

# Théorie et applications de la fouille d'associations

Groupe 40

Jeudi, de 13h00 à 16h00 SH-3120 (cours)

---

## Responsable(s) du cours

**Nom du coordonnateur :** VALTCHEV, Petko**Nom de l'enseignant :** VALTCHEV, Petko**Local :** PK-4415**Téléphone :** (514) 987-3000 #1919**Courriel :** valtchev.petko@uqam.ca**Site Web :** [http://intra.info.uqam.ca/Members/valtchev\\_p](http://intra.info.uqam.ca/Members/valtchev_p)

---

## Description du cours

La découverte d'associations est un aspect fondamental de la fouille de données. Ce cours met l'accent sur les bases théoriques de l'approche et sur les liens avec des problématiques de la théorie de la normalisation en bases de données, l'analyse formelle de concepts et les fonctions Booléennes. - Problème générique de découverte d'associations et de la fouille de motifs fréquents. - Variantes : motifs fermés, motifs maximaux, motifs clés ou générateurs. - Structures algébriques mises en jeu : treillis Booléen, classes d'équivalence, correspondances de Galois, treillis de concepts, contextes. - Approches de fouille de motifs : algorithmes par niveaux, algorithmes verticaux, algorithmes hybrides. - Représentations compactes pour les associations : base canonique, bases génériques et informatives. - Famille réduites de motifs : motifs indériverables, motifs delta-libres, motifs sans disjonction, motifs k-libres. - Applications de la fouille d'associations.

---

## Objectifs du cours

Ce cours vise à approfondir les connaissances de l'étudiant sur un domaine en pleine expansion qu'est la fouille de données. Le cours se focalise sur la découverte d'associations et de motifs fréquents qui est une discipline fondamentale de la fouille de données. L'accent est mis sur la présentation des diverses instanciations du problème général de la fouille ainsi que sur les fondements théoriques de l'approche et leurs liens avec des problématiques de la théorie de la normalisation en bases de données et l'analyse formelle de concepts.

Le but du cours est de permettre à l'étudiant de se familiariser avec la fouille de données en général à travers l'étude plus approfondie d'une des disciplines pertinentes, soit la fouille d'associations. Les objectifs concrets du cours peuvent être énoncés comme suit :

- d'approfondir la maîtrise de certains concepts fondamentaux en bases de données;
- de familiariser l'étudiant avec la démarche générale en fouille d'associations;
- de lui fournir des connaissances exploitables en conception de méthodes de fouille;
- de faire connaître à l'étudiant les plus récents développements dans le domaine;
- de permettre l'approfondissement d'un des thèmes de recherche dans le domaine;
- d'initier l'étudiant à la recherche à travers la rédaction d'un rapport sur un sujet d'actualité.

---

## Contenu du cours

---

Ce cours est destiné aux étudiants ayant suivi au préalable un ou des cours de bases de données au niveau baccalauréat et, de préférence, un cours d'introduction à l'intelligence artificielle. Les sujets abordés sont :

- Introduction à la problématique de la fouille de données
- Principales tâches de la fouille et solutions de principe
- Problème générique de la fouille d'associations et ses défis techniques
- Approche de référence pour l'extraction d'associations
- Approches alternatives et familles réduites de motifs et d'associations
- Ordres partiels et treillis et leur rôle en fouille de motifs
- Analyse de concepts comme cadre fondamental pour fouille de motifs
- Associations, implications, dépendances fonctionnelles
- Motifs et associations structurés : séquences, arbres, graphes
- Fouille en présence de connaissances du domaine : les motifs généralisés
- Principales applications des motifs et associations : médecine, recherche d'information, recommandation, modélisation de l'utilisateur.

## Modalités d'évaluation

Description sommaire	Date	Pondération
Résumé d'article de recherche 1	Semaine	15%
Résumé d'article de recherche 2	Semaine	15%
Travail de session : 1re partie - Proposition de sujet	Semaine	10%
Travail de session : 2e partie - Présentation orale	Semaine	25%
Travail de session : 3e partie - Mémoire sur le sujet choisi	Semaine	35%

Le travail de session est réalisé par groupe de deux étudiants. Une liste de sujets potentiels et des recommandations seront fournis pendant la session. La qualité du français constitue un critère d'évaluation (pour un maximum de 10%). En cas de retard dans la remise des travaux, une pénalité de 5% par jour ouvrable sera appliquée. Un retard de plus d'une semaine ne sera pas accepté.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, consultez le site suivant : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html>

### Politique d'absence aux examens

**L'autorisation de reprendre un examen en cas d'absence est de caractère exceptionnel. Pour obtenir un tel privilège, l'étudiant-e doit avoir des motifs sérieux et bien justifiés.**

Il est de la responsabilité de l'étudiant-e de ne pas s'inscrire à des cours qui sont en conflit d'horaire, tant en ce qui concerne les séances de cours ou d'exercices que les examens. **De tels conflits d'horaire ne constituent pas un motif justifiant une demande d'examen de reprise.**

Dans le cas d'une absence pour raison médicale, l'étudiant-e doit joindre un certificat médical original et signé par le médecin décrivant la raison de l'absence à l'examen. Les dates d'invalidité doivent être clairement indiquées sur le certificat. Une vérification de la validité du certificat pourrait être faite. Dans le cas d'une absence pour une raison non médicale, l'étudiant-e doit fournir les documents originaux expliquant et justifiant l'absence à l'examen &ndash; par exemple, lettre de la Cour en cas de participation à un jury, copie du certificat de décès en cas de décès d'un proche, etc. Toute demande incomplète sera refusée. Si la direction du programme d'études de l'étudiant-e constate qu'un étudiant a un comportement récurrent d'absence aux examens, l'étudiant-e peut se voir refuser une reprise d'examen.

L'étudiant-e absent-e lors d'un examen doit, dans les cinq (5) jours ouvrables suivant la date de l'examen, présenter une demande de reprise en utilisant le formulaire prévu, disponible sur le site Web du département à l'adresse suivante : <http://info.uqam.ca/politiques/>

L'étudiant-e doit déposer le formulaire dûment complété au secrétariat de la direction de son programme d'études : PK-3150 pour les programmes de premier cycle, PK-4150 pour les programmes de cycles supérieurs. Pour plus de détails sur la politique d'absence aux examens du Département d'informatique, consultez le site web suivant : <http://info.uqam.ca/politiques>

## Intégrité académique

### PLAGIAT Règlement no 18 sur les infractions de nature académique. (extraits)

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes;
- l'utilisation totale ou partielle du texte d'autrui en la faisant passer pour sien ou sans indication de référence;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant;
- l'obtention par vol, manoeuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche;
- Les sanctions liées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, veuillez consulter les sites suivants : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html> et <http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

## Médiagraphie

VC TAN, P.-N., STEINBACH, M., et KUMAR, V. -- *Introduction to Data Mining* -- **Pearson (2005)** <http://www-users.cs.umn.edu/~kumar/dmbook/>

VC GODIN, R. -- *Systèmes de gestion de bases de données par l'exemple* -- **Loze-Dion (2006)** <http://www.info2.uqam.ca/~godin/livreEd2.html>

VC HAN J. et KAMBER, M. -- *Data Mining : Concepts and Techniques, 2nd éd.* -- **Morgan Kaufmann (2006)**.

VC BERRY, M. et LINOFF, G. -- *Mastering Data Mining* -- **John Wiley & Sons (2000)**.

VC HAND, D., MANNILA, H. et SMYTH, P. -- *Principles of Data Mining* -- **MIT Press (2000)**.

VC CARPINETO, C. et ROMANO, G. &ndash; -- *Concept Data Analysis : Theory and Applications* -- **Wiley (2004)**.

VC GANTER, B. et WILLE R. &ndash; -- *Formal Concept Analysis, Mathematical Foundations.* -- **Springer-Verlag (1999)**.

AC FAYYAD, U., PIATETSKY-SHAPIRO, G. et SMYTH, P. &ndash; -- *From Data Mining to Knowledge Discovery in Databases* -- **AI Magazine 17(3): 37-54, 1996**.

AC AGRAWAL, R., IMIELINSKI, T. et SWAMI, A. -- *Mining association rules between sets of items in large databases* -- **Proceedings of the ACM SIGMOD International Conference on the Management of Data, Washington (DC), USA, pages 207&ndash;216, 1993.**

AC AGRAWAL, R., MANNILA, H., SRIKANT, R., TOIVONEN, H. et VERKAMO, A. -- *Fast Discovery of Association Rules* -- **U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, édts, Advances in Knowledge Discovery and Data Mining, pages 307&ndash;328. AAAI Press, Menlo Park (CA), USA, 1996.**

AC D'autres références (selon les sujets abordés) seront remises durant la session.

A : article - C : comptes rendus - L : logiciel  
S: Standard - U : uri - V : volume

C : complémentaire - O : Obligatoire - R : recommandé