

GROUPE	20 VALTCHEV, Petko	valtchev.petko@uqam.ca	(514) 987-3000 1919	PK-4415
Mardi, de 17h30 à 20h00				

DESCRIPTION Analyse ordinale de données et fondements théoriques en théorie des treillis et des ordres partiels, algèbre universelle et logique. Rapports avec la théorie des bases de données relationnelles. Constructions fondamentales de l'analyse formelle : contextes, concepts, treillis, implications. Algorithmique de la construction des treillis et des bases d'implications. L'analyse du panier de la ménagère et la problématique de la fouille de règles d'association. Approches classiques d'extraction de motifs fréquents et de règles d'association. Représentations compactes des motifs et des règles basées sur l'analyse formelle, propriétés structurelles. Algorithmique de la fouille de sous-familles de motifs fréquents et de bases de règles.

- OBJECTIFS**
- Le cours est une double introduction, d'une part, au paradigme d'analyse de données ordinales basée sur la construction de treillis et d'autre part, à la discipline de fouille de règles d'association qui est centrale et exemplaire du vaste domaine de la fouille de données.
 - Durant le cours, les étudiants seront confrontés à un large spectre de concepts, allant des constructions mathématiques fondamentales au développement d'outils efficaces de fouille de données et la fouille de données réelles à l'aide de ces outils, en passant par l'étude des propriétés desdites structures et la conception d'algorithmes qui les exploitent.
 - Les aptitudes développées dans le cadre du cours rendront l'étudiant capable :
 - de modéliser un problème d'analyse sous forme d'un ou plusieurs contextes formels, en appliquant les transformations appropriées,
 - d'apprécier les besoins algorithmiques d'un ensemble spécifique de données à analyser et d'effectuer un choix de méthode de fouille en conséquence,
 - de concevoir et réaliser, en cas de besoin, une méthode originale de fouille, adaptée à un profil de données particulier,
 - de façon plus fondamentale, d'appréhender les difficultés inhérentes, d'ordre combinatoire, que rencontrent les méthodes de fouille de motifs et de règles d'association.

ÉVALUATION	Description sommaire	Date	Pondération
	Résumés d'articles		3 X 10%
	Synthèse bibliographique avec présentation orale		40%
	Projet de conception et de réalisation d'une méthode de fouille		30%

CONTENU

Bases de données : rappels sur les dépendances fonctionnelles, formes normales, axiomes de Armstrong et couvertures non-redondantes et minimales d'une famille de dépendances fonctionnelles.

Théorie des treillis et des ordres partiels : ensembles ordonnés, pré-ordres, quasi-ordres et ordres partiels; treillis généraux, treillis booléens/distributifs/modulaires; représentation, éléments inf- et sup-irréductibles, idéaux d'ordre et anti-chaînes, complétion d'ordres partiels en treillis ; décomposition, congruences, produits direct et sous-direct de treillis ; correspondances de Galois, fonctions adjointes, opérateurs de fermeture et systèmes de fermeture, représentations minimales et correspondance avec les treillis.

Fondements de l'analyse formelle : contextes formels, opérateurs de dérivation et correspondance de Galois ; concepts formels, ordre de généralité et treillis de concepts ; théorème fondamental de l'AFC ; implications et fermeture implicative ; implications partielles ; familles remarquables d'ensembles associées à un jeu d'implications et bases non-redondantes d'implications, base de Duquenne-Guigues ; correspondance entre implications et contextes.

Algorithmique des treillis de concepts et des implications : construction du treillis, calcul de l'ensemble des concepts, l'algorithme NextClosure ; calcul de l'ordre de généralité entre concepts, l'algorithme de Bordat ; calcul de la base de Duquenne-Guigues ; maintien de structures, calcul incrémental des concepts et de l'ordre, algorithmes de Norris et de Godin ; fusion de structures de treillis.

Découverte de connaissances et fouille de données : analyse du panier de la ménagère et fouille de règles d'association; espaces de recherche des motifs fréquents et règles; approches classiques de la fouille de règles : algorithmes des familles Apriori et FP-growth; complexité du problème de la fouille de motifs; limites des approches classiques.

Représentations compactes des motifs et des règles : motifs fermés et maximaux ; générateurs minimaux; bases non-redondantes de règles d'association; algorithmes de fouille utilisant le cadre de l'analyse de concepts : AClose, Close, Closet, ChARM et Titanic.

RÉFÉRENCES

- VR B. A. Davey and H. A. Priestley – *Introduction to Lattices and Order* – Cambridge University Press, 2nd edition, 2002.
- VR B. Ganter and R. Wille – *Formal Concept Analysis, Mathematical Foundations* – Springer-Verlag, 1999.
- VR C. Carpineto and G. Romano – *Concept Data Analysis: Theory and Applications* – Wiley, 2004.
- VR J. Han and M. Kamber – *Data Mining : Concepts and Techniques* – Morgan Kaufmann, 2001.
- VR R. Godin – *Systèmes de gestion de bases de données par l'exemple* – Loze-Dion, 2003.
- VC K. Denecke, M. Ern , and S. Wismath, editors – *Galois connections and Applications* – Kluwer Academic Press, 2004.

A : article – C : comptes rendus – L : logiciel – N : notes – R : revue –
S : standard – U : uri – V : volume

C : compl mentaire – O : obligatoire – R : recommand 