

Séminaire de maîtrise en informatique II

Groupe 30

Jeudi, de 13h30 à 16h30 PK-4360 (cours)

Responsable(s) du cours

Nom du coordonnateur : SADAT, Fatiha**Nom de l'enseignant :** SADAT, Fatiha**Local :** PK-4625**Téléphone :** (514) 987-3000 #3885**Courriel :** sadat.fatiha@uqam.ca

Description du cours

Ce séminaire a pour but de favoriser l'accès à divers domaines spécialisés ou nouveaux en informatique ainsi que l'initiation à la recherche.

Objectifs du cours

Ce cours apportera à l'étudiant les bases théoriques nécessaires à la compréhension des objectifs et des défis actuels du traitement automatique du langage naturel (TALN) pour le Web et plus spécifiquement la recherche d'information sous tous ses aspects, monolingue, multilingue, etc.

Ainsi, ce cours tentera de présenter des problématiques de base du TALN à travers quelques sujets et de discuter des techniques proposées pour les résoudre.

Contenu du cours

1. Introduction au traitement automatique du langage naturel;
2. Recherche d'information (RI) et fouille de textes (Information Retrieval);
3. Recherche d'information translinguistique (RIT) - en Ang. Cross-Language Information Retrieval: problèmes théoriques, approches, évaluations;
4. Campagnes d'évaluations en RI et RIT
5. Traduction automatique : problèmes théoriques, différentes approches;
4. Pré-traitement, analyse et outils linguistiques;
5. Corpus parallèles vs. comparables : étude de cas;
6. Le TALN pour les données massives (*big data*)

Modalités d'évaluation

Description	Pondération
-------------	-------------

Projet et présentation du projet	60%
Présentation d'un article de recherche	30%
Participation en classe	10%

- 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 – 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, constitue 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 reliées 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

- Gregory Grefenstette: Cross-Language Information Retrieval. Springer; édition Mar 31 1998.
- M. Boughanem, J. Savoy: Recherche d'information : Etat des lieux et perspectives. Lavoisier, Paris, 2008.
- Christopher D. Manning, Hinrich Schütze: Foundations of Statistical Natural Language Processing. Disponible sur le web : <http://www-nlp.stanford.edu/fsnlp/>
- Manning Christopher D., Raghavan Prabhakar, and Schütze Hinrich: Introduction to Information Retrieval.
- Ricardo Baeza-Yates, Berthier Ribeiro-Neto : Modern Information Retrieval ,
- C. J. van RIJSBERGEN: Information Retrieval. Butterworths, London (UK), 1979. (<http://www.iro.umontreal.ca/~nie/IR-book/Preface.html>)
- (<http://www.dcs.gla.ac.uk/Keith/Preface.html>)
- G. Salton: Automatic Text Processing. Addison-Wesley, Reading (MA), 1989.
- W. B. Frakes, R. Baeza-Yates: Information Retrieval: Data Structures & Algorithms. Prentice Hall, Englewood Cliffs (NJ), 1992.
- Ayse Goker , John Davies: Information Retrieval: Searching in the 21st Century.
- Mitkov, R. 2005: The Oxford handbook of computational linguistics. Oxford : Oxford University Press.
- Salton, G. et McGill, M. 1983 : Introduction to Modern Information Retrieval. New-York: McGraw-Hill.
- D. Grossman and O. Frieder: Information Retrieval - Algorithms and Heuristics. Springer, 2004 (second edition).
- FUCHS C.: Linguistique et traitements automatiques des langues, 1993.
- (<http://cat.inist.fr/?aModele=afficheN&cpsid=3726134>)
- Gerald Gazdar, Chris Mellish: Natural Language Processing in Prolog. An Introduction to Computational Linguistics. Addison-Wesley, Workingham (UK).
- Pierre Nugues: An Introduction to Language Processing with Perl and Prolog. Springer, Berlin, 2006.
- N. Indurkha, F. Damerou: Handbook of Natural Language Processing. Chapman & Hall/CRC Press, 2nd Ed., Boca Raton, 2010
- Erik Hatcher, Otis Gospodnetic: Lucene in Action. Manning.
- Ian H. Witten, Alistair Moffat, Timothy C. Bell: Managing Gigabytes. Compressing and Indexing Documents and Images. Morgan Kaufmann, 1989.
- Daniel Jurafsky and James H. Martin : SPEECH and LANGUAGE PROCESSING - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition 2009 (<http://www.cs.colorado.edu/~martin/slp2.html>)
- Erik Hatcher, Otis Gospodnetic: Lucene in Action. Manning.
- Ian H. Witten, Alistair Moffat, Timothy C. Bell: Managing Gigabytes. Compressing and Indexing Documents and Images. Morgan Kaufmann, 1989.