



## Teaching Guide

Identifying Data					2023/24
Subject (*)	Text Mining	Code	614544011		
Study programme	Máster Universitario en Intelixencia Artificial				
Descriptors					
Cycle	Period	Year	Type	Credits	
Official Master's Degree	1st four-month period	Second	Optional	3	
Language	English				
Teaching method	Face-to-face				
Prerequisites					
Department					
Coordinador	Cabalar Fernandez, Jose Pedro	E-mail	pedro.cabalar@udc.es		
Lecturers	Cabalar Fernandez, Jose Pedro	E-mail	pedro.cabalar@udc.es		
Web					
General description	<p>La asignatura introduce al alumno en la derivación de información y conocimiento a partir del análisis de una colección de documentos en lengua natural, lo que refiere a la práctica totalidad de datos generados y almacenados. Se le capacitará en el análisis del contenido sobre modelos de representación documental enriquecidos, con el fin de abordar aplicaciones concretas sobre distintos dominios. Merecerán especial atención la extracción de información relevante, la determinación de la polaridad contextual (sentimiento) deducible a partir de un contenido y la respuesta automática a preguntas planteadas directamente en lengua natural. Se trata en definitiva de dar respuesta a cuestiones fundamentales en el desarrollo de interfaces, entornos de ayuda a la decisión y acceso a nuevo conocimiento.</p>				

## Study programme competences

Code	Study programme competences
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		
Knowing, understanding and analysing the current techniques for web search and mining	AC1 AC2 AC3	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8
Designing, implementing and knowing algorithms and structures for recommender systems	AC1 AC2	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8
Knowing how to apply different models of information retrieval and extraction, sentiment analysis and other possible applications of text mining	AC1 AC2 AC3	BC1 BC3 BC4 BC6 BC7 BC10	CC2 CC3 CC7 CC8

Contents	
Topic	Sub-topic
Análisis de documentos.	Conceptos y definiciones. Estructura argumental, coherencia y co-referencias.
Recuperación y extracción de información	Conceptos y definiciones. Técnicas y herramientas de RI (Recuperación de Información) Técnicas y herramientas de EI (Extracción de Información)
Análisis del sentimiento	Conceptos y definiciones. Técnicas y herramientas. Tendencias actuales.
Búsqueda de respuestas.	Conceptos y definiciones. Arquitecturas típicas, técnicas y herramientas. Tendencias actuales.
Otras aplicaciones de la minería de textos.	Problemas emergentes. Text mining en dominios específicos.

Planning				
Methodologies / tests	Competencias	Ordinary class hours	Student?s personal work hours	Total hours
Laboratory practice	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	5	15	20
Supervised projects	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	5	29	34
Objective test	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	1	0	1



Guest lecture / keynote speech	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	10	10	20
Personalized attention		0		0
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description
Laboratory practice	Realización de problemas de carácter práctico que incluyen el empleo de herramientas específicas y la programación de software relacionado con los contenidos de la materia. EVALUACION CONTINUA Caracter: Obligatorio Asistencia: No obligatoria EVALUACION GLOBAL Caracter: Obligatorio
Supervised projects	Conjunto de uno o más trabajos teórico-prácticos individuales, entregables y evaluables, sobre los aspectos teóricos presentados en la materia y trabajados en las actividades prácticas desarrolladas por los alumnos. Se trata de una tarea autónoma que contará con la tutorización puntual del profesorado. El resultado se plasmará en una o más memorias con la estructura que se determine. EVALUACION CONTINUA Caracter: Obligatorio Asistencia: No obligatoria EVALUACION GLOBAL Caracter: Obligatorio
Objective test	Examen escrito
Guest lecture / keynote speech	Exposición de los contenidos teóricos de la materia. Con el fin de facilitar la comprensión de la misma y aumentar el interés del alumno, se incluirán diversos ejemplos y ejercicios en los que se puede requerir la participación activa del alumno. Se promoverá una actitud activa, fomentando la realización de preguntas y proponiendo cuestiones abiertas para la reflexión del alumno,

Personalized attention	
Methodologies	Description
Laboratory practice Supervised projects	El profesor tutorizará al alumno en el laboratorio para la realización de los proyectos que se evaluarán al final de la materia, respondiendo dudas individualmente  Seguimiento del trabajo de los alumnos, resolución de dudas generales y puesta en común de problemas específicos de carácter teórico/práctico relacionados con la materia

Assessment			
Methodologies	Competencies	Description	Qualification
Laboratory practice	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	Evaluación de las prácticas de laboratorio propuestas mediante la entrega de memoria y/o del código desarrollado. La entrega de estas prácticas es obligatoria. Tendrán una fecha de entrega y, opcionalmente, de defensa. - PUNTUACION MINIMA: 4 puntos sobre 10 - RESULTADOS APRENDIZAJE EVALUADOS: RA1	40



Supervised projects	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	Evaluación de la memoria del trabajo (o trabajos) de investigación tutelado, de caracets teórico-práctico, asignado a cada alumno. Se evaluará la capacidad de síntesis y la completitud y adecuada presentación de las ideas y conceptos relativos al tema escogido. La entrega de estos trabajos es obligatoria. Tendrán una fecha de entrega y, opcionalmente, de defensa. - PUNTUACION MINIMA: 4 puntos sobre 10 - RESULTADOS APRENDIZAJE EVALUADOS: RA1	40
Objective test	A2 A3 A4 B1 B3 B4 B6 B7 B10 C2 C3 C7 C8	Prueba escrita donde se evaluarán los contenidos y competencias revisados en las sesiones magistrales y los aspectos teóricos de su puesta en práctica llevada a cabo en las sesión prácticas. El tipo de prueba consistirá en un conjunto de preguntas tipo test o cuestiones de respuesta corta sobre conceptos concretos. Se realizará en la fecha oficial prevista en el calendario de la titulación. - PUNTUACION MINIMA: no hay mínimo exigido - RESULTADOS APRENDIZAJE EVALUADOS: RA1	20

Assessment comments



(1) SISTEMA DE EVALUACIÓN CONTÍNUA PRUEBA 1: Prácticas de laboratorio Descripción: Evaluación de la memoria de las prácticas de laboratorio entregadas en las fechas estipuladas. Metodología(s): Prácticas de laboratorio % Calificación: 40% % Mínimo: 4 puntos sobre 10 Páxina 4 de 6 Competencias evaluadas: A2, B1, C1, C2, C3, D3, D8 Resultados aprendizaje evaluados: RA1 PRUEBA 2: Trabajo tutelado Descripción: Evaluación de la memoria del trabajo(s) entregados en las fechas estipuladas y de supresentación en clase. Metodología(s): Trabajo tutelado % Calificación: 40% % Mínimo: 4 puntos sobre 10 Competencias evaluadas: A2, A5, B3, B4, C1, C2, C3, D2, D8 Resultados aprendizaje evaluados: RA1 PRUEBA 3: Examen final Descripción: Examen tipo test o de conceptos sobre los contenidos teóricos de la materia Metodología(s): Examen de preguntas objetivas % Calificación: 20% % Mínimo: no hay mínimo Competencias evaluadas: A1, A2, B1, C1, C2, C3 Resultados aprendizaje evaluados: RA1 ACLARACIONES ADICIONALES Para superar la materia es preciso alcanzar los mínimos indicados en las pruebas anteriores y sumar en la nota final ponderada un mínimo de 5 puntos sobre 10. En el caso de constatar un comportamiento no ético (copia, plagio) en alguna de las entregas realizadas (total o parcial), se anulará la totalidad de la contribución del correspondiente elemento de evaluación sobre la calificación final.

(2) SISTEMA DE EVALUACIÓN GLOBAL Procedimiento para la elección de la modalidad de evaluación global: Se asume por defecto la modalidad de evaluación continua. Los alumnos que opten por la evaluación global deberán comunicarlo via Moovi, mediante los mecanismos que se habiliten y en el plazo estipulado, una vez superado el plazo de un mes desde el comienzo del cuatrimestre PRUEBA 1: Prácticas de laboratorio Descripción: Evaluación de la memoria de las prácticas de laboratorio entregadas en las fechas estipuladas. Metodología(s): Prácticas de laboratorio % Calificación: 40% % Mínimo: 5 puntos sobre 10 Páxina 5 de 6 Competencias evaluadas: A2, B1, C1, C2, C3, D3, D8 Resultados aprendizaje evaluados: RA1 PRUEBA 2: Trabajo tutelado Descripción: Evaluación de la memoria del trabajo(s) entregados en las fechas estipuladas y de supresentación en clase. Metodología(s): Trabajo tutelado % Calificación: 40% % Mínimo: 5 puntos sobre 10 Competencias evaluadas: A2, A5, B3, B4, C1, C2, C3, D2, D8 Resultados aprendizaje evaluados: RA1 PRUEBA 3: Examen final Descripción: Examen tipo test o de conceptos sobre los contenidos teóricos de la materia Metodología(s): Examen de preguntas objetivas % Calificación: 20% % Mínimo: 5 puntos sobre 10 Competencias evaluadas: A1, A2, B1, C1, C2, C3 Resultados aprendizaje evaluados: RA1 ACLARACIONES ADICIONALES Para superar la materia es preciso alcanzar los mínimos indicados en las pruebas anteriores y sumar en la nota final ponderada un mínimo de 5 puntos sobre 10. En el caso de constatar un comportamiento no ético (copia, plagio) en alguna de las entregas realizadas (total o parcial), se anulará la totalidad de la contribución del correspondiente elemento de evaluación sobre la calificación final.

(3) CRITERIOS DE EVALUACIÓN PARA CONVOCATORIA EXTRAORDINARIA Y FIN DE CARRERA Se emplearán los sistemas de evaluación continua y global expuestos anteriormente. En estas convocatorias, los alumnos sólo deberán realizar las pruebas en las cuales no hubieran obtenido la calificación mínima indicada.

(4) PROCESO DE CALIFICACIÓN DE ACTAS En el caso de los alumnos que superen parte de los elementos evaluados, pero no alcancen el mínimo preciso para aprobarla materia completa, la calificación a incluir en las respectivas actas se calculará como el mínimo entre el promedio ponderado de las partes superadas y 4,9.

(5) FECHAS DE EVALUACIÓN El calendario de pruebas de evaluación aprobado oficialmente por la Junta de Centro de la ESEI se encuentra publicado en la página web <http://www.esei.uvigo.es>

(6) EMPLEO DE DISPOSITIVOS MÓVILES Se recuerda a todo el alumnado la prohibición del uso de dispositivos móviles en ejercicios y prácticas, en cumplimiento del artículo 13.2.d) del Estatuto del Estudiante Universitario, relativo a los deberes del estudiantado universitario, que establece el deber de "Abstenerse de la utilización o cooperación en procedimientos fraudulentos en las pruebas de evaluación, en los trabajos que se realicen o en documentos oficiales de la universidad."

(7) CONSULTA/SOLICITUD DE TUTORÍAS Las tutorías pueden consultarse a través de la página personal del profesorado, accesible a través de <https://esei.uvigo.es/docencia/profesorado/>

## Sources of information

<b>Basic</b>	- Berry, M. W., & Kogan, J. (Eds.), (2010). Text mining: applications and theory. John Wiley & Sons - Taeho Jo (2019). Text Mining: Concepts, Implementation, and Big Data Challenge (Studies in Big Data Book 45). Springer
<b>Complementary</b>	

## Recommendations

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

Natural Language Understanding/614544008

Language Modelling/614544009

Web Intelligence and Semantic Technologies/614544010

### Subjects that are recommended to be taken simultaneously



Subjects that continue the syllabus
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.