



Guide pédagogique

Module « Traitement Automatique de la Langue Naturelle »

Option IASD, Spécialité "Connaissance et Texte" – 10.2

(3 crédits ECTS)

Place du module et enjeux

Les langues naturelles nous entourent : nous les parlons, nous les écrivons, nous les lisons et les manipulons quotidiennement avec aisance. Cependant, le développement d'applications informatique étant capable de manipuler les langues naturelles (analyser, comprendre, générer, traduire) constitue un problème de recherche très difficile depuis plus de 7 décennies, à l'intersection de la linguistique, de l'informatique et des sciences cognitives. C'est ce que l'on appelle le Traitement Automatique des Langues Naturelles (TALN). De nos jours, avec l'avènement de l'apprentissage profond et des applications innovantes (assistants vocaux, traducteurs automatiques, *chatbots*, résumé automatique, détection du plagiat, etc.), il y a une recrudescence d'intérêt pour le TALN à la fois dans les laboratoires et dans l'industrie, qui est considéré comme le prochain grand défi pour l'IA.

Ce module s'intéresse aux traitements automatiques d'analyse et fournit aux étudiant de l'option « Intelligence Artificielle et Science des Données » (IASD) qui ont choisi la spécialité « Connaissance et Texte » des bases solides pour appliquer à un large éventail de cas réels les modèles théoriques.

Teaching guide and syllabus

***« Natural Language Processing » module, DSAI option,
specialization "Knowledge and Text" – 10.2 (3 ECTS credits)***

Subject matter importance and associated issues

Natural Language is all around us: we speak it, we write it, we read it, we wield it every day with ease. However, being able to develop computer applications able to work with natural language (analyze, understand, generate, translate) has been a challenge for researcher for over 7 decades, at the intersection of computer science, linguistics and cognitive science: Natural Language Processing (NLP). Nowadays, with the advent of deep learning and of innovative language technology applications (voice assistants, machine translation, chatbots, automatic summarizers, plagiarism detection, etc.), there is renewed significant interest in NLP both in academic labs and for industrial applications as the next big challenge for AI. This module gives students a sound understanding of linguistics and NLP concepts and allow them to operationalize this theoretical knowledge for a broad range of tasks. This module is proposed to students from the option « Data Science and Artificial Intelligence » (DSAI) of the « Computer Science and Artificial Intelligence » department (CSAI), this module is optional and will only be followed by students who have chosen the speciality "Knowledge and Text".

Responsable : Andon Tchechmedjiev

Téléphone : 04 34 24 62 16

Courriel : andon.tchechmedjiev@mines-ales.fr



IMT Mines Alès
École Mines-Télécom

ENSEIGNEMENTS ACADÉMIQUES	Volume horaire	Détail des coefficients	Crédits
Traitement automatique du langage naturel	50 h	1	3

Matière 1

Titre de la matière : Traitement automatique du langage naturel	
Code : 2IA-iasd-con 10.2	Titre du module : «Traitement Automatique de la Langue Naturelle »
Semestre : S10	Cursus de rattachement : département 2IA, options IASD (spécialité Connaissance et Texte)

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
50h	70h	9h		27h	12h	2h	20h	1	3

Résumé	Une introduction au traitement automatique du langage naturel (TALN) allant des bases de la linguistique à un panorama des tâches et techniques en lien avec le TALN (symboliques mais aussi basées sur l'apprentissage automatique), suivies d'une introduction à la recherche d'information.
---------------	--

Responsable	Andon Tchechmedjiev – LGI2P/IMT Mines Alès
Équipe enseignante	Sébastien Harispe – LGI2P/IMT Mines Alès Pierre-Antoine Jean – LGI2P/IMT Mines Alès Andon Tchechmedjiev – LGI2P/IMT Mines Alès Intervenants extérieurs universitaires et industriels Doctorants pour l'encadrement de TP.

Mots-clés	Traitement automatique de la langue naturelle, Recherche d'informations, Apprentissage Automatique
Prérequis	Introduction à l'IA, Introduction à l'Apprentissage Automatique (2IA-8.3)

Contexte et objectif général :

Les technologies en lien avec le langage naturel jouent un rôle de plus en plus prépondérant dans nos vies quotidiennes : que ce soit avec des assistants vocaux, des *chatbots*, des applications de traduction automatique, et au-delà. De nombreuses entreprises, grandes comme petites investissent des sommes importantes dans ces technologies du traitement automatique de la langue naturelle et cherchent de nouveaux talents ayant la capacité de naviguer parmi les subtilités d'un domaine aux frontières de l'informatique, de la linguistique et des sciences cognitives. Cette enseignement vous introduira aux bases de la linguistique ainsi qu'aux principales tâches, outils et problématiques relatives aux applications du TAL. Vous apprendrez alors comment mobiliser vos connaissances d'apprentissage automatique au TAL, pour finir sur une introduction à la recherche d'information, une problématique connexe centrale aux cas d'usages industriels.

Programme et contenu :

- Introduction au Traitement Automatique des Langues Naturelles (Ressources, Tâches, Problèmes) (9h)
- Apprentissage automatique et NLP (ingénierie des caractéristiques, classification, apprentissage profond) (21h)
- Introduction à la Recherche d'Information (6h)
- Projet (12h)
- Examen (2h)

Méthode et organisation pédagogique :

Les exercices pratiques seront réalisés sur les ordinateurs personnels des étudiants. La première partie sera principalement composée de cours magistraux et de TP, dont une part sera assurée par un intervenant extérieur. Le reste sera composé de cours/TP et seront suivies d'un projet par binôme pour l'évaluation.

Acquis d'apprentissage visés :

Être capable d'invoquer des connaissances élémentaires en linguistique pour approcher des problèmes de TAL. Identifier les problèmes de TAL et être capable de proposer des stratégies de résolution afin d'aller vers l'implémentation de solutions (avec les bons outils et les bonnes ressources) alliant à la fois approches symboliques ou d'apprentissage automatique. Être capable de différencier des problèmes de recherche d'information de problèmes TAL et implémenter une solution adéquate à de tels problèmes.

Évaluation :

Devoir sur table / QCM (coef 2), Projet (coef 1)

Retour sur l'évaluation fait à l'élève :

Retours personnalisés sur le travail de TP à la demande des élèves. Corrections de l'examen et évaluation du projet sous trois semaines au maximum.

Support pédagogique et références :

Des diapositives de cours, des notebooks Anaconda, des ressources en ligne.

Méthode et organisation pédagogique

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP et Projets. Toutefois pour certaines parties du cours, une forme de pédagogie inversée est envisagée.

Modalité d'évaluation

Le niveau d'acquisition des compétences sera évalué selon les exigences suivantes :

N° indicateur	Indicateur
1	Connaitre les savoirs formels et pratiques du socle des fondamentaux
2	Exploiter les savoirs théoriques et pratiques
3	Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre

Répartition

Matière	Contrôle	Coefficients	Type de notation	Indicateurs évalués	Chapitres
Traitement automatique du langage naturel	Projet	1	Binôme	Tous	Tous
Traitement automatique du langage naturel	Devoir sur table	2	Individuelle	Tous	Tous

Engagement de l'étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l'établissement. Chaque étudiant s'engage à en prendre connaissance et à la respecter.

Obligation des cours : Présence obligatoire pour tous à chaque séance

Nombre d'heures estimées de travail personnel : pour acquérir les compétences demandées, il est nécessaire que l'étudiant consacre minimum 30 min de travail personnel de compréhension et d'approfondissement par séance de cours auquel on peut rajouter 20h de travail personnel pour définir le périmètre et finaliser les projets.

Nombre d'heures estimées de préparation aux travaux dirigés (TD) : 1 à 2h.

Pénalité pour retard : Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard (notation effectuée sur 20).

Équipe enseignante

Nom	Domaine d'expertise	Courriel/Téléphone
Andon TCHECHMEDJIEV	Apprentissage automatique, Traitement automatique de la Langue, IA et développement informatique	andon.tchechmedjiev@mines-ales.fr 04 34 24 62 16
Sébastien HARISPE	Intelligence Artificielle, Représentation des connaissances, Apprentissage Automatique	sebastien.harispe@mines-ales.fr 04 34 24 62 82
Pierre-Antoine JEAN	Intelligence Artificielle, Traitement automatique de la langue, Développement informatique	Pierre-antoine.jean@mines-ales.fr 04 34 24 62 30

ACADEMIC TEACHING	Teaching hours	Coefficients	Credits
Natural Language Processing	50 h	1	3

Class 1

Class title: Natural Language Processing	
Code: 2IA-iasd-con 10.2	Module title: Natural Language Processing
Semester: S10	Classification: CSAI department, DSAI option (specialization: Knowledge and Text)

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
50h	70h	9h		27h	12h	2h	20h	1	3

Title	Natural Language Processing
Summary	An introduction to natural language processing (NLP) from the basics of linguistics to an overview of the main tasks and techniques involved in NLP (both symbolic and based on machine learning), followed by an initiation to information retrieval.

Head	Andon Tchechmedjiev – LGI2P/IMT Mines Alès
Teaching team	Sébastien Harispe – LGI2P/IMT Mines Alès Pierre-Antoine Jean – LGI2P/IMT Mines Alès Andon Tchechmedjiev – LGI2P/IMT Mines Alès External academic or industrial speaker Teaching assistant for practical work

Key words	Natural Language Processing, Machine Learning, Information Retrieval
Prerequisites	Introduction to AI, Introduction to Machine Learning (Module 2IA-8.3)
Context and general objective: Natural language technologies play an ever-increasing role in our everyday lives, between voice assistants, and chatbots to machine translation and beyond. Companies large and small are investing heavily into natural language processing and looking for fresh talent able to navigate the intricacies of a field that stands at the intersection of computer science, linguistics and cognitive science. This module will introduce you the basics of linguistics and the main tasks, resources and issues involved in NLP applications. Then you will learn how to apply your knowledge of Machine Learning to NLP. Finally, we will introduce you to related notions in the field of information retrieval that are central to many modern industrial use-cases.	
Programme and contents: - Introduction to Natural Language Processing (Resources, Tasks, Problems) (9h) - Machine learning and NLP (Feature Engineering, Classification, Deep Learning) (21h) - Introduction to Information Retrieval (6h) - Project (12h) - Exam (2h)	
Method and pedagogic organisation: The first introductory part will consist of mostly lectures from an external lecturer along with a few practical sessions. The remainder of the module will involve mixed lecture/lab sessions, followed by a group project.	
Targeted skills or knowledge: To be able to muster up basic knowledge of linguistics to be able to tackle NLP problems. Identify NLP problems and devise concrete strategies to implement a solution (which resources, which tasks, which rules) based on both symbolic approaches and machine learning. Be able to distinguish between information retrieval problems and NLP problems and implement an appropriate solution.	
Evaluation: 2h written exam – project (coef 1)	
Feedback made to the student: Personalized feedback will be given about lab work to each student upon request. Grades will be given to students a maximum of three weeks after the exam and projects have taken place.	
Teaching material and references: Slides, anaconda notebooks, online resources	

Method and teaching organisation

Teachings will consist of lectures, and labs for project realisation.

Testing procedures

The student's level of knowledge acquisition will be evaluated according to the following points :

N° Indicator	Indicator
1	To know the formal and practical knowledge constituting the foundation of a given field
2	Exploit theoretical and practical knowledge
3	Analyse, interpret, model, hypothesize and solve problems

Grading scheme

Class	Exam	Coefficients	Administration mode	Evaluated Indicators	Chapters
Natural Language Processing	Project	1	Binôme	All	All
Natural Language Processing	Written examination	2	Individual	All	All

Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment's code of conduct. Each student is expected to know and respect the code of conduct.

Obligatory presence in classes: Students must attend all courses, seminars and labs.

Estimated hours of personal study: in order to acquire the required learning level, the student is expected (must) to spend a minimum of 30min of personal study time per class session.

Estimated hours of preparation required for labs/Work Shop: 1 to 2h.

Late penalties: Late works are subject to penalties as follows: 3 points per day (ratings between 0 and 20).

Teaching team

(Title) Name	Field of expertise	Email/phone
Andon TCHECHMEDJIEV	Computational Linguistics, Machine Learning, Knowledge representation, Artificial Intelligence	andon.tchechmedjiev@mines-ales.fr 04 34 24 62 16
Sébastien HARISPE	Artificial Intelligence, Knowledge Representation, Machine Learning	sebastien.harispe@mines-ales.fr 04 34 24 62 82
Pierre-Antoine JEAN	Artificial Intelligence, Natural language processing, Software development	Pierre-antoine.jean@mines-ales.fr 04 34 24 62 30

Approbation

Ce guide pédagogique entre en vigueur à compter du 07 janvier 2019

Il est porté à la connaissance des élèves par une publication sur le site Internet de l'école

Rédaction	Vérification	Validation
L'enseignant responsable du module : Andon TCHECHMEDJIEV	Le responsable d'UE / de département : Sylvie RANWEZ	Le directeur de l'école, Pour le directeur et par délégation, Le directeur de la DFA / de la DE : Michel FERLUT