UAB
Universitat Autònoma de Barcelona

Aprenentatge i Processament del Llenguatge Natural

Codi: 106585
Crèdits: 6

2024/2025

Titulació	Tipus	Curs
2504392 Intel·ligència Artificial / Artificial Intelligence	OT	3
2504392 Intel·ligència Artificial / Artificial Intelligence	OT	4

Professor/a de contacte

Nom: Joaquin Cerdà Company

Correu electrònic: joaquin.cerda@uab.cat

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

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.

Objectius

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.

Competències

Intel·ligència Artificial / Artificial Intelligence
- Analitzar i resoldre problemes de manera efectiva, i generar propostes innovadores i creatives per aconseguir els objectius.
- Conèixer i aplicar les tècniques del processament del llenguatge natural per a lexplotació de dades de naturalesa lingüística i per a la creació i avaluació dels components dels sistemes dIA basats en el llenguatge.
- Conèixer i utilitzar de manera eficient les tècniques i eines de representació, manipulació, anàlisi i gestió de dades a gran escala.
- Dissenyar, implementar, analitzar i validar solucions algorítmiques eficients i robustes per a problemes computacionals derivats del disseny de sistemes intel·ligents.

- Elaborar estratègies per formular i solucionar diferents problemes daprenentatge de manera científica, creativa, crítica i sistemàtica, coneixent les capacitats i limitacions dels diferents mètodes i eines existents.
- Identificar, analitzar i avaluar limpacte ètic i social, el context humà i cultural i les implicacions legals del desenvolupament daplicacions dintel·ligència artificial i de manipulació de dades en diferents àmbits.
- Identificar, comprendre i aplicar els conceptes i tècniques fonamentals de representació del coneixement, raonament i aprenentatge computacional més adequats per a la solució de problemes dintel·ligència artificial.
- Introduir canvis en els mètodes i els processos de làmbit de coneixement per donar respostes innovadores a les necessitats i demandes de la societat.
- Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de lequip.

Resultats d'aprenentatge

1. Analitzar i resoldre problemes de manera efectiva, i generar propostes innovadores i creatives per aconseguir els objectius.
2. Conèixer i comprendre les solucions actuals a tasques de processament del llenguatge natural per a extracció dinformació, traducció i resum automàtics, i sistemes de diàleg.
3. Conèixer, aplicar i dissenyar tècniques daprenentatge computacional per a problemes de processament del llenguatge natural.
4. Dissenyar i aplicar tècniques de raonament semàntic.
5. Dissenyar i avaluar solucions basades en aprenentatge per a la detecció i extracció de patrons lingüístics i semàntics en col·leccions massives de textos.
6. Dissenyar i implementar tècniques eficients de cerca de patrons lingüístics i semàntics en col·leccions massives de textos.
7. Dissenyar models daprenentatge per a sistemes de processament del llenguatge natural a partir de la teoria i de tècniques generals daprenentatge computacional.
8. Dissenyar, implementar i analitzar solucions algorítmiques per al tractament massiu de dades textuals amb anotacions lingüístiques.
9. Identificar, analitzar i avaluar el biaix de models predictius de processament del llenguatge natural.
10. Ponderar els riscos i les oportunitats de les propostes de millora tant pròpies com alienes.
11. Proposar nous mètodes o solucions alternatives fonamentades.
12. Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de lequip.

Continguts

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

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de teoria	15	0,6	9, 2, 3

Sessió d'exercicis	25	1	1, 4, 5, 8, 7, 6, 9, 3, 12
Sessió de projecte	6	0,24	1, 7, 9, 11, 2, 3, 10, 12
Tipus: Supervisades			
Treball en el projecte	50	2	1, 4, 7, 9, 11, 2, 3, 10, 12
Tipus: Autònomes			
Estudi individual	25	1	1, 7, 2, 3
Resolució d'exercicis	25	1	1, 5, 8, 7, 6, 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.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge

Portfoli	49%	0	0	1, 4, 5, 8, 7, 6, 2, 3, 10, 12
Projecte	30%	2	0,08	1, 7, 9, 11, 2, 3, 10, 12
Test	21%	2	0,08	1, 5, 8, 7, 6, 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 the defined 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.

Bibliografia

- 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.

Programari

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.

Llista d'idiomes

Nom	Grup	Idioma	Semestre	Torn
(PAUL) Pràctiques d'aula	1	Anglès	segon quadrimestre	tarda
(PLAB) Pràctiques de laboratori	1	Anglès	segon quadrimestre	tarda
(TE) Teoria	1	Anglès	segon quadrimestre	tarda