

UE TAMD : Traitement et Analyse des Masses de Données

Master sciences et technologies
Mention Mathématiques-Informatique
Spécialité MatIS

Semestre : 2ème année, 1er semestre

Parcours : optionnelle en MIS, IMOI et SIRES

Volume horaire : 60h – **Crédits** : 6 ECTS

Intervenants : Laurent Amanton, Claude Duvallat, Bruno Sadeg, Moustafa Nakechbandi, Dominique Fournier

Objectifs Les applications actuelles deviennent de plus en plus complexes. Quelques aspects soulignant cette complexité sont la distribution des données et/ou des traitements, et les contraintes temporelles auxquelles ils sont soumis. Les systèmes d'information, notamment les SGBD, qui sont à la base de ces applications doivent donc tenir compte de ces nouvelles caractéristiques. Dans ce cours, nous nous intéressons particulièrement à la manière dont s'expriment ces contraintes temporelles, sur quoi elles portent, par quels mécanismes elles peuvent être respectées. Le cours de cette option a pour objectif de fournir aux étudiants les éléments de base pour poursuivre des recherches dans le cadre d'un doctorat. – Techniques de fouille de données.

Contenu de l'UE

- Systèmes d'information : caractéristiques
- Systèmes temps réel/SGBD temps réel
- Bases de données distribuées : partitionnement, validation des transactions
- Le Produit Oracle : mécanisme de répartition, gestion d'intégrité.
- Ordonnancement et ordonnançabilité
- Exemple de systèmes à temps contraint : BeeHive
- Prise en compte des contraintes temporelles : contrôle de concurrence distribué, les modèles de transactions étendus et validation des transactions distribuées
- Ordonnancement causal et COMMIT temps réel
- Qualité de service et Qualité des données dans les SGBD multimédia
- Problèmes ouverts
- Fouille de données :
 - 1. ECD : définitions, généralités, exemples d'application
 - 2. Pré-traitement des données
 - 3. Fouille de données
 - a. Introduction : apprentissage automatique
 - b. Fouille de données supervisée
 - induction d'arbres de décision
 - induction de règles
 - réseaux de neurones
 - c. Fouille de données non supervisée
 - "clustering" (cartes de Kohonen)
 - recherche d'associations
 - 4. Optimisation de la fouille de données
 - a. Evaluation des résultats de la fouille

- b. Gestion de l'accès aux bases de données
- 5. Visualisation
- 6. Systèmes d'ECD

Bibliographie

- A. Bestavros, K-J. Lin and S.H. Son, Real-Time Database Systems: Issues and Applications, Kluwer Academic Publishers, 1997.
- A. Bestavros and V. Wolfe eds., Real-Time Database and Information Research Advances, Kluwer Academic Publishers, 1997.
- C. Duvallet, Z. Mammeri et B. Sadeg, Les SGBD temps réel, revue Technique et Science Informatique (TSI), Vol. 18, Num. 5, pp. 479-517, 1999.
- K-Y. Lam and T-W. Kuo eds., Real-Time Database Systems: Architecture and Techniques, Kluwer Academic Publishers, 2001.
- K. Ramamritham, [S.H. Son](#), [L. Cingiser DiPippo](#): Real-Time Databases and Data Services. [Real-Time Systems 28](#)(2-3): 179-215 (2004)

Bibliographie fouille de données

- René Lefébure et Gilles Venturi, "Data Mining", Eyrolles, 2001.
- J. Han and M. Kamber, "Data mining - Concepts and Techniques", Morgan Kaufmann, 2001.
- Sites Internet : www.kdnuggets.com