COURSE DATA

DATA SUBJECT

Code: 46578
Name: Natural language processing
Cycle: Master's Degree
ECTS Credits: 6
Academic year: 2026-27

STUDY (S)

Degree	Center	Acad. year	Period
2262 - Master's Degree in Data Science	Escola Tècnica Superior d'Enginyeria	1	Second quarter

SUBJECT-MATTER

Degree	Subject-matter	Character
2262 - Master's Degree in Data Science	Natural language processing	COMPULSORY

COORDINATION

VIVES GILABERT YOLANDA

SUMMARY

Currently, a large portion of the data available for analysis consists of unstructured information in the form of natural language texts. This information includes web pages (Wikipedia, digital newspapers, blogs) or social media (Facebook, Twitter). Being able to analyse these texts using Natural Language Processing (NLP) algorithms is highly beneficial for organizations to make better decisions.

Machine learning algorithms are not capable of understanding text or characters, which is why NLP performs all the necessary pre-processing to convert this text data into a machine-readable format (numbers) and enable various types of subsequent analysis to extract relevant information. Some common applications of NLP include text classification, information retrieval and extraction, summarization, machine translation, and automated question-answering systems, among others.

This is a theoretical and practical course that includes theory sessions, practical sessions, and mixed sessions. The theory classes will be taught in Spanish.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

There are no specified enrollment restrictions with other subjects of the curriculum.

OTHER REQUIREMENTS

The recommended prerequisite is to have passed the Machine Learning course (first semester).

COMPETENCES / LEARNING OUTCOMES

2262 - Master's Degree in Data Science

Be able to assess the need to complete their technical, scientific, language, computer, literary, ethical, social and human education, and to organise their own learning with a high degree of autonomy.

Capacidad para trabajar en equipo para llegar a soluciones de problemas interdisciplinarios usando técnicas de análisis de datos.

Extraer conocimiento de conjuntos de datos en diferentes formatos.

Modelar la dependencia entre una variable respuesta y varias variables explicativas, en conjuntos de datos complejos, mediante técnicas de aprendizaje máquina, interpretando los resultados obtenidos.

Ser capaces de acceder a herramientas de información (bibliográficas y de empleo) y utilizarlas apropiadamente.

Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio, aplicando los conocimientos adquiridos en la identificación de salidas profesionales y yacimientos de empleo.

Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.

Students should demonstrate self-directed learning skills for continued academic growth.

Students should possess and understand foundational knowledge that enables original thinking and research in the field.

Usar las técnicas de procesado de lenguaje natural para analizar textos extrayendo conocimiento útil de ellos.

DESCRIPTION OF CONTENTS

1. Introduction to the Natural Language Processing

1.1. What is NLP?
1.2. The importance of text

1.3. Historical approaches to NLP
1.4. Applications and workflow

2. Text Use and Capture

2.1. Text strings in Python
2.2. Regular expressions
2.3. Text loading
2.4. Web content scraping
2.5. Optical Character Recognition (OCR)

3. Text pre-processing

3.1. Text division
3.2. Text cleaning and normalization
3.3. Morphological analysis
3.4. Semantic analysis
3.5. Grammatical analysis

4. Features extraction

4.1. Simple features
4.2. Bag of Words model
4.3. TF-IDF model
4.4. Word vectors (word embeddings)
4.5. Document vectors

5. NLP applications

5.1. Classification
5.2. Information extraction
5.3. Text mining
5.4. Information retrieval
5.5. Sequential models

6. Deep learning in NLP

6.1 Recurrent Neural Networks in NLP
6.2 Attention in deep learning
6.3 Transformers
6.4 Models and Applications

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	28,00
Theoretical and practical classes	4,00
Laboratory	28,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	2,00
Individual or group project	18,00
Independent study and work	20,00
Preparation of lessons	20,00
Preparation for assessment activities	15,00
Resolution of case studies	15,00
Total hours	**90,00**

TEACHING METHODOLOGY

Theoretical activities.

- Expository development of the subject with student participation in resolving specific questions.

- Completion of individual evaluation questionnaires.

Practical activities.

Learning through problem-solving, exercises, and case studies through which competencies on different aspects of the subject are acquired.

Laboratory and/or computer classroom work.

Learning through the completion of activities carried out individually or in small groups, conducted in computer classrooms.

EVALUATION

Objective test, consisting of one or more exams that will include both theoretical-practical questions and problems (45%). This test can be retaken in the second examination session.

Assessment of practical activities based on the completion of assignments/reports, oral presentations, and the use of the University¿s e-learning tools (45%). This test can be retaken in the second examination session.

Assessment based on student participation and level of engagement in the teaching-learning process, taking into account regular attendance to scheduled in-person activities and the resolution of proposed questions and problems on a regular basis (10%). This test cannot be retaken.

In order to pass the course, the average score of the three assessments must be greater than 5, either in the first or second examination session.

REFERENCES

- Sohom Ghosh, Dwight Gunning. Natural Language Processing Fundamentals. Packt Publishing, 2019
- Akshay Kulkarni, Adarsha Shivananda. Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python. Apress, 2019 (disponible e-libro)
- Dipanjan Sarkar. Text Analytics with Python: A Practitioner's Guide to Natural Language Processing. Apress 2019 (disponible e-libro)
- Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. OReally Media, 2009
- Jacob Eisentein. Natural Language Processing. 2018 (disponible bajo licencia CC-BY-NC-ND)
- Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. Practical Natural Language Processing. OReally Media, 2020.
- Build a Large Language Model (From Scratch). Sebastian Raschka. Editorial Manning. Septiembre 2024.