



## DATOS IDENTIFICATIVOS

### Modelado del lenguaje

Asignatura	Modelado del lenguaje			
Código	O06M193V01204			
Titulación	Máster universitario en Inteligencia artificial			
Descriptores	Creditos ECTS	Seleccione	Curso	Cuatrimestre
	3	OP	1	2c
Lengua	Inglés			
Impartición				
Departamento				
Coordinador/a	Darriba Bilbao, Víctor Manuel			
Profesorado	A0075-Ax2tc-1 A0075-Ax2tc-1, A0075-Ax2tc-1 A0075-Ax2tc-2 A0075-Ax2tc-2, A0075-Ax2tc-2 Darriba Bilbao, Víctor Manuel			
Correo-e	darriba@uvigo.es			
Web	<a href="http://guiadocente.udc.es/guia_docent/index.php?centre=614&amp;ensenyament=614544&amp;consulta=assignatures&amp;ny_academic=2023_24">http://guiadocente.udc.es/guia_docent/index.php?centre=614&amp;ensenyament=614544&amp;consulta=assignatures&amp;ny_academic=2023_24</a>			
Descripción general	La asignatura introduce al estudiante en la modelización del lenguaje humano, es decir en la generación de modelos que permiten estimar la verosimilitud de un texto, una fase esencial en el diseño de cualquier aplicación basada en la explotación de sus mecanismos comunicativos. Se capacitará al estudiante en el dominio de los principios teóricos y las técnicas que permiten su construcción, tanto las basadas en representaciones discretas de los elementos de la lengua como en representaciones continuas. Especial atención recibirá la modelización en contextos de escasez de recursos lingüísticos, tanto si ésta viene motivada por el tratamiento de dominios del conocimiento novedosos como por la explotación de lenguas de reducida difusión.			

## Resultados de Formación y Aprendizaje

Código	
A1	CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
A2	CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
A5	CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
B1	Mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo de la Inteligencia Artificial.
B3	Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo.
B4	Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables en el campo.
C1	Comprensión y dominio de técnicas para el procesado de textos en lenguaje natural.
C2	Comprensión y dominio de los fundamentos y técnicas de procesamiento semántico de documentos enlazados, estructurados y no estructurados, y de la representación de su contenido.
C3	Comprensión y conocimiento de las técnicas de representación y procesado de conocimiento mediante ontologías, grafos y RDF, así como de las herramientas asociadas a las mismas.
D2	Dominar la expresión y la comprensión de forma oral y escrita de un idioma extranjero.
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	Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social.
D8	Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad

## Resultados previstos en la materia

Resultados previstos en la materia	Resultados de Formación y Aprendizaje
Saber usar las técnicas y métodos del procesamiento del lenguaje natural para resolver problemas reales de análisis de textos en lenguaje natural	A2 A5 B1 B3 B4 C1 C3 D2 D3 D7
Conocer, comprender y analizar las técnicas de aprendizaje profundo aplicadas al procesamiento del lenguaje natural	A1 A2 A5 B1 B3 C1 C2 D2 D3 D7 D8
Saber usar las técnicas y métodos del aprendizaje profundo para resolver problemas prácticos de procesamiento del lenguaje natural	A1 A2 A5 B1 B3 B4 C1 C2 D2 D3 D7 D8
Conocer y comprender los problemas medioambientales que plantea el coste computacional de las técnicas de aprendizaje profundo cuando son aplicadas al análisis de textos.	A1 B1 C1 C2 D2 D8

## Contenidos

Tema	
Modelos de lenguaje	Modelos de lenguaje basados en n-gramas Modelos de lenguaje basados en redes neuronales
Modelos de semántica distribucional	Hipótesis lingüísticas sobre el significado distribucional Modelos distribucionales clásicos Modelos neuronales de significado estático (word embeddings) Modelos neuronales de significado dinámico-contextual Modelos composicionales
Etiquetado secuencial	Uso y adaptación de modelos para el etiquetado secuencial
Modelos Texto a Texto	Modelos Texto a Texto

## Planificación

	Horas en clase	Horas fuera de clase	Horas totales
Lección magistral	10	10	20
Prácticas de laboratorio	5	15	20
Aprendizaje basado en proyectos	6	28	34
Examen de preguntas objetivas	1	0	1

\*Los datos que aparecen en la tabla de planificación son de carácter orientativo, considerando la heterogeneidad de alumnado

## Metodologías

	Descripción
Lección magistral	El profesorado presenta un tema al alumnado con el objetivo de facilitar un conjunto de información con un alcance concreto.
Prácticas de laboratorio	El profesorado de la materia presenta al alumnado uno o varios problemas de carácter práctico que requieren la comprensión y la aplicación de los contenidos teórico-prácticos incluidos en el programa de la materia. El alumnado puede trabajar la solución a los problemas planteados de manera individual o en equipo. El trabajo podrá ser autónomo, aunque guiado por el profesorado de la asignatura.
Aprendizaje basado en proyectos	Se presentan al alumnado proyectos prácticos que requieren una parte importante de su dedicación total a la materia. Además, y debido al alcance de los trabajos a realizar, es preciso que la/el estudiante aplique no solo competencias de gestión sino también de carácter técnico.

### Atención personalizada

Metodologías	Descripción
Prácticas de laboratorio	El profesorado atenderá al alumnado en sesiones de tutorías individualizadas, dedicadas a la orientación en el estudio y a la resolución de dudas sobre los contenidos, trabajos y actividades de la asignatura.
Aprendizaje basado en proyectos	El profesorado atenderá al alumnado en sesiones de tutorías individualizadas, dedicadas a la orientación en el estudio y a la resolución de dudas sobre los contenidos, trabajos y actividades de la asignatura.

### Evaluación

	Descripción	Calificación	Resultados de Formación y Aprendizaje			
Lección magistral	Seguimiento continuado de la asistencia a clase y participación del alumno	5	A1	B1	C1	D2
				B4	C2	D8
					C3	
Aprendizaje basado en proyectos	Las entregas de los proyectos deben realizarse dentro del plazo establecido en el campus virtual y deben seguir las especificaciones indicadas en el enunciado tanto para su presentación como su defensa. Carácter Obligatorio.	50	A1	B1	C1	D2
			A2	B3	C2	D3
			A5	B4	C3	D7
						D8
Examen de preguntas objetivas	Se evaluará el dominio de los conocimientos teóricos y operativos de la asignatura. Carácter Obligatorio.	45	A1	B1	C1	D2
				B4	C2	D8
					C3	

### Otros comentarios sobre la Evaluación

#### CRITERIOS DE EVALUACIÓN PARA TODOS LOS ESTUDIANTES EN TODAS LAS CONVOCATORIAS

Cada alumno deberá alcanzar un mínimo del 40% de la nota máxima de las partes "Prácticas de Laboratorio" y "Prueba Objetiva", y en todo caso la suma de las tres partes deberá ser superior a 5 para aprobar la asignatura. De no cumplirse alguno de los requisitos anteriores, la calificación de la convocatoria se establecerá en función de la puntuación más baja obtenida.

En caso de no obtener el mínimo en las "Prácticas de Laboratorio" o "Prueba Objetiva", el alumno tendrá una segunda oportunidad en la que únicamente se exigirá la entrega de esa parte.

No se guardarán calificaciones entre años académicos.

La presentación de las prácticas deberá realizarse en el plazo establecido en el campus virtual, y seguirá las especificaciones detalladas en el enunciado tanto para la presentación como para la defensa.

Tendrán la condición de "Presentados" quienes presenten todas las prácticas obligatorias o realicen la prueba objetiva en el periodo oficial de evaluación.

En el caso de realización fraudulenta de ejercicios o pruebas, se aplicará la Normativa para la evaluación del rendimiento académico de los estudiantes y de revisión de calificaciones. En aplicación de la normativa correspondiente sobre plagio, la copia total o parcial de algún ejercicio de prácticas o teoría supondrá la suspensión en las dos oportunidades del curso, con la calificación de 0,0 en ambos casos.

#### FECHAS DE EVALUACIÓN

Las fechas oficiales de examen de las diferentes convocatorias, se encuentran publicadas en la página web de la

ESEI: <https://esei.uvigo.es/docencia/exames/>

## **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/>

---

### **Fuentes de información**

#### **Bibliografía Básica**

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

#### **Bibliografía Complementaria**

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

---

### **Recomendaciones**

#### **Asignaturas que continúan el temario**

Inteligencia web y tecnologías semánticas/O06M193V01205

Minería de textos/O06M193V01302

#### **Asignaturas que se recomienda cursar simultáneamente**

Aprendizaje Automático II/O06M193V01207

Aprendizaje profundo/O06M193V01206

#### **Asignaturas que se recomienda haber cursado previamente**

Aprendizaje automático I/O06M193V01105

Comprensión del lenguaje natural/O06M193V01104

### **Otros comentarios**

Se recuerda a todo el alumnado la prohibición del uso de dispositivos móviles en ejercicios, prácticas y exámenes, 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."