

Traitement Automatique des Langues et Littérature
Natural Language Processing and Literature

Responsable du cours : Pascal Amsili

Autre(s) enseignante(s) / enseignant(s) : Jean Barré, Mathieu Dehouck

Descriptif du cours :

- Introduction à la discipline du traitement automatique des langues (TAL), qui se concentre sur la manipulation informatique du langage humain (texte, parole), avec des méthodes symboliques (formelles) mais surtout statistiques et neuronales.
- Présentation des outils élaborés pour le traitement des données textuelles (segmentation, étiquetage, analyse syntaxique), et des méthodes utilisées pour les tâches courantes dans le domaine (traduction automatique, recherche d'information, résumé automatique...).
- Introduction aux études littéraires computationnelles, discipline récente qui s'est construite avec la disponibilité de vastes corpus numérisés (BNF, HathiTrust...) et des méthodes d'IA performantes.
- Utilisation des méthodes du TAL pour l'opérationnalisation des concepts de la théorie littéraire, ce qui donne la possibilité de mener des études empiriques à large échelle.

Objectifs pédagogiques et compétences développées :

- Connaissance des applications et méthodes du domaine du TAL, et capacité d'appliquer et de développer ces méthodes en python.

Contenu détaillé du cours :

- 1 - Introduction aux études littéraires computationnelles
- 2 - Introduction au traitement automatique des langues
- 3 - Normalisation de textes
- 4 - Modèles de langue simples
- 5 - Distances et comparaisons de documents
- 6 - Sémantique distributionnelle
- 7 - Plongements Lexicaux
- 8 - Classification (1)
- 9 - Classification (2)
- 10 - Étiquetage en parties-du-discours et catégories morphosyntaxiques
- 11 - Apprentissage non supervisé
- 12 - Perceptron, réseaux neuronaux récurrents, convolution, seq2seq
- 13 - Attention et architecture transformers
- 14 - Modèles de langue
- 15 - Examen

Langue d'enseignement : Français / ~~Anglais~~

Type de cours : Cours magistral & TP

Modalités d'évaluation :

- Examen Final - 50 %
- TP noté - 25 %
- Mini Projet - 25 %

Année : L2

Semestre : Semestre 1

Lectures obligatoires : /

Lectures recommandées : Daniel Jurafsky and James H. Martin. 2025. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 24, 2025. <https://web.stanford.edu/~jurafsky/slp3>.