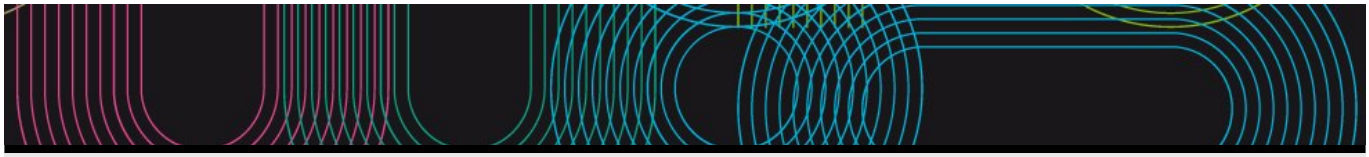

| IDENTIFYING DATA | | | | |
|---|---|---|---|---|
| **Language modelling** | | | | |
| Subject | Language modelling | | | |
| Code | O06M193V01204 | | | |
| Study programme | Máster universitario en Inteligencia artificial | | | |
| Descriptors | ECTS Credits | Choose | Year | Quadmester |
| | 3 | Optional | 1st | 2nd |
| Teaching language | English | | | |
| Department | | | | |
| Coordinator | Darriba Bilbao, Víctor Manuel | | | |
| Lecturers | A0075-Ax2tc-1 A0075-Ax2tc-1, A0075-Ax2tc-1 A0075-Ax2tc-2 A0075-Ax2tc-2, A0075-Ax2tc-2 Darriba Bilbao, Víctor Manuel | | | |
| E-mail | darriba@uvigo.es | | | |
| Web | http://guiadocente.udc.es/guia_docent/index.php?centre=614&ensenyament=614544&consulta=assignatures&any_academic=2023_24 | | | |
| General description | The course introduces the student to the modeling of human language, i.e. the generation of models that allow estimating the plausibility of a text, an essential phase in the design of any application based on the exploitation of its communicative mechanisms. The student will be trained to master the theoretical principles and techniques that allow their construction, both those based on discrete representations of the elements of the language and on continuous representations. Special attention will be given to modeling in contexts of scarce linguistic resources, whether this is motivated by the treatment of novel domains of knowledge or by the exploitation of languages of limited diffusion. | | | |

| Training and Learning Results | |
|---|---|
| Code | |
| A1 | CB6 - Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context |
| A2 | CB7 - Students should be able to apply their acquired knowledge and problem-solving skills in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study. |
| A5 | CB10 - That students possess the learning skills that will enable them to continue studying in a manner that will be largely self-directed or autonomous. |
| B1 | Maintain and extend sound theoretical approaches to enable the introduction and exploitation of new and advanced technologies in the field of Artificial Intelligence. |
| B3 | Search and select useful information needed to solve complex problems, handling with fluency the bibliographic sources of the field. |
| B4 | Elaborate adequately and with certain originality written compositions or motivated arguments, write plans, work projects, scientific articles and formulate reasonable hypotheses in the field. |
| C1 | Understanding and mastering techniques for text processing in natural language |
| C2 | Understanding and mastery of the fundamentals and techniques of semantic processing of linked, structured, and unstructured documents, and of the representation of their content. |
| C3 | Understanding and knowledge of the techniques of representation and processing of knowledge through ontologies, graphs, and RDF, as well as the tools associated with them. |
| D2 | Master the oral and written expression and comprehension of a foreign language. |
| D3 | Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida. |
| D7 | Develop the ability to work in interdisciplinary or transdisciplinary teams to offer proposals that contribute to sustainable environmental, economic, political and social development. |
| D8 | Value the importance of research, innovation and technological development in the socioeconomic and cultural progress of society. |

| Expected results from this subject | |
|---|---|
| Expected results from this subject | Training and Learning Results |
| To know how to use the techniques and methods of natural language processing to solve real problems of analysis of texts in natural language. | A2<br>A5<br>B1<br>B3<br>B4<br>C1<br>C3<br>D2<br>D3<br>D7 |
| To know, understand and analyze deep learning techniques applied to natural language processing. | A1<br>A2<br>A5<br>B1<br>B3<br>C1<br>C2<br>D2<br>D3<br>D7<br>D8 |
| To know how to use deep learning techniques and methods to solve practical problems in natural language processing. | A1<br>A2<br>A5<br>B1<br>B3<br>B4<br>C1<br>C2<br>D2<br>D3<br>D7<br>D8 |
| To know and understand the environmental problems posed by the computational cost of deep learning techniques when applied to text analysis | A1<br>B1<br>C1<br>C2<br>D2<br>D8 |

| Contents | |
|---|---|
| Topic | |
| Language models | N-gram based language models<br>Neural based language models |
| Distributional semantics models | Linguistic hypothesis about distributional meaning<br>Classic models of distributional semantics<br>Neural models representing static meaning (word embeddings)<br>Neural models representing dynamic-contextual meaning<br>Compositional models |
| Sequence labeling | Use and fine-tuning of models for sequence labeling |
| Text-To-Text models | Text-To-Text models |

| Planning | | | |
|---|---|---|---|
| | Class hours | Hours outside the classroom | Total hours |
| Lecturing | 10 | 10 | 20 |
| Laboratory practical | 5 | 15 | 20 |
| Project based learning | 6 | 28 | 34 |
| Objective questions exam | 1 | 0 | 1 |

*The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

| Methodologies | |
|---|---|
| | Description |

| | |
|---|---|
| Lecturing | Teachers present a topic to students with the aim of providing a set of information with a specific scope. |
| Laboratory practical | The teachers of the discipline present to the students one or more practical problems that require the comprehension and application of the theoretical and practical contents included in the syllabus of the subject. Students can work on solving problems individually or as a team. These activities may require autonomous work, although guided by the teacher of the subject. |
| Project based learning | Students are presented with practical projects that require an important part of their total dedication to the topic. In addition, and due to the scope of the work to be performed, it is necessary for the student to use not only management skills, but also technical skills. |

**Personalized assistance**

| Methodologies | Description |
|---|---|
| Laboratory practical | The teachers will attend the students in individualized mentoring sessions, dedicated to the orientation in the study and to the resolution of doubts on the contents, duties and activities of the discipline. |
| Project based learning | The teachers will attend the students in individualized mentoring sessions, dedicated to the orientation in the study and to the resolution of doubts on the contents, duties and activities of the discipline. |

**Assessment**

| | Description | Qualification | Training and Learning Results | | | |
|---|---|---|---|---|---|---|
| Lecturing | Continuous monitoring of class attendance and student participation | 5 | A1 | B1<br>B4 | C1<br>C2<br>C3 | D2<br>D8 |
| Project based learning | Deliveries of proyects must be made within the period established in the virtual campus and must follow the specifications indicated in the assigment both for their presentation and their defense.<br>Mandatory. | 50 | A1<br>A2<br>A5 | B1<br>B3<br>B4 | C1<br>C2<br>C3 | D2<br>D3<br>D7<br>D8 |
| Objective questions exam | Mastery of theoretical and operational knowledge of the subject will be assessed.<br>Mandatory. | 45 | A1 | B1<br>B4 | C1<br>C2<br>C3 | D2<br>D8 |

**Other comments on the Evaluation**

**EVALUATION CRITERIA FOR ALL STUDENTS IN ALL OPPORTUNITIES**

Students must achieve a minimum of 40% of the maximum mark of the "Laboratory Practices" and "Objective Test" parts, and in any case the sum of the three parts must be greater than 5 to pass the subject. If any of the above requirements is not met, the grade for the course will be established according to the lowest grade obtained.

In case of not reaching the minimum score in the "Laboratory Practices" or "Objective Test" parts, the student will have a second opportunity in which only the delivery of the failed part will be required.

Grades will not be saved between academic years.

The delivery of the practicals must be done within the deadline established in the virtual campus and must follow the specifications indicated in the statement for both its presentation and defense.

The student who submits all the compulsory practicals or attends the objective test in the official evaluation period will be considered "Presented".

In the case of fraudulent completion of exercises or tests, the Regulations for the evaluation of the academic performance of students and review of grades will be applied. In application of the corresponding regulations on plagiarism, the total or partial copy of any practice or theory exercise will result in suspension on both occasions of the course, with a grade of 0.0 in both cases.

**EXAM DATES**

The official exam dates for the different opportunities, will be published on the ESEI

website: https://esei.uvigo.es/docencia/exames/

**CONSULTATION/REQUEST OF TUTORING SESSIONS**

Tutoring sessions schedules can be consulted through the faculty's personal page, available at

## Sources of information

### Basic Bibliography

Jurafsky, Daniel & James H. Martin, **N-gram Language Models.**, https://web.stanford.edu/~jurafsky/slp3/, 2022

Jurafsky, Daniel & James H. Martin, **Vector Semantics and Embeddings**, https://web.stanford.edu/~jurafsky/slp3/, 2022

Jurafsky, Daniel & James H. Martin, **Neural Networks and Neural Language Models**, https://web.stanford.edu/~jurafsky/slp3/, 2022

Jurafsky, Daniel & James H. Martin, **Sequence Labeling for Parts of Speech and Named Entities**, https://web.stanford.edu/~jurafsky/slp3/, 2022

### Complementary Bibliography

Baroni, Marco, Raffaella Bernardi & Roberto Zamparelli, **Frege in space: A program for compositional distributional semantics**, Linguistic Issues in Language Technologies 9(6): 5-110, University of Colorado Boulder, 2014

Baroni, Marco, Georgiana Dinu & Germán Kruszewski, **Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors**, Proceedings of the 52nd Annual Meeting of the ACL (Vol. 1), Association for Computational Linguistics, 2014

Church, Kenneth Ward, Zeyu Chen & Yanjun Ma, **Emerging trends: A gentle introduction to fine-tuning**, Natural Language Engineering, 27, Cambridge University Press, 2021

Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova, **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**, Proceedings of the 2019 Conference of the NACL-HLT (Vol. 1), Association for Computational Linguistics, 2019

Erk, Katrin, **Vector space models of word meaning and phrase meaning: A survey**, Language and Linguistics Compass 6.10, Wiley & Sons, 2012

Linzen, Tal, **Issues in evaluating semantic spaces using word analogies**, Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, Association for Computational Linguistics, 2016

Hirschberg, Julia & Manning, Christopher D., **Advances in natural language processing**, Science 349.6245, AAAS, 2015

Mikolov, Tomas, Wen-tau Yih & Zweig Geoffrey, **Linguistic Regularities in Continuous Space Word Representations**, Proceedings of the 2013 Conference of the NACL-HLT, Association for Computational Linguistics, 2013

Taher Pilehvar, Mohammad & Jose Camacho-Collados, José, **Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning**, Computational Linguistics, 47(3), MIT Press, 2021

## Recommendations

### Subjects that continue the syllabus

Web intelligence and semantic technologies/O06M193V01205
Text mining/O06M193V01302

### Subjects that are recommended to be taken simultaneously

Machine learning II/O06M193V01207
Deep learning/O06M193V01206

### Subjects that it is recommended to have taken before

Machine learning I/O06M193V01105
Natural language understanding/O06M193V01104

### Other comments

All students are reminded of the prohibition of the use of mobile devices in exercises, practices and exams, in compliance with article 13.2.d) of the Statute of the University Student, regarding the duties of the university student body, which establishes the duty to "Refrain from using or cooperating in fraudulent procedures in assessment tests, in the work carried out or in official university documents."