



Atelier EGC2005 :
Qualité des Données et des Connaissances

18 Janvier 2005, Paris, France

Laure Berti-Equille, IRISA-CNRS, Rennes
Fabrice Guillet, LINA-CNRS, Nantes



Historique

- 1990 : travaux précurseurs de R. Gras et I.C. Lerman
- 2002-2003 : Groupe *GafoQualité* (R. Gras, F. Guillet)
 - Contexte de l'AS GafoDonnées STIC-CNRS (R. Cicchetti et M. Sebag), 5 groupes de travail
 - 30 membres, 10 laboratoires (Brest , Genève, Lyon, Nantes, Paris, Rennes)
 - Etude des propriétés des mesures, état de l'art
 - Un rapport d'activité debut 2003
- 2003-2004 : Numéro spécial revue RNTI Cépaduès,
« Mesures de qualité pour la fouille de données », Briand *et al. eds*
Recueil de travaux récents, publié en juin 2004.
- 2005 : Session Invitée «Quality in Data Mining » à la conférence ASMDA, Brest, 2005
- 2005 : Livre « Knowledge Quality in Data Mining », F. Guillet and H. Hamilton, editors
 - Appel à chapitres : mars 2005
 - date de parution : fin 2005
- Janvier 2005 : atelier DKQ ...

DKQ 2005

Atelier Qualité des Données et des Connaissances

associé à EGC 2005, 18 Janvier 2005, Paris, France

Motivations

- *Prolonger l'action GafoQualité*
- *Réunir*
 - *Communauté BD travaillant sur la qualité des données*
 - *Communauté EGC travaillant sur la qualité des connaissances*
 - *Opérationnels*

- *Favoriser les échanges et les savoir-faire*
- *Initier un RDV annuel (DKQ 2006?)*
- *Relayer l'information sur d'autres évènements relatifs à la qualité*
 - *2nd ACM SIGMOD workshop IQIS'05, 17 juin 2005, Baltimore*
 - *Conférence int. ICIQ, MIT, 4-6 Novembre 2005, Boston*



Programme de la journée

9h-9h30 : Accueil et ouverture par F. Guillet et L. Berti-Équille

9h30-10h30: Session 1 - Qualité des données dans les BD

- Data Freshness Evaluation in Different Application Scenarios, Verónika Peralta Mokrane Bouzeghoub (PRISM, Versailles St-Quentin)
- Nettoyage de données XML : combien ça coûte ? Laure Berti-Equille (IRISA, Rennes)

pause

10h45- 12h15 : Session 2 - Qualité des connaissances extraites sous forme de règles d'association

- Extraction de règles en incertain par la méthode implicative, Régis Gras, Raphaël Couturier, Fabrice Guillet, Filippo Spagnolo (École Polytechnique de Nantes, IUT de Belfort, Université de Palerme)
- *IPEE* : Indice Probabiliste d'Écart à l'Équilibre pour l'évaluation de la qualité des règles, Julien Blanchard, Fabrice Guillet, Henri Briand, Régis Gras (École Polytechnique de Nantes), Cyril Nortet
- Le rôle de l'utilisateur dans un processus d'extraction de règles d'association, Ansaf Salleb, Teddy Turmeaux, Christel Vrain (LIFO Orléans, IRISA Rennes)



Atelier Qualité des Données et des Connaissances

associé à ECG 2005, 18 Janvier 2005, Paris, France

14 h-15h : Session 3 – Qualité et classification

- Gilbert Ritschard (Université de Genève), Arbre BIC optimal et taux d'erreur
- Jérôme David, Fabrice Guillet, Vincent Philippé, Henri Briand, Régis Gras (Ecole Polytechnique de Nantes, PerformanSE SA), Validation d'une expertise textuelle par une méthode de classification basée sur l'intensité d'implication

15h- 16h : Session 4 - Plateformes d'évaluation de la qualité des connaissances

- Xuan-Hiep Huynh, Fabrice Guillet, Henri Briand, (Ecole Polytechnique de Nantes), ARQAT: une plateforme d'analyse exploratoire pour la qualité des règles d'association
- Benoît Vaillant, Patrick Meyer, Elie Prudhomme, Stéphane Lallich, Philippe Lenca (ENST Bretagne, Université du Luxembourg, ERIC - Université de Lyon 2), Mesurer l'intérêt des règles d'association

pause

16h15-17h45 : Session 5 - Pratiques opérationnelles

- Mireille Cosquer, Béatrice Le Vu, Alain Livartowski (Institut Curie), Mise en place d'un plan d'Assurance et Contrôle Qualité du Dossier Patient
- Gilles Amat, Brigitte Labois (BDQS), B.D.Q.S. Une gestion opérationnelle de la qualité de données
- David Graveleau (DGA/CTSN), SILURE, mise en oeuvre d'un méta-modèle associant traçabilité et qualité des données pour la constitution d'une base de référence multi-sources en veille technologique