

Sigle : INF5153 Gr. 01

Titre : Traitement automatique du langage naturel

Session : Hiver 2025 Horaire et local

Professeur : Davoust, Alan et Allili, Mohand Said

1. Description du cours paraissant à l'annuaire :

Objectifs

Connaitre les principales problématiques liées au traitement automatisé des langages naturels. Savoir appliquer les techniques de pré-traitement nécessaires à l'utilisation de données textuelles dans les systèmes intelligents.

Contenu

Traitement linguistique: morphologie et syntaxe, parsing et étiquetage grammatical. Identification d'entités nommées, collocations de mots, coréférences. Introduction aux modèles statistiques des langues. Modèles sémantiques : analyse sémantique latente, vectorisation de mots. Apprentissage machine sur données textuelles, utilisation de réseaux de neurones. Applications : recherche d'informations, analyse de sentiments et d'émotions, traduction automatique, synthèse d'annotations et de texte, systèmes conversationnels et de questions-réponses.

Descriptif – Annuaire

2. Stratégies pédagogiques :

- La matière sera principalement dispensée à travers des cours magistraux (3h / semaine, en présentiel) comprenant des activités interactives : la participation des étudiant.e.s sera fortement sollicitée
- Lectures et étude autonomes : afin de pouvoir consacrer plus de temps à des activités interactives, il est important que les étudiant.e.s préparent chaque séance de cours en utilisant les ressources fournies
- Devoirs et projet : les devoirs et projet sont des instruments d'apprentissage plus que d'évaluation. Le but est de permettre aux étudiants d'approfondir et de mettre en pratique certaines parties du cours, selon leurs intérêts.

4. Heures de disponibilité ou modalités pour rendez-vous :

Sur rendez-vous : mohandsaid.allili@uqo.ca, alan.davoust@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

		Dates
1	<p>Introduction au cours</p> <ul style="list-style-type: none"> • Introduction au traitement du langage naturel (TLN). TLN vs. Linguistique. • Présentation des principales tâches techniques (parsing, étiquetage grammatical, reconnaissance d'entités nommées, etc.) et applications (transcription, traduction, classification, génération, etc.). 	16 janv. 2025
2	<p>Mots</p> <ul style="list-style-type: none"> • Éléments de linguistique: Phonologie, Morphologie. • Corpus de textes, propriétés statistiques des langues. Collocations. • Découpage des mots et des phrases, utilisation d'expressions régulières. Lemmatisation. <p>Travail pratique I (Utilisation de bibliothèques NLP de Python)</p>	23 janv. 2025
3	<p>Phrases</p> <ul style="list-style-type: none"> • Éléments de linguistique: syntaxe, grammaires. • Parsing et étiquetage grammatical. 	30 janv. 2025
4	<p>Modèles de langage</p> <ul style="list-style-type: none"> • Rappels de probabilité. • Modèles de langage n-gram. 	06 fév. 2025

5	Classification de textes <ul style="list-style-type: none"> • Modèles <i>bag-of-words</i>, séquences, tf-idf, caractéristiques linguistiques. • Classification de texte par apprentissage automatique. • Application: analyse de sentiments. Travail pratique II (Classification de texte par le modèle n-gram)	13 fév. 2025
6	Sémantique <ul style="list-style-type: none"> • Éléments de linguistique: sémantique • Sémantique logique • Plongements de mots 	20 fév. 2025
7	Classification de textes (2) <ul style="list-style-type: none"> • Apprentissage profond pour les données textuelles : LSTM, CNN, Transformeurs. Travail pratique III (Classification de texte par apprentissage profond)	27 fév. 2025
8	Semaine d'études	06 mars 2025
9	Applications <ul style="list-style-type: none"> • Traduction automatique • Systèmes de dialogue (chatbots) 	13 mars 2025
10	Applications <ul style="list-style-type: none"> • Recherche d'informations • Analyse thématique • Indexation sémantique Travail pratique IV : (Analyse thématique de documents)	20 mars 2025
11	Modèles de langage (2) <ul style="list-style-type: none"> • Modèles de langage neuronaux 	27 mars 2025
12	Modèles de langage (3) <ul style="list-style-type: none"> • Modèles de langage neuronaux avancés pour la génération de textes • Travail pratique V : (Génération de textes simples)	03 avr. 2025
13	Modèles de langage (4) <ul style="list-style-type: none"> • LLM, ingénierie des <i>prompt</i> • Retrieval-Augmented Generation (RAG) 	10 avr. 2025
14	Examen	17 avr. 2025
15	Présentations des projets	24 avr. 2025

6. Évaluation du cours :

- Examen final : 40%
- Projet : 40%
- Devoirs : 20%

• 7. Politiques départementales et institutionnelles :

- Politiques relatives à la tenue des examens
- Note sur le plagiat et les fraudes
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

Tolérance **ZÉRO** en matière de violence à caractère sexuel.

Le Bureau d'intervention et de prévention en matière de harcèlement (BIPH) a pour mission d'accueillir, soutenir et guider toute personne vivant une situation de harcèlement, de discrimination ou de violence à caractère sexuel. Le BIPH oriente ses actions afin de prévenir les violences à caractère sexuel pour que nous puissions étudier, travailler et s'épanouir dans un milieu sain et sécuritaire.

Vous vivez ou êtes une personne témoin d'une situation de violence à caractère sexuel ? Vous êtes une personne membre de la communauté étudiante ou une personne membre du personnel, autant à Gatineau qu'à Ripon et St-Jérôme, l'équipe du BIPH est là pour vous, sans jugement et en toute confidentialité.

Ensemble, participons à une culture de respect.

Pour de plus amples renseignements consultez [UQO.ca/biph](https://uqo.ca/biph) ou écrivez-nous au Biph@uqo.ca

8. Principales références :

9. Page Web du cours :

<https://moodle.uqo.ca>