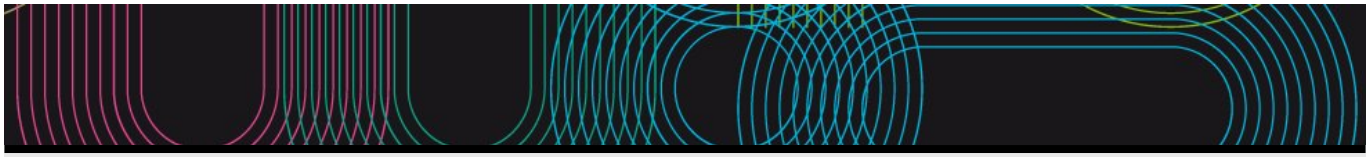

IDENTIFYING DATA				
Text mining				
Subject	Text mining			
Code	O06M193V01302			
Study programme	Máster universitario en Inteligencia artificial			
Descriptors	ECTS Credits	Choose	Year	Quadmester
	3	Optional	2nd	1st
Teaching language	English			
Department				
Coordinator	Ribadas Pena, Francisco José			
Lecturers	A0075-Ax2tc-1 A0075-Ax2tc-1, A0075-Ax2tc-1 A0075-Ax2tc-2 A0075-Ax2tc-2, A0075-Ax2tc-2 Ribadas Pena, Francisco José			
E-mail	ribadas@uvigo.es			
Web	http://moovi.uvigo.gal			
General description	The course introduces the student to the derivation of information and knowledge from the analysis of a collection of documents in natural language, which refers to almost all generated and stored data. The student will be trained in content analysis on enriched document representation models to address specific applications in different domains. Special attention will be paid to the extraction of relevant information, the determination of the contextual polarity (sentiment) of a content, and the automatic response to questions posed directly in natural language. In short, the goal is to answer fundamental questions in the development of interfaces, decision support environments, and access to new knowledge.			

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
New	A1 A2 A5 B1 B3 B4 C1 C2 C3 D2 D3 D7 D8

Contents	
Topic	
Document analysis	Concepts and definitions. Plot structure, coherence, and co-references.
Information retrieval and extraction.	Concepts and definitions. IR (Information Retrieval) techniques and tools. IE (Information Extraction) techniques and tools.
Sentiment analysis	Concepts and definitions. Techniques and tools. Current trends.
Question answering	Concepts and definitions. Typical architectures, technical and tools. Current trends.
Other text mining applications.	Emerging tasks. Text mining in specific domains.

Planning			
	Class hours	Hours outside the classroom	Total hours
Lecturing	10	10	20
Laboratory practical	5	15	20
Mentored work	5	29	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	Presentation of the theoretical content of the course. In order to facilitate the understanding of the same and to increase the interest of the student, various examples and exercises will be included, in which the active participation of the student may be required. Various examples and exercises in which the active participation of the student may be required. An active attitude is promoted by encouraging questions and proposing open-ended questions for the student's reflection.
Laboratory practical	Hands-on problems that involve the use of specific tools and the programming of software related to the course content. CONTINUOUS ASSESSMENT Character: mandatory Attendance: not mandatory GLOBAL ASSESSMENT Character: mandatory

Mentored work	One or more individual theoretical-practical works, deliverable and evaluable, on the theoretical aspects presented in the course and worked on in the practical activities developed by the students. This is an autonomous task that will have occasional guidance from the teacher. The result will be expressed in one or more reports with a structure to be determined. CONTINUOUS ASSESSMENT Character: mandatory Attendance: not mandatory GLOBAL ASSESSMENT Character: mandatory

Personalized assistance

Methodologies	Description
Laboratory practical	The teacher will guide the student in the laboratory for the realization of the projects that will be evaluated at the end of the course, answering doubts and questions individually.
Mentored work	Follow-up of students' work, solving general doubts and sharing specific theoretical/practical problems related to the course topics.

Assessment

	Description	Qualification	Training and Learning Results			
Laboratory practical	Evaluation of the proposed hands-on exercises by submitting a writen report and/or the developed code. The submission of these exercises is mandatory. They will have a delivery date and, optionally, a defense date. - MINIMUM SCORE: 4 points out of 10 - LEARNING OUTCOMES: RA1	40	A2	B1	C1 C2 C3	D3 D8
Mentored work	Evaluation of the writen report of the supervised research work (or works) of theoretical-practical nature assigned to each student. The ability to synthesize, completeness and adequate presentation of ideas and concepts related to the chosen topic will be evaluated. The submission of these papers is compulsory. They will have a due date and, optionally, a defense date - MINIMUM SCORE: 4 points out of 10 - LEARNING OUTCOMES: RA1	40	A1 A5	B3 B4	C1 C2 C3	D2 D8
Objective questions exam	Written test that evaluates the content and competencies reviewed in the lectures and the theoretical aspects of their implementation in the practical sessions. The type of test will consist of a series of multiple choice or short answer questions on specific concepts. It will take place on the official date indicated in the academic calendar. - MINIMUM SCORE: no minimum score required - LEARNING OUTCOMES: RA1	20	A1 A2	B1	C1 C2 C3	

Other comments on the Evaluation

(1) CONTINUOUS ASSEMENT SYSTEM

TEST 1: Practical hand-on exercises

Description: Evaluation of the writen reports and the code of the laboratory hand-on exercises delivered on the

stipulated dates.

Applied methodology: Laboratory practical

% Qualification: 40%

Minimum %: grade equal to or greater than 4 points (out of 10)

Evaluated learning results: A2, B1, C1, C2, C3, D3, D8

Expected results: RA1

TEST 2: Tutored work/essay

Description: Evaluation of the writen report of the supervised research work (or works) of theoretical-practical nature

assigned to each student.

Applied methodology: Mentored work

% Qualification: 40%

Minimum %: grade equal to or greater than 4 points (out of 10)

Evaluated learning results: A2, A5, B3, B4, C1, C2, C3, D2, D8

Expected results: RA1

TEST 3: Final exam

Description: Multiple-choice final test on the theoretical contents of the subject.

Applied methodology: Objective questions exam

% Qualification: 20%

Minimum %: no minimum score required

Evaluated learning results: : A1, A2, B1, C1, C2, C3

Expected results: RA1

ADDITIONAL CLARIFICATIONS

- To pass the subject it is necessary to reach the minimums indicated in the previous tests and to add in the final weighted grade a minimum of 5 points out of 10.
- In the case of finding unethical behavior (copying, plagiarism) in any of the deliveries made (total or partial), the total contribution of the corresponding evaluation element on the final grade will be annulled.

(2) GLOBAL ASSEMENT SYSTEM

Procedure for the choice of the global assessment modality:

- The continuous assessment modality is assumed by default.
- Students who opt for the global evaluation must communicate it via Moovi, using the mechanisms that are enabled and within the stipulated period, once the period of one month from the beginning of the term has passed.

TEST 1: Practical hand-on exercises

Description: Evaluation of the writen reports and the code of the laboratory hand-on exercises delivered on the

stipulated dates.

Applied methodology: Laboratory practical

% Qualification: 40%

Minimum %: grade equal to or greater than 5 points (out of 10)

Evaluated learning results: A2, B1, C1, C2, C3, D3, D8

Expected results: RA1

TEST 2: Tutored work/essay

Description: Evaluation of the writen report of the supervised research work (or works) of theoretical-practical nature

assigned to each student.

Applied methodology: Mentored work

% Qualification: 40%

Minimum %: grade equal to or greater than 5 points (out of 10)

Evaluated learning results: A2, A5, B3, B4, C1, C2, C3, D2, D8

Expected results: RA1

TEST 3: Final exam

Description: Multiple-choice final test on the theoretical contents of the subject.

Applied methodology: Objective questions exam

% Qualification: 20%

Minimum %: grade equal to or greater than 5 points (out of 10)

Evaluated learning results: : A1, A2, B1, C1, C2, C3

Expected results: RA1

ADDITIONAL CLARIFICATIONS

- To pass the subject it is necessary to reach the minimums indicated in the previous tests and to add in the final weighted grade a minimum of 5 points out of 10.
- In the case of finding unethical behavior (copying, plagiarism) in any of the deliveries made (total or partial), the total contribution of the corresponding evaluation element on the final grade will be annulled.

(3) ASSESSMENT CRITERIA FOR EXTRAORDINARY AND FINAL CALLS

- The continuous and global evaluation systems described above will be used.

- In these calls, students must only take the tests in which they have not obtained the minimum grade indicated.

(4) GRADING PROCESS

In the case of students who pass part of the evaluated elements, but do not reach the minimum required to pass the whole subject, the grade to be included in the respective minutes will be calculated as the minimum between the weighted average of the parts passed and 4.9.

(5) EVALUATION DATES

The official exam dates of the different calls, officially approved by the Xunta de Centro of the ESEI, are published on the ESEI website https://esei.uvigo.es/docencia/horarios/.

(6) USE OF MOBILE DEVICES

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

(7) TUTORING SCHEDULE AND PERSONAL TUTORING REQUEST

The tutoring schedule, and the way to request a personal tutoring, is published in the personal page of the teaching staff, accessible through https://esei.uvigo.es/docencia/profesorado/.

Sources of information

Basic Bibliography

Apuntes y material proporcionado por el profesorado.,

Berry, M. W., & Kogan, J. (Eds.)., **Text mining: applications and theory.**, 978-0-470-74982-1, 1, John Wiley & Sons., 2010

Complementary Bibliography

Taeho Jo, **Text Mining: Concepts, Implementation, and Big Data Challenge (Studies in Big Data Book 45)**, 978-3319918143, 1, Springer, 2019

Recommendations

Subjects that it is recommended to have taken before

Natural language understanding/O06M193V01104
Language modelling/O06M193V01204

Other comments

Course coordinated by the University of Vigo