

## **Chapitre 8**

### **Fouille de données complexes dans un processus d'extraction des connaissances**

L'atelier sur la fouille de données complexes dans un processus d'extraction de connaissances est organisé à l'instigation du groupe de travail « Fouilles de Données Complexes » (<http://morgon.univ-lyon2.fr/GT-FDC/>) et s'inscrit dans le cadre de la conférence EGC. La première édition de cet atelier avait pris place au sein d'EGC2004 à Clermont-Ferrand.

L'objectif de cet atelier est de dresser un panorama des travaux récents dans le domaine de la fouille de données complexes par des articles présentant un état de l'art ou des perspectives ouvertes dans ce domaine. Avec cette deuxième édition nous avons voulu reprendre et étendre cet objectif. Ainsi, en plus des communications longues ou courtes, nous avons proposé une session exposant des problématiques de recherche (position papers). Cette session a permis d'offrir un tour d'horizon sur des recherches récemment enclenchées dans le domaine de la fouille de données complexes.

Enfin, nous avons voulu que la grande variété de thèmes abordés par les papiers sélectionnés pour cette édition de l'atelier permette de couvrir un large spectre de ce que l'on peut entendre par « fouille de données complexes ». Cette diversité fut un élément constructif pour la réunion du groupe de travail qui s'est tenue à la fin de cette journée.

---

Responsables du chapitre : P. Gançarski, F. Masseglia

## **Le recueil d'articles proposé dans cet ouvrage**

Sélectionner les articles à paraître dans cet ouvrage était une tâche délicate en raison de la qualité des papiers retenus lors de l'atelier, mais aussi de leur grand nombre. Dans le présent ouvrage, nous avons proposé d'inscrire les articles qui ont été acceptés en « version longue » pour une communication orale afin d'offrir un spectre aussi large que possible sur les travaux actuels du domaine.

## **Quelques mots sur la fouille de données complexes**

Dans tous les domaines, les données à traiter pour y extraire de la connaissance utilisable sont de plus en plus complexes et volumineuses. Ainsi est-on amené à devoir manipuler des données :

- Souvent non structurées
- Issues de diverses provenances comme des capteurs ou sources physiques d'informations variées ;
- Représentant la même information à des dates différentes ;
- Regroupant différents types d'informations (images, textes) ou encore de différentes natures
- (logs, contenu de documents, ontologies, etc.).

De plus, la fouille de données complexes ne doit plus être considérée comme un processus isolé mais davantage comme une des étapes du processus plus général d'extraction de connaissances à partir de données (ECD). En effet, les difficultés liées à la complexité des données se répercutent sur toutes les étapes du processus d'ECD. Les thèmes liés à la fouille de données complexes peuvent donc comprendre :

- Pré-traitement, structuration et organisation des données ;
- Données manquantes ou erronées ;
- Fusion de données et de résultats ;
- Processus et méthodes de fouille de données complexes ;
- Utilisation des connaissances (domaine analysé ou domaine de l'analyste) en ECD complexe ;
- Post-traitement (tri, filtres, lisibilité) ;
- Visualisation et aide à l'interprétation des résultats ;
- Applications et retours d'expériences (Web, sciences du vivant).

Comité de programme et d'organisation :

Aufaure Marie-Aude (SUPELEC)  
Bouet Marinette (LIMOS)  
Boussaid Omar (ERIC)  
Briand Henri (IRIN)

Masseglia Florent (INRIA)  
Morin Annie (IRISA)  
Napoli Amedeo (LORIA)  
Nugier Sylvaine (EDF)

Djeraba Chabane (LIFL)  
Elfaouzi Nour-Eddin (INRETS)  
Fertil Bernard (IMED)  
Gallinari Patrick (LIP6)  
Gancarski Pierre (LSIIT)  
Hacid Mohand-Said (LIRIS)  
Hebrail Georges (ENST)  
Lebart Ludovic (CNRS & ENST)  
Lechevallier Yves (INRIA)  
Martin Arnaud (ENSIETA)

Petit Jean-Marc (LIMOS)  
Philipp-Foliguet Sylvie (ENSEA)  
Poncelet Pascal (LGI2P)  
Saidi-Glandus Alexandre (LIRIS)  
Teisseire Maguelonne (LIRMM)  
Trousse Brigitte (INRIA)  
Vrain Christel (LIFO)  
Wemmert Cedric (LSIIT)  
Zeitouni Karine (PRISM)  
Zighed Djamel (ERIC)

#### Remerciements

Les responsables de l'atelier tiennent à remercier :

- Les auteurs pour la qualité de leurs contributions.
- Les membres du comité de lecture pour leur travail indispensable à la qualité de cet atelier.
- Florence CLOPPET et Chabane DJERABA, responsables des ateliers pour EGC 2005.
- Nicole VINCENT présidente du comité d'organisation d'EGC 2005.
- Fabrice ROSSI : relecteur additionnel qui a travaillé dans des délais très courts.