

## [Formation] Traitement automatique du langage naturel – NLP Text Mining

### OBJECTIFS

- Présenter les outils de traitement de langue, qu'ils soient basés sur des méthodes statistiques ou sur de méthodes symboliques
- Décrire le fonctionnement et identifier les atouts et les faiblesses des grands modèles de langage LLM comme GPT-4
- Évaluer les techniques et les adapter à chaque type de problème
- Comparer et combiner les approches : exploration d'outils statistiques (approches fréquentistes, similarité sémantique, plongements) et formels (langages formels, logiques de premier ordre et de description, lambda-calcul, ontologies)?

### PROGRAMME

#### Introduction à la linguistique

#### Approche neuronales

#### Approches statistiques

- Désambiguïsation de mot
- Classification supervisée de textes
- Similarité et parenté sémantiques
- Pré-traitement du texte
- Modèles fréquentistes : Représentation Bag-of-words, modèles de langue n-gram, et dérivés.
- Deep learning et modèles de langue neuronaux
- Plongements et applications
- Modèles séquentiels et mécanisme d'attention
- Transformers
- Représentations contextuelles
- Apprentissage par transfert et Large Language Models
- Utilisation de SentiWordNet pour la classification des critiques
- Utilisation de réseaux de neurones sur le même corpus de textes, comparaison des résultats ; possibilité d'approche hybride (plongement d'arbres



### DATES ET LIEUX

Du 24/04/2024 au 26/04/2024 à Paris

Du 09/10/2024 au 11/10/2024 à Paris

### PUBLIC / PREREQUIS

Ingénieurs, chefs de projet devant traiter des données textuelles.

Des connaissances du langage Python sont requises afin de tirer pleinement profit de la formation.

### COORDINATEURS

#### Fabian SUCHANEK

Enseignant-chercheur à Télécom Paris. Il a fait ses recherches à l'Institut Max Planck en Allemagne, chez Microsoft Research Cambridge/UK, chez Microsoft Research Silicon Valley/USA, et à l'INRIA Saclay. Il est l'auteur principal de YAGO, une des plus grandes bases de connaissances publiques dans le monde.

#### Matthieu LABEAU

Enseignant-chercheur à Télécom Paris. Son activité de recherche en traitement automatique du langage, concerne principalement l'apprentissage de représentations et la

syntaxiques)

## Travaux pratiques

### Approches symboliques

- Langages formels Graphes conceptuels, ontologies, bases de connaissances
- Extraction d'informations
- Désambiguïsation
- Détection d'entités

## Travaux pratiques

### Synthèse et conclusion

modélisation du langage.

### MODALITES PEDAGOGIQUES

La formation comprend des travaux pratiques qui permettent d'appliquer les notions abordées.

Appelez le 01 75 31 95 90  
International : +33 (0)1 75 31 95 90

[contact.exed@telecom-paris.fr](mailto:contact.exed@telecom-paris.fr) / [executive-education.telecom-paris.fr](http://executive-education.telecom-paris.fr)