

GROUPE	30 SADAT, Fatiha	sadat.fatiha@uqam.ca	(514) 987-3000 3885	PK-4625
Mercredi, de 13h00 à 16h00				

**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
	Participation en classe		15%
	Deux travaux pratiques (travail d'équipe de deux personnes ou individuel)		30%
	Une présentation en classe (présentation individuelle)		15%
	Examen final		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, veuillez consulter les sites suivants :

<http://www.sciences.uqam.ca/decanat/reglements.php>

<http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

### Politique d'absence aux examens

Un étudiant absent à un examen se verra normalement attribuer la note zéro pour cet examen. Cependant, si l'étudiant était dans l'impossibilité de se présenter à l'examen pour un motif valable, certains arrangements pourront être pris avec son enseignant. Pour ce faire, l'étudiant devra présenter à son enseignant l'un des formulaires prévus à cet effet accompagné des pièces justificatives appropriées (par ex., attestation d'un médecin que l'étudiant était dans l'impossibilité de se présenter à l'examen pour des raisons de santé, lettre de la Cour en cas de participation à un jury).

Une absence pour cause de conflit d'horaires d'examen n'est pas considérée comme un motif valable d'absence, à moins d'entente préalable avec la direction du programme et l'enseignant durant la période d'annulation des inscriptions avec remboursement : tel qu'indiqué dans le guide d'inscription des étudiants, il est de la responsabilité d'un étudiant de ne s'inscrire qu'à des cours qui ne sont pas en conflit d'horaire.

Pour plus de détails sur la politique d'absence aux examens du Département d'informatique et pour obtenir les formulaires appropriés, consultez le site web suivant :

<http://www.info.uqam.ca/enseignement/politiques/absence-examen>

- CONTENU**
1. Introduction à la recherche d'information et fouille de textes (Information Retrieval);
  2. Recherche d'information translinguistique (Cross-Language Information Retrieval) : problèmes théoriques, approches;
  3. Traduction automatique (statistique) : problèmes théoriques, approches;
  4. Pré-traitement et étude morphologique de langue;
  5. Corpus parallèles vs. comparables : étude de cas;
  6. Extraction de lexique, de terminologie bilingue et d'entités nommées.

### 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 Schütze – *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&cpsidt=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