
INF600H

Traitement automatique du langage naturel

Plan de cours

Responsable(s) du cours

Coordination : SADAT, Fatiha
PK-4625
sadat.fatiha@uqam.ca
<https://sites.google.com/site/fatihasadat/home/>

Enseignement :

MALENFANT, Bruno
PK-4115
malenfant.bruno@uqam.ca
Groupes : 020

Description du cours

Objectifs

Ce cours présente les fondements du traitement automatique de la langue naturelle (abr. TALN) ou des langues (abr. TAL), une discipline de l'intelligence artificielle, se trouvant à la frontière entre l'informatique, les sciences cognitives et la linguistique.

Sommaire du contenu

Principes fondamentaux du traitement automatique du langage naturel (TALN). Architecture typique d'une application de TALN. Pré-traitement et normalisation de textes (corpus). Analyse lexicale (tokenisation), lemmatisation, racinisation (stemming). Représentation de texte : sac de mots, N-Gramme, TF-IDF, plongements lexicaux (word embeddings). Méthodes pour l'analyse syntaxique et l'analyse sémantique de textes. Techniques de classification, de recherche et d'extraction d'information. Modèle de langage. Modèles de Markov. Applications : extraction et recherche d'information, traduction automatique, agent conversationnel, analyse des sentiments, détection d'émotions, synthèse de résumé de texte. Les outils du TALN seront présentés selon les applications étudiées.

Modalité d'enseignement

Ce cours comporte une séance obligatoire de laboratoire (2 heures).

Préalables académiques

- INF3105 - Structures de données et algorithmes
- MAT4681 - Statistique pour les sciences

Formule pédagogique

Les exposés magistraux (3h) et les séances de laboratoires (2h) se dérouleront de façon synchrone avec le logiciel Zoom. Ils seront enregistrés dans la mesure du possible et mis à disposition des étudiants inscrits au cours via la page Moodle.

Contenu détaillé

Semaine 1

- Historique du TALN et présentation des applications
- Rappel : outils mathématiques de base (vecteur, probabilité)

Semaine 2

- Corpus (acquisition, annotation : un peu de XML)
- Lexiques et dictionnaires
- Prétraitement : Présentation sommaire des expressions régulières en Java

Semaine 3

- Prétraitement (suite)
- Tokenisation des mots (distance d'édition)
- Séparation des phrases
- Lemmatisation et racinisation
- Introduction à la désambiguïsation

Semaine 4

- Présentation du premier devoir

- Modèles de mots
- TF-IDF (exemple : recherche de document)
- Contexte sémantique
- Sac de mots (espace de vecteur, plongement lexical)

Semaine 5

- Modèles de mots (suite)
- N-gram (collocation)
- Mesure de similarité cosinus (homonymie, synonymie)

Semaine 6

- Modèle de suite de mots
- Modèle de Markov (lissage)
- Partie du discours (POS)

Semaine 7

- Modèle de suite de mots (suite)
- Arbre syntaxique (analyse syntaxique superficielle)
- Classification et types d'algorithmes
- Génération de texte

Semaine 8

- Examen intra

Semaine 9

- Extraction d'information
- Reconnaissance d'entités nommées

Semaine 10

- Extraction d'information (suite)
- Détection de relations et d'événements
- Rappel, précision et f-mesure

Semaine 11

- Présentation du deuxième devoir
- Système de questions-réponses

Semaine 12

- Génération automatique de résumé
- Métrique ROUGE

Semaine 13

- Traduction automatique
- Métrique BLEU

Semaine 14

- Autre sujet à choisir, s'il reste du temps

Semaine 15

- Examen final

Modalités d'évaluation

Outil d'évaluation	Pondération	Échéance
TP1	16%	22 juin 2021
TP2	16%	10 août 2021
Examen Intra	32%	29 juin 2021
Examen Final	32%	17 août 2021
Mini quiz1	2%	1 juin 2021
Mini quiz2	2%	20 juillet 2021

Remarques

Les travaux pratiques peuvent être réalisés seul ou en équipe. La qualité du français constitue un critère d'évaluation (pour un maximum de 10%).

L'étudiant doit obtenir une moyenne cumulée de 50% aux examens ; dans le cas contraire, il se verra attribuer un échec.

L'étudiant ne remettant aucun des travaux se verra attribuer un échec.

Deux mini Quiz seront offerts afin de donner une évaluation plus continue aux étudiants. Un mini quiz est une question ou 2 sur Moodle, de type choix de réponse ou avec réponse simple. (20 minutes)

Médiagraphie

Recommandé

- Natural Language Processing in Action [<https://livebook.manning.com/book/natural-language-processing-in-action/about-this-book/>]
- Practical Natural Language Processing [<https://www.oreilly.com/library/view/practical-natural-language/9781492054047/>] [<https://github.com/practical-nlp/practical-nlp>]
- Chris Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press. Cambridge, MA : May 1999.
- Daniel Jurafsky and James H. Martin. Speech and Language Processing. Pearson Prentice Hall, 2009.
- François-Régis Chaumartin et Pirmin Lemberger. Traitement automatique des langues. Comprendre les textes grâce à l'intelligence artificielle. Collection InfoPro, Dunod. Fev. 2020.
- Patrice Enjalbert. Sémantique et traitement automatique du langage naturel. Hermes Science Publications, 2005.

Ressources complémentaires

- Cours ailleurs :
- Stanford / CS 224N / Ling 284 — Natural Language Processing (ancienne version) [<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1162/syllabus.shtml>]
- Stanford / CS224n : Natural Language Processing with Deep Learning <http://web.stanford.edu/class/cs224n/>
- UAlberta [<https://www.ualberta.ca/computing-science/graduate-studies/course-directory/courses/index.html>] :
- CMPUT 501 Introduction to Natural Language Processing
- CMPUT 650 Natural Language Processing
- CMPUT 651 Deep Learning for Natural Language Processing

Participation à un cours ou à une activité d'enseignement en ligne

- Lors d'un cours ou d'une activité d'enseignement en ligne, le personnel enseignant peut décider, selon le cas, de procéder à l'enregistrement audio ou audiovisuel du cours ou de l'activité d'enseignement. Le personnel enseignant peut partager l'enregistrement uniquement à son groupe-cours.
- En cas d'enregistrement, l'étudiante, l'étudiant sera informé au début de la séance.
- Il est de la responsabilité de l'étudiante, de l'étudiant de désactiver son microphone et/ou sa caméra s'il ne souhaite pas être enregistré.
- À défaut de désactiver son microphone et/ou sa caméra, l'étudiante, l'étudiant consent à l'enregistrement audio ou audiovisuel, à la conservation, à la rediffusion et à l'utilisation de l'enregistrement de son nom, de sa voix et de son image dans le cadre du cours ou de l'activité en ligne. L'étudiante, l'étudiant reconnaît ne détenir aucun droit dans l'enregistrement.
- Sauf avec l'autorisation expresse écrite du personnel enseignant, il est interdit de reproduire, d'enregistrer, de publier, de diffuser, de communiquer ou de partager, par quelque moyen que ce soit, tout ou partie de l'enregistrement d'un cours ou d'une activité d'enseignement en ligne de même que tout matériel pédagogique s'y rattachant.
- Une étudiante, un étudiant qui contrevient à ce qui précède s'expose aux sanctions prévues dans les règlements et politiques de l'UQAM ou à tout recours légal, notamment en vertu de la Loi sur le droit d'auteur.

Monitorat de programme

Le département d'informatique offre un service de monitorat gratuit s'adressant plus particulièrement aux étudiant.e.s du baccalauréat et du certificat en informatique. Il concerne principalement les cours de base comme INF1070, INF1120, INF1132, INF2120 et INF2171, mais, selon la connaissance du moniteur ou de la monitrice, un support dans d'autres cours peut également être offert.

Objectifs. Permettre aux étudiant.e.s de :

- Bénéficier d'un encadrement par les pairs ;
- Recevoir un suivi personnalisé en cas de difficulté ;
- Profiter d'un soutien supplémentaire à la matière vue en classe ;
- Obtenir un support technique sur les technologies, les outils, les bibliothèques et les logiciels utilisés dans les cours (installation, configuration, utilisation)

Informations.

- Voir <https://info.uqam.ca/aide/> pour la grille horaire et tous les détails
- Le service est généralement disponible à partir de la deuxième semaine
- D'autres plages horaires pourraient être ajoutées en cours de session selon les besoins
- Clavardage en direct : `~monitorat-de-programme` (Mattermost)

Politique d'absence aux examens

Reprise d'examen. 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.

Conflits d'horaire. 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.**

Procédure. 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 <http://info.uqam.ca/repriseexamen/>.

Pièces justificatives. 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.e étudiant.e a un comportement récurrent d'absence aux examens, l'étudiant.e peut se voir refuser une reprise d'examen.

Pour plus d'informations. Consulter la page <http://info.uqam.ca/politiques>.

Règlement numéro 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 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 :

- <http://www.infosphere.uqam.ca/rediger-un-travail/eviter-plagiat>
- <http://r18.uqam.ca/>

Politique no 16 visant à prévenir et combattre le sexisme et les violences à caractère sexuel

Les violences à caractère sexuel se définissent comme étant des comportements, propos et attitudes à caractère sexuel non consentis ou non désirés, avec ou sans contact physique, incluant ceux exercés ou exprimés par un moyen technologique, tels les médias sociaux ou autres médias numériques. Les violences à caractère sexuel peuvent se manifester par un geste unique ou s'inscrire dans un continuum de manifestations et peuvent comprendre la manipulation, l'intimidation, le chantage, la menace implicite ou explicite, la contrainte ou l'usage de force.

Les violences à caractère sexuel incluent, notamment :

- la production ou la diffusion d'images ou de vidéos sexuelles explicites et dégradantes, sans motif pédagogique, de recherche, de création ou d'autres fins publiques légitimes ;
- les avances verbales ou propositions insistantes à caractère sexuel non désirées ;
- la manifestation abusive et non désirée d'intérêt amoureux ou sexuel ;
- les commentaires, les allusions, les plaisanteries, les interpellations ou les insultes à caractère sexuel, devant ou en l'absence de la personne visée ;
- les actes de voyeurisme ou d'exhibitionnisme ;
- le (cyber) harcèlement sexuel ;
- la production, la possession ou la diffusion d'images ou de vidéos sexuelles d'une personne sans son consentement ;
- les avances non verbales, telles que les avances physiques, les attouchements, les frôlements, les pincements, les baisers non désirés ;
- l'agression sexuelle ou la menace d'agression sexuelle ;
- l'imposition d'une intimité sexuelle non voulue ;
- les promesses de récompense ou les menaces de représailles, implicites ou explicites, liées à la satisfaction ou à la non-satisfaction d'une demande à caractère sexuel.

Pour consulter la politique no 16

https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_16.pdf

Pour obtenir de l'aide, faire une divulgation ou une plainte

Bureau d'intervention et de prévention en matière de harcèlement
514-987-3000, poste 0886

Pour obtenir la liste des services offerts à l'UQAM et à l'extérieur de l'UQAM

<https://harcelement.uqam.ca>

Soutien psychologique (Services à la vie étudiante)

514-987-3185
Local DS-2110

CALACS Trêve pour Elles – point de services UQAM

514 987-0348
calacs@uqam.ca
<http://trevepourelles.org>

Service de la prévention et de la sécurité

514-987-3131

Politique no 44 d'accueil et de soutien des étudiant.e.s en situation de handicap

Politique. Par sa politique, l'Université reconnaît, en toute égalité des chances, sans discrimination ni privilège, aux étudiant.e.s en situation de handicap, le droit de bénéficier de l'ensemble des ressources du campus et de la communauté universitaire, afin d'assurer la réussite de leurs projets d'études, et ce, dans les meilleures conditions possibles. L'exercice de ce droit est, par ailleurs, tributaire du cadre réglementaire régissant l'ensemble des activités de l'Université.

Responsabilité de l'étudiant.e. Il incombe aux étudiant.e.s en situation de handicap de rencontrer les intervenant.e.s (conseiller.ère.s à l'accueil et à l'intégration du Service d'accueil et de soutien des étudiant.e.s en situation de handicap, professeur.e.s, chargé.e.s de cours, direction de programmes, associations étudiantes concernées, etc.) qui pourront faciliter leur intégration à la communauté universitaire ou les assister et les soutenir dans la résolution de problèmes particuliers en lien avec les limitations entraînées par leur déficience.

Service d'accueil et de soutien aux étudiant.e.s en situation de handicap. Le Service d'accueil et de soutien aux étudiant.e.s en situation de handicap (SASESH) offre des mesures d'aménagement dont peuvent bénéficier certains étudiant.e.s. Il est fortement recommandé aux de se prévaloir de ces services afin de réussir ses études, sans discrimination. Pour plus d'information, visiter le site de ce service : <https://vie-etudiante.uqam.ca/etudiant-situation-handicap/nouvelles-ressources.html> et celui de la politique institutionnelle d'accueil et de soutien aux étudiant.e.s en situation de handicap : https://instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_44.pdf

Il est important d'informer le SASESH de votre situation le plus tôt possible :

- En personne : 1290, rue Saint-Denis, Pavillon Saint-Denis, local AB-2300
- Par téléphone : 514 987-3148
- Par courriel : situation.handicap@uqam.ca
- En ligne : <https://vie-etudiante.uqam.ca/>