

# Fouille, extraction et visualisation



## Présentation

**Code interne :** EI9IS329

---

### Description

Le but de ce cours est double et s'articulera autour de deux projets.

Le premier projet a pour but de présenter et de mettre en oeuvre quelques techniques d'extraction d'information issues de données textuelles.

Nous verrons dans un premier temps comment des algorithmes éprouvés tels que le modèle Bag-of-Word et le TF-IDF permettent d'extraire des données pertinentes de documents.

Nous aborderons ensuite les méthodes d'embedding vectoriels, en étudiant le modèle Word2Vec, qui permet d'extraire des données contextuelles.

Enfin nous verrons comment exploiter ces informations pour identifier des textes sémantiquement proches ou encore les catégoriser au moyen d'algorithmes de clustering.

Le second projet s'intéressera à un problème similaire, mais dans le cadre de données visuels.

Les cours sont accompagnés de TDs/TPs permettant la mise en oeuvre effective des algorithmes présentés ci-dessus.

Deux projets, l'un sur les données textuelles et l'autre sur les données visuelles, adossés à des données réelles, permettront aux élèves de mettre en application les algorithmes vu en cours, tout en mettant en oeuvre leurs compétences en calcul distribué pour traiter la volumétrie du jeu de données en un temps raisonnable.

---

### Pré-requis obligatoires

Notions de python, d'algorithmie et d'algèbre linéaire

---

### Informations complémentaires

L'analyse et le traitement du langage naturel (NLP) est l'un des grands challenges actuel en Intelligence Artificielle. Les avancées dans ce domaine sont utilisées au quotidien dans les moteurs de recherches, les chatbots ou encore les boites mails (détection de spam, ciblage publicitaire, ...).



D'autres part, le traitement d'un grand nombre d'image pour en extraire de l'information est un autre grand challenge actuel. Détection automatique de personnes, de panneaux, reconnaissance d'objets, ... les domaines sont nombreux et variés.

---

## Modalités de contrôle des connaissances

### Évaluation initiale / Session principale - Épreuves

Type d'évaluation	Nature de l'épreuve	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'épreuve	Note éliminatoire de l'épreuve	Remarques
Projet	Contrôle Continu			1		

---

## Infos pratiques

---

### Contacts

Mathieu Faverge

✉ [Mathieu.Faverge@bordeaux-inp.fr](mailto:Mathieu.Faverge@bordeaux-inp.fr)