UAB
Universitat Autònoma de Barcelona

Aprendizaje y Procesamiento del Lenguaje Natural

Código: 106585
Créditos ECTS: 6

2024/2025

Titulación	Tipo	Curso
2504392 Inteligencia Artificial / Artificial Intelligence	OT	3
2504392 Inteligencia Artificial / Artificial Intelligence	OT	4

Contacto

Nombre: Joaquin Cerdà Company

Correo electrónico: joaquin.cerda@uab.cat

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

There are no official prerequisites, but it is recommended to have completed the subjects of Fundamentals of Programming I and II, Fundamentals of Mathematics I and II, Probability and Statistics, Data Engineering, Fundamentals of Machine Learning, and Fundamentals of Natural Language.

Objetivos y contextualización

This course provides an overview of the Natural Language Processing (NLP) applications, from classical approaches for text processing to advanced methods for person-computer interaction. This course covers both machine learning and deep learning techniques for NLP, considering both text and speech processing.

By the end of this course, students will be able to:

- Understand the fundamental concepts and techniques used in NLP.
- Implement and evaluate various NLP techniques using Python and popular NLP libraries.
- Apply NLP methods to real-world problems and interpret the results.

Competencias

Inteligencia Artificial / Artificial Intelligence
- Analizar y resolver problemas de forma efectiva, generando propuestas innovadoras y creativas para alcanzar los objetivos.
- Conocer y aplicar las técnicas del procesamiento del lenguaje natural para la explotación de datos de naturaleza lingüística y la creación y evaluación de los componentes de los sistemas de IA basados en el lenguaje.
- Conocer y utilizar de forma eficiente las técnicas y herramientas de representación, manipulación, análisis y gestión de datos a gran escala.
- Diseñar, implementar, analizar y validar soluciones algorítmicas eficientes y robustas a problemas computacionales derivados del diseño de sistemas inteligentes.

- Elaborar estrategias para formular y solucionar diferentes problemas de aprendizaje de manera científica, creativa, crítica y sistemática, conociendo las capacidades y limitaciones de los diferentes métodos y herramientas existentes.
- Identificar, analizar y evaluar el impacto ético y social, el contexto humano y cultural, y las implicaciones legales del desarrollo de aplicaciones de inteligencia artificial y de manipulación de datos en diferentes ámbitos.
- Identificar, comprender y aplicar los conceptos y técnicas fundamentales de representación del conocimiento, razonamiento y aprendizaje computacional para la solución de problemas de inteligencia artificial.
- Introducir cambios en los métodos y los procesos del ámbito de conocimiento para dar respuestas innovadoras a las necesidades y demandas de la sociedad.
- Trabajar cooperativamente para la consecución de objetivos comunes, asumiendo la propia responsabilidad y respetando el rol de los diferentes miembros del equipo.

Resultados de aprendizaje

1. Analizar y resolver problemas de forma efectiva, generando propuestas innovadoras y creativas para alcanzar los objetivos.
2. Conocer y comprender las soluciones actuales a tareas de procesamiento del lenguaje natural para extracción de información, traducción y resumen automáticos, y sistemas de diálogo.
3. Conocer, aplicar y diseñar técnicas de aprendizaje computacional para problemas de procesamiento del lenguaje natural.
4. Diseñar e implementar técnicas eficientes de búsqueda de patrones lingüísticos y semánticos en colecciones masivas de textos.
5. Diseñar modelos de aprendizaje para sistemas de procesamiento del lenguaje natural en base a la teoría y técnicas generales de aprendizaje computacional.
6. Diseñar y aplicar técnicas de razonamiento semántico.
7. Diseñar y evaluar soluciones basadas en aprendizaje para la detección y extracción de patrones lingüísticos y semánticos en colecciones masivas de textos.
8. Diseñar, implementar y analizar soluciones algorítmicas para el tratamiento masivo de datos textuales con anotaciones lingüísticas.
9. Identificar, analizar y evaluar el sesgo de modelos predictivos de procesamiento de lenguaje natural.
10. Ponderar los riesgos y las oportunidades de las propuestas de mejora tanto propias como ajenas.
11. Proponer nuevos métodos o soluciones alternativas fundamentadas.
12. Trabajar cooperativamente para la consecución de objetivos comunes, asumiendo la propia responsabilidad y respetando el rol de los diferentes miembros del equipo.

Contenido

1. Fundamentals of NLP
2. Semantic Analysis
3. Pragmatic Analysis
4. Recurrent Neural Networks for NLP Applications
5. Transformers for NLP Applications
6. Foundation Models for NLP Applications

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			

Clases de teoría	15	0,6	9, 2, 3
Sesión de ejercicios	25	1	1, 6, 7, 8, 5, 4, 9, 3, 12
Sesión de proyecto	6	0,24	1, 5, 9, 11, 2, 3, 10, 12
Tipo: Supervisadas			
Trabajo en el proyecto	50	2	1, 6, 5, 9, 11, 2, 3, 10, 12
Tipo: Autónomas			
Estudio individual	25	1	1, 5, 2, 3
Resolución de ejercicios	25	1	1, 7, 8, 5, 4, 2, 3, 10, 12

Sessions will combine three types of teaching activities: theory classes, solving project-based exercises, and a project development. Students will work in small groups of two or three people to solve both the exercises and the project.

1. Theory lectures, which will contain the theoretical background needed to solve the exercises and the project will be presented using a presentation. These presentations will contain theoretical concepts, and mathematical formulation, as well as the corresponding algorithmic solutions.

2. Project-based exercises will be developed using Jupyter notebooks to be able to comment, point by point, the coding solutions. These exercises will be submitted regularly through Campus Virtual, explaining the proposed solution and showing the obtained results. These reports could be presented to the whole classroom in the subsequent session. The group/s responsible for presentation will be chosen randomly in class. The exercise's submissions will comprise the portfolio.

3. A project will be carried out during the semester, where students will have to solve a real-world problem. The project will be solved in small groups of two or three students, where each member of the group must contribute a part and put it together with the rest to obtain the final solution. These working groups must be maintained until the end of the semester and must be self-managed in terms of distribution of roles, work planning, assignment of tasks, management of available resources, conflicts, etc. To develop the project, the groups will work autonomously, while the practical sessions will be used (1) for the teacher to present the project theme and discuss possible approaches, (2) for monitoring the status of the project and (3) for the teams to present their final results.

The above activities will be complemented by a system of tutoring and consultations outside class hours.

All the information of the subject and the related documents that the students need will be available at the virtual campus (cv.uab.cat).

Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Nota:es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Portfolio	49%	0	0	1, 6, 7, 8, 5, 4, 2, 3, 10, 12
Proyecto	30%	2	0,08	1, 5, 9, 11, 2, 3, 10, 12
Test	21%	2	0,08	1, 7, 8, 5, 4, 9, 2, 3

To assess the level of student learning, a formula is established that combines knowledge acquisition, the ability to solve problems and the ability to work as a team, as well as the presentation of the results obtained.

Final grade

The final grade is calculated in the following way and according to the different activities that are carried out:

$$\text{Final grade} = 0.7 * \text{Exercises Grade} + 0.3 * \text{Project Grade}$$

This formula will be applied as long as the Exercises and Project grades, are higher than 4. If the formula yields >= 5, but the student does not reach the minimum required in any of the evaluation activities, then a final grade of 4.5 will be assigned.

Exercises Grade

The aim of the exercises is to become familiar with the use of the theoretical concepts, and to apply them to a real-world problem. The regular submission of problem solutions will be used as evidence of this work.

In order to obtain a grade for exercises, it is necessary that more than 50% of the exercises are submitted during the semester. In the contrary, the portfolio grade will be 0.

A test about the exercises will be performed individually at the end of the semester. The final problems grade will be the combination of the exercise's portfolio and this test.

$$\text{Problems Grade} = 0.7 * \text{Portfolio evaluation} + 0.3 * \text{Test}$$

The formula will be applied as long as the Test grade is higher than 4.

At the end of the semester, students will have the opportunity to re-submit two different deliveries for retaking, and they will also be able to retake the Test. After the retakes, the maximum grade which can be obtained is 8.

Project Grade

The project carries an essential weight in the overall mark of the subject. Developing the project requires that the students work in groups and design an integral solution to thedefined challenge. In addition, the students must demonstrate their teamwork skills and present the results to the class.

The project is evaluated through its report, an oral presentation that students will present in class, and an individual-evaluation process. The participation of students in all three activities (preparing the report, presentation and individual evaluation) is necessary in order to obtain a projects grade. The grade is calculated as follows:

$$\text{Project Grade} = 0.6 * \text{Report} + 0.3 * \text{Presentation} + 0.1 * \text{Individual evaluation}$$

If performing the above calculation yields >= 5 but the student did not participate in any of the activities (report, presentation, individual evaluation), then a final grade of 4.5 will be given to the corresponding project.

There will be a retake of the project in case the final project grade does not reach the minimum of 4. In case of copy, there will be no recovery and the subject will be considered failed. The maximum project grade that can be obtained in case of retake is 7.

Bibliografía

- D. Jurafsky, J.H. Martin. *Speech and Language Processing.* Third Edition. 2021 < https://web.stanford.edu/~jurafsky/slp3/>
- R.S.T. Lee. *Natural Language Processing.* 2024. Springer
- G. Paab. *Foundation Models for Natural Language Processing.* 2023. Springer
- J. Eisenstein. *Natural Language Processing.* 2018. MIT Press
- H. Lane, C. Howard, H. M. Hapke. *Natural Language Processing in Action.* 2019. Manning Publications
- Kenny, Dorothy, ed. *Machine translation for everyone.* Lanugage Science Press, 2022. < https://langsci-press.org/catalog/book/342>
- Rowe, Bruce M., and Diane P. Levine. *A concise introduction to linguistics.* Routledge, 2018.

Software

For the problems and projects of the course we will use Python, along with some Python libraries for NLP that will be specified during the course.

Lista de idiomas

Nombre	Grupo	Idioma	Semestre	Turno
(PAUL) Prácticas de aula	1	Inglés	segundo cuatrimestre	tarde
(PLAB) Prácticas de laboratorio	1	Inglés	segundo cuatrimestre	tarde
(TE) Teoría	1	Inglés	segundo cuatrimestre	tarde