

Licence Informatique

Traitement automatique des langues

Responsables	Descriptions	Informations
Carlos RAMISCH carlos.ramisch@univ-amu.fr	Code : S04IN6M2	Composante : Faculté des Sciences
Marie helene STEFANINI marie-helene.stefanini@univ-amu.fr	Nature : Unité d'enseignement Domaines : Sciences et Technologies	

LANGUE(S) D'ENSEIGNEMENT

Français

CONTENU

L'objectif de cette UE est de présenter le domaine du traitement automatique des langues. Il s'agit d'un domaine porteur qui permet de développer des applications innovantes pour traiter le langage humain, telles que les assistants personnels, agents conversationnels, traducteurs automatiques. Chaque module du cours abordera une famille de modèles et/ou applications permettant de résoudre un problème concret du traitement automatique des langues.

- Analyse morphologique et syntaxique (16,7%) : notions élémentaires de linguistique, ambiguïtés, automates finis et grammaires hors contexte pour l'analyse du langage naturel
- Représentation et manipulation de textes (16,7%) : encodage de caractères, segmentation de phrases, tokénisation, loi de Zipf, vocabulaire, TTR.
- Recherche d'informations et classification de textes (16,7%) : vecteurs de co-occurrences, similarité du cosinus, classification de textes, algorithme des k plus proches voisins
- Dictionnaires électroniques (16,7%) : représentation d'informations lexicales, valence verbale, sens des mots, sémantique lexicale
- Analyse morphosyntaxique probabiliste (16,6%) : parties du discours, probabilité conditionnelle, n-grammes, probabilité de séquences, modèle d'étiquetage en parties du discours
- Grammaires logiques (16,6%) : variables, définition de grammaires logiques, unification

COMPÉTENCES À ACQUÉRIR

- Utiliser les concepts fondamentaux de l'informatique (langages formels, logique, et graphes) pour la programmation et la modélisation.
- Concevoir le traitement informatisé d'informations de différentes natures, telles que du texte, des images et des nombres.
- Concevoir et implémenter une application/bibliothèque mêlant plusieurs technologies en expliquant les choix de mise en œuvre.
- Rédiger des documents écrits en tenant compte des exigences contextuelles et des conventions sociales en la matière (manuel d'utilisation, documentation, rapport de projet).
- Faire un exposé oral convaincant en utilisant les techniques modernes de communication.

MODALITÉS D'ORGANISATION

Cours magistraux suivis de travaux dirigés pour fixer les notions, suivi de travaux pratiques sur le même thème. Certains travaux pratiques seront notés. De plus, les étudiants devront élaborer un projet au choix parmi les thèmes proposés. Chaque thème consistera en un problème pratique à résoudre (par exemple Classification de la langue d'un texte). Il s'agira de programmer une solution utilisant les techniques vues en cours, puis évaluer sur des jeux de données fournis. Ensuite ils devront synthétiser le développement et les résultats obtenus dans un compte-rendu écrit et les présenter à l'oral.

BIBLIOGRAPHIE, LECTURES RECOMMANDÉES

- Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.
- Daniel Jurafsky and James H. Martin, Speech and Language Processing, Pearson Prentice Hall, 2009.

PRÉREQUIS RECOMMANDÉS

- Programmation et conception
- Langages formels

VOLUME HORAIRE

- Volume total: 30 heures
- Cours magistraux: 9 heures
- Travaux dirigés: 9 heures
- Travaux pratiques: 12 heures

CODES APOGÉE

- SIN6U20A [ELP]

POUR PLUS D'INFORMATIONS

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



Dernière modification le 05/12/2022