

# 3MD3060 - Traitement automatique du langage naturel

Responsables : **Fragkiskos MALLIAROS , Matthias GALLE**

Nombre d'heures d'études élèves (HEE) : **40**

Nombre d'heures présentielles d'enseignement (HPE) : **24**

Année académique : **2024-2025**

Niveau avancé : **non**

## Présentation, objectifs généraux du cours :

Ce cours donnera un aperçu des sujets d'étude du domaine du traitement du langage naturel. Il adoptera une approche centrée sur les problèmes, en présentant des problèmes de plus en plus complexes, en commençant par des éléments de base comme la modélisation du langage, l'étiquetage et l'analyse syntaxique, et en progressant vers des problèmes complexes comme l'extraction d'opinions, la traduction automatique, les questions-réponses et le dialogue. Les méthodes historiques importantes seront mentionnées (et étudiées si elles sont toujours pertinentes), mais l'accent sera mis sur l'état actuel de la technique, qui implique des avancées souvent récentes dans la formation des réseaux neuronaux et les nouvelles architectures.

## Période(s) du cours (n° de séquence ou hors séquence) :

SM10

## Prérequis :

- 
- Compétences de codage. Tous les exercices seront effectués en python
- La connaissance des bibliothèques pertinentes (scikit-learn, spacy, kears, pytorch, tensorflow) n'est pas obligatoire mais utile.
- Algèbre linéaire
- Une connaissance de base de l'apprentissage automatique sera utile
- Concepts de base en informatique

## Plan détaillé du cours (contenu) :

### PREMIÈRE PARTIE

#### 1. Introduction

#### 2. Modèle linguistique

#### 3. Représentation des mots et des documents

#### 4. Balisage, reconnaissance d'entités nommées

#### 5. Analyse syntaxique

### DEUXIÈME PARTIE

**6. Extraction de contenu généré par l'utilisateur**

**7. Traduction automatique et génération de langage naturel**

**8. Lecture automatique**

**9. Dialogue**

## **Déroulement, organisation du cours :**

Lectures, lab sessions and project. The course consist of two sections. The first one introduces typical NLP tasks, while the second focuses on end application.

## **Organisation de l'évaluation :**

L'évaluation du cours se fera par trois exercices, notés de manière égale. L'évaluation sera principalement automatique (performance sur un ensemble de tests), mais l'évaluation qualitative (fichier texte contenant des explications, qualité du code, etc) sera également prise en compte.

Tous les projets seront réalisés en groupes de 4 étudiants maximum.

## **Acquis d'apprentissage visés dans le cours :**

L'objectif de ce cours est de fournir une introduction au domaine du traitement du langage naturel. À la fin du cours, l'étudiant devra :

savoir quelles sont les principales applications, pourquoi elles sont difficiles et comment elles peuvent être utilisées.

avoir une bonne compréhension des méthodes éprouvées et de leur fonctionnement interne

être exposé aux directions de recherche actuelles, et avoir les bases fondamentales pour être capable de suivre le domaine dans les années à venir.

## **Bibliographie :**

The main reference will be Speech and Language Processing. Jurafsky & Martin. Draft of 3rd edition online at <https://web.stanford.edu/~jurafsky/slp3/>