALTEREN, JAKUB ZAVREL, WALTER DAELEMANS*
- The Need for Accurate Alignment in Natural Language System Evaluation, *ANDY KEHLER, JOHN BEAR, DOUGLAS APPELT*
- Probabilistic Top-Down Parsing and Language Modeling, *BRIAN ROARK*
- Nonminimal Derivations in Unification-Based Parsing, *NORIKO TOMURO, STEVEN L. LYTIMEN*

Book review

- James Rogers, A Descriptive Approach to Language-Theoretic Complexity, *PHILIP MILLER, GEOFFREY K. PULLUM*
- John F. Sowa, Representation: Logical, Philosophical, and Computational Foundations, *STUART C. SHAPIRO*
- Edited by Lucja M. Iwanska and Stuart C. Shapiro, Natural Language Processing and Knowledge Representation, *ROBERT E. MERCER*
- Ehud Reiter and Robert Dale, Building Natural Language Generation Systems, *HELMUT HORACEK*
- Edited by W. Bruce Croft, Advances in Information Retrieval, *SANDA HARABAGIU*
- Leonard Talmy, Toward a Cognitive Semantics, *KEITH ALLAN*
- Bruce Tesar and Paul Smolensky, Learnability in Optimality Theory, *WALTER DAELEMANS*

NATURAL LANGUAGE ENGINEERING

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

NATURAL LANGUAGE ENGINEERING V7 N°2 JUNE 2001

- Randomized rule selection in transformation-based learning: a comparative study, *SANDRA CARBERRY, K. VIJAY-SHANKER, ANDREW WILSON, KEN SAMUEL*
- A natural language system for retrieval of captioned images, *DAVID ELWORTHY, TONY ROSE, AMANDA CLARE, AARON KOTCHEFF*
- Real-time automatic insertion of accents in French text, *MICHEL SIMARD, ALEXANDRE DESLAURIERS*
- Compound noun segmentation based on lexical data extracted from corpus, *JUNTAE YOON*

Book Review

- Sake Jager, John Nerbonne and Arthur van Essen (eds). Language Teaching and Language Technology. Swets & Zeitlinger, 1998, *MAYUMI MASUKO*

NATURAL LANGUAGE ENGINEERING V7 N°3 SEPTEMBER 2001

- Scalable generation of texts using causal and temporal expansions of sentences, *YLLIAS CHALI*
- Applied morphological processing of English, *GUIDO MINNEN, JOHN CARROLL, DARREN PEARCE*

SOMMAIRES DES REVUES

• ILEX: an architecture for a dynamic hypertext generation system, *M. O'DONNELL, C. MELLISH, J. OBERLANDER, A. KNOTT*

• A corpus-based approach for Korean nominal compound analysis based on linguistic and statistical information, *JUN-TAE YOON, KEY-SUN CHOI, MANSUK SONG*

Book Review

• Ehud Reiter and Robert Dale. Building Natural Language Generation Systems. Cambridge University Press, 2000, *ADVAITH SIDDHARTHAN*

• Inderjeet Mani and Mark T. Maybury (eds). Advances in Automatic Text Summarization. MIT Press, 1999, *ADVAITH SIDDHARTHAN*

USER MODELING AND USER-ADAPTED INTERACTION

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

USER MODELING AND USER-ADAPTED INTERACTION V11 N°3 AUGUST 2001

• Information Filtering: Overview of Issues, Research and Systems, *URI HANANI, BRACHA SHAPIRA, PERETZ SHOVAL*

Book review

• M. Kyng and L. Mathiassen (eds.), Computers and Design in Context, *CAROL STROHECKER*

USER MODELING AND USER-ADAPTED INTERACTION V11 N°4 NOVEMBER 2001

• Preface: Towards Adaptation of Interaction to Affective Factors, *FIORELLA DE ROSIS*

• How Convincing is Mr. Data's Smile: Affective Expressions of Machines, *CHRISTOPH BARTNECK*

• Modeling Emotion and Attitude in Speech by Means of Perceptually Based Parameter Values, *SYLVIE J. L. MOZZI-CONACCI*

COMPUTER SPEECH AND LANGUAGE

<http://www.idealibrary.com/links/toc/csla/latest>

COMPUTER SPEECH & LANGUAGE V15 N°3 JULY 2001

• Enhanced speech recognition using an articulatory production model trained on X-ray data, *C. SIMON BLACKBURN, STEVE YOUNG*

• Union: a model for partial temporal corruption of speech, *JJ MING, F. JACK SMITH*

• Applying dynamic context into MLP/HMM speech recognition system, *PETRI SALMELA*

• Maximum likelihood stochastic transformation adaptation for medium and small data sets, *CONSTANTINOS BOULIS, VASSILIOS DIAKOLOUKAS, VASSILIOS DIGALAKIS*

• Normalization of non-standard words, *RICHARD SPROAT, ALAN W. BLACK, STANLEY CHEN, SHANKAR KUMAR, MARI OSTENDORF, CHRISTOPHER RICHARDS*

MACHINE LEARNING

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

MACHINE LEARNING V44 N°3 SEPTEMBER 2001

• Editorial: Inductive Logic Programming is Coming of Age, *PETER FLACH, SASO DZEROSKI*

• On Exact Learning of Unordered Tree Patterns, *THOMAS R. AMOTH, PAUL CULL, PRASAD TADEPALLI*

• Parameter Estimation in Stochastic Logic Programs, *JAMES CUSSENS*

• Approximate Match of Rules Using Backpropagation Neural Networks, *BOONSERM KIJSIRIKUL, SUKREE SINTHUPINYO, KONGSAK CHONGKASEMWONGSE*

• Extracting Context-Sensitive Models in Inductive Logic Programming, *ASHWIN SRINIVASAN*

MACHINE LEARNING V45 N°1 OCTOBER 2001

• Random Forests, *LEO BREIMAN*

• On the VC Dimension of Bounded Margin Classifiers, *DON HUSH, CLINT SCOVEL*

• Iterated Phantom Induction: A Knowledge-Based Approach to Learning Control, *MARK BRODIE, GERALD DEJONG*

• Optimizing Epochal Evolutionary Search: Population-Size Dependent Theory, *ERIK VAN NIMWEGEN, JAMES P. CRUTCHFIELD*

MACHINE LEARNING V45 N°2 NOVEMBER 2001

• Learning with Maximum-Entropy Distributions, *YISHAY MANSOUR, MARIANO SCHAIN*

• Robust Learning with Missing Data, *MARCO RAMONI, PAOLA SEBASTIANI*

• A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems, *DAVID J. HAND, ROBERT J. TILL*

• Predicting the Future of Discrete Sequences from Fractal Representations of the Past, *PETER TINO, GEORG DORFFNER*

• Reinterpreting the Category Utility Function, *BORIS MIRKIN*

MACHINE LEARNING V45 N°3 DECEMBER 2001

• Efficient Construction of Regression Trees with Range and Region Splitting, *YASUHIKO MORIMOTO, HIROMU ISHII, SHINICHI MORISHITA*

• Using Iterated Bagging to Debias Regressions, *LEO BREIMAN*

• Accelerating EM for Large Databases, *BO THIESSON, CHRISTOPHER MEEK, DAVID HECKERMAN*

• Relative Loss Bounds for Multi-dimensional Regression Problems, *J. KIVINEN, M. K. WARMUTH*

NEURAL NETWORKS

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

NEURAL NETWORKS V14 N°4-5 MAY 2001

• Quantitative examinations for multi joint arm trajectory planning using a robust calculation algorithm of the minimum

commanded torque change trajectory, *YASUHIRO WADA, YUICHI KANEKO, ERI NAKANO, RIEKO OSU, MITSUO KAWATO*

- Solving the binding problem of the brain with bi-directional functional connectivity, *MASATAKA WATANABE, KOUSAKU NAKANISHI, KAZUYUKI AIHARA*
 - Coronado Learning from noisy information in FasArt and FasBack neuro-fuzzy systems, *JOSÉ MANUEL CANO IZQUIERDO, YANNIS A. DIMITRIADIS, EDUARDO GÓMEZ SÁNCHEZ, JUAN LÓPEZ*
 - Comparing Bayesian neural network algorithms for classifying segmented outdoor images, *FRANCESCO VIVARELLI, CHRISTOPHER K. I. WILLIAMS*
 - Three learning phases for radial-basis-function networks, *FRIEDHELM SCHWENKER, HANS A. KESTLER, GÜNTHER PALM*
 - Noise suppression in training examples for improving generalization capability, *AKIKO NAKASHIMA, HIDEIMITSU OGAWA*
 - Networks with trainable amplitude of activation Functions, *EDMONDO TRENTIN*
 - A model with an intrinsic property of learning higher order Correlations, *MARIFI GÜLER*
 - S-TREE: self-organizing trees for data clustering and online vector quantization, *MARCOS M. CAMPOS, GAIL A. CARPENTER*
 - The constraint based decomposition (CBD) training Architecture, *SORIN DRGHICI*
 - Life-long learning Cell Structures-continuously learning without catastrophic interference, *FRED H. HAMKER*
- ∑ Pattern classification by a condensed neural Network, *A. MITICHE, M. LEBIDOFF*

NEURAL NETWORKS V14 N°8 OCTOBER 2001

- Global exponential stability of delayed Hopfield neural Networks, *TIANPING CHEN*
 - A neurodynamical model for selective visual attention using oscillators, *SILVIA CORCHS, GUSTAVO DECO*
 - Information processing in dendrites - I. Input pattern Generalisation, *KEVIN N. GURNEY*
 - Information processing in dendrites - II. Information theoretic complexity, *KEVIN N. GURNEY*
 - A Modified General Regression Neural Network (MGRNN) with new, efficient training algorithms as a robust 'black box'-tool for data analysis, *DIRK TOMANDL, ANDREAS SCHOBER*
 - An approach to guaranteeing generalisation in neural networks, *J. GARY POLHILL, MICHAEL K. WEIR*
 - Algebraic geometrical methods for hierarchical learning Machines, *SUMIO WATANABE*
 - A hybrid learning network for shift-invariant recognition, *RUYE WANG*
 - A dynamical model for the analysis and acceleration of learning in feedforward networks, *NIKOLAOS AMPAZIS, STAVROS J. PERANTONIS, JOHN G. TAYLOR*
 - Estimates of average complexity of neurocontrol algorithms, *TOMAS HRYCEJ*
- vFuzzylot: a novel self-organising fuzzy-neural rule-based

pilot system for automated vehicles, *MICHEL PASQUIER, CHAI QUEK, MARY TOH*

vReal time distributed processing of multiple associated pulse pattern sequences, *A.J.B. TRAVIS*

- Myopotential denoising of ECG signals using wavelet thresholding methods, *VLADIMIR CHERKASSKY, STEVEN KILTS*

NEURAL NETWORKS V14 N°9 NOVEMBER 2001

- Neuronal integration mechanisms have little effect on spike auto-correlations of cortical neurons, *YUTAKA SAKAI*
- Best estimated inverse versus inverse of the best estimator, *AMIR KARNIEL, RON MEIR, GIDEON F. INBAR*
- How to be a gray box: dynamic semi-physical modelling, *YACINE OUSSAR, GÉRARD DREYFUS*
- An Infomax-based learning rule that generates cells similar to visual cortical neurons, *K. OKAJIMA*
- On the stability analysis of delayed neural networks systems, *CHUNHUA FENG, RÉJEAN PLAMONDON*
- A two-level Hamming network for high performance associative memory, *NOBUHIKO IKEDA, PAUL WATTA, METIN ARTIKLAR, MOHAMAD H. HASSOUN*
- A closed-form neural network for discriminatory feature extraction from high-dimensional data, *ASHIT TALUKDER, DAVID CASASENT*
- The enhanced LBG algorithm, *GIUSEPPE PATANÉ, MARCO RUSSO*
- Reconstruction of chaotic dynamics by on-line EM Algorithm, *S. ISHII, M.-A. SATO*
- Novelty detection using products of simple experts a potential architecture for embedded systems, *ALAN F. MURRAY*
- A new algorithm to design compact two-hidden-layer artificial neural networks, *MD. MONIRUL ISLAM, K. MURASE*
- Cross-validation in Fuzzy ARTMAP for large databases, *ANNA KOUFAKOU, MICHAEL GEORGIPOULOS, GEORGE ANAGNOSTOPOULOS, TAKIS KASPARIS*
- Fingerprints classification using artificial neural networks: a combined structural and statistical approach, *KHALED AHMED NAGATY*
- Bi-directional computing architecture for time series prediction, *HIROSHI WAKUYA, JACEK M. ZURADA*

Book review

- Advances in Synaptic Plasticity: A Compact Account of the New, the Important, and the Interesting - Edited by Michel Baudry, Joel L. Davis, and Richard F. Thompson. The MIT Press, 1999, *MURAT OKATAN*

ARTIFICIAL INTELLIGENCE IN MEDECINE

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

ARTIFICIAL INTELLIGENCE IN MEDICINE V22 N°3 JUNE 2001

- A new virtual reality approach for planning of cardiac interventions, *THOMAS SANGILD SØRENSEN, SØREN VORRE THERKILDSEN, PIOTR MAKOWSKI, JØRGEN LINDSKOV KNUDSEN, ERIK MORRE PEDERSEN*

- Data mining for indicators of early mortality in a database of clinical records, *G. RICHARDS, V.J. RAYWARD-SMITH, P.H. SÖNKSEN, S. CAREY, C. WENG*
- Using Bayesian networks in the construction of a bi-level multi-classifier. A case study using intensive care unit patients data, *BASILIO SIERRA ET AL.*
- Logarithmic simulated annealing for X-ray diagnosis, *A. ALBRECHT, K. STEINHÖFEL, M. TAUPITZ, C.K. WONG*
- A formal framework of knowledge to support rational psychoactive drug selection, *DIRK VAN HYFTE, ARNOUD VAN DER MAAS, TIKMA TJANDRA-MAGA, PIETER DE VRIES ROBBÉ*

Book review

- Building Natural Language Generation Systems – Ehud Reiter and Robert Dale (Eds.), University of Aberdeen and Macquarie University, Cambridge University Press (Studies in Natural Language Processing), *HELMUT HORACEK*

ARTIFICIAL INTELLIGENCE IN MEDICINE V23 N°1 AUGUST 2001

- AIM portraits: tracing the evolution of artificial intelligence in medicine and predicting its future in the new millennium, *ELPIDA KERAVALOU, NADA LAVRA*
- AI in medicine on its way from knowledge-intensive to data-intensive systems, *WERNER HORN*
- Proactive health computing, *TOOMAS TIMPKA*
- The socio-organizational age of artificial intelligence in medicine, *MARIO STEFANELLI*
- Multiple representations and multi-modal reasoning in medical diagnostic systems, *PIETRO TORASSO*
- Medical informatics: reasoning methods, *WILLIAM J. LONG*
- Machine learning for medical diagnosis: history, state of the art and perspective, *IGOR KONONENKO*
- AIM: a personal view of where I have been and where we might be going, *ALAN RECTOR*

ARTIFICIAL INTELLIGENCE IN MEDICINE V23 N°2 OCTOBER 2001

- Managing uncertainty in diagnosis of acute coronaric ischemia, *P. BARONI, G. GUIDA, M. ZANELLA*
- Knowledge discovery approach to automated cardiac SPECT diagnosis, *LUKASZ A. KURGAN, KRZYSZTOF J. CIOS, RYSZARD TADEUSIEWICZ, MAREK OGIELA, LUCY S. GOODENDAY*
- Case-based reasoning for antibiotics therapy advice: an investigation of retrieval algorithms and prototypes, *RAINER SCHMIDT, LOTHAR GIERL*
- Feature subset selection by genetic algorithms and estimation of distribution algorithms - A case study in the survival of cirrhotic patients treated with TIPS, *I. INZA ET AL.*

Book review

- Foundations of Rational Agency - M. Wooldridge, A. RAO (Eds.), Kluwer Academic Publishers, Dordrecht, 1999, *HANS WEIGAND*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V27 N°1 JUNE 2001

- Fuzzy automaton induction using neural network, *A. BLANCO, M. DELGADO, M.C. PEGALAJAR*
- Information sharing between heterogeneous uncertain reasoning models in a multi-agent environment: a case study, *XUDONG LUO, CHENGQI ZHANG, HO-FUNG LEUNG*
- Fuzzy modeling by hierarchically built fuzzy rule bases, *OSCAR CORDÓN, FRANCISCO HERRERA, IGOR ZWIR*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V27 N°2 AUGUST 2001

- Special Issue on causal networks: papers from the Second CaNew Workshop, *RAMÓN SANGÜESA, ULISES CORTÉS*
- Bayesian model-based diagnosis, *PETER J.F. LUCAS*
- Accelerating chromosome evaluation for partial abductive inference in Bayesian networks by means of explanation set absorption, *LUIS M. DE CAMPOS, JOSÉ A. GÁMEZ, SERAFÍN MORAL*
- Feature subset selection by Bayesian networks: a comparison with genetic and sequential algorithms, *IÑAKI INZA, PEDRO LARRAÑAGA, BASILIO SIERRA*
- Learning Bayesian network parameters from small data sets: application of Noisy-OR gates, *AGNIESZKA ONIKO, MAREK J. DRUZDZEL, HANNA WASYLUK*
- Estimating probability values from an incomplete dataset, *SILVIA ACID, LUIS M. DE CAMPOS, JUAN F. HUETE*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V27 N°3 SEPTEMBER 2001

- Spatial relations between indeterminate regions, *A.J. ROY, J.G. STELL*
- A hybrid methodology for learning belief networks, *BENEDICT, SILVIA ACID, LUIS M. DE CAMPOS*
- Partial abductive inference in Bayesian belief networks by simulated annealing, *LUIS M. DE CAMPOS, JOSÉ A. GÁMEZ, SERAFÍN MORAL*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V28 N°1 OCTOBER 2001

- Automatic derivation of probabilistic inference rules, *MANFRED JAEGER*
- Performance evaluation of compromise conditional Gaussian networks for data clustering, *J.M. PEÑA, J.A. LOZANO, P. LARRAÑAGA*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V28 N°2-3 NOVEMBER 2001

- A methodology for computing with words, *JONATHAN LAWRY*
- Belief decision trees: theoretical foundations, *ZIED ELOUEDI, KHALED MELLOULI, PHILIPPE SMETS*
- Fuzzy decision support system knowledge base generation using a genetic algorithm, *LUC BARON, SOFIANE ACHICHE, MAREK BALAZINSKI*

Interaction et programmation par Catherine Letondal

Thèse de l'Université de Paris-Sud soutenue le 27 septembre 2001 à l'Institut Pasteur (Paris).

Résumé.

Le manque de flexibilité des nombreux outils d'analyse disponibles pour les biologistes limite souvent leur utilité. La solution la plus connue à ce problème est la programmation. Mais comment donner accès à cette discipline complexe à ceux des biologistes pour qui cela n'est pas un but en soi ?

Pour répondre à cette question nous avons d'abord constaté que la nature du problème - la programmation, son apprentissage et son utilisation effective dans le cadre de travail - n'était pas bien définie. L'objectif que nous devions nous fixer était-il de transformer des biologistes en informaticiens, ou bien de leur permettre de bénéficier de la programmation pour faciliter leur travail lorsque cela est nécessaire ? Nous pensons qu'à force de considérer la programmation d'une part comme un but en soi et d'autre part comme une activité parfaitement définie par la théorie informatique (Turing) et par les pratiques professionnelles (programmer c'est construire un logiciel), à force de constater qu'un logiciel est quelque chose qui s'utilise, mais qu'il est quasiment interdit d'en faire autre chose, il y a peut-être des solutions au problème que nous posons ici qui n'apparaissent pas aisément. Il nous a notamment semblé utile de re-situer l'activité de programmation comme une activité secondaire et non-professionnelle, ayant sa place dans le cadre d'une utilisation avancée et des interactions normales avec le logiciel, en tenant compte de l'idée que le travail de construction du logiciel ne peut qu'être partagé entre professionnels et experts du domaine de la biologie. C'est pour ces raisons et afin de répondre au problème de manière concrète et située dans un contexte que nous avons adopté une démarche de conception participative à travers l'organisation d'ateliers de brainstorming et maquettage (papier, vidéo, ...).

Nous sommes ainsi arrivés à l'idée d'application complètement programmable à partir de son interface, principe ayant pour avantage de constituer d'abord un outil de travail utilisable sans programmation, tout en donnant aux utilisateurs une liberté de décision, guidée mais totale, sur la manière dont l'outil fonctionne, c'est-à-dire son code, ainsi qu'un environnement d'apprentissage riche d'exemples centrés sur le travail réel de l'utilisateur et permettant une programmation incrémentale.

Un prototype, biok (*), a été réalisé. Il comporte des composants pour l'analyse de séquences comme un éditeur d'alignement, un outil d'affichage de courbes, etc... L'éditeur d'alignement fonctionne comme un tableur, et dispose d'un mécanisme programmable d'étiquetage graphique pour visualiser des propriétés biologiques. L'architecture de cet environnement repose sur la notion d'objet graphique, permettant la composition d'objets biologiques par des formules et l'accès structuré au code de l'application grâce à des techniques réflexives. L'originalité de ce prototype est d'ouvrir plusieurs niveaux d'utilisation/programmation au biologiste, allant de l'utilisation interactive à la programmation par objets en passant par la programmation de formules ou de scripts et de fonctions d'étiquetage graphique.

Mots clés : programmation par l'utilisateur, programmation

dans l'interface, conception participative, psychologie de la programmation, apprentissage de la programmation

Contact : Catherine Letondal, Institut Pasteur, Service d'Informatique Scientifique,
28, rue du Docteur Roux, 75724 Paris Cedex 15

Email :

letondal@pasteur.fr - <http://www-alt.pasteur.fr/~letondal/>

Adresse Web du document :

<http://www-alt.pasteur.fr/~letondal/>

Apprentissage supervisé pour la généralisation cartographique par Sébastien Mustière

Thèse d'informatique de l'université Paris VI, soutenue le 8 juin 2001

Résumé.

Cette thèse a pour contexte l'automatisation de la généralisation cartographique, processus de création d'une carte à partir d'une base de données géographique trop détaillée. Pour réaliser cela, de nombreux algorithmes existent pour transformer la géométrie des objets géographiques à représenter sur la carte, mais aucun d'entre eux n'est générique : chacun possède un but et un champ d'application propre. Dans cette thèse nous adoptons une approche pas à pas et focalisée, où le traitement d'un objet nécessite l'application de plusieurs algorithmes sur des espaces de travail adéquats. Dans ce contexte, il faut définir des règles permettant de choisir quels algorithmes appliquer sur un objet donné à partir de la description de celui-ci par un ensemble de mesures numériques. Un processus d'enchaînement des algorithmes est mis au point empiriquement pour la généralisation des routes. Ce processus, nommé GALBE, a en particulier été utilisé dans le processus de création des nouvelles cartes régionales de l'IGN. L'efficacité et les limites de ce processus conduisent à envisager l'utilisation de techniques développées en Intelligence Artificielle pour acquérir les connaissances nécessaires à un système expert cartographique : l'apprentissage automatique à partir d'exemples. Notre problème d'apprentissage se caractérise par la recherche de règles compréhensibles à partir d'exemples peu nombreux, bruités et avec une description riche. Un apprentissage classique produit alors des règles de faible qualité. Pour améliorer cela, nous guidons l'apprentissage par les connaissances du domaine cartographique en décomposant notre problème d'apprentissage en plusieurs sous-problèmes plus simples : nous apprenons tour à tour à abstraire puis à choisir la transformation à appliquer sur les objets géographiques manipulés. La phase d'abstraction consiste à transformer la représentation des objets sous la forme d'un ensemble restreint d'attributs symboliques. La phase de choix de transformation consiste à déterminer quel algorithme géométrique choisir en fonction de la description abstraite de l'objet. L'introduction de cette phase d'abstraction permet d'apprendre des règles cartographiques à la fois plus efficaces et plus compréhensibles qu'un apprentissage direct.

Contact : Sébastien Mustière

Institut Géographique National, Laboratoire COGIT,
2-4, avenue Pasteur,
94165 St-Mandé.

Tél : (33) 01 43 98 80 03 - Fax : (33) 01 43 98 81 71

Email : Sebastien.Mustiere@lip6.fr

Logique de la détermination d'objets : concepts de base et mathématisation en vue d'une modélisation-objet

Par Anca Pascu

Thèse de Doctorat de l'Université Paris-Sorbonne (Paris IV), soutenue le 30 juin 2001 en Sorbonne.

Résumé.

L'utilisation des modélisations-objet pour représenter les connaissances a connu un important développement en informatique. Cette évolution provoque un regain d'intérêt pour la notion d'objet, dans le domaine de la logique.

Dans ce travail nous proposons une nouvelle approche de cette notion s'appuyant sur la Logique de la Détermination d'Objets (LDO). La LDO est un système de catégorisation, comprenant une théorie de la quantification. Du point de vue formel la LDO est une logique applicative typée, dans le sens de Curry. Il s'agit d'une contribution théorique à la notion d'objet par une nouvelle approche logique constructive, fondée sur l'opération de détermination. Cette logique rend compte des problèmes cognitifs et déductifs rencontrés par l'héritage de propriétés par des objets typiques et atypiques. La quantification de la LDO prend en compte la typicalité.

Notre travail présente notamment le système conceptuel de la LDO, sa formalisation comme logique applicative, le système d'opérateurs de quantification, une sémantique en termes de classes d'objets et une comparaison entre la LDO et les réseaux, sémantiques. Ce travail représente les bases conceptuelles et la mathématisation d'une logique applicable à la modélisation-objet.

Mots-clés : objet, détermination d'objet, objet typique, objet atypique, quantification, catégorisation.

Contact : Anca Pascu, Département d'Informatique, Université de Bretagne Occidentale, Faculté de Lettres et Sciences Sociales, 20, rue Duquesne BP 814 29285 Brest Cedex (Tél : 02 98 01 63 54).

Email : Anca.Pascu@univ-brest.fr

Adresse Web du document :

<http://geai.univ-brest.fr/recherche/index.html#publis>

Techniques d'intelligence artificielle pour le diagnostic automatique de systèmes industriels par Marc Porcheron

Habilitation à Diriger les Recherches soutenue le 26 septembre 2001 à Paris (Université de Paris 13, LIPN, Institut Galilée)

Résumé.

Les travaux présentés témoignent d'une dizaine d'années d'activités de recherche dont la plupart se situent dans le domaine de l'Intelligence Artificielle (IA). Ils ont été réalisés à la Direction des Études et Recherches d'Électricité de France, en collaboration étroite avec des laboratoires Universitaires et du CNRS.

Les travaux menés dans le domaine de l'IA ont porté principalement sur l'application des *techniques de diagnostic automatique à base de modèles* à des systèmes industriels

complexes. Ces techniques, qui exploitent des connaissances descriptives du système diagnostiqué permettant la simulation de ses comportements, ont vu le jour à la fin des années quatre-vingts en réaction aux faiblesses mises en évidence dans les systèmes experts dits de « première génération » qui reposaient essentiellement sur l'exploitation de connaissances heuristiques de résolution de problèmes.

La mise en œuvre des techniques à base de modèles sur des applications de taille industrielle présente des difficultés importantes : taille, complexité et hétérogénéité des modèles exploités ; prise en compte du temps dans le diagnostic ; mise en relation des modèles, qui sont symboliques et qualitatifs par nature, avec les informations délivrées par les systèmes d'observations des machines, etc.

Entre 1991 et 1999, nous avons mené et co-encadré des recherches dans le but de mettre au point des solutions concrètes à ces problèmes. Ces travaux ont conduit à la réalisation d'un *atelier logiciel de diagnostic automatique* supportant l'implémentation et l'expérimentation de méthodes opérationnelles prenant en compte les difficultés listées ci-dessus.

Les solutions mises au point dans le cadre de cet atelier ont été expérimentées avec succès sur plusieurs matériels des ensembles de production d'électricité d'EDF et certaines d'entre elles industrialisées dans des logiciels mis à la disposition des centrales de production.

En complément de ces travaux spécifiquement consacrés au diagnostic, nous avons également conduit entre 1997 et 1999 des recherches sur le thème des mémoires d'entreprise et plus généralement de la *capitalisation* et de la *gestion des connaissances*. Ces recherches, appliquées à la problématique de la gestion de connaissances de diagnostic et de maintenance de matériels ont également conduit à des réalisations industrielles opérationnelles.

Enfin nous avons engagé plus récemment des travaux dans le domaine du *génie logiciel*, visant à définir une architecture applicative adaptée au développement de logiciels implémentant des méthodes complexes d'optimisation, permettant en particulier la définition déclarative des énoncés des problèmes résolus.

Mots-clés : Diagnostic automatique, Ingénierie des Connaissances, Applications Industrielles

Contact : Marc Porcheron,

EDF R&D, ER/MOS, 1, avenue du General De Gaulle, 92141 Clamart Cedex

Email : Marc.Porcheron@edf.fr

Conception d'un atelier d'expérimentation de logiciels éducatifs. Application en géométrie par Marilyne Rosselle

Thèse de l'Université de Henri Poincaré (Nancy 1), spécialité informatique, soutenue le 21 septembre 2001 à Nancy

Résumé.

Cette recherche se situe dans le contexte des EIAO (Environnements Interactifs d'Apprentissage avec Ordinateur). Les nombreux prototypes développés en recherche implantent une ou plusieurs fonctionnalités requises pour la formation (simulation, explication, etc.) mais jamais l'ensemble de ces fonctionnalités. Au lieu de chercher à déve-

opper un nouvel outil qui proposerait cet ensemble, nous proposons de faire coopérer divers prototypes offrant des fonctionnalités complémentaires. L'objectif du travail est donc de définir des critères, une architecture et des outils permettant cette coopération. Nous focalisons notre proposition sur la coopération de prototypes existants et nous restreignons l'application à l'enseignement de la géométrie plane. Nous proposons un atelier logiciel qui permet à un enseignant ou un chercheur d'utiliser des fonctionnalités implantées dans des prototypes différents, à travers une interface unificatrice, à peu près comme s'ils étaient disponibles dans le même logiciel. Pour le définir, nous avons modélisé une activité d'apprentissage du point de vue de l'exécution de logiciels et de l'échanges de données. Notre modèle comprend des scénarios découpés en étapes munies de transitions, des fonctionnalités offertes par un prototype, et la notion d'observable construite à partir de traces d'interaction et d'événements sémantiques. Nous proposons la notion de macro-définition avec les grammaires et interprètes associés pour adapter aussi bien des données du domaine que des observables. Notre atelier est implanté dans une maquette en Java et toutes les propositions sont faites avec un objectif de généralité qui confère aux propositions un caractère générique.

Mots clés : coopération, EIAO, architecture, enseignement, géométrie.

Contact : Marilyne Rosselle, LORIA, Campus scientifique - B.P. 239,

F54506 Vandoeuvre lès Nancy

Email :

Marilyne.Rosselle@loria.fr - <http://www.loria.fr/~rosselle>

Tél. 03 83 59 20 61 (direct) - Fax. 03 83 41 30 79

Adresse Web du document :

<http://www.loria.fr/~rosselle/recherches.html>

Modèles pour la conception et la réalisation de formations au travers de réseaux : l'environnement ECSAIWeb. par Charun Sanrach

Thèse de l'Institut National Polytechnique de Lorraine (INPL), spécialité informatique, soutenue le 14 septembre 2001, à Nancy.

Résumé.

Ce travail porte sur la conception et la réalisation de formation au travers de réseaux. L'une des difficultés rencontrées est que les solutions proposées sont rarement acceptées par les

enseignants car ces systèmes Ce travail porte sur la conception et la réalisation de formation au travers de réseaux. L'une des difficultés rencontrées est que les solutions proposées sont rarement acceptées par les enseignants car ces systèmes leur sont fermés : très peu de personnalisation des contenus et de prise en compte de sa pédagogie. L'objectif du travail est donc de produire des modèles généraux et de les implanter dans un environnement de création de cours et de déploiement de formation pour le Web, appelé ECSAIWeb. Cet environnement sera ouvert aux enseignants pour la création de matériaux et de stratégies pédagogiques et tirera au mieux partie des possibilités offertes sur le Web, tout en restant simple à déployer.

Le modèle est issu de l'environnement existant ECSAI. Il décrit le domaine à partir de concepts appelés items, organisés en réseaux et servant à modéliser en termes de contenus et de caractéristiques pédagogiques les modules d'apprentissage, appelés unités logiques. Ces unités logiques (modèles des unités physiques directement exécutables) sont aussi organisées en réseau dynamique qui évolue au fur et à mesure des actions de l'apprenant et permet de calculer le parcours d'apprentissage en mode guidé. Les items servent également de base à un modèle d'apprenant utilisant la technique classique de recouvrement, mais nécessitant une exploitation selon deux points de vue (construction dynamique du parcours apprenant ou adaptation des présentations). L'adaptation des présentations a nécessité également une modélisation de la présentation des unités à l'aide de balises de type XML, on parle de modèle des unités physiques.

Nous avons créé un environnement auteur ouvert. L'enseignant-auteur peut construire et organiser des matériaux pédagogiques dans sa matière. Il peut observer l'utilisation de ces matériaux afin de les améliorer. Nous avons créé un environnement élève qui permet l'exploration de ces matériaux pédagogiques, et un environnement partagé. Le système fonctionne selon un mode client-serveur sous Windows, utilise un navigateur Web standard. Il est développé majoritairement en Perl et comprend 20500 lignes de code.

Mots clés : environnement d'apprentissage adaptatif basé sur le Web, environnement auteur basé sur le Web, système tuteur intelligent basé sur le Web, formation sur le Web, système de distribution de cours sur le Web.

Adresse : Charun Sanrach, LORIA, Campus scientifique - B.P. 239

F54506 Vandoeuvre lès Nancy CEDEX

Email :

Charun.Sanrach@loria.fr - <http://www.loria.fr/~sanrach>

Tél. 03 83 59 20 61 (direct) - Fax. 03 83 41 30 79

Adresse Web du document :

<http://www.loria.fr/~sanrach/html/these/>

Pour l'envoi d'un résumé de thèse ou d'habilitation :

Envoyer à Amedeo Napoli (adresse page 2 de ce Bulletin) un fichier texte ou un document word attaché contenant les informations suivantes :

Titre de la thèse
Prénom et nom du candidat
Type de la thèse
(Thèse de l'Université de xxx, ou Habilitation à diriger
des recherches de l'Université de xxx)
Date et lieu de soutenance de la thèse

Mots clefs (*5 maximum*)
Adresse postale du candidat
Mél
Adresse Web
Téléphone et télécopie (facultatif)

Résumé (une demi-page)

Adresse où le document peut être récupéré (le cas échéant)
L'ensemble de ces informations doit tenir en 400 mots environ.

Dossier « IA et Médecine »

Date de parution prévue : janvier 2001 (bulletin numéro 48)

Coordonnateurs : Vincent Corruble (LIP6), Jean Charlet (DIAM, DSI/AP-HP)

Objectifs

La médecine a été l'un des premiers champs d'investigation des systèmes experts au point que, pendant longtemps, le vocable « IA & Médecine » ne faisait référence qu'aux systèmes experts. Actuellement, l'IA dans le domaine médical, c'est beaucoup plus que cela et on y trouve des travaux sur les systèmes à base de connaissance, sur les systèmes d'aide à la décision, sur la représentation des connaissances (e. g. ontologies), sur les systèmes de « suivi intelligent » de patients, sur la fouille de données médicales, sur la compréhension des langues, sur l'interprétation d'images, sur l'analyse, l'interprétation et la classification d'images, sur la découverte et la gestion des connaissances, sur les EIAH en médecine, etc.

Plus récemment, l'irruption des sciences et technologies de l'information et de la communication ont bouleversées l'offre informatique dans le domaine médical, que ce soit du point de vue de l'échange d'informations comme des nouvelles potentialités qui se font jour en termes de dossiers médicaux virtuels, de carnets de santé sur le Web, etc. L'IA a-t-elle des réponses à proposer aux problèmes de représentation des connaissances qui se posent dans ce contexte ? L'aide à la décision est-elle une réalité en médecine ? La médecine est-elle un domaine privilégié de l'IA ?

Dans le dossier IA et médecine, nous souhaitons faire le point sur les recherches menées dans ce domaine aussi bien que sur les axes de réflexion proposés. Pour cela, nous mettons en place une procédure un peu différente des celle des dossiers habituels et demandons deux types de contribution : des descriptions standard d'activité d'équipes d'un côté et, pour ceux qui le souhaitent, un « article court » sur la présentation ou la défense d'une idée ou d'un projet. Les descriptions d'activités seront, comme d'habitude, retenues obligatoirement pour être intégrées dans le dossier. Les « articles courts » seront relus et retenus en fonction des idées ou projets innovants qui y sont proposés.

Qui peut soumettre ?

Tout laboratoire ou industriel de la communauté francophone impliqué dans des actions de recherche et développement directement liées au domaine.

Nous appelons les équipes qui sont impliquées dans des travaux d'IA et médecine à nous faire parvenir leurs contributions : ***** Avant le 30 novembre 2001 *****

à

Vincent Corruble

Université Pierre et Marie Curie (Paris VI), LIP6,
Pole IA

8, rue du Capitaine Scott

75015 Paris

Tél. : +33.1.44.27.88.00 ; Fax : +33.1.44.27.70.00

Mél : Vincent.corruble@lip6.fr

Les descriptions d'activité pourront être organisées suivant le schéma suivant :

- identification du chercheur ou de l'équipe, adresse, téléphone, courriel
- cadre général, mission et composition de l'équipe,
- bref historique des travaux,
- types de problèmes traités,
- panorama des travaux effectués, présentation d'un ou deux projets représentatifs,
- méthodes et outils employés, résultats obtenus,
- position par rapport à d'autres travaux, nationaux ou internationaux,
- bibliographie.

Les « articles courts » seront organisés suivant le schéma suivant :

1. titre
2. auteur (contact)
3. description
4. courte (3) bibliographie

Présentation des contributions

Les contributions feront **1400** mots maximum (en raison de l'organisation spécifique de ce dossier, merci de respecter scrupuleusement cette contrainte), **que ce soit pour la description d'activité ou pour les « articles courts »** et devront être envoyées à Vincent Corruble avant le **30 novembre 2001**, 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.

Adhésion et abonnement		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
Nom : Prénom : Affiliation : Adresse postale : N° de téléphone : N° de télécopie : Adresse électronique : Activités (à titre professionnel / à titre privé (<i>rayez la mention inutile</i>)) :			
	Consultation du bulletin sur WEB (un accès)	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	/	90 Euros	
<input type="checkbox"/> Abonnement au bulletin sans adhésion	/	55 Euros	
<input type="checkbox"/> Adhésion au collègue :ajouter 7,5 Euros pour les étudiants, 15 Euros pour les autres <input type="checkbox"/> Adhésion au collègue <i>Cafe (Apprentissage)</i> : gratuit			
Veillez trouver un règlement (à l'ordre de l'AFIA) de Euros			

Adhésion Personne morale		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
Organisme : Adresse postale (commune aux bénéficiaires couverts par cette adhésion) :			
Le tarif d'hadésion comprend une partie fixe et une partie par bénéficiaire Coordonnées des bénéficiaires (10 maximum) :			
NOM et prénom	Adresse électronique	N° téléphone	N° 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	/	
<input type="checkbox"/> j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA <input type="checkbox"/> j'accepte que les renseignements ci-dessus soient transmis à l'ECCAI pour constituer un fichier européen.			
MODE D'HADÉSION : Sur le site web de l'AFIA : http://www.afia.polytechnique.fr à défaut ce bulletin d'adhésion doit être retourné au trésorier : Marc AYEL LIA-Université de Savoie 73376 Le Bourget du Lac cedex Société Générale, Résidence du Val de Seine, 78430 LOUVECIENNES code banque30003, code guichet 01902, numéro de compte 00037283856 clef RIB 39			

Les dossiers du *Bulletin de l'AFIA*

Fouille de données	Bulletin n°46/47	I.A. et fusion de données.....	Bulletin n°24
IA et document.....	Bulletin n°44	I.A. et musique	Bulletin n°23
IA et document.....	Bulletin n°44	Apprentissage.....	Bulletin n°22
IA et connexionnisme.....	Bulletin n°43	Les explications dans les SBC	Bulletin n°20
IA et Vie Artificielle	Bulletin n°42	Pétrole-Chimie	Bulletin n°19
IA et CHM.....	Bulletin n°41	Le raisonnement à partir de cas.....	Bulletin n°18
IA et EIAH	Bulletin n°40	I.A. et temps-réel	Bulletin n°17
Plates-formes multi-agents	Bulletin n°39	Planification et action.....	Bulletin n°16
IA et WEB	Bulletin n°38	Traitement automatique des langues.....	Bulletin n°15
Mémoires d'entreprises.....	Bulletin n°36	I.A. et médecine	Bulletin n°14
IA et logique	Bulletin n°35	Diagnostic à base de modèles.....	Bulletin n°13
Ingénierie des connaissances	Bulletin n°34	Validation des SBC	Bulletin n°12
IA et Télécommunications	Bulletin n°33	Le connexionnisme.....	Bulletin n°11
IA et Terminologie.....	Bulletin n°32	I.A. et jeux	Bulletin n°10
Décision et IA.....	Bulletin n°31	E.I.A.O.....	Bulletin n°9
Raisonnement IA et Image	Bulletin n°30	I.A. et gestion.....	Bulletin n°8
Raisonnement temporel et spatial.....	Bulletin n°29	Conception et I.A.....	Bulletin n°7
Systèmes Multi-agents.....	Bulletin n°28	Intelligence artificielle distribuée	Bulletin n°6
IA et robotique	Bulletin n°27	Acquisition des Connaissances	Bulletin n°5
I.A. et biologie moléculaire.....	Bulletin n°26	I.A. et ordonnancement	Bulletin n°4
I.A. et droit.....	Bulletin n°25		

SOMMAIRE DU BULLETIN N° 46/47

La vie de l'A.F.I.A.	4
Débats : l'évaluation scientifique	6
L'IA au Québec	12
Synthèse	14
Présentation de Laboratoires	18
Dossier	30
Livres	46
Compte rendu de conférences	48
Calendrier	51
Sommaire des revues	55
Comptes rendus habilitations et thèses	63
Appels à dossiers	66

CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Eté</i>
Réception des contributions:	Réception des contributions:
15 décembre	15 juin
Sortie le 31 janvier	Sortie 31 juillet

<i>Printemps</i>	<i>Automne</i>
Réception des contributions:	Réception des contributions:
15 mars	30 septembre
Sortie le 30 avril	Sortie le 31 octobre