Teaching Guide				
Identifying Data				2022/23
Subject (*)	Natural Language Understanding		Code	614544008
Study programme	Máster Universitario en Intelixencia Artificial			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	First	Obligatory	6
Language	English			
Teaching method	Face-to-face			
Prerequisites				
Department	Ciencias da Computación e Tecnoloxías da Información			
Coordinador	Gómez Rodríguez, Carlos	E-mail	carlos.gomez@udc.es	
Lecturers	Gómez Rodríguez, Carlos Vilares Calvo, David Vilares Ferro, Jesus	E-mail	carlos.gomez@udc.es david.vilares@udc.es jesus.vilares@udc.es	
Web	campusvirtual.udc.es			
General description	The course introduces the basic concepts and techniques associated with natural language processing, the starting point for the design of information exploitation and dialogue environments based on human language, both at the lexical and syntactic, semantic and pragmatic levels. The objective is to introduce the student to the complexity inherent to the analysis of human natural language, mainly associated to the ambiguity and contextual dependencies it presents, and to the design of data structures and algorithms that allow its practical treatment.			

Study programme competences / results	
Code	Study programme competences / results
A2	CE01 - Understanding and command of techniques for lexical, syntactic and semantic processing of text in natural language
A3	CE02 - Understanding and command of fundamentals and techniques for processing linked documents, both structured and unstructured, and of the representation of their contents
A4	CE03 - Understanding and knowledge of the techniques for knowledge representation and processing for ontologies, graphs and RDF, together with their associated tools
B1	CG01 - Maintaining and extending theoretical foundations to allow the introduction and exploitation of new and advanced technologies in the field of AI
B3	CG03 - Searching and selecting that useful information required to solve complex problems, with a confident handling of bibliographical sources in the field
B4	CG04 - Suitably elaborating written essays or motivated arguments, including some point of originality, writing plans, work projects, scientific papers and formulating reasonable hypotheses in the field
B6	CB01 - Acquiring and understanding knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, frequently in a research context
B7	CB02 - The students will be able to apply the acquired knowledge and to use their capacity of solving problems in new or poorly explored environments inside wider (or multidisciplinary) contexts related to their field of study
B10	CB05 - The students will acquire learning abilities to allow them to continue studying in way that will mostly be self-directed or autonomous
C2	CT02 - Command in understanding and expression, both in oral and written forms, of a foreign language
C3	CT03 - Use of the basic tools of Information and Communications Technology (ICT) required for the student's professional practice and learning along her life
C7	CT07 - Developing the ability to work in interdisciplinary or cross-disciplinary teams to provide proposal that contribute to a sustainable environmental, economic, political and social development
C8	CT08 - Appreciating the importance of research, innovation and technological development in the socioeconomic and cultural progress of society

Learning outcomes			
Learning outcomes	Study programme competences / results		
To know, understand and analyze the formal representation of diverse lexical, syntactic and semantic phenomena of natural language.	AC1	BC1 BC3 BC4 BC6 BC10	CC2 CC8
To know, understand and know how to use the technologies, frameworks and libraries for the construction of natural language processing systems.	AC1 AC2	BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7
To design, implement and know how to use algorithms and data structures to treat and support the various phenomena characteristic of natural language.	AC1 AC2 AC3	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8
To know, understand and analyze natural language processing techniques for processing and disambiguation at the lexical, syntactic and semantic levels.	AC1 AC2 AC3	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8
To know and understand the problems posed by ambiguity and imprecision in natural language data sources and techniques to solve them.	AC1 AC2	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8

Contents	
Topic	Sub-topic
Introduction.	Levels of analysis. Ambiguity and contextual dependencies.
Lexical analysis.	Segmentation. Dictionaries and thesauri. Part-of-speech tagging.
Syntactic parsing.	Algebraic grammars. Mildly context-sensitive grammars. Dependency grammars. Probabilistic grammars.
Semantic parsing.	Lexical semantics. Semantic dependencies. Semantic graphs.

Planning

Methodologies / tests	Competencies / Results	Teaching hours (in-person & virtual)	Student?s personal work hours	Total hours
Guest lecture / keynote speech	A2 A3 A4 B1 B3 B6 B7 B10 C2 C8	21	21	42
Laboratory practice	A2 A3 A4 B3 B4 B6 B7 B10 C2 C3 C7 C8	14	48	62
Problem solving	A2 A3 A4 B3 B4 B6 B7 B10 C2	7	25	32
Objective test	A2 A3 A4 B1 B6 B7 C2	3	9	12
Personalized attention		2	0	2
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description
Guest lecture / keynote speech	Theoretical classes, in which the content of each topic is exposed. The student will have copies of the slides beforehand and the professor will promote an active attitude, asking questions to clarify specific aspects and leaving open questions for the student's reflection.
Laboratory practice	Practical classes with the use of computers, which allow the student to familiarize himself/herself from a practical point of view with the issues presented in the theoretical classes.
Problem solving	Problem-based learning, seminars, case studies and projects.
Objective test	The mastery of the theoretical and operating knowledge of the subject will be evaluated.

Personalized attention	
Methodologies	Description
Guest lecture / keynote speech Laboratory practice Problem solving Objective test	The development of the master classes, as well as of the problem solving classes and the practical laboratories, will be carried out according to the progress of the students in the comprehension and assimilation of the contents taught. The general progress of the class will be combined with a specific attention to those students who present greater difficulties in the task of learning and with an additional support to those who present greater fluency and wish to broaden their knowledge. With regard to individual tutorials, given their personalized nature, they should not be devoted to extend the contents with new concepts, but to clarify the concepts already exposed. The teacher will use them as an interaction that will allow them to draw conclusions regarding the degree of assimilation of the subject by the students.

Assessment			
Methodologies	Competencies / Results	Description	**Qualification**
Laboratory practice	A2 A3 A4 B3 B4 B6 B7 B10 C2 C3 C7 C8	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 presentation and defense.	40
Objective test	A2 A3 A4 B1 B6 B7 C2	Compulsory realization. The mastery of the theoretical and operative knowledge of the subject will be evaluated.	60

Assessment comments

Students must achieve at least 40% of the maximum grade for each part (theory, practice) and in any case the sum of both parts must reach a 5 to pass the course. 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 grade in one of the parts, the student will have a second opportunity in which only the delivery of that part will be required. Grades will not be kept 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 performance of exercises or tests, the Regulations for the evaluation of students' academic performance and review of qualifications will be applied. In application of the corresponding regulations on plagiarism, the total or partial copy of any practical or theory exercise will result in failure in both opportunities of the course, with a grade of 0.0 in both cases.

Sources of information	
Basic	- Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT Press - Goldberg, Y. (2017). Neural network methods for natural language processing. Synthesis lectures on human language technologies. Morgan Claypool - Jacob Eisenstein (2019). Introduction to Natural Language Processing. MIT Press - Jurafsky, D. & Martin, J. H. (2022). Speech and Language Processing (3rd ed. draft). Disponible en: https://web.stanford.edu/~jurafsky/slp3/ - Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT Press
Complementary	- Chollet, F. (2018). Keras: The python deep learning library. Astrophysics Source Code Library - Stuart Russell, Peter Norvig (2020). Artificial Intelligence: A Modern Approach, 4th Edition. Pearson - Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency Parsing. Synthesis lectures on human language technologies. Morgan Claypool

Recommendations
Subjects that it is recommended to have taken before
Subjects that are recommended to be taken simultaneously
Machine Learning I /614544012
Subjects that continue the syllabus
Text Mining/614544011 Language Modelling/614544009 Web Intelligence and Semantic Technologies/614544010
Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.