

# Bilan : rapport d'activité

## 2005-2008

Laboratoire ERIC  
Université Lumière Lyon 2  
5, avenue Pierre Mendès-France  
Bât L.  
69600 Bron France

Tel. +33 4 78 77 23 76  
Fax. +33 4 78 77 23 75  
Web. <http://eric.univ-lyon2.fr>



# Sommaire

## **1 Note de synthèse**

## **2 Travaux scientifiques**

2.1 Positionnement scientifique de ERIC

2.2 Présentation de l'ECD

2.3 Avancées scientifiques 2005-2008

## **3 Valorisation scientifique**

3.1 Publications

3.2 Activités Editoriales

3.3 Animations scientifiques

3.4 Projets de recherche appliquée

3.5 Développement de logiciels

3.6 Synergie entre enseignement et recherche

## **4 Ressources**

4.1 Bilan financier

4.2 Ressources humaines au 31 décembre 2008

## **5 Formation des personnels**

## **6 Hygiène et sécurité**



# 1 NOTE DE SYNTHÈSE

Au cours de l'année 2009, le laboratoire ERIC souhaitait accueillir de nouveaux chercheurs issus de deux autres établissements : l'Université Claude Bernard Lyon 1 et l'École Pratique des Hautes Études de Paris. Le but était de réunir les forces de trois équipes de recherche :

- EA 3083 ERIC de l'université Lyon 2, dirigée par le Pr. Zighed ;
- Equipe LaISC de l'EPHE de Paris (issue de l'EA 4004), dirigée par le Pr. Bui ;
- Equipe MA<sup>2</sup>D de l'université Lyon 1, dirigée par le Pr. Lamure et issue du LIRIS<sup>1</sup> UMR 5205 CNRS-Lyon1-Lyon2-INSA-ECL.

Cependant, compte tenu de la réorganisation en cours dans les universités parisiennes, seuls les collègues de l'université Lyon 1 ont pu intégrer le laboratoire ERIC.

Le regroupement du laboratoire ERIC avec l'équipe MA<sup>2</sup>D de l'université Lyon 1 se justifie par :

- Les complémentarités thématiques des deux équipes qui, mises en commun, permettent d'envisager la prise en main de problèmes scientifiques complexes notamment pour la fouille de données volumineuses et non structurées, la modélisation et la simulation de processus complexes en particulier dans le domaine de la décision ;
- La forte sensibilité de chacune des équipes aux applications dans les domaines des SHS (Sciences Humaines et Sociales) et de la santé ainsi que leur expertise acquise et reconnue dans ces domaines ;
- L'existence de longues et nombreuses collaborations à la fois scientifiques et pédagogiques entre les équipes.

Le bilan demandé ne couvre que la période 2005-2008. Par conséquent nous n'y intégrerons pas celui de la nouvelle équipe récemment intégrée. Un rapport complémentaire et succinct, la concernant et couvrant la même période, est joint à celui-ci.

Le laboratoire ERIC est localisé sur le campus Porte des Alpes à Bron et partage les locaux du Département d'Informatique et de Statistique de la Faculté de Sciences Économiques et de Gestion.

---

<sup>1</sup> LaISC: Laboratoire d'Informatisation des Systèmes Complexes; MA2D Méthodes & Algorithmes pour l'Aide à la Décision; LIRIS Laboratoire Informatique des Réseaux de l'Image et des Systèmes.

Les travaux menés au sein du laboratoire s'articulent autour du processus d'Extraction des Connaissances à partir des Données (ECD) et s'attaquent à des verrous tant scientifiques que technologiques comme :

- La prise en compte des données complexes : hétérogènes (tableaux, multimédia, graphiques...), peu structurées, volumineuses, pouvant présenter ou non un caractère temporel ou spatial ... ;
- La prise en compte du caractère empirique de la fouille de données et son impact sur les mesures de qualité et leur optimisation en apprentissage automatique, l'identification d'espaces de représentation efficaces, la combinaison et l'agrégation de classifieurs...
- La prise en compte des connaissances du domaine soit pour l'enrichissement sémantique des corpus de données soit pour le déploiement des systèmes de prise de décision à base de connaissances.

La mise en perspective de ces problématiques dans un processus d'ECD permet d'une part de les unifier et d'autre part de faire émerger des questions et des approches nouvelles.

Quantitativement, le bilan d'ERIC, pour les quatre dernières années universitaires (2005-2008), est :

- 11 chercheurs ont pu effectuer ou achever leur doctorat au sein d'ERIC ;
- 3 collègues ont réalisé leur Habilitation à Diriger les Recherches ;
- 5 projets de création d'entreprise ont été ou sont incubés au sein du laboratoire ;
- Une participation à des projets de recherche nationaux ou européens qui a généré plus de 210K€ ;
- Un partenariat en termes de contrats avec les entreprises privées qui a généré 81,7K€ ;
- La qualité et la diversité des publications d'ERIC montre que l'équipe, 11 enseignants-chercheurs sur 2005-2008, est active et présente à tous les niveaux : publications dans des revues internationales (17) ou nationales (5), communications dans des conférences internationales (98) ou nationales (72), chapitres (15) ou ouvrages ou directions d'ouvrages (7), diffusion de logiciels, organisation de conférences majeures, contacts avec des universités étrangères... L'expertise développée par les chercheurs d'ERIC est reconnue comme en témoignent les nombreux contrats qui représentent près de 50% de ses ressources.
- ERIC maintient un lien fort et explicite avec l'enseignement, notamment, à travers l'animation du Master d'Informatique de Lyon 2 et la candidature pour l'obtention d'un Master Erasmus Mundus en partenariat avec 6 universités de 4 pays européens (Italie, Espagne, Roumanie et France).

- Les enseignants chercheurs d'ERIC sont présents dans les comités de programme des principales conférences internationales du domaine : PKDD-ECML, ISMIS, ICDM, COMPSTAT etc.

Afin de garder une dynamique de recherche et garantir une qualité et une transparence dans sa gouvernance, ERIC organise depuis sa création, tous les deux ans, une évaluation de son activité par des experts indépendants, français et étrangers couvrant l'ensemble des aspects, académiques, industriels, de valorisation, d'implication pédagogique etc.

Dans la suite de ce rapport figurent le bilan avec :

- Organigramme fonctionnel de l'équipe : § 4.2.1
- Enseignement et formation par la recherche : § 3.6
- Valorisation : §3 et annexes
- Diffusion de l'information et de la culture scientifique et technique : §3 et annexes
- Formation des personnels : §5
- Hygiène et sécurité : §6

Pour le futur, nous souhaitons :

- renforcer les thématiques sur lesquelles ERIC est reconnu, à savoir l'Extraction des Connaissances à partir des Données (ECD) tout en élargissant grace aux nouveaux arrivant, vers des questions jusque là effleurée : comme les systèmes complexes en sciences sociales ou en santé, la simulation et la décision ;
- maintenir une activité de recherche à trois niveaux : travaux à caractère théorique, développement de logiciels, recherche de terrains d'application en particulier dans le domaine des Sciences Humaines, Sociales, Économiques et Santé sur lesquels l'équipe a acquis une expertise;
- renforcer la politique éditoriale et d'animation scientifique de l'équipe par plus de publications dans les journaux internationaux spécialisés ;
- développer nos relations avec la communauté scientifique locale, nationale et internationale autour à la fois d'activités de recherche mais aussi d'enseignement telles que les co-tutelles de thèses, les doubles diplômes ou le master européen ;
- valoriser nos recherches sur le plan industriel par l'incubation de projets préindustriels.



## 2 TRAVAUX SCIENTIFIQUES

### 2.1 Positionnement scientifique de ERIC

Le positionnement scientifique des recherches d'ERIC peut se faire selon deux points de vue que nous présentons de manière succincte.

#### 2.1.1 Point de vue historique

Une étude<sup>1</sup> de la société américaine *Disk/Trend* basée en Californie et spécialisée dans la veille industrielle a montré que le coût moyen de stockage d'un mégaoctet sur disque dur est passé de \$11.54 en 1988 à \$0.04 en 1998 et à \$0.003 en 2007. Cette décroissance exponentielle du coût de stockage, que chacun de nous a pu observer au quotidien, a eu pour effet d'accroître de manière exponentielle le volume des données stockées par les entreprises. En moyenne, selon les mêmes études, ce volume doublerait tous les 9 mois. Cette tendance à l'accumulation des données ne semble pas avoir atteint son niveau asymptotique et se trouve même accélérée sous l'effet du développement des réseaux de transmission comme Internet, qui sont de plus en plus puissants. En effet, ces derniers peuvent, de nos jours, atteindre les 10 gigaoctet par seconde et leur coût ne cesse de décroître. Ainsi, depuis 1975, le coût, au Mbit/s.km, a été divisé par 1000. Aujourd'hui, on peut affirmer que l'accès aux données et leur stockage reposent sur des technologies fiables et peu coûteuses. Autrement dit, le défi des années 60-70 visant la maîtrise des systèmes d'information a été pleinement relevé et gagné. Où se situent alors les nouveaux enjeux ? Ils se sont déplacés vers l'accès à l'information et à la connaissance cachée dans les immenses bases de données. Pour s'en convaincre, il suffit d'observer la plus grande réussite industrielle de l'informatique de la dernière décennie, Google et son moteur de recherche d'information sur le web. Ainsi, il est assez aisé de comprendre que ce qui pose problème de nos jours, ce n'est ni l'accès aux données, ni leur stockage mais l'accès au contenu sémantique de celles-ci.

Les activités du laboratoire ERIC se situent dans ce vaste champ de recherche en forte croissance et dont les objectifs sont de définir des méthodologies et de proposer des outils informatiques permettant l'accès au contenu sémantique des grandes bases de données. C'est ce que nous appelons Extraction des Connaissances à partir des Données (ECD).

---

<sup>1</sup> Gamze Zeytinci, CSIS-550 History of Computing Spring-2001 ; Evolution of the Major Computer Storage Devices From Early Mechanical Systems to Optical Storage Technology.

## 2.1.2 Point de vue pragmatique

La prise de décision est au cœur de toutes les activités qu'elles soient humaines, sociales, biologiques, économiques... Elle repose sur l'identification d'une situation et, en fonction d'un état désiré que l'on peut considérer comme l'objectif, d'entreprendre les actions adéquates.

Considérons trois exemples simples :

- Un praticien qui examine un patient et qui observe une anomalie va prescrire une action thérapeutique dont l'objectif est le soin de la pathologie en vue d'améliorer l'état de santé du malade.
- Un garde côte maritime qui, au moyen de ses radars, observe un mouvement suspect en mer, lance un processus de vérification pour savoir s'il s'agit d'un bateau qu'il faut intercepter.
- Un juge, au vu du récit des faits reconnus par un prévenu, identifie le type d'infraction et décide de la sanction.

Dans de nombreux domaines, une large partie du processus de décision, notamment, le processus d'identification, est confiée à l'ordinateur. Les raisons de cette délégation sont diverses comme la réduction des coûts, l'accroissement de l'efficacité ou la rapidité, une meilleure maîtrise de la complexité, ... En effet, considérons des services de sécurité en charge d'Internet et dont l'objectif est par exemple d'intercepter les communications jugées « sensibles ». Comment envisager cette tâche quand on sait qu'il y a près de 700 millions d'internautes et qu'en moyenne, un internaute émet une dizaine de messages (e-mail, blog, requête...) par jour ? Dans ce contexte, il est impossible d'imaginer une organisation humaine capable d'assurer une surveillance sur les contenus de 7,5 milliards de messages journaliers sans parler des sites Web dont le contenu peut être ciblé.

On peut également observer cette situation dans d'autres domaines comme le marketing ou la santé. Par exemple, la généralisation du dépistage des cancers du sein à toutes les femmes dans la tranche d'âge 50-74 ans exigerait une infrastructure radiologique capable de traiter correctement près de 11 millions d'exams par an pour seulement 2000 radiologues en France qui sont déjà au bord de la saturation alors qu'ils ne traitent qu'environ 60%.

Dans de telles situations, le recours à la puissance des ordinateurs semble naturel pour assurer un passage à l'échelle. Mais pour pouvoir recourir à l'aide des ordinateurs, il faut être en mesure de lui « expliquer » comment reconnaître un échange de nature suspecte parmi des millions de messages et comment identifier ou bien localiser des cas pathologiques à partir d'un dossier médical comportant des images (mammographie), des examens cliniques et/ou biologiques, des comptes-rendus, etc.

Supposons que nous souhaitons mettre au point un système informatique capable d'aider les responsables de la sécurité publique dans l'identification des messages « sensibles ».

Le concepteur du système d'aide à l'identification collectera, auprès des experts en analyse des contenus, l'ensemble des règles conduisant au diagnostic. Ces règles décrivent le procédé d'analyse du contenu d'un message et les règles de déduction pour décider enfin de la catégorie sensible ou non d'un message. Pour y parvenir, il faut deux conditions qui reposent sur des hypothèses fortes :

- La première postule que le processus d'identification peut être décrit comme une suite d'opérations sur des structures symboliques qu'on appelle des règles d'inférence et que l'on assimile aux connaissances qu'utilisent les experts pour identifier les catégories de message.
- La seconde postule que l'expert est en mesure d'explicitier ses connaissances sous forme de règles dans un formalisme précis qui pourrait être codé en machine.

Si ces conditions sont réunies, alors les connaissances sont introduites dans la machine sous forme de règles pour en constituer la base de connaissances. On dote ensuite la machine d'un programme capable d'interpréter les règles et d'enchaîner les calculs qui en découlent. Ce programme est appelé « moteur d'inférence » et il appliquera les règles dont il dispose en vue d'inférer sur des faits qui lui sont soumis. Si le système Expert (Base de connaissance + moteurs d'inférence) est jugé pertinent, on peut alors le dupliquer en une population d'agents dits « intelligents » qui seraient ensuite déployés sur Internet pour identifier et signaler la présence de tout message suspect. Dans la mesure où cette approche vise à mimer le raisonnement de l'expert face à des cas concrets, on pourrait la qualifier de démarche psycho-mimétique.

Malheureusement cette approche se heurte à deux difficultés majeures qui remettent en cause, au moins en partie, les hypothèses que l'on vient d'énoncer. En effet, il arrive que, sur des champs d'application, il n'y ait pas d'experts, donc aucune connaissance ne peut être transférée sur l'ordinateur. Que faut-il faire alors ? Faut-il attendre que des personnes confrontées à des situations de surveillance finissent par acquérir l'expertise nécessaire par un processus d'essai-erreur pour ensuite la communiquer à la machine ? Il arrive également qu'un expert sache parfaitement identifier les situations requises, mais qu'il soit en revanche incapable d'explicitier le procédé cognitif qu'il met en œuvre pour y parvenir. Par exemple, on sait tous reconnaître une personne dans une foule, mais sommes nous pour autant capables d'explicitier la manière dont nous y parvenons ?

C'est pour combler ce genre de déficience que l'on fait appel aux méthodes d'Extraction des Connaissances à partir des Données. Les connaissances ne sont pas fournies par l'expert mais engendrées par la machine suite à un apprentissage automatique sur des situations passées. Par exemple, le médecin fournit un ensemble de données relatives à des patients cancéreux et non cancéreux déjà traités et, grâce aux méthodes d'ECD, on cherchera à déterminer les règles de diagnostic qui pourraient être appliquées sur les nouveaux cas à diagnostiquer. Une fois validées, ces connaissances pourraient, à leur tour, être insérées dans le Système Expert. Ce système Expert pourra

également incorporer des fragments de connaissances venant d'experts humains. Cette démarche peut s'appliquer dans tous les domaines de la décision.

Le laboratoire ERIC travaille sur le processus d'ECD que nous allons décrire de manière un peu plus détaillée.

## 2.2 Présentation de l'ECD

Sans trop s'étendre sur les détails techniques, on peut dire que l'ECD fait appel à des méthodes et à des outils issus de différents domaines de l'informatique : bases de données, intelligence artificielle, statistique, optimisation etc. en vue d'explorer des données volumineuses et hétérogènes à la recherche d'éléments structurants, d'invariants qui, une fois extraits et validés, pourraient être considérés comme des connaissances.

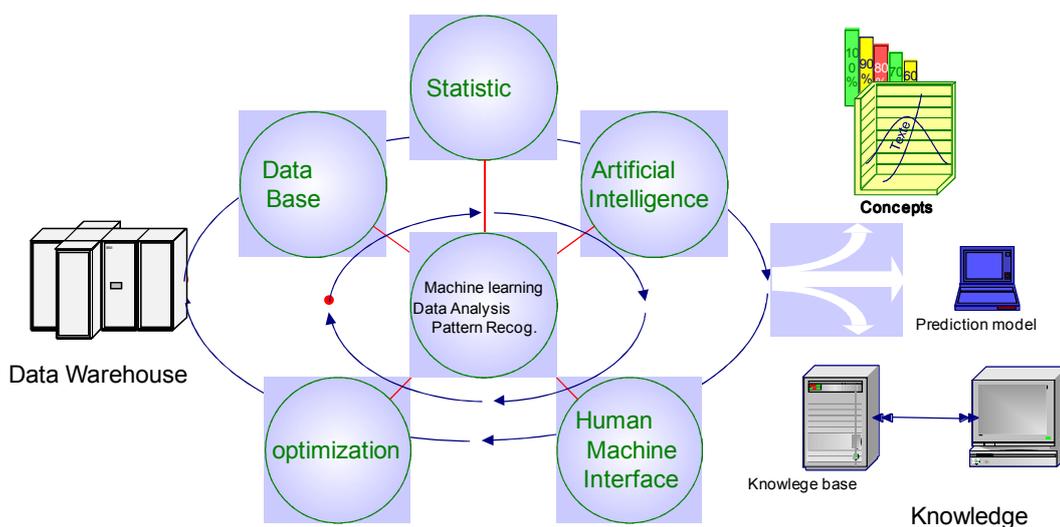


Figure 1 : KDD : From Data to Knowledge, Technologies involved

L'agencement de ces méthodes obéit à un processus logique en quatre étapes que nous décrivons très brièvement :

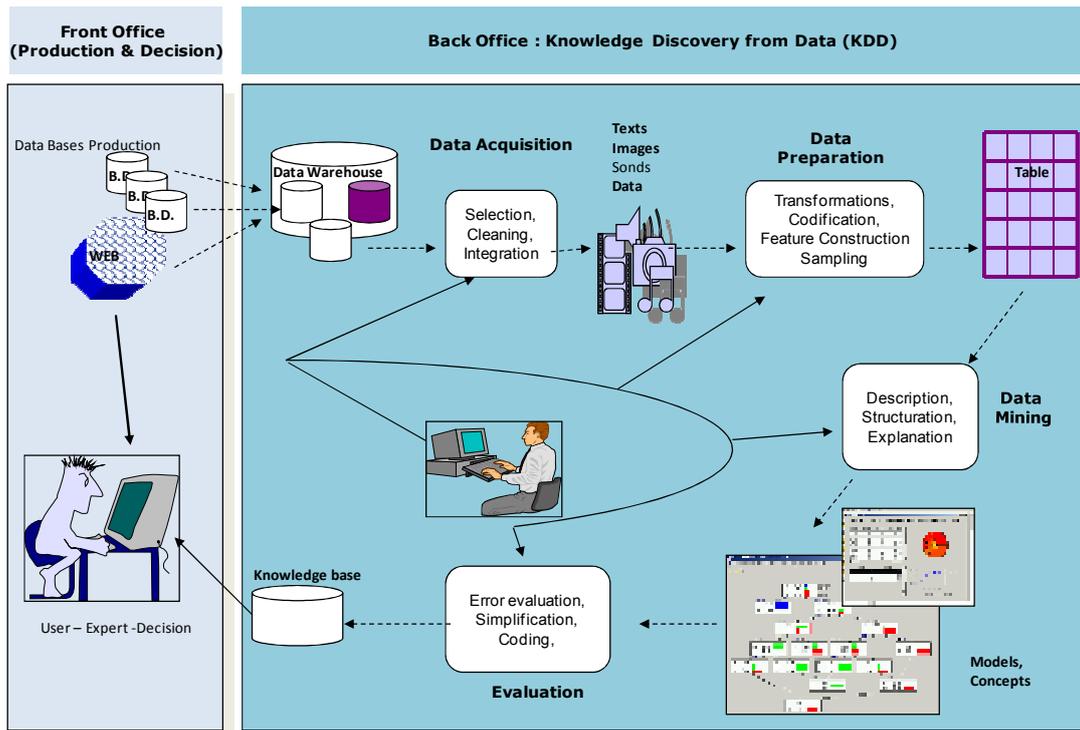


Figure 2: KDD workflow

**Acquisition :** elle a pour but de récupérer dans les entrepôts de données sources celles que l'on estime susceptibles d'aider dans la réalisation d'une tâche d'ECD. L'acquisition va s'effectuer sur des sources très volumineuses, distribuées géographiquement, enregistrées dans des environnements informatiques différents (bases de données relationnelles, bases de données XML, fichiers plats ou formats spécialisés comme le DICOM ou MPEG7). Outre l'accès et la sélection, il s'agit également d'organiser et d'intégrer les données dans un environnement local adapté à la fouille de données : bases de données XML, Entrepôts OLAP...

**Préparation des données :** les données acquises peuvent être de différents types : tableaux de chiffres, données textuelles, images etc. La phase de préparation a pour but de les structurer pour rendre possible la mise en œuvre des méthodes de fouille de données. La forme généralement la plus appropriée est un tableau de données à double entrée; cela peut être un tableau d'observations individus-variables, un tableau de contingence ; un tableau de similarité, etc. Il convient de préciser que cette opération est loin d'être simple et constitue généralement un véritable goulot d'étranglement. Par exemple si les données sources sont des textes, il convient de définir une série de prétraitements linguistiques comme la lemmatisation, la suppression de la ponctuation ou de la casse, le recours à une ontologie pour unifier le vocabulaire, l'extraction de concepts, etc. Il convient également d'identifier les individus statistiques : s'agit-il des textes, des paragraphes ou des concepts etc. Ce travail exige, généralement, une expertise et il est fortement lié au domaine d'application. Les mêmes difficultés apparaissent quand les données sont des images. Quels attributs faut-il extraire

pour décrire ces données sans structure mathématique évidente ? Quelles unités faut-il prendre en compte : images entières, imagerie issue d'un découpage ou d'une segmentation automatique pour obtenir des régions d'intérêt, etc. Quand la base d'exemples comporte des données de différents types : images, texte, courbes, tableaux de chiffres..., on parle de données complexes. Il faut non seulement définir un codage *ad hoc* pertinent pour ces données mais également les unifier. Le dossier médical d'un patient est une parfaite illustration de ce cas de figure car il contient souvent des radiographies, des courbes d'électrocardiogramme, des données quantitatives sur des mesures biologiques, des données textuelles qui décrivent un bilan clinique par exemple etc. Il faut par conséquent être en mesure de construire des mesures de similarité entre individus en considérant la totalité des données. C'est à ce niveau également que se pose le problème du traitement des données incomplètes et/ou imprécises, comment les prendre en compte en fouille de données. Cette phase est cruciale. Avec la phase de sélection, elle représente 80% du temps passé dans un cycle d'ECD.

**Fouille de données :** cette étape est le cœur de la démarche d'ECD. On y fait appel à une variété d'algorithmes soit pour la description des lignes et/ou colonnes des tableaux, soit pour la structuration des lignes et/ou colonnes du tableau, soit, enfin, pour établir un modèle de prédiction au moyen des méthodes d'explication ou de classification. On peut citer pêle-mêle : les méthodes d'analyse factorielles des données, les méthodes de classification dites d'apprentissage non supervisé, ou les méthodes de prédiction comme celles issues de l'apprentissage supervisé ou des règles d'association, etc. La nature des données et du codage introduisent ensuite des variantes d'algorithmes qui enrichissent la gamme des méthodes de fouille. Par exemple si les données sont floues ou symboliques, il conviendrait de proposer des algorithmes spécifiques dans chacune des trois catégories d'algorithmes cités.

**Validation :** A ce stade, il convient d'apprécier les résultats selon leur fiabilité, leur intérêt pour l'utilisateur et le cas échéant de se poser le problème de leur intégration dans une base de connaissances auquel cas, il faut procéder à leur codage dans un formalisme approprié pour ensuite les déployer en situation concrète.

## 2.3 Avancées scientifiques 2005-2008

A la lecture des fiches d'activité des chercheurs et en parcourant la liste des publications sur la période considérée, on observe que les travaux d'ERIC s'étalent sur un spectre de sujets relativement large qui couvre tout le cycle de l'ECD. Mais, leur mise en perspective permet néanmoins d'observer qu'une attention particulière a été accordée à la prise en compte des données complexes dans le processus d'ECD. Ce cadre spécifique a permis de faire émerger de nouveaux problèmes et des défis tant théoriques que technologiques pour répondre aux besoins des applications réelles. Nous allons

donc, surtout, mettre en avant cette originalité pour présenter les contributions des chercheurs d'ERIC dans le domaine de l'Extraction des Connaissances à partir des Données Complexes (ECDC). Nous verrons ensuite, à travers le cycle complet d'ECD, quels sont les problèmes abordés, où se situent les apports théoriques, les contributions méthodologiques et les applications traitées.

### **2.3.1 Contributions à l'entreposage de données complexes**

Les entrepôts de données classiques ont été développés autour du modèle des bases de données relationnelles. Ils ont donné naissance à des technologies telles que l'OLAP (*On Line Analysis Process*) permettant de naviguer dans la grande masse données qu'ils contiennent et d'en extraire des synthèses. Quels seraient alors les défis scientifiques et techniques à relever dans le contexte des données complexes ? Outre le problème de la représentation des données peu ou pas structurées, peut-on imaginer des modèles d'interrogation et d'exploration équivalents aux données tabulaires ? Quelles seraient ensuite les performances de tels systèmes une fois déployés ? Dans ce qui suit nous allons décrire les travaux menés pour tenter de répondre à ces questions.

#### **2.3.1.1 Représentation et navigation dans les entrepôts de données complexes**

Pour s'attaquer à ces verrous scientifiques, nous proposons un processus complet d'entreposage et d'analyse en ligne des données complexes. L'intégration et la modélisation physique consistent en l'intégration de données complexes dans une base de données agissant comme un ODS (*Operational Data Storage*). Dans la phase d'intégration, nous définissons des modèles conceptuels, logiques et physiques. Nous utilisons XML comme formalisme pour décrire les modèles logiques et physiques. Le modèle conceptuel est traduit au niveau logique sous la forme d'une DTD ou d'un schéma XML. Du modèle logique obtenu, nous générons une collection de documents XML comme modèle physique. Les documents XML générés sont valides et peuvent être stockés dans une base de données XML native ou une base de données relationnelle par mapping. De plus, nous avons proposé une approche pour construire un cube OLAP décrit par un schéma XML. Ce cube XML est généré automatiquement à partir des besoins de l'utilisateur exprimés par le modèle conceptuel multidimensionnel (MCM) et à partir d'un corpus de données complexes représentées par des documents XML. Le MCM et les documents XML sont exprimés en utilisant des schémas XML (XSD), puis ils sont transformés en arbres d'attributs afin d'être comparés. Certains algorithmes rendent cette comparaison possible grâce à des opérateurs de fusion par élagage ou greffe afin de traiter les arbres d'attributs et de générer un schéma XML du cube et des documents XML. Ce cube XML fournit un contexte d'analyse qui peut être analysé par des opérateurs OLAP ou par des méthodes de fouille de données.

Nous avons également développé une autre approche d'intégration de données basée sur un système de médiation utilisant des ontologies pour décrire chaque source de données. A partir des ontologies locales, l'objectif est de construire une ontologie globale qui permet au médiateur de proposer les données pertinentes pour la construction d'un cube OLAP. Cette ontologie globale est construite en utilisant une classification de l'ensemble des termes des ontologies locales.

Comme résultats de cette recherche, deux logiciels ont été développés. (1) SMAIDoC est un système multi-agents, articulé autour de cinq agents, afin d'intégrer les données complexes dans une base de données relationnelle ou une base de données XML native. (2) X-Warehousing est une plate-forme Java dédiée à la génération automatique de cubes XML. Les cubes XML sont obtenus à partir des besoins de l'utilisateur exprimés par le biais du modèle conceptuel multidimensionnel et à partir des documents XML contenant les données complexes.

### **2.3.1.2 Optimisation et évaluation des performances des entrepôts de données complexes**

Dans ce contexte d'utilisation du langage XML comme support pour l'entreposage des données complexes, la performance des entrepôts de données reste plus que jamais une question cruciale. Les principales structures physiques des données utilisées pour optimiser les temps d'accès aux données lors de l'exécution de requêtes complexes d'analyse sont les index, les vues matérialisées et les partitions. Sélectionner un ensemble optimal de ces objets est un problème NP-complet qui a été très largement traité. Toutefois, le passage à l'échelle demeure un problème car les approches existantes nécessitent soit une expertise humaine soit des structures de données coûteuses. En outre, les relations entre les index et les vues matérialisées ne sont jamais prises en compte, alors que ces structures de données vont mutuellement s'enrichir.

Pour répondre à ces questions, nous avons conçu une approche, générique, automatique, qui utilise des techniques de fouille de données, pour, à partir d'une charge (ensemble de requêtes) représentative de l'utilisation de l'entrepôt de données, d'en déduire une configuration quasi-optimale des index et / ou des vues matérialisées. Cette approche réduit considérablement l'espace de recherche des index des vues et donc améliore le passage à l'échelle. Ensuite, les modèles de coût aident à choisir les index et les vues matérialisées les plus efficaces en termes de gain de performance. Ces modèles prennent en compte les relations interdépendantes entre les index et les vues afin d'obtenir le meilleur compromis. Notre recherche d'optimisation des performances a été supportée et appliquée à deux projets pour optimiser l'accès aux données complexes : MAP (entrepôt de données biomédicales) et CLAPI (entrepôt XML de corpus de langue parlée).

De plus, afin d'évaluer et de comparer l'efficacité des techniques d'optimisation des performances, il est nécessaire de les tester avec différents jeux de données. Cette tâche est généralement réalisée à l'aide de bancs d'essai. Le « Transaction Processing Performance Council » (TPC) fournit des bancs d'essai

standard, mais le schéma de bases de données et la charge sont fixes (seule la taille de l'entrepôt varie). Ces bancs d'essai présentent donc peu d'intérêt dans un contexte d'ingénierie et de conception. Par conséquent, pour valider expérimentalement notre approche d'optimisation des performances, nous avons conçu plusieurs bancs d'essai génériques. Leur principale caractéristique de conception est la capacité d'adaptation : nos bancs d'essai permettent la production d'entrepôts de données avec des configurations différentes, ainsi que les charges associées. DWEB (Data Warehouse Engineering Benchmark) est le plus mature des outils développés. Il est actuellement le seul banc d'essai opérationnel, pour évaluer les performances des entrepôts de données, disponible en ligne.

### **2.3.2 Contributions à la recherche d'information dans les entrepôts de données complexes**

La Recherche d'Information dans les Entrepôts de données Complexes (RIEC) pose des problèmes de nature spécifique. En effet, dans tous les processus de recherche d'information il est nécessaire de se doter soit d'une mesure de similarité entre les objets soit d'une structure topologique sur ces mêmes objets sans avoir à expliciter la mesure de similarité sous jacente. Dans le cas où les données sont tabulaires, il existe une multitude d'indices de similarité. Mais, dans le cas où les données sont non structurées, comme par exemple pour les molécules chimiques, les textes, les images, les séries temporelles, les vidéos ou les documents multimédia, définir une proximité entre objets n'est pas aisé. Les nombreuses solutions proposées jusque là ne prennent qu'un seul type de données à la fois : soit textuel, soit image, soit la structure de la molécule en chimie,... mais peu de travaux ont été consacrés aux mesures qui prennent en compte cette hétérogénéité dans les données.

Nous avons exploré différentes stratégies pour construire une mesure de similarité entre objets complexes. Par exemple, par agrégation des mesures de similarité classiques issues de chaque type de données. L'agrégation est alors réalisée par une combinaison linéaire des similarités partielles qui résulte de chaque type de données. Nous avons également adopté d'autres approches qui combinent à la fois un point de vue topologique et un point de vue probabiliste. Pour cela, nous construisons par exemple une structure topologique dans chaque sous-ensemble homogène de données en utilisant des mesures de similarité classiques. La similarité globale entre deux objets sera d'autant plus forte que ces objets sont voisins dans les sous-espaces spécifiques. Par exemple deux patients seront d'autant plus voisins (semblables) qu'ils sont proches dans l'espace des données cliniques, proches dans l'espace des données image, proches dans l'espace des données biologiques etc. Ainsi, deux patients seront déclarés voisins s'ils sont voisins dans chacun des sous espaces topologiques induits. De là, on peut construire un indicateur qui une fois normalisé s'apparenterait à la probabilité pour que deux individus soient voisins. Si cette probabilité était égale à 1 alors les deux individus pourraient être considérés comme identiques

ou quasi identiques. Ces approches, encore en cours, ont été mises en œuvre et testées avec efficacité dans des applications réelles, parmi lesquelles on peut citer :

- Le corpus langage parlé qui contient des données textuelles, audio et vidéo. Ce travail de recherche a été réalisé dans le cadre d'un projet conjoint avec le laboratoire ICAR<sup>1</sup> (UMR Lyon 2 ENS lettres), et a bénéficié d'un soutien financier du ministère de l'enseignement supérieur.
- Le corpus de textes juridiques liés au droit international du travail. Cette recherche a été conduite en collaboration avec l'Université de Genève et le Bureau International du travail.
- Un corpus de bases d'images indexées par des textes disponibles comme benchmark dans la communauté de Recherche d'Information.

### 2.3.3 Contributions à la préparation des données

Dans l'ECDC, plus qu'ailleurs, la préparation des données pour la fouille s'avère ardue. Le principal problème qui se pose est celui du choix de l'espace de représentation. En effet, à l'origine, les instances d'objets enregistrés dans la base de données sont exprimées dans un formalisme qui ne se prête pas ou peu aux traitements mathématiques qui fondent la plus part des méthodes de fouille de données. Le passage par un codage unifié, généralement sous forme vectorielle, s'avère souvent incontournable. Comment alors passer de données textuelles, images, vidéo, temporelles vers des vecteurs ? Faut-il « vectoriser » et tout aligner dans un tableau unique ? Ces choix ne sont pas neutres car ils pourraient engendrer, à leur tour, d'autres difficultés. Par exemple, le choix de coder par un vecteur numérique les textes, exige des procédés linguistiques qui sont souvent très sophistiqués et où l'intuition, les connaissances *a priori* du domaine et même les choix arbitraires sont couramment mis à contribution. Et cela, sans, pour autant, être certain du bon choix. Parmi les problèmes qui peuvent surgir, par exemple, le vecteur résultant peut s'avérer d'une grande dimension ce qui impacterait fortement les temps de calcul mais également agirait sur la cohérence de l'interprétation des proximités entre points observations. Outre le cas des données textuelles et/ou image, on peut citer également les données génomiques où l'espace des variables est généralement nettement plus grand que celui des individus. Comment alors réduire la dimensionnalité avec une perte minimale d'information ? Par sélection ? Par élimination ? Par projection ? Et, surtout, comment évaluer la pertinence du nouvel espace de représentation ? De plus, les données qui arrivent sont parfois entachées de bruits et incomplètes. Par exemple, le contenu des courriers électroniques réunit bon nombre de ces problèmes. A défaut de redresser les anomalies, peut-on au moins prendre en compte l'incomplétude, l'incertitude et l'imprécision dans nos analyses ?

---

<sup>1</sup> Interaction Corpus Apprentissage Représentation

Nous avons réalisé de nombreux travaux autour d'applications réelles dans les domaines du marketing, de l'exploitation de textes juridiques, de l'identification de plancton à partir d'images, d'identification de structures de courbes dans les flots de séries chronologiques... où ces questions ont été centrales.

L'expertise acquise par le laboratoire dans ce domaine est conséquente et a produit des résultats théoriques et méthodologiques particulièrement intéressants. Par exemple, un test statistique non paramétrique pour mesurer la séparabilité des classes en vue d'un apprentissage supervisé, des stratégies de détection d'objets atypiques dans des espaces multidimensionnels, des mesures de similarités sur des textes permettant de s'affranchir d'une vectorisation explicite ou encore le recours à des approches de taxinomie comme les cartes de Kohonen pour réduire la dimensionnalité.

### **2.3.4 Contributions à la fouille de données**

La fouille de données opère, le plus souvent, sur des structures tabulaires, préparées à la phase précédente. C'est, généralement, la partie la plus visible du processus de fouille car, c'est à ce stade que l'on produit les connaissances sous la forme de modèles : règles logiques, algébriques, probabilistes, topologiques etc. Et, pour cela, on fait appel aux méthodes d'apprentissage, qu'elles soient supervisées ou non, aux méthodes exploratoires comme les algorithmes de recherche des règles d'association, aux analyses factorielles, ou aux méthodes de modélisation comme les réseaux bayésiens etc. Nous allons décrire quelques uns des travaux réalisés.

- La mise en œuvre de méthodes de fouille dans le cadre de l'ECDC a fait apparaître, à la fois, des problèmes théoriques et des problèmes de mise en œuvre pratique sur machine. En effet, même si, au stade de la fouille, les données sont structurées sous forme tabulaire, leur volume peut être très grand et par conséquent peut poser des problèmes de temps de calcul. Dans ce cadre, nous avons été amenés à réfléchir dans deux directions. La première part du constat qu'un fichier, même de taille très grande, reste un échantillon issu d'une population plus large. Par conséquent, si le même traitement était effectué de façon itérative sur ce même fichier enrichi, itérativement, par de nouveaux cas, le modèle qui en résulterait serait très probablement différent par sa structure et par son taux d'erreur. D'où l'idée d'exploiter à fond cette piste en travaillant sur de petits échantillons dont on fait croître la taille par ajout aléatoire de cas jusqu'à ce que la variance du taux d'erreur par exemple devienne quasi nulle. On utilisera ainsi, de manière efficiente l'information disponible sans subir son poids massif. La seconde direction visait à mieux exploiter les technologies informatiques disponibles. Il s'agit de voir comment faire migrer les méthodes de fouille de données vers des structures systèmes et logiciels qui supportent l'aspect massif ? Dans ce contexte, nous avons réalisé des couplages

forts entre les systèmes de gestion de bases de données capables de travailler sur des vues (tableaux) de taille quasi illimitée et les algorithmes de fouille comme les arbres de décision. On peut ainsi bénéficier des structures de données, notamment les indexes *bitmap*, pour implémenter de manière plus efficace et scalable de nombreux algorithmes de fouille. On s'oriente ainsi vers l'introduction de nouveaux opérateurs de fouille de données aux côtés des opérateurs SQL classiques au sein des systèmes de gestion des entrepôts de données ce qui peut, par là-même, conduire à des plates formes logicielles intégrées.

- En apprentissage automatique, nous travaillons sur des données réelles et nous avons été souvent confrontés à la sous-représentation de certaines classes d'intérêt. Dans ce cas, la mise en œuvre des méthodes d'apprentissage supervisé nécessite une prise en compte et un contrôle de l'asymétrie des classes. A ce propos, nous avons effectué un travail quasi exhaustif sur les mesures généralement utilisées dans les arbres de décision, dans l'extraction de règles d'association etc. L'étude à la fois des propriétés théoriques de ces mesures ainsi que leurs performances sur des Benchmarks a débouché sur de nouvelles mesures d'entropie généralisées. Nous avons également proposé de nouvelles axiomatiques pour ces mesures qui donnent par ailleurs des résultats plus probants sur des cas pratiques.
- A l'issue du processus de fouille de données, les utilisateurs préfèrent disposer de modèles de prédiction intelligibles comme ceux qui sont issus des arbres ou des graphes de décision et qui s'expriment sous la forme de règles logiques. Mais, l'utilisation de ces algorithmes se heurte à des difficultés quand par exemple les variables possèdent une large distribution qui nécessite des groupements de modalités ou quand la variable à prédire est d'un genre particulier par exemple une courbe de survie ou un vecteur. Nous avons proposé plusieurs extensions aux méthodes basées sur les arbres de décision.

### 2.3.5 Validation-intégration et déploiement

Les modèles issus de l'apprentissage doivent être validés avant d'être utilisés comme connaissance par l'utilisateur ou par un système de décision. La plupart des méthodes d'apprentissage proposent des procédures d'évaluation de la qualité des modèles qui sont généralement basées sur les taux d'erreur en resubstitution ou sur échantillon test. La réduction de ces taux d'erreur a conduit à de nouvelles stratégies d'apprentissage comme le *bagging*, le *boosting* ou encore l'apprentissage semi-supervisé. Les évaluations qui ont été conduites au sein du laboratoire ont montré, de façon évidente, l'intérêt d'exploiter ces techniques de ré-échantillonnage autour des algorithmes d'apprentissage et des extensions que nous avons ajoutées.

Les connaissances produites de façon automatique, ne constituent souvent qu'un fragment de connaissances pour bâtir de vrais systèmes d'aide à la décision. Dans ce cadre, et afin d'accroître les performances et l'intérêt de ces systèmes, nous avons développé et testé des méthodologies pour intégrer dans une même base de connaissances, celles qui proviennent du domaine et/ou de l'expert. Cette intégration se fait au moyen d'ontologies. Ces ontologies servent à la fois d'outil d'expansion et d'unification de fragments de connaissances provenant de sources différentes et de réceptacle de connaissances.



### 3 VALORISATION SCIENTIFIQUE

#### 3.1 Publications

Le tableau ci-dessous résume le bilan scientifique quantitatif sur la période 2005-2008. Une liste complète des publications est donnée en annexe.

<b>Publications</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>Total</b>
Ouvrages et actes	2	1	2	2	<b>7</b>
Revue internationale	4	4	5	4	<b>17</b>
Chapitres d'ouvrage international	2	3	5	5	<b>15</b>
Revue nationale	2	2	1	0	<b>5</b>
Chapitres d'ouvrage national	0	0	1	1	<b>2</b>
Conférences internationales	21	33	22	22	<b>98</b>
Conférences nationales	16	20	19	17	<b>72</b>
<b>Total</b>	<b>47</b>	<b>63</b>	<b>55</b>	<b>51</b>	<b>216</b>
Chercheurs statutaires	11	11	11	11	/
Thèses soutenues	1	3	1	5	<b>10</b>
HDR soutenues	0	2	1	0	<b>3</b>
Etudiants en DEA/Master ECD	30	25	35	36	<b>126</b>

Tableau 1 : Bilan scientifique 2005-2008

#### 3.2 Activités Editoriales

- D. A. Zighed (ERIC, Lyon2) et G. Venturini (LI, Tours) sont co-directeurs de la Revue des Nouvelles Technologies de l'Information (RNTI) publiée par Cépaduès. Les publications RNTI (<http://www.antsearch.univ-tours.fr/rnti>) sont des numéros spéciaux autour des problématiques de la fouille de données et de l'Extraction des Connaissances à partir des Données. La liste des numéros parus de 2005 à 2008 se trouve en annexe.
- J. Darmont est membre des comités éditoriaux suivants :
  - International Journal of Biomedical Engineering and Technology (IJBET, see <http://www.inderscience.com/browse/index.php?journalCODE=ijbet>)
  - Idea Groupe Inc. Editorial Advisory Review Board

- Editorial Review Board of the Advances in Data Warehousing and Mining (ADWM, see [http://users.monash.edu.au/%7Edtaniar/book\\_series\\_warehousing.html](http://users.monash.edu.au/%7Edtaniar/book_series_warehousing.html))

- F. Bentayeb, O. Boussaid, J. Darmont et S. Loudcher font partie du comité de pilotage de la conférence “Entrepôts de Données et Analyse en ligne (EDA)”.

### 3.3 Animations scientifiques

Les membres du laboratoire ERIC sont régulièrement impliqués dans de nombreuses manifestations scientifiques, conférences, séminaires, groupes de travail, ...

#### 3.3.1 Conférences et ateliers

- Conférence “Extraction et Gestion des Connaissances” (EGC 2000-2008).
- Conférence “Entrepôts de Données et Analyse en Ligne” (EDA 2005-2008).
- Atelier “Mining Complex Data” en association avec ICDM IEEE International Conference (2005-2006) et PKDD International conference (2007).
- Atelier “Qualité des Données et des Connaissances”, en association avec la conférence “Extraction et Gestion des Connaissances” (2007-2008).
- Atelier “Fouille de Données Complexes”, en association avec la conférence “Extraction et Gestion des Connaissances” (2007-2008).
- Atelier “Mesure de similarité sémantique” en association avec la conférence “Extraction et Gestion des Connaissances” (2007-2008).
- Atelier “Systèmes Décisionnels” (ASD 2006-2008).

Le détail de ces manifestations est en annexe.

#### 3.3.2 Groupes de travail

Le laboratoire ERIC a créé et organise régulièrement le groupe de travail sur la “Fouille de Données Complexes” (<http://eric.univ-lyon2.fr/~gt-fdc/>).

#### 3.3.3 Séminaires

Le laboratoire ERIC organise, deux fois par mois, des séminaires avec des intervenants d’horizons différents (<http://eric.univ-lyon2.fr/index.php?section=6&soussection=14> et <http://dea-eed.univ-lyon2.fr/?page=seminaire&section=0>). Les séminaires ont pour objectifs de :

- mettre en relation les membres du laboratoire avec d’autres chercheurs dont certains viennent de l’étranger,

- obtenir un point de vue différent sur des problèmes qui entrent dans le cadre de l'activité de recherche du laboratoire,
- permettre aux chercheurs, notamment les doctorants, de présenter leurs travaux récents,
- mieux connaître des sujets connexes aux préoccupations du laboratoire.

La liste complète des séminaires figure en annexe.

### 3.4 Projets de recherche appliquée

Les collaborations des membres d'ERIC (voir la liste complète en annexe) sont de différents ordres.

On citera pour l'essentiel les collaborations avec :

- des équipes de recherche locales en Sciences Humaines et Sociales à l'Université Lyon 2 : DDL, CED, ICAR ;
- des équipes nationales ou dans le cadre de projets nationaux ACI : TELECOM, LSIIT et LIV, IRC;
- l'industrie : Crédit Lyonnais, Védior Bis, l'hôpital Léon Berard, ... ;
- des équipes internationales, notamment avec les universités de Genève (Suisse), d'Oklahoma (USA), de la Mer Egée (Grèce) qui ont conduit à de nombreuses publications conjointes.

Le laboratoire ERIC est également impliqué dans l'incubation et la création d'entreprises innovantes. Il leur offre l'expertise scientifique par le biais de différents types de collaborations.

Avec l'incubateur CREALYS, ERIC est actuellement ou a été impliqué dans la création de sociétés dans différents secteurs :

- TradingBots (2006-2007) : outils pour l'analyse des données financières et l'aide à la décision boursière ;
- Tapeo (2007-2008) : gestion de portefeuilles virtuels d'actions détenues par des communautés d'utilisateurs sur le Web.
- Rickshaw (2007-2008) : guides collaboratifs sur Internet pour rassembler les informations et opinions les plus pertinentes sur un sujet, un produit ou un service.
- ProXan (2008-2009) : évaluation de zones de chalandise avec un outil d'aide à l'implantation des commerces : comparaison de l'environnement socio-économique et démographique.
- Between (2008-2009) : modèle de représentation et d'analyse des débats en ligne sur Internet.

### 3.5 Développement de logiciels

TANAGRA est un logiciel (*open source*) de fouille de données destiné à l'enseignement et à la recherche. Le projet a commencé en Janvier 2004. Le logiciel est disponible gratuitement sur le Web<sup>1</sup>. TANAGRA met en œuvre diverses méthodes statistiques et d'apprentissage automatique. A ce jour, il y a environ 130 méthodes implémentées.

Le principal objectif du projet est de proposer aux chercheurs et aux enseignants un outil qui respecte les normes du domaine de la fouille de données. Les utilisateurs utilisent le logiciel pour des études universitaires, pour leurs activités de recherche, mais aussi pour leurs publications. TANAGRA est maintenant reconnu par la communauté. Il est référencé dans les études comparatives et dans les projets avec des données réelles (par exemple, X. Chen, Y. Ye, G. Williams, X. Xu, "A Survey of Open Source Data Mining Systems", Industrial Track Workshop, PAKDD-2007, 3-14.).

Le logiciel est également sur un site avec de nombreux tutoriels (en français et en anglais) sur l'extraction de données et l'analyse exploratoire des données. Il y a environ 70 tutoriels en ligne. Les pages Web concernant les didacticiels sont les pages les plus visitées du site. Au cours de l'année 2008, il y a eu en moyenne près de 4000 visiteurs par mois (# 130 visiteurs par jour).

### 3.6 Synergie entre enseignement et recherche

ERIC a toujours connecté ses activités d'enseignement avec ses travaux de recherche, plus particulièrement avec des masters. Nous avons mis en place une palette complète de formations afin de répondre aux besoins des entreprises et de fournir au laboratoire des chercheurs et des doctorants.

Au niveau de la licence, les deux voies principales sont :

- Informatique décisionnelle et économétrie appliquée (IDEA) à la faculté de Sciences Economiques et de Gestion ;
- Mathématiques, Informatique et Statistiques Appliquées aux Sciences Humaines et Sociales (MISASHS).

Au niveau du master, nous avons mis en place des formations qui s'articulent autour des systèmes d'aide à la décision et de la statistique. Quatre spécialisations sont possibles pour la deuxième année de master :

- Statistique, Informatique et Socio-Economique (SISE). Son contenu forme les étudiants au traitement statistique des données dans les domaines du marketing ou de l'industrie

---

<sup>1</sup> <http://eric.univ-lyon2.fr/~ricco/Tanagra/>

(pharmacie, etc.) Cette spécialité est également réalisée dans le contexte d'un master commun avec l'Université de Kharkov (Ukraine).

- Ingénierie Informatique pour la Décision et l'Évaluation Économiques (IIDEE) dont le contenu est plus ciblé sur le développement d'outils pour les systèmes d'aide à la décision. Cette spécialité forme chaque année une deuxième promotion en cours du soir pour des professionnels.
- Organisation et Protection des Systèmes d'Information dans les Entreprises (OPSIE) dont le programme a trois facettes : technologies de l'information, la gestion et le droit. Cette spécialité forme également chaque année une deuxième promotion en cours du soir pour des professionnels.
- Extraction des Connaissances à partir des Données (ECD) qui met plus l'accent sur nos activités de recherche et qui forme la plus grande partie de nos futurs doctorants. Cette spécialité est co-habilitée depuis sa création par l'École Polytechnique de l'Université de Nantes et jusqu'en 2006 par l'université d'Orsay Paris 11. Les cours sont assurés en visioconférence ce qui permet à certains de nos étudiants de suivre les cours depuis leur lieu de résidence en France ou à l'étranger. C'est le cas, par exemple, pour les étudiants roumains de Bucarest et les étudiants vietnamiens de Cantho.

Le détail des cours est disponible sur le site de l'université et en particulier sur celui du département d'informatique et de statistique de la faculté de sciences économiques et de gestion (<http://dis.univ-lyon2.fr/>).

Nous poursuivons nos efforts d'ouverture pour établir des partenariats avec d'autres universités, y compris européennes, et attirer des bons étudiants. Dans cette perspective, et pour donner un impact plus fort, nous allons soumettre à la commission européenne, un projet de création d'un master européen dans le cadre du programme Erasmus Mundus. Il sera positionné dans le domaine de l'extraction de données et de la gestion des connaissances. Il sera mis en place avec 6 universités dont 3 étrangères (Italie, Espagne et Roumanie). D'autres actions sont également prévues dans le domaine professionnel avec des écoles d'ingénieurs ou de commerce, ou avec des universités étrangères pour des formations délocalisées.



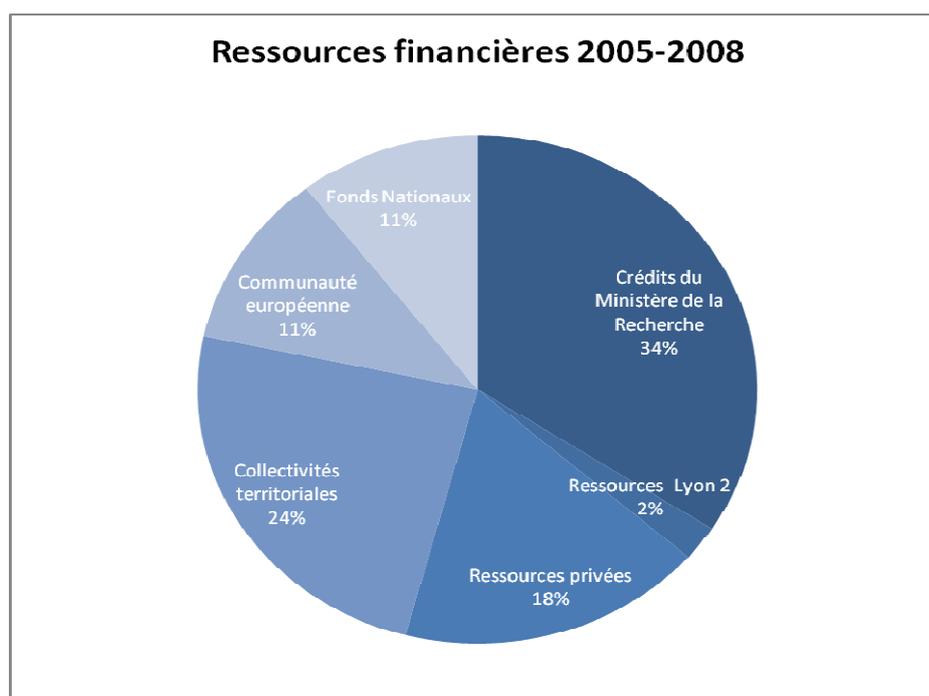
## 4 RESSOURCES

### 4.1 Bilan financier

Les tableaux et graphiques suivants présentent les recettes et les dépenses pour la période 2005-2008.

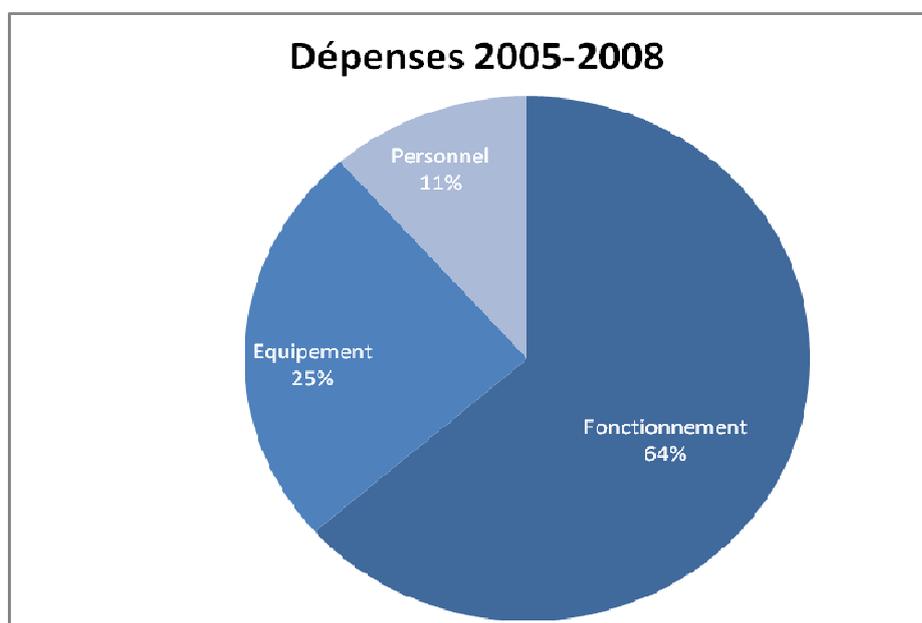
Ressources annuelles (HT)	2005	2006	2007	2008
Crédits du Ministère de la Recherche (BQR déduit)	41 650 €	41 650 €	37 400 €	37 400 €
Ressources supplémentaires provenant de Lyon 2	738 €	1 000 €	1 000 €	7 500 €
Ressources privées (contrats de recherche, prestations, ...)	13 593 €	33 784 €	25 510 €	8 895 €
Collectivités territoriales	4 250 €	4 500 €	33 850 €	69 300 €
Communauté européenne	20 898 €	30 602 €		
Fonds Nationaux pour la Science (ACI, ANR)	20 700 €	27 700 €		
<b>Total</b>	<b>101 829 €</b>	<b>139 236 €</b>	<b>97 760 €</b>	<b>123 095 €</b>

Tableau 2 : Recettes financières pour 2005-2008



<b>Dépenses annuelles (TTC)</b>	2005	2006	2007	2008
Fonctionnement	81 275 €	83 944 €	62 558 €	68 203 €
Equipement	18 958 €	39 620 €	12 636 €	45 169 €
Personnel	25 204 €	13 993 €	6 512 €	7 878 €
<b>Total</b>	<b>125 437 €</b>	<b>137 557 €</b>	<b>81 705 €</b>	<b>121 250 €</b>

Tableau 3 : Dépenses pour 2005-2008

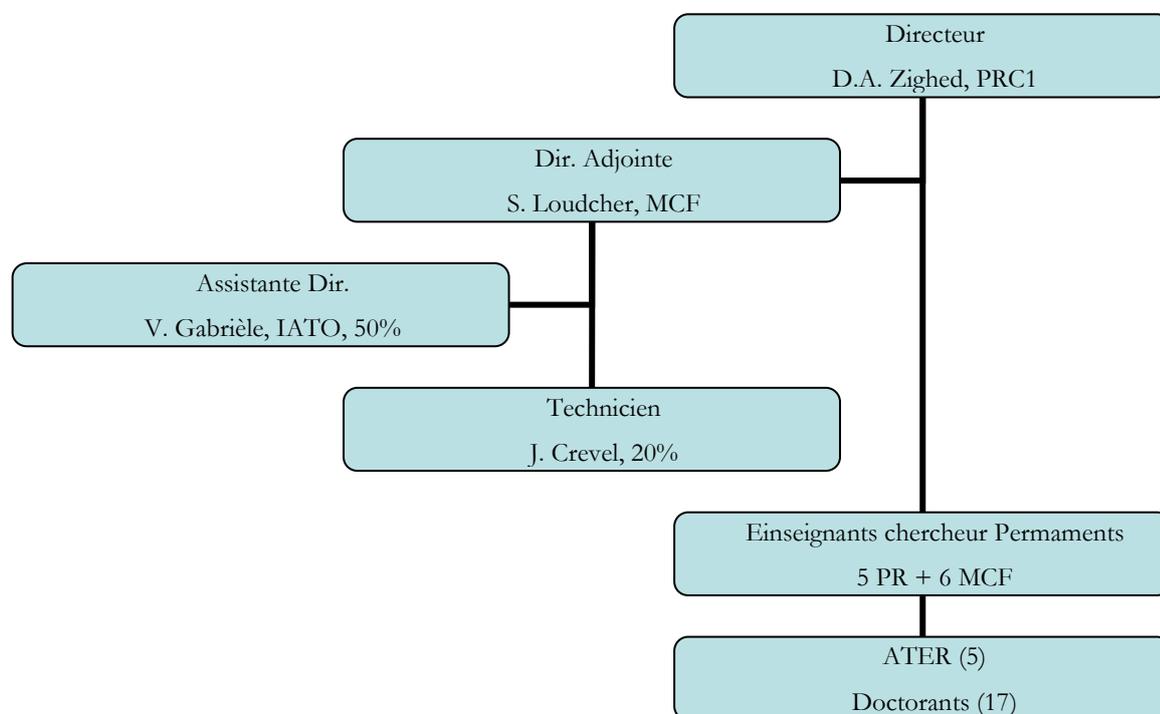


Le décalage qui peut apparaître, certaines années, entre les dépenses et les recettes s'explique par plusieurs faits :

- En 2005, les réserves financières issues de l'année 2004 s'élevaient à 18311€.
- Dans les projets nationaux, européens, ... les soldes des subventions sont versés à la fin des projets après justification des dépenses et parfois l'année suivante.

## 4.2 Ressources humaines au 31 décembre 2008

### 4.2.1 Organigramme de l'équipe



### 4.2.2 Enseignants-chercheurs statutaires

Nom, Prénom, Date de Naissance	Corps grade	Section CNU	Date d'arrivée
Bentayeb Fadila, 15 mai 1966	MCF	27	oct-01
Bousaïd Omar, 2 juin 1954	PR2	27	janv-95
Chauchat Jean-Hugues, 6 juillet 1946	PR2	27	juin-97
Darmont Jérôme, 15 janvier 1972	PR2	27	oct-99
Harbi Nouria, 27 août 1961	MCF	27	oct-05
Lallich Stéphane, 20 septembre 1947	PR2	27	juin-97
Loudcher Rabaséda Sabine, 27 octobre 1969	MCF	27	oct-98
Rakotomalala Ricco, 19 juillet 1967	MCF	27	oct-98
Velcin Julien, 09 mars 1978	MCF	27	oct-07
Viallaneix Jacques, 6 juillet 1963	MCF	27	janv-95
Zighed Abdelkader, 12 mars 1955	PR1	27	janv-95

### 4.2.3 ATER

Nom, Prénom	Année universitaire
Bimonte Sandro	2008-2009
Favre Cécile	2007-2008 2008-2009
Mahboubi Hadj	2007-2008 2008-2009
Maïz Nora	2007-2008 2008-2009
Elie Prudhomme	2008-2009

### 4.2.4 Thèses en cours

Nom, Prénom	Début	Directeur	Co directeur	Financement
Bahri Emma	2006	S. Lallich		Bourse MENRT
Ben Hassine Soumaya	2008	J.H. Chauchat et J. Darmont		Bourse CIFRE
Bouatour Sonia	2007	O. Boussaid (Co-tutelle avec la Tunisie)		Ressources propres
Forestier Mathilde	2008	D. Zighed	J. Velcin	Bourse MENRT Moniteur
Gaudin Rémi	2004	D. Zighed		Bourse MENRT Moniteur
Gicquel Quentin	2008	S. Lallich		Bourse CIFRE
Grabova Oksana	2008	J.H. Chauchat et J. Darmont (Co-tutelle avec l'Ukraine)		Bourse du gouvernement ukrainien
Hachicha Marouane	2007	J. Darmont		Bourse MENRT Moniteur
Maïz Nora	2005	O. Boussaid	F. Bentayeb	Ressources propres
Mavrikas Efthimios	2002	D. Zighed S. Dascalopoulos (Co-tutelle avec l'Université de l'Egée)		Bourse de la Grèce
Pisetta Vincent	2008	D. Zighed		Bourse CIFRE
Prudhomme Elie	2005	S. Lallich		Bourse MENRT
Qureshi Taimur	2006	D. Zighed		Bourse du gouvernement pakistanais
Ralaivao Jean-Christian	2003	S. Lallich V. Manantsoa (Co-tutelle avec l'Université de Fianarantsao)	J. Darmont	Bourse de l'Ambassade de France
Salem Rashed Kh.	2007	O. Boussaid et J. Darmont		Bourse du gouvernement égyptien

Stavrianou Anna	2005	JH. Chauchat		Bourse MENRT
Wei Zhihua	2006	JH. Chauchat		Bourse du gouvernement chinois

#### 4.2.5 Thèses soutenues

Nom, Prénom	Année	Directeur	Co directeur	Devenir
Aouiche Kamel	2005	D. Zighed	J. Darmont	Post Doc au Canada
Ben Messaoud Riadh	2006	N. Nicoloyannis	O. Boussaid S. Loudcher	MCF en Tunisie
Clerc Frédéric	2006	N. Nicoloyannis	R. Rakotomalala	Privé
Charbel Julien	2008	D. Zighed		Privé
El Sayed Ahmad	2008	D. Zighed	F. Bentayeb	Privé
Fangseu Badjio Edwige	2006	D. Zighed	F. Poulet	Privé
Favre Cécile	2007	O. Boussaid	F. Bentayeb	ATER
Hacid Hakim	2008	D. Zighed		Privé
Mahboubi Hadj	2008	J. Darmont		ATER
Marcellin Simon	2008	D. Zighed		Privé
Thomas Julien	2009	D. Zighed		Privé

#### 4.2.6 Habilitations à diriger des recherches

Nom, Prénom	Année	Directeur	Devenir
Lenca Philippe	2007	D. Zighed	
Boussaid Omar	2006	D. Zighed	Lyon 2
Darmont Jérôme	2006	D. Zighed	Lyon 2

#### 4.2.7 Personnel administratif

Nom, Prénom	Corps grade	Quotité recherche	Date d'arrivée
Gabrièle Valérie	IATOS	0,5	Sept-00
Crevel Julien	Technicien	0,2	Sept-07

#### 4.2.8 Récapitulatif au 31 décembre 2008

Catégorie	Effectif
Enseignants-chercheurs statutaires	11
ATER	5
Thèses en cours	17
Thèses soutenues	11
HDR soutenues	3
Personnel administratif	2

#### 4.2.9 Personnes ayant terminé leur contrat ou quitté le laboratoire

##### Enseignants-chercheurs statutaires

Nom, Prénom	Corps grade	Section CNU	Date d'arrivée	Date de départ
Viallefont Anne	MCF	26	oct-00	sept-06

##### ATER

Nom, Prénom	Année universitaire
Arigon Anne-Muriel	2006-2007 et 2007-2008
Ben Messaoud Riadh	2006-2007
Effantin Dit Toussaint Brice	2004-2005
Kouomou Choupo Anicet	2005-2006
Legrand Gaëlle	2004-2005
Lefort Virginie	2006-2007 et 2007-2008
Suchier Maxime	2006-2007
Walid Erray	2003-2004 et 2004-2005

##### Post Doc rattaché au laboratoire

Nom, Prénom	Année universitaire
Jouve Pierre	2004-2005

## **5 FORMATION DES PERSONNELS**

Les doctorants et/ou enseignants-chercheurs du laboratoire ERIC suivent régulièrement des formations en anglais soit sous forme annuelle (2h par semaine) soit sous forme intensive (6h par jour pendant 5 jours de suite). Chaque année le laboratoire ERIC compte 5 à 6 chercheurs en formation.

Plusieurs enseignants-chercheurs ont bénéficié ou bénéficient d'un congé de recherche (CRCT) afin de développer leurs recherches notamment avec des universités étrangères.

Le laboratoire ERIC encourage chaque année le personnel administratif et technique à suivre des formations, notamment en :

- Finances publiques et logiciel spécifique
- Création de site Web sous SPIP
- Bureautique
- Hygiène et sécurité

## **6 HYGIENE ET SECURTIE**

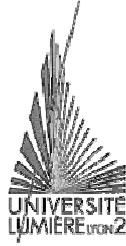
Les locaux actuels sur le Campus Porte des Alpes de l'université Lyon 2 n'assurent plus les conditions de sécurité et d'hygiène nécessaires à l'accueil des personnels. L'université est en train d'étudier comment reloger les chercheurs du laboratoire ERIC rattachés à Lyon 2.



# Table des matières

<b>1</b>	<b>Note de synthèse</b>	<b>5</b>
<b>2</b>	<b>Travaux scientifiques</b>	<b>9</b>
2.1	Positionnement scientifique de ERIC	9
2.1.1	Point de vue historique	9
2.1.2	Point de vue pragmatique	10
2.2	Présentation de l'ECD	12
2.3	Avancées scientifiques 2005-2008	14
2.3.1	Contributions à l'entreposage de données complexes	15
2.3.2	Contributions à la recherche d'information dans les entrepôts de données complexes	17
2.3.3	Contributions à la préparation des données	18
2.3.4	Contributions à la fouille de données	19
2.3.5	Validation-intégration et déploiement	20
<b>3</b>	<b>Valorisation scientifique</b>	<b>23</b>
3.1	Publications	23
3.2	Activités Editoriales	23
3.3	Animations scientifiques	24
3.3.1	Conférences et ateliers	24
3.3.2	Groupes de travail	24
3.3.3	Séminaires	24
3.4	Projets de recherche appliquée	25
3.5	Développement de logiciels	26
3.6	Synergie entre enseignement et recherche	26
<b>4</b>	<b>Ressources</b>	<b>29</b>
4.1	Bilan financier	29
4.2	Ressources humaines au 31 décembre 2008	31
4.2.1	Organigramme de l'équipe	31
4.2.2	Enseignants-chercheurs statutaires	31
4.2.3	ATER	32
4.2.4	Thèses en cours	32
4.2.5	Thèses soutenues	33
4.2.6	Habilitations à diriger des recherches	33
4.2.7	Personnel administratif	33

4.2.8	Récapitulatif au 31 décembre 2008.....	34
4.2.9	Personnes ayant terminé leur contrat ou quitté le laboratoire.....	34
<b>5</b>	<b>Formation des personnels .....</b>	<b>35</b>
<b>6</b>	<b>Hygiène et sécurité.....</b>	<b>35</b>



# Bilan ERIC : annexes

2005-2008

Laboratoire ERIC  
Université Lumière Lyon 2  
5, avenue Pierre Mendès-France  
Bât L.  
69600 Bron France

Tel. +33 4 78 77 23 76  
Fax. +33 4 78 77 23 75  
Web. <http://eric.univ-lyon2.fr>



## Table des annexes

<b>I. Publications 2005-2008 .....</b>	<b>43</b>
a. Ouvrages et actes (Eds.).....	43
b. Revues internationales.....	43
c. Chapitres d'ouvrage international.....	44
d. Revues nationales.....	45
e. Chapitres d'ouvrage national.....	46
f. Conférences internationales.....	46
g. Conférences nationales.....	53
<b>II. Activite éditoriale.....</b>	<b>61</b>
<b>III. Organisation de manifestations scientifiques.....</b>	<b>63</b>
a. Conférences, ateliers et groupes de travail.....	63
b. Séminaires du master ECD.....	65
c. Séminaires du laboratoire ERIC.....	67
<b>IV. Projets de recherche appliquée.....</b>	<b>69</b>
<b>V. Collaborations internationales .....</b>	<b>75</b>



## I. PUBLICATIONS 2005-2008

Les publications sont numérotées de 1 à 215. De plus, chaque numéro est suivi du code du type de publications. Les auteurs appartenant (ou ayant appartenu) au laboratoire ERIC sont soulignés.

### a. Ouvrages et actes (Eds.)

1. OS - A. Ciampi, D. Zighed, G. Ritschard, *Prediction Trees*, John Wiley & Sons, Hoboken, USA, 2008.
2. DO - S. Lallich, P. Lenca, F. Guillet, Eds., *Actes du 4ème Atelier Qualité des Données et des Connaissances (QDC-EGC 08)*, Nice, Janvier 2008; INRIA Sophia antipolis.
3. DO - S. Lallich, D. Pastor, Eds., *Special Issue on the ASMDA International Symposium on Applied Stochastic Models and Data Analysis, Communications in Statistics - Theory and Methods*, Vol. 36(14), Taylor & Francis, January 2007 (Special issue).
4. DO - S. Lallich, P. Lenca, F. Guillet, Eds., *Actes du 3ème Atelier Qualité des Données et des Connaissances (QDC-EGC 07)*, Namur, Belgique, Janvier 2007.
5. DO - J. Darmont, O. Boussaïd, Eds., *Managing and Processing Complex Data for Decision Support*, Idea Group Publishing, April 2006.
6. DO - F. Bentayeb, O. Boussaïd, J. Darmont, S. Loudcher-Rabaseda, Eds., *Actes de la 1ère journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 05)*, Juin 2005, *Revue des Nouvelles Technologies de l'Information*, Vol. B-1, Cépaduès, Toulouse.
7. DO - O. Boussaïd, P. Gançarski, F. Masegla, B. Trousse, Eds., *Actes de l'atelier Fouille de Données Complexes (FDC-EGC 05)*, Janvier 2005, *Revue des Nouvelles Technologies de l'Information*, Vol. 3, Cépaduès, Toulouse.

### b. Revues internationales

8. ACL - P. Lenca, P. Meyer, B. Vaillant, S. Lallich, "On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid", *European Journal of Operational Research*, Vol. 184, No. 2, 2008, 610–626.
9. ACL - P. Lenca, S. Lallich, B. Vaillant, "Construction of an off-centered entropy for the supervised learning of imbalanced classes: Some first results", *Communications in Statistics - Theory and Methods*, 2008.
10. ACL - F. Bentayeb, C. Favre, O. Boussaïd, "A User-driven Data Warehouse Evolution Approach for Concurrent Personalized Analysis Needs Integrated Computer-Aided Engineering", *Journal of Integrated Computer-Aided Engineering*, Vol. 15, No. 1, 2008, 21-36.
11. ACL - O. Boussaïd, J. Darmont, F. Bentayeb, S. Loudcher-Rabaseda, "Warehousing complex data from the Web", *International Journal of Web Engineering and Technology*, Vol. 4, No. 4, 2008, 408-433.
12. ACL - F. Bentayeb, J. Darmont, C. Favre, C. Udréa, "Efficient On-Line Mining of Large Databases", *International Journal of Business Information Systems*, Vol. 2, No. 3, 2007, 328-350.

13. ACL - A. Stavriano, P. Andritsos, N. Nicoloyannis, "Overview and Semantic Issues of Text Mining", *SIGMOD Record*, Vol. 36, No. 3, September 2007, 23-34.
14. ACL - S. Lallich, P. Lenca, B. Vaillant, "Probabilistic framework towards the parametrisation of association rule interestingness measures", *Methodology and Computing in Applied Probability*, Vol. 9, No. 3, 2007, 447-463.
15. ACL - J. Darmont, F. Bentayeb, O. Boussaïd, "Benchmarking Data Warehouses", *International Journal of Business Intelligence and Data Mining*, Vol. 2, No. 1, 2007, 79-104.
16. ACL - O. Boussaïd, A. Tanasescu, F. Bentayeb, J. Darmont, "Integration and Dimensional Modelling Approaches for Complex Data Warehousing", *Journal of Global Optimization*, Vol. 37, No. 4, April 2007, 571-591.
17. ACL - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "A Data Mining-Based OLAP Aggregation of Complex Data: Application on XML Documents", *International Journal of Data Warehousing and Mining*, Vol. 2, No. 4, Oct.-Dec. 2006, 1-26.
18. ACL - F. Mhamdi, R. Rakotomalala, M. Elloumi, "A Compromise Between N-gram Length and Classifier Characteristics for Protein Classification", *International Journal of Computer Science and Network Security*, Vol. 6, No. 4, 2006, 82-87.
19. ACL - F. Clerc, D. Farrusseng, R. Rakotomalala, N. Nicoloyannis, C. Mirodatos, "Meta Modeling for Combinatorial Catalyst Optimization", *International Journal of Computer Science and Network Security*, Vol. 6, No. 10, 2006, 256-262.
20. ACL - R. Rakotomalala, F. Mhamdi, "Supervised and Unsupervised Feature Reduction for Protein Classification", *WSEAS Transactions on Information Science and Applications*, Vol. 3, No. 12, 2006, 2448-2455.
21. ACL - M. Scuturici, J. Clech, V. Scuturici, D. Zighed, "Topological representation model for image databases query", *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 17, No. 1-2, 2005, 145-160.
22. ACL - Z. He, J. Darmont, "Evaluating the Dynamic Behavior of Database Applications", *Journal of Database Management*, Vol. 16, No. 2, April-June 2005, 21-45.
23. ACL - D. Zighed, G. Ritschard, W. Erray, V. Scuturici, "Decision tree with optimal join partitioning", *Journal of Intelligent Information Systems*, Vol. 20, 2005, 1-26.
24. ACL - D. Zighed, S. Lallich, F. Muhlenbach, "A statistical approach of class separability", *Applied Stochastic Models in Business and Industry*, Vol. 21, No. 2, 2005, 187-197.

### c. Chapitres d'ouvrage international

25. OS - R. Rakotomalala, A. Morineau, "The TVpercent principle for the counterexamples statistic", *Statistical Implicative Analysis: theory and applications*, Vol. 127, Springer, Heidelberg, Germany, 2008, 449-462.
26. OS - D. Boukraa, O. Boussaïd, "Dimensional Modeling for Complex Data", *Encyclopedia of Data Warehousing and Mining, Second Edition*, IGI Publishing, Hershey, PA, USA, 2008.
27. OS - R. Rakotomalala, F. Mhamdi, "Using the Text Categorization Framework for the Protein Classification", *Handbook of Research on Text and Web Mining Technologies*, Information Science Reference, September 2008, 297-311.

28. OS - B. Vaillant, S. Lallich, P. Lenca, "On the behaviour of the generalisations of the intensity of implication: a data-driven comparative study", *Statistical Implicative Analysis: theory and applications*, Springer, Heidelberg, Germany, 2008.
29. OS - J. Darmont, E. Olivier, "Biomedical Data Warehouses", *Encyclopaedia of Healthcare Information Systems*, Idea Group Publishing, May 2008, 149-156.
30. OS - R. BenMessaoud, S. Loudcher-Rabaseda, "OLEMAR: an On-Line Environment for Mining Association Rules in Multidimensional Data", *Advances in Data Warehousing and Mining*, Vol. 2, Idea Group Publishing, 2007.
31. OS - P. Lenca, B. Vaillant, P. Meyer, S. Lallich, "Association rule interestingness measures: experimental and theoretical studies", *Quality Measures in Data Mining*, Springer, Heidelberg, Germany, 2007, 51-76.
32. OS - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "A multiple correspondence analysis to organize data cubes", *Databases and Information Systems IV - Frontiers in Artificial Intelligence and Applications*, Vol. 155(1), IOS Press, 2007, 133-146.
33. OS - D. Zighed, "Induction Graphs for Data Mining", *Studies in Classification, Data Analysis and Knowledge Organisation*, Springer, Heidelberg, Germany, 2007, 419-430 (In Selected Contributions in Data Analysis and Classification).
34. OS - M. Bouet, M. Aufaure, P. Gançarski, O. Boussaïd, "Pattern Mining and Clustering on Image Databases", *Successes and New Directions in Data Mining*, Idea Group Publishing, 2007, 187-212.
35. OS - H. Hacid, D. Zighed, "A Machine Learning Based Model For Content Based Image Retrieval", 2006.
36. OS - Z. He, J. Darmont, "Evaluating the Performance of Dynamic Database Applications", *Advanced Topics in Database Research*, Vol. 5, Idea Group Publishing, 2006, 294-319.
37. OS - S. Lallich, O. Teytaud, E. Prudhomme, "Association rules interestingness: measure and validation", *Quality Measures in Data Mining*, Springer, Heidelberg, Germany, 2006.
38. OS - F. Muhlenbach, R. Rakotomalala, "Discretization of Continuous Attributes", *Encyclopedia of Data Warehousing and Mining, Second Edition*, Idea Group Publishing, 2005, 397-402.
39. OS - J. Darmont, "Object Database Benchmarks", *Encyclopedia of Information Science and Technology*, Vol. 1, Idea Group Publishing, January 2005, 2146-2149.

#### **d. Revues nationales**

40. ACL - R. Rakotomalala, "Data Mining : Spécificités et outils", *Actes de Chimométrie*, Novembre 2007, 108 - 110 (Lyon).
41. ACL - J. Labarère, J. Bosson, D. Farrusseng, B. Crémilleux, R. Rakotomalala, C. Robert, "Arbres d'Induction : méthodes et exemple d'application", *Journal d'Economie Médicale*, Vol. 24, No. 2, 2006, 115-129.
42. ACL - M. Raimbault, R. Rakotomalala, X. Morandi, P. Jannin, "Mise en évidence d'invariants dans une population de cas chirurgicaux", *Revue des Nouvelles Technologies de l'Information*, Vol. E-5, 2006, 339-348.
43. ACL - R. Rakotomalala, "TANAGRA, une plate-forme d'expérimentation pour la fouille de données", *MODULAD*, No. 32, 2005, 70-85.

44. ACL - K. Aouiche, J. Darmont, O. Boussaïd, F. Bentayeb, "Auto-administration des entrepôts de données complexes", *Revue des Nouvelles Technologies de l'Information*, Vol. E-4, Septembre 2005, 47-70.

## e. Chapitres d'ouvrage national

45. OS - R. Rakotomalala, F. Mhamdi, "Evaluation des méthodes supervisées pour la discrimination de protéines", *Revue des Nouvelles Technologies de l'Information*, Vol. C-2, Cépaduès, Toulouse, 2008, 255-266.

46. OS - R. Rakotomalala, T. LeNouvel, "Interactive Clustering Tree : Une méthode de classification descendante adaptée aux grands ensembles de données", *Revue des Nouvelles Technologies de l'Information*, Vol. A1, Cépaduès, Toulouse, 2007, 75-94 (In Data Mining et apprentissage statistique : application en assurance, banque et marketing).

## f. Conférences internationales

47. ACTI - N. Pham, T. Do, P. Lenca, S. Lallich, "Using local node information in decision trees: coupling a local labeling rule with an off-centered entropy", *International Conference on Data Mining (DMIN'08)*, Las Vegas, Nevada, USA, July 2008; CSREA Press.

48. ACTI - E. Prudhomme, S. Lallich, "Self-Organized map ensemble adapted to large high dimensional data, an application to semi-supervised learning", *17th International Symposium on Methodologies for Intelligent Systems (ISMIS 08)*, Toronto, Canada, May 2008; *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, Germany.

49. ACTI - M. Hachicha, H. Mahboubi, J. Darmont, "Expressing OLAP operators with the TAX XML algebra", *3rd International Workshop on Database Technologies for Handling XML Information on the Web (DataX-EDBT 08)*, Nantes, France, March 2008.

50. ACTI - S. Bouattour, R. BenMessaoud, O. Boussaïd, H. Abdallah, J. Feki, "A Framework for Active Data Warehouses", *International Arab Conference on Information Technology (ACIT 08)*, Hammamet, Tunisie, 2008.

51. ACTI - A. ElSayed, H. Hacid, D. Zighed, "Exploring Validity Indices in Agglomerative Clustering of Documents", *17th International Symposium on Methodologies for Intelligent Systems (ISMIS 08)*, Toronto, Canada, May 2008; *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, Germany.

52. ACTI - A. ElSayed, H. Hacid, D. Zighed, "A New Framework for Taxonomy Discovery from Text", *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 08)*, Osaka, Japan, 2008; *Lecture Notes in Computer Science*, Springer, Heidelberg, Germany.

53. ACTI - N. Harbi, M. Meuke-Fante, F. Bentayeb, O. Boussaïd, "Un méta modèle multidimensionnel générique pour la conception des entrepôts de données", *3ème Atelier sur les Systèmes Décisionnels (ASD 08)*, Mohammédia, Maroc, 2008.

54. ACTI - Z. Wei, J. Chauchat, D. Miao, "Comparing different text representation and feature selection methods on Chinese text classification using character n-grams", *Journées Internationales d'Analyse des Données Textuelles*, Lyon, France, Mars 2008.

55. ACTI - E. Prudhomme, S. Lallich, "Optimization of Self-Organizing Maps Ensemble in Prediction", *International Conference on Data Mining (DMIN'08)*, Las Vegas, Nevada, USA, July 2008; CSREA Press.

- 56. ACTI** - G. Ritschard, V. Pisetta, D. Zighed, "Inducing and evaluating classification trees with statistical implicative criteria", *Statistical Implicative Analysis: Theory and Applications*, R. Gras, E. Suzuki, F. Guillet, F. Spagnolo, Eds., 2008; *Studies in Computational Intelligence*, Springer, Heidelberg, Germany.
- 57. ACTI** - N. Harbi, G. Maaroufi, O. Boussaïd, "Sécurité des entrepôts de données -état de l'art-", *3ème Atelier sur les Systèmes Décisionnels (ASD 08)*, Mohammédia, Maroc, 2008.
- 58. ACTI** - R. Billot, H. Suchier, S. Lallich, "Toward an Unsupervised Boosting-Like Approach", *Société Francophone de Classification and the Classification and Data Analysis Group of the Italian Society of Statistics (SFC CLADAG 2008)*, Caserta, Italie, June 2008.
- 59. ACTI** - A. Stavrianou, E. Bahri, N. Nicoloyannis, "Text Mining Issues and Noise Handling in Health Care Systems", *9th International Conference on System Science in Health Care*, Lyon, France, September 2008.
- 60. ACTI** - C. Julien, L. Saitta, "Image Database Browsing by Unsupervised Learning", *17th International Symposium on Methodologies for Intelligent Systems (ISMIS 08)*, Toronto, Canada, May 2008; *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, Germany.
- 61. ACTI** - J. Darmont, "Entreposage de données complexes pour la médecine d'anticipation personnalisée", *9th International Conference on System Science in Health Care (ICSSHC 08)*, Lyon, France, September 2008.
- 62. ACTI** - H. Mahboubi, K. Aouiche, J. Darmont, "A Join Index for XML Data Warehouses", *2008 International Conference on Information Resources Management (Conf-IRM 08)*, Niagara Falls, Canada, May 2008.
- 63. ACTI** - Z. Wei, D. Miao, J. Chauchat, C. Zhong, "Feature Selection on Chinese Text Classification Using Character N-Grams", *3rd International Conference on Rough Sets and Knowledge Technology (RSKT 08)*, Chengdu, China, May 2008; *Lecture Notes in Computer Science*, Springer, Heidelberg, Germany.
- 64. ACTI** - Q. Gicquel, M. Metzger, A. Lepape, E. Bahri, S. Lallich, "Risk factors analysis for nosocomial infections in intensive care unit: a data mining approach", *IXth International Conference on System Science in Health Care*, Lyon, 2008.
- 65. ACTI** - C. Julien, L. Saitta, "Automatic Handling of Digital Image Repositories: A Brief Survey", *17th International Symposium on Methodologies for Intelligent Systems (ISMIS 08)*, Toronto, Canada, May 2008; *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, Germany.
- 66. ACTI** - H. Mahboubi, J. Darmont, "Data Mining-based Fragmentation of XML Data Warehouses", *ACM 11th International Workshop on Data Warehousing and OLAP (CIKM/DOLAP 08)*, Napa Valley, USA, October 2008, 9-16; ACM.
- 67. ACTI** - P. Lenca, S. Lallich, T. Do, N. Pham, "A comparison of different off-centered entropies to deal with class imbalance for decision trees", *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 08)*, Osaka, Japan, May 2008.
- 68. ACTI** - A. ElSayed, J. Velcin, D. Zighed, "Word Clustering with Validity Indices", *21st Conference of the Canadian Society for Computational Studies of Intelligence (Canadian AI 08)*, Windsor, Canada, May 2008; *Lecture Notes in Computer Science*, Vol. 5032, Springer, Heidelberg, Germany, 259-270.
- 69. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "Mining semantic distance between corpus terms", *1st Ph.D. Workshop, 16th ACM Conference on Information and Knowledge Management (PIKM-CIKM 07)*, Lisbon, Portugal, A. Varde, J. Pei, Eds., November 2007, 49-54; ACM.

- 70. ACTI** - C. Favre, F. Bentayeb, O. Boussaïd, "Evolution of data warehouses' optimization: a workload perspective", *9th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2007)*, 2007; *Lecture Notes in Computer Science*, Vol. 4654, Springer, Heidelberg, Germany, 13 - 22.
- 71. ACTI** - J. Azé, P. Lenca, S. Lallich, B. Vaillant, "A Study of the Robustness of Association Rules", *2007 International Conference on Data Mining (DMIN 07)*, Las Vegas, USA, 2007, 163–169; CSREA Press.
- 72. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "A Multisource Context-Dependent Semantic Distance Between Concepts", *18th International Conference on Database and Expert Systems Applications (DEXA 07)*, Regensburg, Germany, R. Wagner, N. Revell, G. Pernul, Eds., September 2007; *Lecture Notes in Computer Science*, Vol. 4653, Springer, Heidelberg, Germany, 54-63.
- 73. ACTI** - J. Thomas, P. Jouve, N. Nicoloyannis, "Asymmetric measure for supervised learning models assessment, application to breast cancer detection", *International Conference on Industrial Engineering and Systems Management (IESM 07)*, Beijing, China, May 2007.
- 74. ACTI** - S. Bouattour, R. BenMessaoud, O. Boussaïd, "Modélisation de Règles d'Analyse Dédiées aux Entrepôts de Données Actifs", *2ème Atelier Systèmes Décisionnels (ASD 07)*, Sousse, Tunisie, Octobre 2007.
- 75. ACTI** - J. Chauchat, A. Morin, R. Rakotomalala, "Correcting the error rate estimation bias in Data Mining when the dataset comes from a two-stage sampling", *Statistics for Data Mining, Learning and Knowledge Extraction (IAST 07)*, Aveiro, Portugal, August 2007.
- 76. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "Using Semantic Distance in a Content-based Heterogeneous Information Retrieval System", *3rd International Workshop on mining complex data (MCD 07)*, Warsaw, Poland, 2007; *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, Germany, 185-197.
- 77. ACTI** - C. Favre, F. Bentayeb, O. Boussaïd, "Dimension Hierarchy Updates in Data Warehouses: a User-driven Approach", *9th International Conference on Enterprise Information systems (ICEIS 07)*, Funchal, Madeira, Portugal, June 2007, 206 - 211.
- 75. ACTI** - N. Maiz, F. Bentayeb, O. Boussaïd, "Ontology based mediation system", *18th Information Ressource Management Association International Conference (IRMA 07)*, Vancouver, Canada, May 2007; IRMA, Hershey, USA.
- 76. ACTI** - J. Ralaivao, J. Darmont, "Knowledge and Metadata Integration for Warehousing Complex Data", *6th International Conference on Information Systems Technology and its Applications (ISTA 07)*, Kharkiv, Ukraine, May 2007; *Lecture Notes in Informatics*, Vol. P-107, GI-Edition, Bonn, Germany, 164-175.
- 77. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "A New Context-Aware Measure for Semantic Distance Using a Taxonomy and a Text Corpus", *IEEE International Conference on Information Reuse and Integration (IRI 07)*, Las Vegas, USA, August 2007, 279-284; IEEE Systems, Man, and Cybernetics Society.
- 78. ACTI** - M. Hachicha, H. Mahboubi, J. Darmont, "Vers une algèbre XML-OLAP : État de l'art", *2ème Atelier Systèmes Décisionnels (ASD 07)*, Sousse, Tunisie, Octobre 2007.
- 79. ACTI** - S. Lallich, P. Lenca, B. Vaillant, "Construction of an off-centered entropy for supervised learning", *XIIth International Symposium on Applied Stochastic Models and Data Analysis (AMSDA 07)*, Chania, Crete, Greece, 2007.

- 80. ACTI** - L. Baumes, M. Moliner, R. Gaudin, N. Nicoloyannis, A. Corma, "Genetic Algorithms in Materials Science and Engineering", *2007 E-MRS Fall Meeting, Warsaw, Poland*, September 2007.
- 81. ACTI** - A. Silic, J. Chauchat, B. Basic, A. Morin, "N-Grams and Morphological Normalization in Text Classification: A Comparison on a Croatian-English Parallel Corpus", *13th Portuguese Conference on Artificial Intelligence (EPIA 2007), Guimaraes, Portugal*, J. Neves, M. Santos, J. Machado, Eds., December 2007; *Lecture Notes in Computer Science*, Vol. 4874, Springer, Heidelberg, Germany, 671-682.
- 82. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "A Context-Dependent Semantic Distance Measure", *19th International Conference on Software Engineering and Knowledge Engineering (SEKE 07), Boston, USA*, July 2007, 432-437; Knowledge Systems Institute Graduate School.
- 83. ACTI** - E. Antajan, R. Rakotomalala, S. Gasparini, M. Picheral, L. Stemmann, G. Gorsky, "Automatic quantification and recognition of major zooplankton groups in a North Sea time series using the Zooscan imaging system", *4th International Zooplankton Production Symposium*, 2007, 189 - 190 (Hiroshima, Japan).
- 84. ACTI** - E. Bahri, N. Nicoloyannis, M. Maddouri, "Improving boosting by exploiting former assumptions", *3rd International Workshop on mining complex data (MCD 07), Warsaw, Poland*, 2007.
- 85. ACTI** - S. Azefack, K. Aouiche, J. Darmont, "Dynamic index selection in data warehouses", *4th International Conference on Innovations in Information Technology (Innovations 07), Dubai, United Arab Emirates*, November 2007; IEEE.
- 86. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "A New Approach Towards Content-based Heterogeneous Information Retrieval", *ECML/PKDD Workshop on Mining Complex Data*, 2007.
- 87. ACTI** - A. ElSayed, H. Hacid, D. Zighed, "Combining Text and Image for Content-Based Information Retrieval", *2007 International Conference on Information and Knowledge Engineering (IKE 07)*, 2007; CSREA Press.
- 89. ACTI** - N. Durand, S. Derivaux, G. Forestier, C. Wemmert, P. Gañçarski, O. Boussaïd, A. Puissant, "Ontology-based Object Recognition for Remote Sensing Image Interpretation", *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 07), Patras, Greece*, October 2007.
- 90. ACTI** - R. Gaudin, N. Nicoloyannis, "An Adaptable Time Warping Distance for Time Series Learning", *5th International Conference on Machine Learning and Applications (ICMLA 06), Orlando, USA*, December 2006.
- 91. ACTI** - H. Hacid, D. Zighed, "Content-Based Image Retrieval Using Topological Models", *12th International MultiMedia Modelling Conference (MMM 06), Beijing, China*, 2006.
- 92. ACTI** - D. Zighed, H. Hacid, "Proximity graphs and separability of classes", *11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 06), Paris, France*, July 2006, 1488-1495; IPMU.
- 93. ACTI** - R. Rakotomalala, F. Mhamdi, "Combining feature selection and feature reduction for protein classification", *2nd WSEAS International Symposium on Data Mining, Lisbon, Portugal*, 2006.
- 94. ACTI** - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "Mining Association Rules in OLAP Cubes", *International Conference on Innovations in Information Technology (ITT 06), Dubai*, November 2006.

- 95. ACTI** - B. Vaillant, S. Lallich, P. Lenca, "Modelling of the counter-examples and association rules interestingness measures behavior", *2nd International Conference on Data Mining (DMIN 06)*, Las Vegas, USA, June 2006, 132-137.
- 96. ACTI** - N. Maiz, F. Bentayeb, O. Boussaïd, "Un système de médiation basé sur les ontologies pour l'entreposage des données", *Atelier Systèmes Décisionnels (ASD 06)*, *9th Maghrebian Conference on Information Technologies (MCSEAI 06)*, Agadir, Maroc, Décembre 2006.
- 97. ACTI** - C. Favre, F. Bentayeb, O. Boussaïd, "A Knowledge-driven Data Warehouse Model for Analysis Evolution", *13th ISPE International Conference on Concurrent Engineering: Research and Applications (CE 06)*, Antibes, France, September 2006; *Frontiers in Artificial Intelligence and Applications*, Vol. 143, IOS Press, 271-278.
- 98. ACTI** - R. Gaudin, S. Barbier, N. Nicoloyannis, M. Banens, "Clustering of Bi-Dimensional and Heterogeneous Time Series: Application to Social Sciences Data", *2006 International Conference on Data Mining (DMIN 06)*, Las Vegas, USA, June 2006, 10-16.
- 99. ACTI** - J. Darmont, E. Olivier, "A complex data warehouse for personalized, anticipative medicine", *17th Information Resources Management Association International Conference (IRMA 06)*, Washington, USA, May 2006, 685-687; Idea Group Publishing.
- 100. ACTI** - R. BenMessaoud, S. Loudcher-Rabaseda, O. Boussaïd, R. Missaoui, "Enhanced Mining of Association Rules from Data Cubes", *9th ACM International Workshop on Data Warehousing and OLAP (DOLAP 06)*, Arlington, USA, November 2006.
- 101. ACTI** - R. Rakotomalala, F. Mhamdi, "Improved Singular Value Decomposition for Supervised Learning in a High Dimensional Dataset", *6th International Workshop on Pattern Recognition in Information Systems (PRIS 06)*, Paphos, Cyprus, May 2006, 38-47.
- 102. ACTI** - C. Favre, F. Bentayeb, O. Boussaïd, "WEDriK : une plateforme pour des analyses personnalisées dans les entrepôts de données évolutifs", *Atelier Systèmes Décisionnels (ASD 06)*, *9th Maghrebian Conference on Information Technologies (MCSEAI 06)*, Agadir, Maroc, Décembre 2006.
- 103. ACTI** - D. Zighed, S. Marcellin, G. Ritschard, "An asymmetric entropy measure for decision trees", *Knowledge Extraction and Modeling Workshop (KNEMO 06)*, Capri, Italy, September 2006.
- 104. ACTI** - P. Lenca, B. Vaillant, S. Lallich, "On the Robustness of Association Rules", *IEEE International Conferences on Cybernetics and Intelligent Systems and Robotics, Automation and Mechatronics (CIS-RAM 06)*, Bangkok, Thailand, June 2006, 596-601.
- 105. ACTI** - N. Maiz, O. Boussaïd, F. Bentayeb, "Ontology-Based Mediation System", *13th ISPE International Conference on Concurrent Engineering: Research and Applications (CE 06)*, Antibes, France, September 2006; *Frontiers in Artificial Intelligence and Applications*, Vol. 143, IOS Press, 181-189.
- 106. ACTI** - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "Using a Factorial Approach for Efficient Representation of Relevant OLAP Facts", *Seventh International Baltic Conference on Databases and Information Systems (DB&IS 06)*, Vilnius, Lithuania, July 2006.
- 107. ACTI** - J. Chauchat, "Microeconomics Forecast: Learning by Doing, A Ten Years Graduate Level Experience", *7th International Conference On Teaching Statistics (ICOTS7)*, Salvador, Bahia, Brazil, July 2006.
- 108. ACTI** - K. Aouiche, P. Jouve, J. Darmont, "Clustering-Based Materialized View Selection in Data Warehouses", *10th East-European Conference on Advances in Databases and Information Systems (ADBIS 06)*, Thessaloniki, Greece, September 2006; *Lecture Notes in Computer Science*, Vol. 4152, Springer, Heidelberg, Germany, 81-95.

- 109. ACTI** - S. Marcellin, D. Zighed, G. Ritschard, "Detection of breast cancer using an asymmetric entropy measure", *Computational Statistics (COMPSTAT 06)*, A. Rizzi, M. Vichi, Eds., 2006; *Computational Statistics*, Vol. XXV, Springer, Heidelberg, Germany, 975-982 (On CD).
- 110. ACTI** - H. Mahboubi, K. Aouiche, J. Darmont, "Materialized View Selection by Query Clustering in XML Data Warehouses", *4th International Multiconference on Computer Science and Information Technology (CSIT 06)*, Amman, Jordan, April 2006, 68-77.
- 111. ACTI** - R. Rakotomalala, J. Chauchat, F. Pellegrino, "Accuracy Estimation with Clustered Dataset", *The Australasian Data Mining Conference (AusDM 06)*, Sidney, Australia, November 2006; *Conferences in Research and Practice in Information Technology*, Vol. 61.
- 112. ACTI** - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "Efficient Multidimensional Data Representation Based on Multiple Correspondence Analysis", *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 06)*, Philadelphia, USA, August 2006.
- 113. ACTI** - O. Teytaud, S. Gelly, S. Lallich, E. Prudhomme, "Quasi-random bootstrap, with applications to rule extraction and (sub)bagging", *International Workshop on Intelligent Information Access (IIA 06)*, Helsinki, Finland, July 2006.
- 114. ACTI** - F. Mhamdi, R. Rakotomalala, M. Elloumi, "A Hierarchical N-Grams Extraction Approach for Classification Problem", *IEEE International Conference on Signal-Image Technology and Internet-Based Systems (SITIS 06)*, Tunisia, 2006, 310-321.
- 115. ACTI** - J. Thomas, P. Jouve, N. Nicoloyannis, "Optimisation and evaluation of random forests for imbalanced datasets", *16th International Symposium on Methodologies for Intelligent Systems (ISMIS 06)*, Bari, Italy, September 2006; *Lecture Notes in Artificial Intelligence*, Vol. 4203, Springer, Heidelberg, Germany, 642-651.
- 116. ACTI** - O. Boussaïd, R. BenMessaoud, R. Choquet, S. Anthoard, "X-Warehousing: an XML-Based Approach for Warehousing Complex Data", *10th East-European Conference on Advances in Databases and Information Systems (ADBIS 06)*, Thessaloniki, Greece, September 2006; *Lecture Notes in Computer Science*, Vol. 4152, Springer, Heidelberg, Germany, 39-54.
- 117. ACTI** - S. Marcellin, D. Zighed, G. Ritschard, "An asymmetric entropy measure for decision trees", *11th Information Processing and Management of Uncertainty in knowledge-based systems (IPMU 06)*, Paris, France, July 2006, 1292-1299.
- 118. ACTI** - A. Morineau, R. Rakotomalala, "The TVpercent Criteria to Eliminate Uninformative Models among Association Rules", *Knowledge Extraction and Modeling IASC-INTERFACE-IFCS Workshop (KNEMO 06)*, Anacapri, Italy, 2006.
- 119. ACTI** - G. Ritschard, D. Zighed, "Implication Strength of Classification Rules", *Foundations of Intelligent Systems (ISMIS 06)*, Bari, Italy, September 2006; *Lecture Notes in Artificial Intelligence*, Vol. 4203, Springer, Heidelberg, Germany, 463-472.
- 120. ACTI** - H. Mahboubi, J. Darmont, "Benchmarking XML data warehouses", *Atelier Systèmes Décisionnels (ASD 06)*, 9th Maghrebien Conference on Information Technologies (MCSEAI 06), Agadir, Maroc, December 2006.
- 121. ACTI** - H. Hacid, D. Zighed, "Content-Based Image Retrieval in Large Image Databases", *IEEE International Conference on Granular Computing (GrC 2006)*, Atlanta, USA, May 2006.
- 122. ACTI** - S. Lallich, O. Teytaud, E. Prudhomme, "Statistical inference and data mining: false discoveries control", *17th COMPSTAT Symposium of the IASC*, Rome, Italy, August 2006, 325-336.

- 123. ACTI** - S. Lallich, B. Vaillant, P. Lenca, "Parametrised measures for the evaluation of association rule interestingness", *International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Brest, France, 2005, 220-229.
- 124. ACTI** - V. Pisetta, H. Hacid, D. Zighed, "Automatic Juridical Texts Classification and Relevance Feedback", *First IEEE International Workshop on Mining Complex Data (IEE MCD05)*, Texas, USA, 2005.
- 125. ACTI** - J. Darmont, O. Boussaïd, J. Ralaivao, K. Aouiche, "An Architecture Framework for Complex Data Warehouses", *7th International Conference on Enterprise Information Systems (ICEIS 05)*, Miami, USA, May 2005, 370-373.
- 126. ACTI** - H. Hacid, D. Zighed, "An Effective Method for Locally Neighborhood Graphs Updating", *16th International Conference on Database and expert Systems Applications (DEXA 05)*, 2005; *Lecture Notes in Computer Science*, Vol. 3588, Springer, Heidelberg, Germany, 930-939.
- 127. ACTI** - E. Prudhomme, S. Lallich, "Quality measure based on Kohonen maps for supervised learning of large high dimensional data", *International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Brest, France, 2005, 246-255.
- 128. ACTI** - R. BenMessaoud, O. Boussaïd, S. Loudcher-Rabaseda, "Evaluation of a MCA-Based Approach to Organize Data Cubes", *ACM Fourteenth Conference on Information and Knowledge Management (CIKM 05)*, Bremen, Germany, 2005.
- 129. ACTI** - A. Morin, A. Kouomou-Choupo, J. Chauchat, "Dimension reduction and clustering for query-by-example in huge image databases", *3rd World Conference on Computational Statistics and Data Analysis (CSDA 05)*, Cyprus, November 2005.
- 130. ACTI** - A. Morin, A. Kouomou-Choupo, J. Chauchat, "Dimension reduction and clustering for query-by-example in huge image databases", *3rd IASC World Conference on Computational Statistics and Data Analysis*, Limassol, Cyprus, October 2005.
- 131. ACTI** - G. Legrand, N. Nicoloyannis, "A new feature selection method", *8th International Conference on Pattern Recognition and Information Processing (PRIP05)*, Minsk Belarus, 2005.
- 132. ACTI** - R. Rakotomalala, F. Mhamdi, M. Elloumi, "Hybrid Feature Ranking for Protein Classification", *1st International Conference on Advanced Data Mining and Applications (ADMA'05)*, 2005; *Lecture Notes in Artificial Intelligence*, Vol. 3584, Springer, Heidelberg, Germany, 610-617.
- 133. ACTI** - K. Aouiche, J. Darmont, O. Boussaïd, F. Bentayeb, "Automatic Selection of Bitmap Join Indexes in Data Warehouses", *7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 05)*, Copenhagen, Denmark, August 2005; *Lecture Notes in Computer Science*, Vol. 3589, Springer, Heidelberg, Germany, 64-73.
- 134. ACTI** - G. Legrand, N. Nicoloyannis, "Feature selection and preferences aggregation", *International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Brest, France, 2005, 305-312.
- 135. ACTI** - F. Mhamdi, R. Rakotomalala, M. Elloumi, "Feature Ranking for Protein Classification", *4th International Conference on Computer Recognition Systems (CORES'05)*, 2005; *Advances in Soft Computing*, Springer, Heidelberg, Germany, 611-617.
- 136. ACTI** - H. Hacid, D. Zighed, "An Incremental Algorithm for Neighborhood Graphs Construction", *3rd World Conference on Computational Statistics & Data Analysis*, Cyprus, October 2005.

- 137. ACTI** - A. Tanasescu, O. Boussaïd, F. Bentayeb, "Preparing Complex Data for Warehousing", *3rd ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 05)*, Cairo, Egypt, January 2005.
- 138. ACTI** - J. Darmont, F. Bentayeb, O. Boussaïd, "DWEB: A Data Warehouse Engineering Benchmark", *7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 05)*, Copenhagen, Denmark, August 2005; *Lecture Notes in Computer Science*, Vol. 3589, Springer, Heidelberg, Germany, 85-94.
- 139. ACTI** - G. Legrand, N. Nicoloyannis, "Feature selection method using preferences aggregation", *International Conference on Machine Learning and Data Mining (MLDM 05)*, Leipzig Germany, 2005; *Lecture Notes in Computer Science*, Vol. 3587, Springer, Heidelberg, Germany, 9-11.
- 140. ACTI** - F. Clerc, R. Rakotomalala, D. Farrusseng, "Learning Fitness Function in a Combinatorial Optimization Process", *International Symposium on Applied Stochastic Models and Data Analysis*, 2005, 535-543.
- 141. ACTI** - H. Hacid, D. Zighed, "Neighborhood Graphs for Image Databases Indexing and Content-Based Retrieval", *First IEEE International Workshop on Mining Complex Data (IEE MCD05)*, Texas, USA, 2005.
- 142. ACTI** - C. Favre, F. Bentayeb, "Bitmap index-based decision trees", *15th International Symposium on Methodologies for Intelligent Systems (ISMIS 05)*, New York, USA, May 2005; *Lecture Notes in Artificial Intelligence*, Vol. 3488, Springer, Heidelberg, Germany, 65-73.
- 143. ACTI** - J. Chauchat, M. Pacaut-Troncin, A. Cuerq, "Model Assessment and Selection : a Case Study on Risk Factors for Acute Suicidality in Psychiatric Patients", *Applied Statistics, Ribno (Bled)*, Slovenia, 2005.

## **g. Conférences nationales**

- 144. ACTN** - E. Bahri, M. Maddouri, "Nouvelle approche du Boosting pour les données bruitées", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, Janvier 2008; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse.
- 145. ACTN** - H. Derrar, O. Boussaïd, M. Ahmed-Nacer, "Une approche de répartition des données d'un entrepôt basée sur l'optimisation par essaim particulière", *4èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2008)*, Toulouse, Juin 2008; *Revue des Nouvelles Technologies de l'Information*, Vol. B-4, Cépaduès, Toulouse, 141-149.
- 146. ACTN** - J. Darmont, "Expérimentations avec le banc d'essais pour entrepôts de données DWEB", *24èmes Journées Bases de Données Avancées (BDA 08)*, Guilhaing-Granges, Octobre 2008 (Démo).
- 147. ACTN** - T. Qureshi, D. Zighed, "Discretization of Continuous Features by Resampling", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, F. Guillet, B. Trousse, Eds., 2008; *Revue des Nouvelles Technologies de l'Information*, Vol. E-8, Cépaduès, Toulouse.
- 148. ACTN** - C. Favre, M. Rougié, F. Bentayeb, O. Boussaïd, "Quels rôles donner aux utilisateurs dans les systèmes d'information ? Retour d'expérience chez LCL-Le Crédit Lyonnais", *Atelier Prise en Compte de l'Utilisateur dans les Systèmes d'Information (PeCUSI 08)*, en conjonction avec le 26ème Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID 08), Mai 2008, 29-42.

- 149. ACTN** - R. Billot, H. Suchier, S. Lallich, "Une approche ensembliste inspirée du boosting en classification non supervisée", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Nice, Janvier 2008; *Revue des Nouvelles Technologies de l'Information*, Vol. E-8, Cépaduès, Toulouse, 361-372.
- 150. ACTN** - A. ElSayed, H. Hacid, D. Zighed, "On Involving Validity Indices in Document Clustering", *5ème Conférence en Recherche d'Information et Applications (CORIA 08)*, Trégastel, 2008.
- 151. ACTN** - V. Lefort, G. Beslon, "Optimisation de réseaux de neurones RBF pour la régression via un algorithme évolutionnaire : RBF-Gene", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, Janvier 2008; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse.
- 152. ACTN** - J. Thomas, P. Jouve, E. Prudhomme, "Echantillonnage adaptatif de jeux de données déséquilibrés pour les forêts aléatoires", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, Janvier 2008; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse.
- 153. ACTN** - S. Bouattour, J. Feki, H. Abdallah, R. BenMessaoud, O. Boussaïd, "Proposition d'une taxonomie d'événement dans les entrepôts de données actifs", *8èmes Journées scientifiques des jeunes chercheurs en génie électrique et informatique (GEI 08)*, Sousse, Tunisie, Mars 2008.
- 154. ACTN** - T. Do, N. Pham, P. Lenca, S. Lallich, "Expérimentation de l'entropie décentrée pour le traitement des données déséquilibrées en induction par arbres", *4ème Atelier Qualité des Données et des Connaissances (QDC-EGC 08)*, Nice, Janvier 2008.
- 155. AFF** - A. Bodin-Niemczuk, R. BenMessaoud, S. Loudcher-Rabaseda, O. Boussaïd, "Vers l'intégration de la prédiction dans l'OLAP", *8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, Janvier 2008; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 203-204 (Poster).
- 156. ACTN** - N. Harbi, O. Boussaïd, F. Bentayeb, "Propriétés d'un modèle conceptuel multidimensionnel pour les données complexes", *5ème atelier Fouille de données complexes dans un processus d'extraction des connaissances (FDC-EGC 08)*, Sophia-Antipolis, Janvier 2008.
- 157. ACTN** - A. ElSayed, H. Hacid, D. Zighed, "Enhancing Semantic Distance with Context Awareness", *Atelier SemSim, 8ème Conférence Extraction et Gestion des Connaissances (EGC 08)*, Sofia Antipolis, 2008; INRIA Sophia antipolis.
- 158. ACTN** - A. Stavrianou, J. Chauchat, "Opinion Mining Issues and Agreement Identification in Forum Texts", *Atelier FOuille des Données d'OPinions (FODOP 08)*, en conjonction avec le *26ème Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID 08)*, Fontainebleau, Mai 2008, 51-58.
- 159. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Maintenance de charges pour l'optimisation des entrepôts de données évolutifs : aide à l'administrateur", *4èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2008)*, Toulouse, Juin 2008; *Revue des Nouvelles Technologies de l'Information*, Vol. B-4, Cépaduès, Toulouse, 115-122.
- 160. ACTN** - N. Maiz, O. Boussaïd, F. Bentayeb, "Fusion des ontologies par classification hiérarchique", *1er Atelier sur les mesures de similarité sémantiques (SimSem-EGC 08)*, Sophia Antipolis, Janvier 2008.
- 161. ACTN** - O. Rakotoarivelo, F. Bentayeb, "Evolution de schéma par classification automatique pour les entrepôts de données", *3èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 07)*, Poitiers, Juin 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. B-3, Cépaduès, Toulouse, 99-112.

- 162. ACTN** - A. ElSayed, H. Hacid, D. Zighed, "Recherche d'Information par le Contenu des Données Hétérogenes", *RIAS*, 2007; IRIT, Université de Toulouse.
- 163. ACTN** - R. Brisson, O. Boussaïd, P. Gançarski, A. Puissant, N. Durand, "Navigation et appariement d'objets géographiques dans une ontologie", *7ème Conférence Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, Janvier 2007; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse.
- 164. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Evolution et personnalisation des analyses dans les entrepôts de données : une approche orientée utilisateur", *XXVème congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 07)*, Perros-Guirec, Mai 2007, 308 - 323.
- 165. ACTN** - E. Prudhomme, S. Lallich, "Ensemble prédicteur fondé sur les cartes auto-organisatrices adapté aux données volumineuses", *7ème Conférence Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, Janvier 2007; *Revue des Nouvelles Technologies de l'Information*, 473-484.
- 166. ACTN** - G. Ritschard, D. Zighed, S. Marcellin, "Données déséquilibrées, entropie décentrée et indice d'implication", *Nouveaux apports théoriques à l'analyse statistique implicative et applications*, 2007, 315-327; Departament de Matemàtiques, Universitat Jaume I; ASI4.
- 167. ACTN** - M. Studer, G. Ritschard, L. Baccaro, I. Georgiou, N. Muller, D.A. Zighed, "Relations entre types de violation des libertés syndicales garanties par les conventions de l'OIT : Une analyse statistique implicative des résultats d'une fouille de texte", *Nouveaux apports théoriques à l'analyse statistique implicative et applications*, 2007, 111-122; Departament de Matemàtiques, Universitat Jaume I.
- 168. ACTN** - S. Lallich, P. Lenca, B. Vaillant, "Construction d'une entropie décentrée pour l'apprentissage supervisé", *3ème Atelier Qualité des Connaissances à partir des Données (QDC-EGC 07)*, Namur, Belgique, Janvier 2007, 45-54.
- 169. ACTN** - D. Zighed, V. Pisetta, D. Ratsimba, "Separability of Classes in a multidimensional Space", *Classification and Data Analysis*, September 2007; *Meeting of the classification and Data Analysis Group of the Italian Statistical Society*, Eum edizioni università di macerata, 147-150.
- 170. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Intégration des connaissances utilisateurs pour des analyses personnalisées dans les entrepôts de données évolutifs", *7èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, Janvier 2007; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 217 - 222.
- 171. ACTN** - J. Thomas, P. Jouve, N. Nicoloyannis, "Mesure non symétrique pour l'évaluation de modèles, utilisation pour les jeux de modèles, utilisation pour les jeux de données déséquilibrés", *7ème Conférence Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, Janvier 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. E-9, Cépaduès, Toulouse, 509-519.
- 172. ACTN** - E. Bahri, N. Nicoloyannis, M. Maddouri, "Amélioration du Boosting par combinaison des hypothèses antérieures", *14èmes Rencontres de la Société Francophone de Classification (SFC 07)*, Paris, Septembre 2007.
- 173. ACTN** - H. Mahboubi, J. Darmont, "Fragmentation des entrepôts de données XML", *3èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 07)*, Poitiers, Juin 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. B-3, Cépaduès, Toulouse, 177-190.
- 174. ACTN** - D. Zighed, S. Marcellin, G. Ritschard, "Mesure d'entropie asymétrique et consistante", *7ème Conférence Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, G. Venturini, M. Noirhomme, Eds., 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. E-9, Cépaduès, Toulouse, 81-86.

- 175. ACTN** - B. Vaillant, S. Menou, S. Moga, P. Lenca, S. Lallich, "Qualité des règles d'association : étude de données d'entreprise", *3ème Atelier Qualité des Connaissances à partir des Données (QDC-EGC 07)*, Namur, Belgique, Janvier 2007, 55-64.
- 176. ACTN** - O. Rakotoarivelo, F. Bentayeb, "Evolution de schéma par classification automatique pour les entrepôts de données", *4ème atelier Fouille de Données Complexes dans un Processus d'Extraction des Connaissances (FDC-EGC 07)*, Namur, Belgique, Janvier 2007.
- 177. ACTN** - N. Maiz, O. Boussaïd, F. Bentayeb, "Clustering method for semi-automatically ontologies fusion", *4ème atelier Fouille de Données Complexes dans un Processus d'Extraction des Connaissances (FDC-EGC 07)*, Namur, Belgique, Janvier 2007.
- 178. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Evolution de modèle dans les entrepôts de données : existant et perspectives", *3èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 07)*, Poitiers, Juin 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. B-3, Cépaduès, Toulouse, 21-36.
- 179. ACTN** - V. Pisetta, G. Ritschard, D. Zighed, "Choix des conclusions et validation des règles issues d'arbres de classification", *7ème Conférence Extraction et Gestion des Connaissances (EGC 07)*, Namur, Belgique, M. Noirhomme, G. Venturini, Eds., 2007; *Revue des Nouvelles Technologies de l'Information*, Vol. E-9, Cépaduès, Toulouse, 485-496.
- 180. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Evolution de schémas dans les entrepôts de données : modèle à base de règles", *2ème journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 06)*, Versailles, Juin 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. B-2, Cépaduès, Toulouse, 175-176.
- 181. ACTN** - A. Brémond, P. Jouve, J. Thomas, J. Clech, "Résultats Préliminaires d'une étude comparative de deux CAD", *Innovations Technologiques et Bonnes Pratiques en Sénologie : Dépistage - Diagnostic - Traitement*, A. Isnard, A. Travade, Eds., Juin 2006, 92-94; Fusium; Sofmis.
- 182. ACTN** - H. Hacid, "Annotation semi-automatique de grandes BD images : Approche par graphes de voisinage", *Conférence en Recherche d'Informations et Applications (CORIA 06)*, Lyon, Mars 2006.
- 183. ACTN** - N. Maiz, O. Boussaïd, F. Bentayeb, "Un système de médiation basé sur les ontologies", *3ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 06*, Lille, Janvier 2006, 27-38.
- 184. ACTN** - D. Zighed, A. Isnard, "Projet NORDOM (Numérisation, Optimisation, Rationalisation du Dépistage Organisé en Mammographie)", *Innovations Technologiques et Bonnes Pratiques en Sénologie : Dépistage - Diagnostic - Traitement*, 2006, 29-34; Fusium; Sofmis.
- 185. ACTN** - O. Boussaïd, R. BenMessaoud, R. Choquet, S. Anthoard, "Conception et construction d'entrepôts XML", *2ème journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 06)*, Versailles, Juin 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. B-2, Cépaduès, Toulouse, 3-22.
- 186. ACTN** - H. Mahboubi, K. Aouiche, J. Darmont, "Un index de jointure pour les entrepôts de données XML", *6èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 06)*, Lille, Janvier 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. E-6, Cépaduès, Toulouse, 89-94.
- 187. ACTN** - R. Gaudin, N. Nicoloyannis, "Séries temporelles : Vers une mesure de distance optimale", *Fouille de données temporelles, 6èmes Journées d'Extraction et de Gestion des Connaissances (EGC 06)*, Lille, Janvier 2006, 67-75.

- 188. ACTN** - R. Rakotomalala, F. Mhamdi, "Evaluation des Méthodes Supervisées pour le Classement de Protéines", *13èmes Rencontres de la Société Française de Classification (SFC 06)*, Metz, Septembre 2006, 181-184.
- 189. ACTN** - V. Pisetta, H. Hacid, D. Zighed, "Multi-catégorisation de textes juridiques et retour de pertinence", *6èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 06)*, Lille, Janvier 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. E-6, Cépaduès, Toulouse, 235-246.
- 190. ACTN** - O. Boussaïd, S. Loudcher-Rabaseda, "Intégration des méta-données dans la fouille de données", *XXIVème Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 06)*, Hammamet, Tunisie, 2006.
- 191. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "Modèle d'entrepôt de données à base de règles", *3ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 06*, Lille, 2006, 39-50.
- 192. ACTN** - F. Mhamdi, M. Elloumi, R. Rakotomalala, "Extraction et Sélection des n-grammes pour le Classement de Protéines", *Atelier Extraction et gestion de connaissances appliquées aux données biologiques (Bio-EGC 06)*, Lille, Janvier 2006, 25-37.
- 193. ACTN** - V. Pisetta, H. Hacid, F. Bellal, G. Ritschard, "Traitement automatique de textes juridiques", *Semaine de la Connaissance (SdC 06)*, Nantes, R. Lehn, M. Harzallah, N. Aussenac-Gilles, J. Charlet, Eds., Juin 2006 (CDrom).
- 194. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, "A Rule-based Data Warehouse Model", *23rd British National Conference on Databases (BNCOD 2006)*, Belfast, Northern Ireland, July 2006; *Lecture Notes in Computer Science*, Vol. 4042, Springer, Heidelberg, Germany, 274-277.
- 195. ACTN** - H. Hacid, D. Zighed, "Graphes de Proximité pour l'Indexation et l'Interrogation d'Images par le Contenu", *6èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 06)*, Lille, Janvier 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. E-6, Cépaduès, Toulouse, 11-22.
- 196. ACTN** - D. Zighed, "Aspects conceptuels : Différentes Méthodologies des Systèmes Experts pour la Détection ou la Caractérisation", *Innovations Technologiques et Bonnes pratiques en Sénologie : Dépistage - Diagnostic - Traitement*, 2006; Sofmis, Fusium, 76-81.
- 197. ACTN** - N. Maiz, K. Aouiche, J. Darmont, "Sélection automatique d'index et de vues matérialisées dans les entrepôts de données", *2ème journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 06)*, Versailles, Juin 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. B-2, Cépaduès, Toulouse, 89-104.
- 198. ACTN** - A. Brémond, A. Isnard, N. Nicoloyannis, D. Zighed, "Numérisation secondaire et Lecture sur Ecran : Evaluation des Performances", *Innovations Technologiques et Bonnes Pratiques en Sénologie : Dépistage - Diagnostic - Traitement*, A. Isnard, A. Travade, Eds., 2006, 22-28; Fusium.
- 199. ACTN** - A. Morineau, R. Rakotomalala, "Critère VT-100 de sélection des règles d'association", *6èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 06)*, Lille, Janvier 2006; *Revue des Nouvelles Technologies de l'Information*, Vol. E-6, Cépaduès, Toulouse, 581-592.
- 200. ACTN** - G. Legrand, N. Nicoloyannis, "Etat de l'art des méthodes de construction de variables", *12èmes Rencontres de la Société Francophone de Classification (SFC 05)*, Montréal, 2005, 182-185.

- 201. ACTN** - R. Gaudin, N. Nicoloyannis, "Apprentissage non supervisé de séries temporelles à l'aide des k-Means et d'une nouvelle méthode d'agrégation de séries", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 201-212.
- 202. ACTN** - C. Favre, F. Bentayeb, "Intégration efficace des arbres de décision dans les SGBD : utilisation des index bitmap", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 319-330.
- 203. ACTN** - G. Legrand, N. Nicoloyannis, "Gestion de la phase de prétraitement des données et coefficient Kappa", *XXXVIIèmes Journées de Statistique*, Pau, 2005, 6-10; SFdS.
- 204. ACTN** - R. BenMessaoud, K. Aouiche, C. Favre, "Une approche de construction d'espaces de représentation multidimensionnels dédiés à la visualisation", *1ère journée sur les Entrepôts de Données et l'Analyse en ligne (EDA 05)*, Lyon, Juin 2005; *Revue des Nouvelles Technologies de l'Information*, Vol. B-1, Cépaduès, Toulouse, 34-50.
- 205. ACTN** - R. Rakotomalala, "TANAGRA : un logiciel gratuit pour l'enseignement et la recherche", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 697-702.
- 206. ACTN** - C. Udréa, F. Bentayeb, "Fouille de données relationnelles dans les SGBD", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 356.
- 207. ACTN** - G. Brunet, S. Lallich, A. Bideau, "Analyse quantitative des réseaux généalogiques ascendants, l'exemple des lignées familiales de la vallée de la Valserine (Jura français)", *XXVe Congrès international de la Population (UIESP)*, Tours, Juillet 2005.
- 208. ACTN** - R. BenMessaoud, S. Rabaseda, O. Boussaïd, "L'analyse factorielle pour la construction de cubes de données complexes", *2ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 05*, Paris, Janvier 2005, 53-56.
- 209. ACTN** - S. Lallich, P. Lenca, B. Vaillant, "Variations autour de l'intensité d'implication", *Colloque Analyse Statistique Implicative (ASI 2005)*, Palerme, Sicile, Octobre 2005, 237-246.
- 210. ACTN** - M. Raimbault, R. Rakotomalala, X. Morandi, P. Jannin, "Mise en évidence d'invariants dans une population de cas chirurgicaux", *2ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 05*, Paris, Janvier 2005, 149-158.
- 211. ACTN** - J. Ralaivao, "Améliorer la performance d'un entrepôt de données complexes par l'utilisation de métadonnées et de connaissances du domaine", *2ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 05*, Paris, Janvier 2005, 81-84.
- 212. ACTN** - P. Jouve, N. Nicoloyannis, "Forage distribué des données : une comparaison entre l'agrégation d'échantillons et l'agrégation de règles", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 31-42.
- 213. ACTN** - B. Vaillant, P. Meyer, E. Prudhomme, S. Lallich, P. Lenca, S. Bigaret, "Mesurer l'intérêt des règles d'association", *Atelier Qualité des Données et des Connaissances (DQK 05)*, *EGC 05*, Paris, Janvier 2005, 69-78.
- 214. ACTN** - C. Favre, F. Bentayeb, O. Boussaïd, N. Nicoloyannis, "Entreposage Virtuel de demandes marketing : de l'acquisition des objets complexes à la capitalisation des connaissances", *2ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 05*, Paris, Janvier 2005, 65-68.

**215. ACTN** - E. Prudhomme, S. Lallich, "Validation statistique des cartes de Kohonen en apprentissage supervisé", *5èmes Journées d'Extraction et de Gestion des Connaissances (EGC 05)*, Paris, Janvier 2005; *Revue des Nouvelles Technologies de l'Information*, Cépaduès, Toulouse, 79-90.



## II. ACTIVITE EDITORIALE

En 2003 et en co-direction D.A. Zighed a créé chez Cépaduès la revue RNTI. Depuis 2005, les numéros suivants ont publiés :

RNTI-B-4, 4e Journées francophones sur les Entrepôts de Données et l'Analyse en ligne, EDA'08, Franck Ravat, Olivier Teste, Gilles Zurfluh (Institut de Recherche en Informatique de Toulouse (UMR5505)), 151 pages, I.S.B.N. : 2.85428.847.6

RNTI-C-2, Classification : points de vue croisés, M. Nadif, ISBN à venir

RNTI-L-2, 2ème Conférence Francophone sur les Architectures Logicielles, Aït-Ameur Yamine, 200 pages, I.S.B.N. : 978.2.85428.826.1

RNTI-L-1, Langages et Modèles à Objets, LMO 2008, M. Blay-Fornarino, Y-G Guéheneuc, H. Sahraoui, 216 pages, I.S.B.N. : 978.2.85428.824.7

RNTI-E-11, Extraction et Gestion des Connaissances EGC'2008, F. Guillet, B. Trousse, 740 pages, I.S.B.N. : 978.2.85428.818.6

RNTI-SM-1, ISoLA 2007, Workshop on leveraging applications of formal methods, verification and validation, Y. Ait-Ameur, F. Boniol, V. Wiels, 234 pages, I.S.B.N. : 978.2.85428.814.8

RNTI-A-2, Apprentissage Artificiel et Fouille de Données, Younès Bennani (LIPN), Emmanuel Viennet (LIPN), 218 pages, I.S.B.N. : 978.2.85428.808.7

RNTI-A-1, Data Mining et apprentissage statistique : application en assurance, banque et marketing, Farid Béninel (UMR CNRS 6086, IUT-STID, Niort), Michel Bera (KXEN Suresnes), Christian Partrat (ISFA Lyon), Gilbert Saporta (CNAM Paris), I.S.B.N. : 978.2.85428.794.3

RNTI-W-1, Fouille du Web, Chantal Reynaud (Université de Paris-Sud, CNRS (LRI) & INRIA (Futurs)), Gilles Venturini (Université François-Rabelais de Tours, LI), 86 pages, I.S.B.N. : 978.2.85428.793.6

RNTI-B-3, Entrepôts de Données et Analyse en ligne 2007, Ladjel Bellatreche, Arnaud Giacometti, Patrick Marcel, 200 pages, I.S.B.N. : 978.2.85428.782.0

RNTI-E-10, Défi Fouille de Textes : reconnaissance automatique des auteurs de discours - Campagne DEFT'05 (TALN'05), V. Prince, Y. Kodratoff, J. Azé, M. Roche, 158 pages, I.S.B.N. : 978.2.85428.7

RNTI-E-9, Extraction et Gestion des Connaissances 2007, M. Noirhomme-Fraiture, G. Venturini, 800 pages, I.S.B.N. : 978.2.85428.7

RNTI-B-2, Entrepôts de Données et Analyse en ligne 2006, D. Grigori, S. Lopes, K. Zeitouni, 194 pages, I.S.B.N. : 2.85428.744.4

RNTI-E-8, Systèmes d'Information pour l'Aide à la Décision en Ingénierie Système, Ali Khenchaf (ENSIETA, E3I2, Brest), 228 pages, I.S.B.N. : 2.85428.734.7

RNTI-E-7, Visualisation en Extraction des Connaissances, François Poulet (ESIEA Recherche), Pascale Kuntz (Laboratoire d'Informatique de Nantes-Atlantique), 182 pages, I.S.B.N. : 2.85428.733.9

RNTI-E-6, Extraction et Gestion des Connaissances 2006, G. Ritschard et C. Djeraba, 782 pages, I.S.B.N. : 2.85428.718.5

RNTI-E-5, Extraction des connaissances : Etat et perspectives (Ateliers de la conférence EGC'2005), Florence Cloppet, Jean-Marc Petit, Nicole Vincent, 442 pages, 2.85428.707.x

RNTI-E-4, Fouilles de données complexes, O. Boussaid, P. Gançarski, F. Masségia, B. Trousse, 300 pages, I.S.B.N. : 2.85428.702.9

RNTI-B-1, Entrepôts de Données et l'Analyse en ligne, Fadila Bentayeb, Omar Boussaid, Jérôme Darmont, Sabine Rabaseda (ERIC, Lyon 2), 198 pages, I.S.B.N. : 2.85428.698.7

RNTI-E-3, Extraction et Gestion des Connaissances 2005, Suzanne Pinson (Lamsade, Université Dauphine Paris IX), Nicole Vincent (Crip5, Université René Descartes Paris 5), 742 pages, I.S.B.N. 2.85428.677.4

RNTI-C-1, Classification et Fouille de données, M. Chavent, M. Langlais, 176 pages, I.S.B.N. : 2.85428.667.7



### III. ORGANISATION DE MANIFESTATIONS SCIENTIFIQUES

#### Conférences, ateliers et groupes de travail

**EGC :** La conférence Extraction et Gestion des Connaissances (EGC) a pour objet de rassembler les chercheurs des disciplines de l'informatique décisionnelle, de l'Extraction de Connaissances dans les Données (ECD, KDD) et de la Gestion des connaissances (GC, KM). Plus précisément, elle se fixe pour objectif de promouvoir les échanges multidisciplinaires (apprentissage, statistiques et analyse et données, systèmes d'information et bases de données, ingénierie des connaissances), en connexion avec 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. Les chercheurs du laboratoire ERIC ont été à l'origine de la création de cette conférence en 2000 et continuent aujourd'hui d'animer cette manifestation d'envergure nationale.

**EDA :** La conférence nationale sur les Entrepôts de Données et l'Analyse en-ligne a été créée par les chercheurs du laboratoire d'ERIC. L'objectif de la 1ère journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 05) est de créer et de pérenniser un cadre exclusivement réservé à ces travaux, afin de favoriser la rencontre des chercheurs, des industriels et des utilisateurs français et francophones afin de discuter de l'avancement de la recherche ainsi que d'expériences de développement dans le domaine des entrepôts de données. Cette journée a pour vocation de devenir un rendez-vous national régulier sur le thème des entrepôts de données. 3 éditions ont déjà eu lieu. La prochaine est prévue à Toulouse le 5 et 6 juin 2008.

EDA 2008: <http://www.irit.fr/EDA08/contact.html>

EDA 2007: <http://eda2007.sir.blois.univ-tours.fr/>

EDA 2006: <http://www.prism.uvsq.fr/~eda06/>

EDA 2005: <http://eric.univ-lyon2.fr/~eda05/>

**L'atelier Qualité des Données et des Connaissances**, en association avec la conférence Extraction et Gestion des Connaissances (2007-2008) a été organisé en collaboration avec P. Lenca (Telecom Bretagne) et F. Guillet (LINA, Nantes).

<http://conferences.enst-bretagne.fr/qdc2007/> et <http://conferences.enst-bretagne.fr/qdc2008/>

**Atelier FDC**, en association avec la conférence Extraction et Gestion des Connaissances, nous avons produit cinq éditions de cet atelier :

- 29 Janvier 2008 : Cinquième Atelier sur la "Fouille de Données Complexes dans un processus d'extraction des connaissances", Nice, Sophia-Antipolis, à l'occasion d'EGC 2008
- 23 Janvier 2007 : Quatrième Atelier sur la "Fouille de Données Complexes dans un processus d'extraction des connaissances", Namur, Belgique, à l'occasion d'EGC 2007
- 16 Janvier 2006 : Troisième Atelier sur la "Fouille de données complexes dans un processus d'extraction des connaissances", Lille, à l'occasion d'EGC 2006
- 18 Janvier 2005, : Deuxième atelier sur la "Fouille de données complexes dans un processus d'extraction des connaissances, Paris, à l'occasion d'EGC 2005."
- 20 Janvier 2004 : Premier Atelier sur la "Fouille de données complexes dans un processus d'extraction des connaissances", Clermont-Ferrand, à l'occasion d'EGC 2004

#### **Atelier Mesure de similarité sémantique (SimSem 2008)**

<http://www-rocq.inria.fr/axis/SimSem/AtelierEGC2008/AtelierEGC.html>

**Journée de travail thématique sur les mesures de similarité sémantiques** : Cette première journée de travail sur les données complexes se donne comme objectif de réfléchir et de discuter autour des travaux présentés sur une première partie d'un bilan sur les approches existantes.

<http://eric.univ-lyon2.fr/%7Enmaiz/cmss07/>

**Atelier sur les Systèmes Décisionnels (ASD)** : L'objectif de ce premier atelier maghrébin sur les systèmes décisionnels est de contribuer, en collaboration avec le laboratoire ERIC, à dynamiser la recherche dans ce domaine et à créer une synergie entre les chercheurs maghrébins travaillant dans leur pays ou dans des laboratoires de recherche à l'étranger. Cet atelier s'adresse à tous les experts de la communauté internationale travaillant sur les entrepôts de données pour venir exposer, discuter leurs travaux de recherche (qu'elle soit fondamentale ou appliquée), s'échanger des points de vue et présenter leurs outils décisionnels. C'est également l'occasion d'encourager l'ensemble des doctorants maghrébins concernés par ces questions à participer à cette manifestation et se faire connaître afin de faire émerger une véritable communauté travaillant dans le domaine des systèmes décisionnels. La prochaine édition aura lieu en 2008 après les 2 précédentes :

- ASD 2008, Mohammadia, Maroc le 10 et 11 octobre 2008 : <http://eric.univ-lyon2.fr/~asd/asd2008/>

- ASD 2007, Sousse, Tunisie le 19 et 20 octobre 2007 : <http://eric.univ-lyon2.fr/~asd/asd2007/>
- ASD 2006, Agadir, Maroc le 6 et 8 décembre 2006 : <http://eric.univ-lyon2.fr/~asd/asd2006/>

**Co-organisation de la 6th International Conference on Flexible Query Answering Systems (FQAS 2004), 24-26 june, 2004 (Lyon).**

## **h. Séminaires du master ECD**

### **2004-2005**

- Sylvie Philipp-Foliguet, ETIS (Equipes Traitement des Images et du Signal), CNRS UMR 8051, ENSEA, Cergy-Pontoise, France, Recherche d'images dans des bases à partir de signatures visuelles, Jeudi 15 octobre 2004
- Dragan Gamberger, Rudjer Boskovic Institute, Division of Electronics, Laboratory for Information Systems, Zagreb, Croatie, Avoiding data overfitting in scientific discovery : Experiments in functional genomics, Jeudi 25 novembre 2004
- Georges Hébrail, LTCI-UMR 5141 CNRS, Département Informatique et Réseaux, ENST Paris, Transformation de longues séries temporelles en descriptions symboliques, Jeudi 13 janvier 2005
- Gilles Venturini, Laboratoire d'Informatique, Université François Rabelais, Tours, Un survol des algorithmes biomimétiques pour la classification, Jeudi 27 janvier 05
- Christian Derquenne, R&D EDF, Clamart, France, Méthodes de fusion mises en oeuvre dans le cadre de l'enrichissement de base de données clientèle EDF , Jeudi 03 février 2005
- Lorenza Saitta, Università del Piemonte Orientale Amedeo Avogadro Dipartimento di Informatica, Complexity of Learning and Phase Transitions, Jeudi 10 février 2005

### **2005-2006**

- Jean-Marc Petit, Laboratoire LIRIS, INSA Lyon, Recherche adaptative de bordures, Jeudi 9 février 06
- Marc Sebban, Laboratoire EURISE, Faculté des Sciences, Université de Saint-Etienne, Apprentissage non biaisé d'une distance d'édition stochastique sous la forme d'un transducteur déterministe, Jeudi 17 novembre 2005

- Jean-Michel POGGI, Laboratoire de Mathématiques, Equipe de Probabilités, Statistique et Modélisation, Université Paris-Sud Orsay, Détection de Données Aberrantes par Boosting, Jeudi 9 mars 2006

## 2006-2007

- Alain Morineau, Directeur de la Revue MODULAD, Préhistoire, histoire et perspectives du DM – le point-de-vue d'un statisticien, Jeudi 5 octobre 2006
- Grégoire de Lassence, Consultant Expert Académique SAS Institute, Exemples d'application de Data Mining et retour d'expérience, Jeudi 12 octobre 2006
- Michel Tenenhaus, HEC School of Management (GRECHEC), Approche PLS et analyse de tableaux multiples, Jeudi 6 octobre 2006
- Abdelaziz Faraj, Ingénieur de recherche, Institut Français du Pétrole, Sélection de modèle en régression PLS, Jeudi 9 novembre 2006
- Marc Boullé, France Telecom R&D, Spécificités du Data Mining dans les Télécoms, Jeudi 16 novembre 2006
- Abderrafih Lehman, PERTINENCE MINING Sarl, Solution de text Mining et Linguistique, Jeudi 23 novembre 2006
- Gilbert Saporta, CEDRIC, CNAM, Paris, Classification supervisée et credit scoring, Jeudi 7 décembre 2006
- Malick Paye, Biomathématicien, bioMérieux, Grenoble, Using Data Mining for Biomarker Identification, Jeudi 14 décembre 2006
- Francois Wahl, Institut Francais du Petrole, Analyse d'incertitude et de sensibilité des modèles, Jeudi 11 janvier 2007
- Christophe Roche, ERT Condillac, LISTIC, Université de Savoie, Introduction aux problématiques des ontologies : état et perspectives en recherche et en applications, Jeudi 18 janvier 2007
- Serge Muller, Ingénieur Principal General Electric, Healthcare, Technologies Applications avancées en mammographie numérique, Jeudi 1er février 2007
- Roland Marion-Gallois, Expert Consultant, Statelis, La biostatistique dans les essais cliniques, Jeudi 8 février 2007
- Alexandre Aussem, Laboratoire PRISMA, Lyon 1, Apprentissage sous contraintes de la structure des réseaux bayésiens : Applications au cancer du Nasopharynx, Jeudi 15 février 2007

- Attilio Giordana, Università del Piemonte Orientale, Dipartimento di Informatica, Modeling Complex events by means of Structured Hidden Markov Models, Jeudi 8 mars 2007
- Jean-Gabriel Ganascia, LIP6 - Université Pierre et Marie Curie (Paris VI), Apprentissage non supervisé sur des données très partiellement décrites, Jeudi 15 mars 2007
- Bertrand Chabbat, CNAF-CNEDI Lyon, L'entreprise informationnelle - Exemple : la Branche Famille de la Sécurité Sociale et les documents réglementaires, Jeudi 5 avril 2007
- Yves Lechevallier, INRIA Paris – Rocquencourt, Autour des données d'intervalles, Jeudi 26 avril 2007

## 2007-2008

- Jean Riondet, Directeur de l'Institut International de Formation des Cadres de Santé, IFSCS, HCL, La statistique administrative et les questionnements sociaux, de Vauban à l'INSEE, Jeudi 13 décembre 2007
- Pablo Jensen, Laboratoire IXXI, ENS Lyon, Analyser la répartition des commerces en ville, 20 décembre 2007
- Gilles Bisson, Laboratoire TIMC-IMAG, Equipe Apprentissage Modèle et Algorithmes, Grenoble, Clustering d'objets structurés, application au traitement des molécules et à celui des données de criblage haut débit, Jeudi 10 janvier 2007
- Christian Derquenne, EDF R&D, Clamart, Méthodes de fusion mises en oeuvre dans le cadre de l'enrichissement de base de données clientèle EDF, Jeudi 17 janvier 2007
- Stefan Trausan-Matu, Equipe RACAI, "POLITEHNICA" University of Bucharest, Extraction de connaissances à partir de conversation chat, Jeudi 23 janvier 2007

## i. Séminaires du laboratoire ERIC

### 2004-2005

- Jerzy Korczak, Fouille interactive de séquences d'images IRMf, 30/05/2005
- Brice Effantin, Extraction de communautés dans le graphe du Web, 14/03/2005
- Chantal Reynaud, Comprendre le Web sémantique, 07/03/2005
- Nicole Vincent, La loi de Zipf en analyse d'images, 14/02/2005
- Sébastien Lefèvre, Introduction à la Morphologie Mathématique : principaux outils et applications 31/01/2005

- Karine Zeitouni, Entreposage et fouille de données spatiales et spatio-temporelles, 29/11/2004
- Zdenko Sonicki, Intelligent Data Analysis and Data Mining – Application in Medicine, 29/11/2004
- Edwige Fangseu Badjio, Qualité des IHM pour la fouille visuelle des données, 27/09/2004

#### 2005-2006

- Michel Simonet, Ontologies, bases de connaissances et bases de données, 10/04/2006
- Ricco Rakotomalala, Les logiciels gratuits de DATA MINING pour l'enseignement, 12/12/2005
- Kamel Aouiche, Techniques de fouille de données pour l'optimisation automatique de performance des entrepôts de données, 28/11/2005
- Silvia Biffignandi, Shift-Share Analysis, 17/10/2005

#### 2006-2007

- Miriam Alvariez, Plans d'expériences pour un modèle de simulation, 18/06/2007
- Rokia Missaoui, Opérateurs algébriques pour la manipulation des treillis de concepts, 23/04/2007
- Anne-Muriel Arigon, Développements d'applications pour l'identification de séquences génomiques, 12/03/2007
- Henri-Maxime Suchier, Nouvelles contributions du boosting en apprentissage automatique, 12/02/2007
- Frédéric Château, Inférence pour la Statistique Structurale, 15/01/2007
- Omar Boussaid, Evolution de l'entrepotage des données complexes, 27/11/2006
- Yvan Bédard, Complexité des données géospatiales et peuplement de cubes de données : problématique, besoins et solutions, 27/11/2006
- Riadh Ben Messaoud, Couplage de l'analyse en ligne et de la fouille de données pour l'exploration, l'agrégation et l'explication des données complexes, 24/11/2006
- Jérôme Darmont, Optimisation et évaluation de performance pour l'aide à la conception et à l'administration des entrepôts de données complexes, 20/11/2006
- Djamel Zighed, Variation autour des mesures d'entropie, 16/10/2006

## IV. PROJETS DE RECHERCHE APPLIQUEE

### Enquête sur le devenir des apprentis de l'enseignement supérieur en Rhône-Alpes

<i>Identification des partenaires</i>	Région Rhône-Alpes Formasup Rhône-Alpes (IPRA) Rectorats de l'académie de Lyon et Grenoble
<i>Objectifs recherchés</i>	Conception, réalisation, exploitation et présentation d'une enquête d'insertion de l'ensemble des apprentis de l'enseignement supérieur en Rhône-Alpes.
<i>Durée et financement</i>	Financement par Formasup Rhône-Alpes : - 15 000 € en 2002-2003 - 3 600 € les années suivantes (depuis 2004)

### Entrepôt virtuel de données bancaires

<i>Identification des partenaires</i>	Crédit Lyonnais. Direction d'Exploitation Rhône-Alpes-Auvergne
<i>Objectifs recherchés</i>	L'objectif de ce projet est d'assurer dans un délai de trois ans la mise au point d'outils méthodologiques pour la gestion et l'analyse de données bancaires, qui se présentent sous forme de données hétérogènes. Du point de vue du Crédit Lyonnais, il s'agit de développer un système d'aide à la décision dans le domaine de ciblage clients. Du point de vue du laboratoire ERIC, il s'agit d'acquérir une expertise dans le domaine de l'entrepôt virtuel de données hétérogènes. Il s'agit de construire des cubes de données à la volée en vue d'analyses (analyse en ligne (OLAP) et fouille de données), qui nécessite une intégration efficace de données.
<i>Durée et financement</i>	Durée : 2004-2007 Financement par le Crédit Lyonnais : - une bourse de thèse CIFRE de 3 ans (Cécile Favre)

### Fouille de Données Multistratégies (FoDoMuSt)

<i>Identification des partenaires</i>	Action Concertée Incitative "Masses de Données" Laboratoire LSIT (Laboratoire des Sciences de l'Images, de l'Informatique et de la Télédétection), Université de Strasbourg I Laboratoire LIV (Laboratoire Image et Ville), Université de Strasbourg I
<i>Objectifs recherchés</i>	Les objectifs du projet, associés à l'imagerie spatiale, sont : d'une part, proposer une méthode d'aide à l'interprétation à partir d'une masse de données images et d'autre part, définir un processus complet de fouilles de données (structuration, construction des « objets », classification et interprétation de l'information)

	permettant une utilisation conjointe et complémentaire des différentes sources. Ce dernier aspect est rarement pris en compte dans les méthodes actuelles d'extraction. Le verrou principal réside dans la nécessité d'utiliser une multi-formalisation à plusieurs niveaux d'abstraction selon une approche multi-stratégie dans le processus de fouilles de données.
<i>Durée et financement</i>	Durée : 2004-2007 Financement par le Ministère de la Recherche (ACI) : 69 000 €

### Système Intelligent pour la Recherche d'Information à l'Usage de la Santé (SIRIUS)

<i>Identification des partenaires</i>	Région Rhône-Alpes Centre anti cancéreux Léon Bérard Lyon
<i>Objectifs recherchés</i>	Le Système Intelligent pour la Recherche d'Information à l'Usage de la Santé (SIRIUS) sera développé et testé avec des usagers (Centre Léon Bérard). Le choix du secteur de la cancérologie résulte à la fois de la longue collaboration que nous entretenons avec le Centre Léon Bérard et de l'intérêt que porte la Région Rhône-Alpes à ce domaine, notamment à travers la création du cancérpôle.
<i>Durée et financement</i>	Durée : 2004-2007 Financement par la Région Rhône-Alpes : - 3 700 € en fonctionnement - bourse de thèse de 30 444 € pour 3 ans (Hakim HACID)

### Interdépendance des marchés immobiliers résidentiels (INTERREG)

<i>Identification des partenaires</i>	Université de Genève
<i>Objectifs recherchés</i>	Etude sur l'interdépendance des marchés immobiliers résidentiels sur le bassin franco-valdo-genevois dans le cadre du programme européen INTERREG. Ce projet concerne l'analyse des marchés fonciers, des marchés résidentiels privés locatifs et des marchés résidentiels de vente d'appartements et de maisons individuelles simultanément sur les différentes zones du bassin. Son objectif est d'améliorer la connaissance du fonctionnement des marchés immobiliers résidentiels privés : - en visualisant l'évolution des prix des biens et des services sur une période de trente ans, - en observant la dynamique de ces marchés immobiliers simultanément dans les quatre zones du bassin, - en mettant en évidence les interdépendances existant entre les marchés immobiliers des différentes zones,

	- en créant des modèles économétriques permettant une analyse prospective.
<i>Durée et financement</i>	Durée : 2004-2006 Financement par le fond européen INTEREG : 55 000 €

### Positionnement relatif des législations du travail

<i>Identification des partenaires</i>	Bureau International du Travail Université de Genève Fondation RUIG
<i>Objectifs recherchés</i>	Ce projet vise à développer des méthodes de fouille de texte visant à étudier et à positionner les législations du travail des différents pays. Le Bureau International du Travail (BIT) souhaite ensuite dresser des cartographies permettant aux représentants des différents pays de se positionner les uns par rapport aux autres. Le laboratoire ERIC assure la partie text mining du projet pour extraire les paramètres descriptifs des corpus juridiques. Il devra pour ce faire exploiter plusieurs centaines de textes relatifs à la législation du travail.
<i>Durée et financement</i>	Durée : 2005-2007 Financement par le BIT : 12 000 €

### Méthodes et logiciels pour l'extraction de règles d'association

<i>Identification des partenaires</i>	Laboratoire d'Ingénierie des Connaissances de l'Université de Prague, République Tchèque
<i>Objectifs recherchés</i>	Nous avons entamé une collaboration scientifique avec l'équipe d'ingénierie des connaissances de l'Université de Prague. L'objectif est de développer en commun des plates formes de data mining. ERIC apportant son expérience et son savoir faire à travers la plate forme SIPINA, l'équipe Tchèque a développé une plate forme pour l'extraction des règles d'association baptisée LispMiner.
<i>Durée et financement</i>	Durée : 2004-2006 Financement par le programme d'échange Franco-Tchèque Barrande : 6 000 €

### Analyse automatique des cours de bourse (Tradingbots)

<i>Identification des partenaires</i>	Nicolas Macherey porteur d'un projet de création d'entreprise accueilli au sein de l'incubateur CREALYS
---------------------------------------	---

<i>Objectifs recherchés</i>	Conception et développement d'un système financier qui permet d'analyser les cours de bourse ou de change afin de prendre des décisions automatiques. Mise en place d'un progiciel.
<i>Durée et financement</i>	Durée : 2006-2007 Financement par la Région Rhône-Alpes : 29 000 €

### Génération de règles d'association

<i>Identification des partenaires</i>	SPAD
<i>Objectifs recherchés</i>	Implémentation d'un module de création de règles d'association dans la dernière version 7.0 du logiciel.
<i>Durée et financement</i>	Durée : 2005 Financement par SPAD : 4 000 €

### PMSI Privé

<i>Identification des partenaires</i>	UMR LIRIS, Université Claude Bernard Lyon 1 EA PRISMA, INSA de Lyon et Université Claude Bernard Lyon 1
<i>Objectifs recherchés</i>	Méthodologie d'analyse à visée décisionnelle des grandes bases de données médico-économiques : le PMSI privé
<i>Durée et financement</i>	Durée : 2005 Financement par la Région Rhône-Alpes : 25 000 €

### Gestion et de visualisation interactive de règles d'association

<i>Identification des partenaires</i>	DEENOV
<i>Objectifs recherchés</i>	Réalisation d'un module dans un logiciel de data mining : gestion et de visualisation interactive de règles d'association
<i>Durée et financement</i>	Durée : 2006-2007 Financement par DEENOV : 8 000 €

### Etudes marketing

<i>Identification des partenaires</i>	DATAEXPRESSO
<i>Objectifs recherchés</i>	Conseil et expertise pour des études dans le domaine du marketing
<i>Durée et financement</i>	Durée : 2005-2007 Financement par DATAEXPRESSO : 25 000 €

### Modélisation du processus de fabrication du vaccin de la coqueluche acellulaire

<i>Identification des partenaires</i>	SANOPHI-PASTEUR
<i>Objectifs recherchés</i>	Méthodologie pour formaliser la connaissance des procédés afin de définir les propriétés requises pour des productions optimales. Utilisation des méthodes de Data Mining et d'Ingénierie des Connaissances pour analyser le processus de fermentation de la production du vaccin « coqueluche acellulaire » ; pour construire un modèle de contrôle du processus ; pour expliquer les dérives observées sur les durées de culture industrielle et pour tenter de maîtriser ce facteur important dans la production des vaccins.
<i>Durée et financement</i>	Durée : 2007 Financement par SANOPHI-PASTEUR : 6 000 €

### Gestion de portefeuilles virtuels d'actions (Tapeo)

<i>Identification des partenaires</i>	Alexis Lewalle et Victorien Leleu porteurs d'un projet de création d'entreprise accueilli au sein de l'incubateur CREALYS
<i>Objectifs recherchés</i>	Conception et développement d'un site Web pour la gestion de portefeuilles virtuels d'actions via des communautés d'utilisateurs sur internet.
<i>Durée et financement</i>	Durée : 2007-2008 Financement par la Région Rhône-Alpes : 30 000 €

### Guides collaboratifs sur Internet (Rickshaw)

<i>Identification des partenaires</i>	Pierre Bizollon porteur d'un projet de création d'entreprise accueilli au sein de l'incubateur CREALYS
---------------------------------------	--

<i>Objectifs recherchés</i>	Conception et développement de guides collaboratifs sur Internet pour rassembler les informations et opinions les plus pertinentes sur un sujet, un produit ou un service.
<i>Durée et financement</i>	Durée : 2007-2008 Financement par la Région Rhône-Alpes : 30 300 €

### Evaluation de zones de chalandise (ProXan)

<i>Identification des partenaires</i>	Romain Nicolas Vallée porteur d'un projet de création d'entreprise accueilli au sein de l'incubateur CREALYS
<i>Objectifs recherchés</i>	Evaluation des zones de chalandise avec un outil d'aide à l'implantation des commerces : comparaison de l'environnement socio-économique et démographique de plusieurs sites potentiels.
<i>Durée et financement</i>	Durée : 2008-2009 Financement par la Région Rhône-Alpes : 30 500 €

### Analyse des débats en ligne (Between)

<i>Identification des partenaires</i>	Robin Coulet porteur d'un projet de création d'entreprise accueilli au sein de l'incubateur CREALYS
<i>Objectifs recherchés</i>	Modèle de représentation et d'analyse des débats en ligne sur Internet. Mise en place d'un progiciel.
<i>Durée et financement</i>	Durée : 2008-2009 Financement par la Région Rhône-Alpes : 20 500 €

## V. COLLABORATIONS INTERNATIONALES

### Université Laval à Québec, Canada

<i>Identification du partenaire</i>	Professeurs Nadir Belkhiter et Guy Mineau
<i>Collaboration en enseignement</i>	Nous entretenons depuis de longues années une collaboration régulière avec l'Université Laval à Québec. Le Professeur Nadir Belkhiter. Est régulièrement invité à l'Université Lyon 2 pour assurer des enseignements en master dans le domaine de la fouille des données et des interfaces de communication homme-machine.
<i>Collaboration en recherche</i>	Grâce aux compétences du Professeur Belkhiter dans le domaine des interfaces de communication homme-machine, nous développons une réflexion méthodologique sur les interfaces et le data mining. En effet, les utilisateurs de ces techniques sont potentiellement très nombreux mais, ces outils ne seront réellement utilisés que s'ils sont facile à appréhender. Cette recherche vise à étudier les modes d'interaction et les techniques de visualisation dans le domaine de l'ECD.

### Université Laval à Québec, Canada

<i>Identification du partenaire</i>	Laboratoire CRG (Centre de Recherche en Géomatique) Professeur Yvan Badard
<i>Collaboration en recherche</i>	Entrepôts actifs de données spatiales.

### Université du Québec en Outaouais, Canada

<i>Identification du partenaire</i>	Laboratoire LARIM Professeur Rokia Missaoui
<i>Collaboration en recherche</i>	Notre collaboration sur le couplage OLAP - Data Mining lors de séjours en France ou au Canada a déjà donné lieu à des publications communes.

### Université de Genève, Suisse

<i>Identification du partenaire</i>	Professeur Gilbert Ritschard
-------------------------------------	------------------------------

<i>Collaboration en enseignement</i>	Le Professeur G. Ritschard intervient depuis 1999 en tant que professeur invité dans un cours en master recherche ECD sur les mesures d'association.
<i>Collaboration en recherche</i>	Nous travaillons avec le Professeur G. Ritschard depuis de nombreuses années et nous avons déjà de nombreuses publications communes.

### Université de Prague, République Tchèque

<i>Identification du partenaire</i>	Professeurs Jan Rauch et Petr Berka
<i>Collaboration en recherche</i>	Développement d'outils communs pour la fouille de données.

### Ecole Nationale d'Informatique de Tunis, Laboratoire RIADI-GDL, Tunisie

<i>Identification du partenaire</i>	Mme Hajer Bazaoui
<i>Collaboration en recherche</i>	Modélisation et analyse de data marts spatio-temporels.
<i>Perspectives</i>	Après avoir construit un modèle générique de data marts spatio-temporels, nous travaillons actuellement sur une démarche exploratoire incluant des analyses descriptives, de l'OLAP et de l'extraction des connaissances.

### Université d'Oklahoma, Norman, USA

<i>Identification du partenaire</i>	Professeur Le Gruenwald
<i>Collaboration en recherche</i>	Un projet de recherche concernant l'utilisation de techniques de fouille de données pour l'auto-administration des entrepôts de données a abouti à plusieurs publications communes (concernant l'auto-indexation, principalement). Nous envoyons régulièrement des étudiants de Master en stage aux Etats-Unis depuis 2001.
<i>Perspectives</i>	Poursuivre et renforcer la collaboration sur le projet d'auto-administration. D'un point de vue scientifique, il s'agit d'une part d'étendre notre approche d'auto-indexation à d'autres techniques d'optimisation de performance (matérialisation de vues, notamment) et, d'autre part, de tester différentes technique de fouille dans ce cadre (motifs fréquents, motifs séquentiels, classification...) pour trouver la plus adaptée à chaque cas.

	Applications prévues aux données complexes.
--	---

### Ecole Nationale d'Informatique, Université de Fianarantsoa, Madagascar

<i>Identification du partenaire</i>	Victor Manantsoa
<i>Collaboration en recherche</i>	Performance des entrepôts de données complexes.
<i>Perspectives</i>	Développer la collaboration entre ERIC et l'ENI. Les deux laboratoires ont des thématiques de recherche très proches et ont la volonté de développer des projets en commun. Le lien entre nos deux structures de recherche est actuellement assuré en grande partie par M. Ralaivao, dont le travail de thèse matérialise cette volonté de collaboration et se renforce à chacun de ses séjours au laboratoire ERIC.

### Université de Zagreb, Croatie

<i>Identification du partenaire</i>	Professeur Bojana DALBELO BASIC. Avec l'aide du Ministère des Affaires Etrangères (programme EGIDE depuis 2004)
<i>Collaboration en recherche</i>	Méthodes de fouille des données médicales Organisation conjointe de « International Workshop on Intelligent Data Analysis and Data Mining, Application in Medicine » depuis plusieurs années. Séjours en France pour fouiller de nouvelles données épidémiologiques et confronter les méthodes.

### Université de Ljubljana, Slovénie

<i>Identification du partenaire</i>	Professeur Blaz ZUPAN Avec l'aide du Ministère des Affaires Etrangères (programme EGIDE depuis 2004)
<i>Collaboration en recherche</i>	Méthodes de fouille des données médicales Organisation conjointe de « International Workshop on Intelligent Data Analysis and Data Mining, Application in Medicine » depuis plusieurs années. Séjours en France pour fouiller de nouvelles données épidémiologiques et confronter les méthodes.

### Université nationale d'Economie de Kharkov, Ukraine

<i>Identification du partenaire</i>	Professeurs Olexandr PUSKAR et Irina ZOLOTORIEVA
<i>Collaboration en enseignement</i>	Mise en place d'un master franco-ukrainien. Financement par le Ministère des Affaires Etrangères depuis 2005-2006.

### Université d'Alessandria, Italie

<i>Identification du partenaire</i>	Professeur Lorenza Saitta
<i>Collaboration en enseignement</i>	Le Professeur Lorenza Saitta intervient depuis plusieurs années en tant que professeur invité dans un cours en master recherche ECD. Partenaire dans le projet de master européen Erasmus Mundus
<i>Collaboration en recherche</i>	Co-encadrement de la thèse de Julien Charbel

### Université Polytechnique de Barcelone, Espagne

<i>Identification du partenaire</i>	Professeur Tomas Aluja
<i>Collaboration en enseignement</i>	Le Professeur Tomas Aluja intervient en 2007-2008 en tant que professeur invité dans un cours en master recherche ECD. Partenaire dans le projet de master européen Erasmus Mundus

### Ecole Polytechnique de Bucarest, Roumanie

<i>Identification du partenaire</i>	Professeurs Eugenia Kalisz et Stefan Trausan
<i>Collaboration en enseignement</i>	Partenaire dans le projet de master européen Erasmus Mundus

# Bilan MA2D : rapport d'activité

2005-2008



## 7 ACTIVITES DE RECHERCHE

### 7.1 Thèmes de recherche

Le processus de décision, dans des environnements complexes, constitue le cœur des travaux de recherche de l'équipe. Ceux-ci tentent de s'intéresser au processus dans son intégralité, c'est à dire du recueil des données jusqu'à la prise de décision elle-même. Par ailleurs, les applications menées dans le domaine des Sciences Humaines et Sociales (SHS), incluant dans notre esprit les systèmes de santé, impliquent la prise en compte de comportements complexes dans les approches proposées.

De manière globale, les travaux de l'équipe concernent les champs suivants :

- analyse des données complexes : description et structuration,
- phénomènes d'apprentissage, extraction et représentation des connaissances,
- modèles d'évaluation de stratégies,
- agrégation de préférences, négociation,
- modélisation de phénomènes complexes.

L'ensemble des travaux concernant ces cinq champs a donné lieu soit à des publications sur des points fondamentaux, via les approches formelles utilisées, soit à des publications à caractère appliqué, essentiellement dans le domaine des SHS.

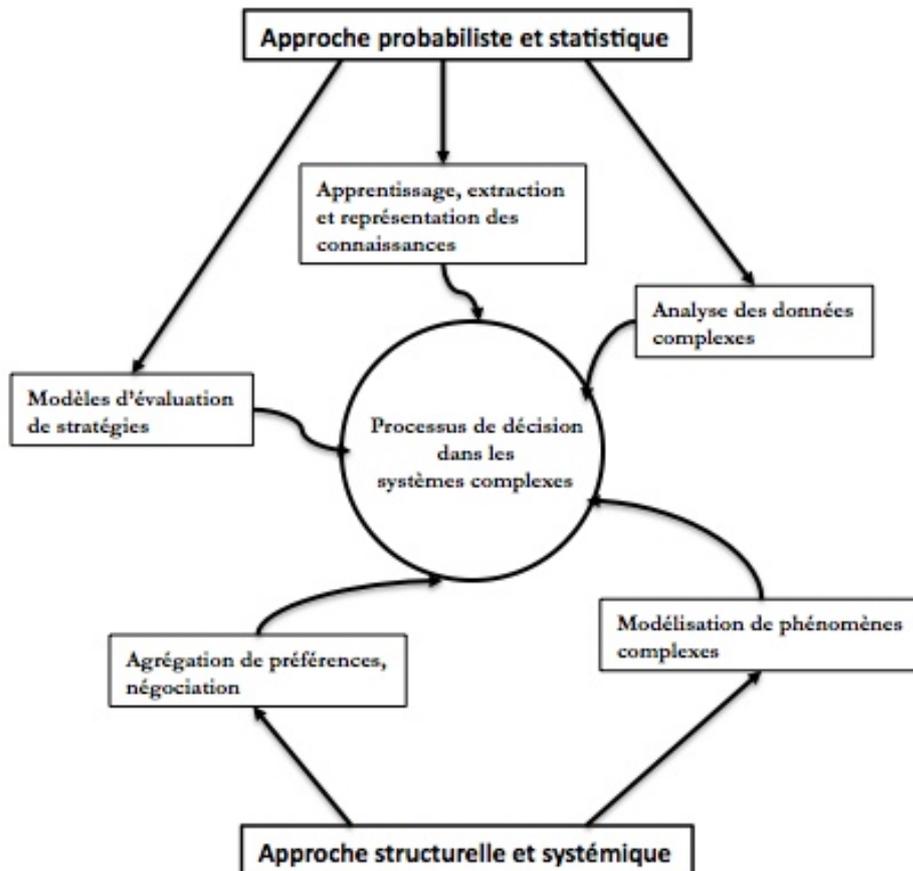
Les approches formelles utilisées dans les travaux menés au sein de l'équipe peuvent être segmentées en deux grands types :

- Les approches de type probabiliste et statistique,
- Les approches de type structuraliste et systémique.

Dans le premier type d'approche, des travaux concernant les ensembles aléatoires, l'économétrie de la santé, l'apprentissage statistique, les modèles graphiques ainsi que l'analyse spatiale ont été menés. Dans le second type d'approche, les travaux menés portent sur la prétopologie, l'analyse des données fonctionnelles, la classification non supervisée, l'agrégation des préférences, la simulation des phénomènes de négociation.

Dans les deux types d'approche, les recherches menées s'appuient à la fois sur des concepts originaux, développés au sein de l'équipe depuis plusieurs années, comme par exemple la prétopologie et la statistique des ensembles aléatoires, et sur des concepts extérieurs à l'équipe qui sont adaptés et/ou modifiés de manière à proposer des solutions originales aux problèmes traités.

Le schéma ci-dessous illustre la démarche de recherche, telle qu'elle est menée au sein de l'équipe.



Dans le cadre de ces travaux, l'équipe a une activité soutenue dans l'animation scientifique notamment :

- Édition de la revue Santé Décision Management (anciennement Santé et Systémique) publiée chez Hermès. Site web : <http://ses.revuesonline.com/>.
- Organisation de la 9<sup>ème</sup> Conférence internationale on Systems Science in Health Care qui s'est tenue à Lyon les 3, 4 et 5 septembre 2008. Cette conférence de niveau internationale a réuni plus de 250 chercheurs de divers horizons et des professionnels de santé. Site web : <http://www.icsshc2008.org/>.
- Co-animation, depuis 2003, du groupe de travail Systèmes Complexes et Décision Distribuée associé à la ROADEF (société Française de Recherche Opérationnelle et d'Aide à la Décision) et au GDR MACS.
- Direction, depuis 2007, de l'association PretopologiCS qui a pour but la promotion, la valorisation et la diffusion de la recherche en prétopologie et modélisation des systèmes complexes.

## 7.2 Collaborations

Dans le cadre des projets et de l'animation des groupes de travail, l'équipe a développé des collaborations nationales et internationales :

- CEDEJ (Centre d'Etudes et de Documentation Economiques, Juridique et sociale) au Caire en Egypte (sous tutelle du CNRS). Contact : *Hala Bayoumi, Ingénieur CNRS*. Collaboration et échanges dans le cadre du projet MOUSSON.
- La Banque Mondiale à Washington DC aux USA. Contact : *Alexander Preker, Health Economist Director*. Collaboration dans le cadre du projet SocialRe.
- Le Bureau International du Travail à Genève en Suisse. Contact : *David Dror, Health Insurance Director*. Collaboration dans le cadre du projet SocialRe.
- School of Computing and Mathematical Sciences à Liverpool. Contact : *Paulo J.G. Lisboa, Professeur*. Collaboration dans le cadre des travaux de valorisation de la recherche menés avec la société ICTA.
- Laboratoire TIMC-IMAG à Grenoble. Contact : *Michel Simonet, Professeur*. Collaboration dans le cadre du projet BC3 (Base de Connaissances Cœur Cerveau).
- Unité Inserm de « neuro-anatomie fonctionnelle du comportement et de ses troubles » à Paris (Hôpital de la Pitié Salpêtrière). Contact : *Bruno Dubois, Directeur de l'Unité INSERM*. Collaboration dans le cadre du projet BC3 (Base de Connaissances Cœur Cerveau)
- *équipe de J.P. Daurès (université de Montpellier 1, CHU de Nîmes), dans le cadre du projet « Méthodologie d'évaluation en Santé Publique », sur les problématiques de modélisation et développement méthodologique d'outils d'aide à la décision en santé.*
- University of Heidelberg Fachhochschule of Heilbronn, pour la mise en place d'un master Franco-Allemand. Contact : Pr. Heinrich Krayl
- Université Cheikh Anta Diop (Dakar – Sénégal) pour le développement de la prétopologie. Contact : Alassane Bah
- Institut de recherche en santé publique CNRST – BurkinaFaso (projet mousson) contact Pr. Blaise Sondo.

Le fait le plus marquant de ce point de vue est l'arrivée au sein de l'équipe de Mondher Toumi, titulaire d'une chaire d'excellence Industrielle en partenariat avec la société Lunbeck SA et l'université Lyon 1, via la fondation partenariale Lyon 1. Le programme scientifique lié à cette chaire exploite les compétences méthodologiques du laboratoire en se concentrant sur le thème de l'aide à la

décision et sur la gestion des organisations de santé à partir de l'information stockée dans les grandes bases de données médico-économiques (PMSI, base SNIIRAM bas PHARMETRIX,..).

## 7.3 Production

L'ensemble de ces travaux a donné lieu, sur la période 2005-2008 à 71 publications dans des revues ou à des congrès et manifestations diverses (cf. tableau ci-dessous).

	Nb
Revues d'audience internationale (ACL)	15
Autres revues avec comité de lecture (ACLN)	17
Actes de conférences internationales (ACTI)	11
Actes de conférences nationales (ACTN)	15
Conférences sans actes avec comité de sélection (COM)	14
Ouvrages scientifiques (OS)	5
Direction d'ouvrage ou de revue (DO)	1
Conférence invité (INV)	1

M. Toumi n'étant arrivé que cette année nous n'avons pas comptabilisé ses publications dans l'équipe (13 revues pour la période considérée) sauf pour les publications communes.

## 7.4 Projets de recherche

Par ailleurs, l'équipe a porté ou porte les projets suivants :

- Projet EGOHID 2 : Projet Européen (European Global Oral Health Indicators Development, phase 2), dans le cadre du 6ème PCRD, coordinateur du projet D Bourgeois.
- Projet « Méthodologie d'évaluation en Santé Publique » financé par l'Institut National de Santé Publique. Coordinateur C. Siani
- Projet ECHOUTCOME : Projet Européen (European Consortium for Patient outcome), dans le cadre du FP7, coordinateur M. Lamure
- Projet « Formalisation de la prétopologie en vue de ses applications au domaine des sciences sociales », dans le cadre des Programmes Exploratoires Pluridisciplinaires (PEPS) conjoints aux départements SHS et ST2I, responsable J.P. Auray.

Elle a également participé aux projets ci-dessous :

- Projet MOUSSON : PIR CNRS
- Projet HRP2 : Hospital Resources Planning, projet régional Rhône-Alpes
- Projet pluridisciplinaire « Negosim », CNRS
- Projet « Trajectoires sous incertitudes : application au contrôle aérien », GDR RO du CNRS.

- Projet « Aménophis » (Application de l'intégrale de Choquet à l'aide à la reconstruction tomographique en vue d'aide à la décision médicale), dans le cadre des Programmes Exploratoires Pluridisciplinaires (PEPS) conjoints aux départements SHS et ST2I.

## **8 ENSEIGNEMENTS ET FORMATION PAR LA RECHERCHE**

### **8.1 Formation par la recherche**

L'équipe a porté le master « Sciences de Systèmes de Santé » de sa création jusqu'en 2007, date à laquelle il a fusionné avec le master AMIV pour donner le master « Santé et Populations » au sein duquel elle reste un acteur majeur (coresponsabilité du master et responsabilité de deux spécialités). Elle participe, par ailleurs, aux enseignements de différents masters de l'université Lyon 1 (Masters MIAGE et Ingénierie Mathématique) ainsi que de l'université Lyon 2 (Master IIIDE). Ces participations favorisent le transfert des connaissances sur les nouvelles générations et permettent de maintenir un effectif de doctorants à un niveau important, en moyenne 3 doctorants par directeur de recherche.

### **8.2 Implication de l'unité dans l'enseignement**

L'équipe s'implique fortement dans différentes unités d'enseignements de l'université. Elle participe notamment à l'UFR d'informatique (l'un des membres a été directeur de l'UFR, un autre est responsable des relations internationales) à l'école d'ingénieur universitaire ISTIL (responsabilité de l'eMiage l'enseignement à distance de la filière Miage).

## **9 PUBLICATIONS 2005-2008**

### **9.1 Revues internationales**

1. E. Hasle-Pham, B. Arnould, H. Spath, A. Fallet, G Duru, P. Marquis. Role of clinical patient-reported outcome and medico economic studies in the public hospital drug formulary decision making process: result of an european survey. *Health Policy* 71(2):205-212. 2005.
2. A. Chateauneuf, A. Rico. A Sugeno integral representation under stone condition. *Fuzzy Sets and Systems* 151(2):261-267. 2005.

3. Marino, P., Siani, C., Roché, H. and Moatti, J.P., “The impact of uncertainty in cost-effectiveness analysis of medical strategies: the case of high dose chemotherapy for breast cancer patients”. *The International Journal of Technology Assessment in Health Care*, Vol. 21, No. 3. 2005.
4. A. Le Lay, N. Despiegel, C. François, G Duru. Can discrete event simulation be of use in modelling major depression?. *Cost Eff Resour Alloc.* 5(4):4-19. 2006.
5. I. Mahé, C. Sollier, G Duru, H. Lamarque, J. Bergman, L. Drouet. Use and monitoring of vitamin K antagonists in everyday medical practice. French results of the international ISAM study of patients with nonvalvular atrial fibrillation. *Presse Med.* 35(12):1797-1803. 2006.
6. F. Jézéquel, F. Rico, J.-M. Chesneaux, M. Charikhi, Reliable computation of a multiple integral involved in the neutron star theory. in *Mathematics and Computers in Simulation*, 71( 1), p. 44-61, 2006.
7. de Peretti, C. and Siani, C., “Decision-making with the incremental cost-effectiveness ratio under uncertainty”. *Health and System Science*, 9(1-2), 111–145, 2006.
8. Siani, C., and de Peretti, C., “Fieller’s Method Performance in Problematic Cases for Decision-Making”. *Health and System Science*, Vol. 9, No. 1-2. Pages 205–226, 2006.
9. S. Debarbieux, G Duru, S. Dalle, O. Béatrix, B. Balme, L. Thomas. Sentinel lymph node biopsy in melanoma: a micromorphometric study relating to prognosis and completion lymph node dissection. *Br J Dermatol.* 157(1):1793-1803. 2007.
10. T.V. Le, N Kabachi, M Lamure. A clustering method associated pretopological concepts and k-means algorithm. Recent advanced in stochastic modelling & data analysis (), Christos Skiadas - World scientific Publishing CoP, ISBN 978-981-270-968-. 2007.
11. Siani, C. and de Peretti, C.. “Analysing the performance of bootstrap neural tests for conditional heteroskedasticity in ARCH-M models using Davidson and MacKinnon graphical methods. *Computational Statistics and Data Analysis*”. Vol. 51, No. 5. Pages 2442-2460, 2007.
12. P.J.G.Lisboa, T.A.Etchells, I.H.Jarman, M.S.H.Aung, S.Chabaud, T.Bachelot, D.Perol, T.Gargi, V.Bourdès, S.Bonnevay, S.Négrier. « Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer ». *Neural Networks*, Vol 21, pages 414-426, 2008.
13. A. Chateauneuf, M. Grabisch, Agnès Rico. Modeling Attitudes toward Uncertainty through the Use of Sugeno Integral. *Journal of Mathematical Economics* 44(11):1084-1099. 2008.
14. Agnès Rico. Sugeno integral in a finite Boolean algebra. *Fuzzy Sets and Systems* 159(13):1709-1718. 2008.

## 9.2 Revues nationales

1. C. Blein, M. Lamure, P. Garassus, D. Clot, Etude de l'impact des évolutions de classification auprès des établissements de santé du secteur privé, *Santé et systémique* vol 9 no3-4/2006, p 169-204
2. C. Blein, D. Clot, M. Lamure, Comparaison internationale des systèmes de classification de patients comme outil du paiement prospectif, *Santé et systémique* vol 9 no3-4/2006, p 65-133
3. C. Blein, D. Clot, M. Lamure, Comparazione internazionale dei sistemi di classificazione dei pazienti, *Mecosan, management ed economia sanitaria* no60, 2006, p 99-119
4. Regnault A., De La Loge C., Bounekkar A., Lamure M. Le fonctionnement différentiel de l'item dans la démarche d'évaluation de la validité transculturelle des questionnaires patients. *Santé et Systémique*, 2006, vol 9 (1-2), pp. 175-203.
5. Le T. V., Lamure M., Kabachi N. Visualisation et analyse de trajectoires patients. Approche méthodologique et outil de visualisation. *Santé et Systémique*, vol 9 (1-2), 2006, pp. 43-71.
6. V. Deslandres, A Bounekkar. Spécification du système d'information hospitalier dans le cadre de regroupement d'établissements. *Santé et systémique* Volume 10, Issue 1-2, p.65-82 (2007)
7. I.B.I. Berthe, A Bounekkar, G Duru. Systèmes d'information et oralité. *Santé et systémique* Volume 10, Issue 1-2, p.65-82 (2007)
8. P. Bernard, A. Dupuy, P. Brun, A. Sasko, G Duru, N. Nicoloyannis, L. Decuypere, Treatment of superficial basal cell carcinomas and multiple actinic keratoses by French dermatologists: a prospective medical cost analysis. *J. Grob. Ann Dermatol Venereol.* 134(6-7):527-533. 2007.
9. J. Benaïm, B. Fuller, B. Briat, M. Bami, M. Aoustin, J. Dubois, P. Rieu, C. Behaghel, C. Jacquelinat, G Duru, JP Auray. Cost of dialysis in France. *Nephrol Ther.* 3(3):96-106. 2007.
10. F.Maunoury, A.Bounekkar, S.Bonnevay, J-P.Auray. « Evaluation médico-économique des stratégies de santé ». *Santé, Décision, Management*, Editions Hermès - Lavoisier, Vol.11, N.3-4, pages 169-181, 2008.
11. Auray, J.P., Brissaud, M., Duru, G., Lamure, M. et Siani, C., 2009, « Eléments de prétopologie généralisée ». *Studia Informatica*, Vol. 7.1, avril 2009, Collection Hermann Informatique,p. 47–77.

## 9.3 Conférences internationales

1. Bonnevay S., Kabachi N., Lamure M., Agents-Based Simulation of Coalition Formation in Cooperative Games. Conférence IEEE/WIC/ACM Joint International Conference on Intelligent Agent Technology, Compiègne, pp. 136-139. 2005.
2. J.-M. Chesneaux, L.-S. Didier, F. Rico : “The SOFA toolbox: a tool for tuning fixed point representations in embedded designs”, *Proceedings of SCAN*, Duisburg, Germany, pp. 39-40 (2006)

3. Le T. V., Lamure M. A pretopological approach for clustering. Proceeding of Knowledge Extraction and Modeling, Workshop, September 2006. Capri, Italie.
4. de Peretti, C. and Siani, C., 2006, "Bootstrapping tests for conditional heteroskedasticity based on artificial neural network", International Association for Mathematics and Computers in Simulations (IMACS) Multiconference on "Computational Ingeneering in Systems Applications" (CESA), ' and SMC Societies, Beijing (Chine), 372–379.
5. Siani, C. and de Peretti, C., 2006, "Algorithm for Making Decision with the Incremental Cost-Effectiveness Ratio handling the Mirror Decision-Making Problem", International Association for Mathematics and Computers in Simulations (IMACS) Multiconference on "Computational Ingeneering in Systems Applications" (CESA), ' and SMC Societies, Beijing (Chine), 366–371.
6. Siani, C. and de Peretti, C., 2006, "Bootstrap Tests for Comparing the Mean Costs Between Two Health Care Strategies", ' national Conference on Service Systems and Service Management (/ICSSSM 06), Troyes (France).
7. S.Bonnevay, J.Champavere, M.Lamure. « Negotiation platform based on game theory ». IEEE International Conference on Research, Innovation and Vision for the future, Hanoi (Vietnam), pages 228-233, ISBN 1-4244-0694-3, March, 2007.
8. P.J.Lisboa, T.A.Etchells, I.H.Jarman, M.H.Aung, S.Chabaud, T.Bachelot, D.Perol, V.Bourdès, S.Bonnevay, S.Négrier. « Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer ». International Joint Conference on Neural Networks, Orlando (USA), pages 2533-2538, August, 2007.
9. G.Thibault, S.Bonnevay, A. Aussem. « Bayesian Network Structure Learning With Estimation of Distribution Algorithms ». IEEE International Conference on Digital Information Management, Lyon (France), pages 127-132, October, 2007.
10. V.Bourdès, S.Bonnevay, P.J.Lisboa, M.H.Aung, S.Chabaud, T.Bachelot, D.Perol, S.Négrier. « Breast Cancer Predictions by Neural Networks Analysis: a Comparison with Logistic Regression ». IEEE Engineering in Medicine and Biology Society, Lyon (France), pages 5424-5427, August, 2007.

## 9.4 Conférences nationales

1. Boubou M., Bounekkar A., Lamure M., Tounissoux D. Utilisation du recuit simulé pour la recherche d'une ultramétrie optimale. 12èmes rencontres de la SFC, Montréal, mai-juin 2005, pp. 71-74
2. A Rico, JP Auray. L'intégrale de Sugeno dans un treillis en vue d'une application pour la mesure de la qualité de vie. Dans Logique Floue et ses Applications, Nimes. 2007.

3. Cynthia Basileu – Ahmed Bounekkar – Nadia Kabachi – Michel Lamure – Jean-Michel Pigeon – Denis Roy. Modèle de diffusion spatiale d'une pandémie basé sur les flux migratoires quotidiens. International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
4. C. Blein, D. Clot, M. Lamure, Sévérité dans les systèmes de classification de patients, ICSSHC 2008 (International conference on system science in health care, Lyon FRANCE), Proceedings of icsshc 2008.
5. C. Blein, D. Clot, M. Lamure, Flux de patients et chaînage PMSI, ICSSHC 2008 (International Conference on System Science in Health Care, Lyon FRANCE), Proceedings of icsshc 2008
6. A. Kornfeld, S Bonnevey, M Lamure, M. Toumi. European Pricing Pilot Model - A decision support tool for optimisation of pricing and reimbursement of medicines in Europe. Dans International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
7. V. Bourdès, S Bonnevey, P.J. Lisboa, M.H. Aung, S. Chabaud, T. Bachelot, D. Perol, S. Négrier. Use of Neural Networks in Cancer Prediction: exploratory results from a French breast cancer cohort. Dans International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
8. Berthé Issa Bara, Ahmed Bounekkar, Gérard Duru. Problématique de réalisation d'un système d'information d'aide à la décision. International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
9. Ahmed Bounekkar, Michel Lamure. Modèle spatial pour la mesure de pollution International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
10. F. Maunoury, A Bounekkar, S Bonnevey, JP Auray. Le processus de décision markovien : Adaptation des routines. International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.
11. D. Clot, C. Blein, Recherche de profils d'activité médicale des établissements de santé MCO, ICSSHC 2008 (International Conference on System Science in Health Care, Lyon FRANCE) Proceedings of icsshc 2008
12. Marcel Egéa « Espaces préuniformes » ICSSHC Lyon 2008
13. Méndez-Hernandez, P., Siani, C., Lamure, M., Dosamantes-Carrasco, L.D., Marin-Martinez, F.H., Cano-Pérez, E.L., Salmeron-Castro, J., "Préférences déclarées sur trois programmes d'activité physique pour empêcher les maladies chroniques au Mexique", 9th Conference on System Science in Health Care, Proceedings of icsshc 2008.

14. de Peretti, C., Siani, C. and Duru, G., "Uncertainty around the Incremental Cost-Utility Ratio. Accounting for Mapping Interpolation in the QALYs", 9th Conference on System Science, in Health Care, Lyon, Proceedings of icsshc 2008.
15. S. BenAmor, S Bonnevey, M. Bui, M Lamure. Un modèle prétopologique stochastique pour la simulation de la pollution aérienne. Dans International Conference on Systems Science in Health Care, Lyon (France). 2008. Proceedings of icsshc 2008.

## 9.5 Conférences sans actes

1. Boubou M., Bounekkar A., Lamure M. Clustering Method based on aggregation of preferences. 3rd World Conference on Computational Statistics & Data Analysis, Limassol, Cyprus, 28-31 October 2005
2. G.Thibault, S.Bonnevey. « Recherche de structure d'un réseau Bayésien par métaheuristiques ». META, Hammamet (Tunisie), November, 2006.
3. Le T. V., Lamure M. Analyse de l'activité des établissements de soins basé sur l'approche prétopologique CALASS 2006, Milan, Septembre 2006.
4. Basileu C., Lamarsalle L., Lamure M., Vainchtock A., Analyse statistique de données de chainage à partir du PMSI Congrès CALASS'07, Marseille
5. Boubou M., Bounekkar A., Lamure M. Classification basée sur l'agrégation d'opinions par la méthode de recuit simulé. XIVèmes Rencontres de la Société Francophone de classification SFC 2007, ENST, Paris.
6. Bounekkar A., Lamure M., Li Y. Modélisation pour la gestion d'une crise sanitaire due à une pandémie Congrès CALASS'07, Marseille
7. A Bounekkar, M Lamure, J.M. PINGEON, D. ROY. Modélisation pour la gestion d'une crise sanitaire due à une pandémie. CALASS'07, Marseille FRANCE. 2007.
8. J.M. Cohen, A Bounekkar. An epidemiological study on noise in Paris area based on GPS practice : methods and preliminary results. 2007 WONCA EUROPE Conference, Paris. 2007.
9. L. D. Dosamantes, M Lamure, A Rico, P. Mendez Hernandez, JP Auray, J.C. Diaz-Montiel. Étude du lien entre les habitudes alimentaires et la variation du poids corporel dans la population urbaine du centre du Mexique. Dans L'évaluation de la santé, Marselleille, France. 2007.
10. G.Thibault, A.Aussem, S.Bonnevey. « Analyse critique des algorithmes EDA dans le cadre de l'apprentissage de structure de réseaux Bayésiens ». 4èmes Journées francophones des réseaux bayésiens, Lyon (France), Mai, 2008.
11. A Bounekkar. Spatial logistic regression based upon contiguity concept. ERCIM'08 - ERCIM Working Group on Computing & Statistics, Neuchatel, Switzerland. 2008.

12. C Wolf, G Gavin. Inference and parameter estimation on belief networks for image segmentation. Dans Journées Francophones sur les Réseaux Bayésiens, Lyon. 2008.

## **9.6 Chapitres d'ouvrage**

1. S Bonnevey. Réseaux Bayésiens. Economie de la santé, par A.Beresniak et G.Duru, Editions MASSON, pages 131-135, 185 et 195 2008.

## **9.7 Ouvrages et actes (Eds.)**

1. A. Bounekkar, A. Guinet Méthodologies pour la gestion des ressources hospitalières. Editions Hermès-Lavoisier, juin 2005, ISBN 978-2-7462-0892-6

2. I. Jars, N Kabachi, M. Lamure ,« Adaptative agents integration in a new environment: interactions as a source of learning, AAMAS, Utrecht. 2005.

3. A. Bounekkar. Systèmes d'informations en santé. Editions Hermès-Lavoisier, Janvier 2007, 2007, ISBN 978-2-7462-1912-0