Syllabus
2024 - 2025

TECHNICAL SHEET OF THE SUBJECT

Data of the subject

Subject name	Natural Language Processing I
Subject code	DTC-IMAT-323
Mainprogram	Bachelor's Degree in Mathematical Engineering and Artificial Intelligence
Involved programs	Grado en Ingeniería Matemática e Inteligencia Artificial [Third year]
Credits	6,0 ECTS
Type	Obligatoria (Grado)
Department	Department of Telematics and Computer Sciencies
Coordinator	Andrés Occhipinti Liberman

Teacher Information

Teacher

Name	Andrés Occhipinti Liberman
Department	Department of Telematics and Computer Sciencies
EMail	aocchipinti@icai.comillas.edu

Profesores de laboratorio

Teacher

Name	Jaime Pizarroso Gonzalo
Department	Department of Telematics and Computer Sciencies
Office	Santa Cruz de Marcenado 26
EMail	jpizarroso@comillas.edu
Phone	2732

SPECIFIC DATA OF THE SUBJECT

Contextualization of the subject

Contribution to the professional profile of the degree

This course is an introduction to Natural Language Processing (NLP). The goal is to provide a solid foundation in the theory and techniques used to perform various NLP tasks, such as natural language understanding and generation. The course covers a combination of techniques, including rule-based, statistical, and machine learning methods.

Prerequisites

Having taken a Machine Learning course.

Competencies - Objectives

Competences	
GENERALES	
CG04	Conocimientos básicos sobre el uso y programación de los ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería.
CG08	Capacidad para identificar, analizar y definir los elementos significativos que constituyen un problema vinculado a la explotación de datos e inteligencia artificial aplicada a las actividades empresariales para resolverlo con criterio y de forma efectiva
ESPECÍFICAS	
CE03	Capacidad para saber aplicar las técnicas matemáticas más adecuadas en la resolución de los diferentes problemas, técnicos y tecnológicos, planteados en el ámbito de la ingeniería y la inteligencia artificial. Aptitud para conocer el rango de aplicabilidad y limitaciones en la resolución de problemas de las diferentes herramientas matemáticas.
CE28	Conocimiento y capacidad para utilizar distintas tecnologías de procesamiento, representación y análisis de lenguaje natural.
CE31	Capacidad para especificar, diseñar e implementar las técnicas de aprendizaje automático y profundo para la resolución de problemas complejos.
CE36	Capacidad para analizar el comportamiento de los sistemas cognitivos y aplicarlos en el dominio artificial. Conocimiento de los principios de los procesos psicológicos básicos

Learning outcomes	
RA1	Conocer y distinguir los niveles del lenguaje natural y la complejidad de cada uno
RA2	Conocer las técnicas y fundamentos aplicados de procesado del lenguaje natural en cada uno de sus niveles
RA3	Conocer las herramientas básicas de procesado de la información para su posterior transformación
RA4	Aplicar los conocimientos aprendidos en otras materias como aprendizaje automático a problemas reales de generación de contenidos
RA5	Estar familiarizado con las herramientas y técnicas que se utilizan dentro de este ámbito

THEMATIC BLOCKS AND CONTENTS

Contents - Thematic Blocks

1. Basic principles of natural language processing.
2. Text classification. Probabilistic and neural classifiers.
3. Language models. The n-gram model. Evaluation of language models.
4. Word embeddings and vector semantics.
5. Analysis of linguistic structure. Dependency parsing and constituency parsing.
6. Logical representation of natural language semantics.

TEACHING METHODOLOGY

General methodological aspects of the subject

In-class Methodology: Activities

Training activities will include:

Explanatory and participatory lectures:

- The teacher will combine the exposition of theoretical content with practical examples, both mathematical and programming.
- The student will have practical code examples generated by the teacher inside and outside the classroom.
- Short tests will be given to assess the understanding of the content, focusing on challenging parts.

Practical exercises and problem-solving:

- Students will solve problems presented by the teacher in person during the second weekly class session, encouraging cooperative work dynamics.
- Occasionally, students (individually or in groups) will present their exercise solutions in class, and discussions will focus on improving or clarifying details.

Practical sessions using software:

- Practical sessions will address questions about the weekly practice, allowing students to complete their tasks.
- The difficulty of the practices will be graded, and students will implement them as they complete each milestone.

Continuous performance assessment activities:

- Tests will be conducted, complementary practices to the weekly ones will be developed, and gamified challenges will be introduced.

CG04, CG08, CE03, CE28, CE31, CE36

Non-Presential Methodology: Activities

Training activities will include:

Practical exercises and problem-solving:

- Students will have specific problems focused on assimilating the theoretical concepts explained in the previous theory session for non-face-to-face development.
- Problem solutions will be uploaded to the platform the following week or presented in class.

Practical sessions using software:

• Once the weekly practice is released after the corresponding theory session, students will work on it non-face-to-face. Students should come to the face-to-face practice session with 80% of the proposed objectives achieved. • In the classroom, the statement will be extended incrementally, covering the proposed milestones progressively. **Personal study:** • The main goal of non-face-to-face work is to understand and comprehend the theoretical concepts of the subject, as well as to be able to apply this knowledge to solve different types of problems. • After each theoretical explanation, the teacher will upload all developed codes to the website, and students should review them and pose "What if" questions to better assimilate theoretical concepts. • Texts for reading at home will be provided one or two weeks before the practical session addressing the topics. When applicable, a short question session about the reading will be prepared.	CG04, CG08, CE03, CE28, CE31, CE36

SUMMARY STUDENT WORKING HOURS

CLASSROOM HOURS				
Clases magistrales expositivas y participativas	Sesiones prácticas con uso de software	Tutorías para resolución de dudas	Ejercicios prácticos y resolución de problemas	Actividades de evaluación continua del rendimiento
28.00	20.00	5.00	10.00	2.00
NON-PRESENTIAL HOURS				
Sesiones prácticas con uso de software	Estudio personal	Trabajos		
30.00	55.00	30.00		
ECTS CREDITS: 6,0 (180,00 hours)				

EVALUATION AND CRITERIA

Evaluation activities	Evaluation criteria	Weight
Written exams evaluating concepts of the subject.	• Midterm exam: 15% • Final exam: 35%	50
Final work	• It will have a weight of 30% of the grade. • To pass the subject, students must obtain at least 5 points out of 10 in the final exam and in the final practice, both in the regular and extraordinary sessions.	30
Continuous assessment.	Continuous assessment assignments (deliverables): 1. Text classification 2. Language models	20

	3. Syntactic and semantic analysis	

Ratings

The final grade in the regular and extraordinary sessions of the subject will depend on the evaluation of the following activities:

- Final exam 35%
- Midterm exam 15%
- Final project 30%
- Continuous assessment assignments (deliverables): 20%
    - Text classification
    - Language models
    - Syntactic and semantic analysis

To pass the subject, students must obtain at least 5 points out of 10 in the final exam and in the final practice, both in the regular and extraordinary sessions. Missing 15% or more of the in-person hours for this subject may result in being unable to participate in both regular and extraordinary sessions. The final practice will be group-based.

BIBLIOGRAPHY AND RESOURCES

Basic Bibliography

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing (2nd ed.)*. Upper Saddle River, NJ: Pearson.

Eisenstein J. *Introduction to Natural Language Processing*. Cambridge Massachusetts: MIT Press; 2019.