

# Recherche d'information



## Composante

École Nationale  
Supérieure  
d'Électrotechnique  
d'Électronique  
d'Informatique  
d'Hydraulique  
et des  
Télécommunications

## En bref

- > **Code Ametys:** N9EN13B
- > **Ouvert aux étudiants en échange:** Oui

## Présentation

### Objectifs

Ce cours introduit tout d'abord les tâches du TAL (Traitement automatique des Langues) puis aborde de manière plus approfondie les concepts fondamentaux de la représentation des textes et de la recherche d'information. Il couvre différentes techniques de modélisation textuelle, notamment les représentations discrètes (Bag of Words, TF-IDF) et continues (Word Embeddings statiques et contextuels), ainsi que les modèles de langage avancés.

Nous explorerons les modèles classiques de recherche d'information (vectoriels, probabilistes, etc.), en détaillant les mécanismes mis en œuvre pour évaluer et sélectionner les documents les plus pertinents en fonction des besoins des utilisateurs. Une attention particulière sera portée aux avancées récentes en deep learning appliquées à la recherche d'information, ainsi qu'à l'utilisation des modèles de langage pré-entraînés (LLMs) dans diverses applications.

La partie pratique du cours permettra aux étudiants de travailler sur différentes approches de représentation des textes en utilisant **scikit-learn**. Nous verrons également comment exploiter les **LLMs via Hugging Face** pour des tâches variées (tokenisation, finetuning des modèles) et utiliser des plateformes comme **LlamaIndex** pour implémenter des systèmes de **Retrieval-Augmented Generation (RAG)**.

---

## Description

1. Introduction au TAL
2. Text processing : représentation des textes en bag of words (tf.idf)
3. Text processing : Modèle statistique de langue (Language model)
4. *Text processing : Représentation continue (LSI, word embedding: W2VEC, BERT, GPT, ...)*
5. *Concepts de base de la RI (Recherche d'Information)*
6. Modèles de RI
7. Modèles de RI basés sur les LLM

---

## Bibliographie

1. M.-R. Amini, E. Gaussier -: Recherche d'information : Applications, modèles et algorithmes - Eyrolles, 2017
2. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. Introduction to Information Retrieval. 2008 <http://nlp.stanford.edu/IR-book/information-retrieval.html>)
3. Baeza-Yates, R. and Ribeiro-Neto, B. (2011). Modern Information Retrieval - the concepts and technology behind search.
4. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval. Addison- Wesley, 1999
5. Baeza-Yates and Ribeiro-Neto, eds. (1999) Modern Information Retrieval Addison- Wesley (site miroir)