

UE Fouille de texte



- > **Langue(s) d'enseignement:** Français
- > **Ouvert aux étudiants en échange:** Non

Présentation

Description

Dans ce cours, nous abordons les problèmes et les méthodes d'accès à l'information disponible dans les textes, qui constituent des données non structurées. La première partie du cours est une introduction générale aux tâches de fouilles de textes et aux architectures neuronales utilisées pour ces tâches : MLP, RNN (en particulier les LSTM), et Transformers. Nous verrons les techniques de représentation vectorielle des éléments textuels (mots, phrases, paragraphes, etc.), en particulier les plongements lexicaux statiques (word2vec ou glove) ou contextuels (par exemple BERT). Nous étudierons ensuite en détail les tâches de classification de textes, de reconnaissance d'entités nommées, d'extractions de relations entre entités, et de fouille d'opinions ciblée par aspect. Nous verrons comment modéliser ces tâches comme des problèmes de classification, et mettre en œuvre les solutions en Python, en utilisant des bibliothèques d'apprentissage automatique (PyTorch, pytorch-lightning, transformers, etc.)

Objectifs

- Connaissance des problèmes spécifiques à l'analyse de données non structurées (textes).
- Connaissance des principales méthodes utilisées, notamment pour l'extraction d'information et la fouille d'opinions dans les textes.
- Capacité à mettre en œuvre ces méthodes en Python.

Heures d'enseignement

UE Ingénierie des connaissances 2 - CM	CM	24h
UE Ingénierie des connaissances 2 - TP	TP	24h

Pré-requis recommandés

Programmation en Python

Contrôle des connaissances

- Langage de programmation / librairies : Python, PyTorch, pytorch-lightning.
- Environnements de développement : PyCharm ou VS Code, et Google Colab.

Période : Semestre 9

Compétences visées

- Connaissance des problèmes spécifiques à l'analyse de données non structurées (textes).
 - Connaissance des principales méthodes utilisées, notamment pour l'extraction d'information et la fouille d'opinions dans les textes.
 - Capacité à mettre en œuvre ces méthodes en Python.
-

Bibliographie

- Daniel Jurafsky & James H. Martin (2021): Speech and Language Processing, Chap. 7: Neural Networks and Neural Language Models - [🔗 https://web.stanford.edu/~jurafsky/slp3/7.pdf](https://web.stanford.edu/~jurafsky/slp3/7.pdf)
- Daniel Jurafsky & James H. Martin (2021): Speech and Language Processing, Chap. 9: Deep Learning Architectures for Sequence Processing - [🔗 https://web.stanford.edu/~jurafsky/slp3/9.pdf](https://web.stanford.edu/~jurafsky/slp3/9.pdf)
- Daniel Jurafsky & James H. Martin (2021): : Speech and Language Processing, [🔗](#) Chap. 11: Transfer Learning with Pre-trained Language Models and Contextual Embeddings
- Daniel Jurafsky & James H. Martin (2021): Speech and Language Processing, Chap. 6: Vector Semantics and Embeddings. - [🔗 https://web.stanford.edu/~jurafsky/slp3/6.pdf](https://web.stanford.edu/~jurafsky/slp3/6.pdf)
- Daniel Jurafsky & James H. Martin (2021): Speech and Language Processing [🔗](#) Chap. 8, Section 8.3: Named Entities and Named Entity Tagging

- Daniel Jurafsky & James H. Martin (2021): [Speech and Language Processing](#) Chap. 17: Information Extraction
- (Zhang et al., 2022): [A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges](#)
- (Brauwers and Frasincar, 2022): [A Survey on Aspect-Based Sentiment Classification](#)

Infos pratiques

Lieu(x) ville

- > Grenoble
-

Campus

- > Grenoble - Domaine universitaire