

Master Informatique

Traitement automatique des langues

Informations

Composante : Faculté des Sciences

Responsable

Frederic BECHET (Enseignant)

Langue(s) d'enseignement

Français

Contenu

But de ce cours

- Donner une culture générale en TAL
- Savoir mener des expériences en TAL

Plan du cours

Partie 1 : Le langage humain est un signal

- d'un signal de parole à un signal de "mots" : modèles paramétriques/statistiques, canal bruité
- modélisation de la "source" d'un texte par chaîne de Markov
- analyse statistique d'un texte : Loi de Zipf, représentation vectorielle, distance cosine

- Partie 2 : Le langage humain est un système formel

- de la compilation d'un langage de programmation vers la compilation du langage humain

- analyse lexicale, syntaxique, sémantique

- Partie 3 : Le langage humain est à la fois un signal et un système formel

- grammaires probabilistes, tagging syntaxique/sémantique - le cas des grands modèles de langue génératifs (chatGPT)

- Mise en pratique

- développement d'un chatbot permettant d'interroger une base de données

- 3 versions pour les 3 "vues" sur le langage

- appliquer un protocole de test et d'évaluation rigoureux ; méthodes d'analyse des résultats

Compétences à acquérir

- Comprendre les principaux paradigmes de représentation du langage naturel par les machines.

- Savoir construire un protocole expérimental.

- Définir une question de recherche.

- Evaluer les résultats obtenus.

Modalités d'organisation

Cours magistraux suivi de TP/TD.

Modes d'évaluation

- Contrôle continu

- Rendu après chaque séance de TP

- évaluation

- note du rendu de TP à la fin des séances

- note du rapport sur les résultats des TP

- note sur la présentation du TP chatbot

Bibliographie, lectures recommandées

Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft online) <https://web.stanford.edu/~jurafsky/slp3/>

Pré-requis obligatoires

- Introduction aux Sciences des Données et Apprentissage Automatique

- Programmation (python, java, C, ..)

Prérequis recommandés

- Théorie des Langages

- Introduction au Traitement de Signal

- Introduction à la Théorie de l'Information

VOLUME HORAIRE

- Volume total: 27 heures
- Cours magistraux: 9 heures
- Travaux dirigés: 6 heures
- Travaux pratiques: 12 heures

Codes Apogée

- SINB44CL [ELP]

Pour plus d'informations

[Aller sur le site de l'offre de formation...](#)



Dernière modification le 13/11/2024