

# UE Fouille de textes



Niveau d'étude  
Bac +5



ECTS  
3 crédits



Crédits ECTS  
Echange  
3.0



Composante  
UFR IM2AG  
(informatique,  
mathématiques  
et  
mathématiques  
appliquées)



Période de  
l'année  
Automne (sept.  
à dec./janv.)

- > **Langue(s) d'enseignement:** Français
- > **Méthodes d'enseignement:** En présence
- > **Forme d'enseignement :** Cours magistral
- > **Ouvert aux étudiants en échange:** Oui
- > **Crédits ECTS Echange:** 3.0
- > **Code d'export Apogée:** GBX9SD12

## 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 également 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 ou Tensorflow/Keras, transformers, etc.)

---

## Heures d'enseignement

CM	CM	12h
TP	TP	12h

---

## Pré-requis recommandés

- Programmation en Python

**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 (2020): Speech and Language Processing, Chap. 7: Neural Networks and Neural Language Models - <https://web.stanford.edu/~jurafsky/slp3/7.pdf>
- Daniel Jurafsky & James H. Martin (2020): Speech and Language Processing, Chap. 9: Deep Learning Architectures for Sequence Processing - <https://web.stanford.edu/~jurafsky/slp3/9.pdf>
- Daniel Jurafsky & James H. Martin (2020): Speech and Language Processing, Chap. 6: Vector Semantics and Embeddings. - <https://web.stanford.edu/~jurafsky/slp3/6.pdf>
- Daniel Jurafsky & James H. Martin (2020): Speech and Language Processing, Chap. 17: Information Extraction. - <https://web.stanford.edu/~jurafsky/slp3/17.pdf>

## Infos pratiques

---

### Lieu(x) ville

› Grenoble



---

## Campus

› Grenoble - Domaine universitaire