

GROUPE	20 SADAT, Fatiha	sadat.fatiha@uqam.ca	(514) 987-3000 3885	PK-4625
Vendredi, de 14h00 à 17h00				

**DESCRIPTION** Systèmes de représentation et de traitement des langues naturelles: traitement et reconnaissance de la parole. Grammaires et théories de parsing syntaxique. Représentation et traitement sémantique: réseaux sémantiques et graphes conceptuels. Pragmatique et communication homme-machine. Apprentissage des langues naturelles et modèles connexionnistes.

**OBJECTIF** 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). Il a pour objectif d'étudier quelques concepts et techniques pour le traitement automatique du langage naturel; dont l'extraction et l'organisation automatiques d'informations, l'analyse de corpus documentaires, la traduction automatique statistique, l'utilités des corpus de documents textuels, etc.

ÉVALUATION	Description sommaire	Date	Pondération
	Projets 1 et 2		60%
	Présentation orale par rapport au sujet d'intérêt		40%

### Contenu attendu

Recherche de fond sur un aspect de la recherche en psycholinguistique en lien avec le traitement des langues naturelles : présentation d'un état de la question organisé; discussion; perspective critique sur les travaux étudiés. (Réflexion conceptuelle sur le sujet choisi, et pas travail empirique de programmation).

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 : SH-4700 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>

CONTENU	<ol style="list-style-type: none"> <li>1. Introduction à la recherche d'information et fouille de textes (Information Retrieval);</li> <li>2. Recherche d'information translinguistique (Cross-Language Information Retrieval) : problèmes théoriques, approches;</li> <li>3. Traduction automatique (statistique) : problèmes théoriques, approches;</li> <li>4. Pré-traitement et étude morphologique de langue;</li> <li>5. Corpus parallèles vs. comparables : étude de cas;</li> <li>6. Extraction de lexique, de terminologie bilingue et d'entités nommées.</li> </ol>
---------	---

### Modalités

- ❑ Le cours sera basé sur la discussion de textes et d'articles scientifiques. Le choix final des articles sera déterminé au début de la session en fonction des profils des étudiants;
- ❑ Les étudiants devront préparer les lectures pour le cours et être à même de les présenter en classe et de les discuter. La participation des étudiants est essentielle. Le professeur n'assumera pas la tâche de faire des exposés magistraux;
- ❑ Les étudiants choisiront dès le début du cours un thème de recherche pertinent au contenu du cours et devront présenter, à intervalles réguliers, un ou plusieurs article(s) développant(s) ce thème.

## RÉFÉRENCES

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

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

C : complémentaire – O : obligatoire – R : recommandé