UNIVERSIDADE DA CORUÑA

Guia docente					
Datos Identificativos					2023/24
Asignatura (*)	Modelado del Lenguaje			Código	614544009
Titulación	Máster Universitario en Intelixencia Artificial				
Descriptores					
Ciclo	Periodo	Curso		Tipo	Créditos
Máster Oficial	2º cuatrimestre	Primero		Optativa	3
Idioma	Inglés				
Modalidad docente	Presencial				
Prerrequisitos					
Departamento	Ciencias da Computación e Tecnoloxías da Información				
Coordinador/a	Vilares Calvo, David		Correo electrónico	david.vilares@udc.es	
Profesorado	Vilares Calvo, David		Correo electrónico	david.vilares@udc.es	
Web	campusvirtual.udc.es				
Descripción general	Proveer conocimientos teóricos que permitan profundizar en el estudio de modelos lingüísticos: modelos de lenguaje, modelos semánticos distribucionales. Asociar el modelado lingüístico y los tipos de modelos con diferentes tareas dentro del área de las tecnologías lingüísticas y del procesamiento de las lenguas naturales. Evaluar diferentes aspectos de los modelos lingüísticos. Proveer conocimiento práctico que permita poder entrenar nuevos modelos y usarlos convenientemente en diferentes tareas de procesamiento de las lenguas naturales.				

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A2	CE01 - Comprensión y dominio de técnicas para el procesamiento léxico, sintáctico y semántico de textos en lenguaje natural
A3	CE02 - Comprensión y dominio de los fundamentos y técnicas de procesamiento de documentos enlazados, estructurados y no estructurados, y de la representación de su contenido
A4	CE03 - 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
B1	CG01 - 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	CG03 - Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo
B4	CG04 - 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
B6	CB01 - 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
B7	CB02 - 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
B10	CB05 - 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
C2	CT02 - Dominar la expresión y la comprensión de forma oral y escrita de un idioma extranjero
C3	CT03 - 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
C7	CT07 - Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social

C8	CT08 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad

Resultados de aprendizaje			
Resultados de aprendizaje	Competencias / Resultados del título		
Saber usar las técnicas y métodos del procesamiento del lenguaje natural para resolver problemas reales de análisis de textos en lenguaje natural	AM1 AM3	BM1 BM3 BM4 BM7 BM10	CM2 CM3 CM7
Conocer, comprender y analizar las técnicas de aprendizaje profundo aplicadas al procesamiento del lenguaje natural	AM1 AM2	BM1 BM3 BM6 BM7 BM10	CM2 CM3 CM7 CM8
Saber usar las técnicas y métodos del aprendizaje profundo para resolver problemas prácticos de procesamiento del lenguaje natural	AM1 AM2	BM1 BM3 BM4 BM6 BM7 BM10	CM2 CM3 CM7 CM8
Conocer y comprender los problemas medioambientales que plantea el coste computacional de las técnicas de aprendizaje profundo cuando son aplicadas al análisis de textos.	AM1 AM2	BM1 BM6	CM2 CM8

Contenidos	
Tema	Subtema
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
Text-To-Text models	Uso e adaptación de modelos para o etiquetado secuencial

Planificación				
Metodologías / pruebas	Competencias / Resultados	Horas lectivas (presenciales y virtuales)	Horas trabajo autónomo	Horas totales
Sesión magistral	A2 A3 A4 B1 B3 B6 B7 B10 C2 C8	10	10	20
Prácticas de laboratorio	A2 A3 B3 B4 B6 B7 B10 C2 C3 C7 C8	5	17	22
Solución de problemas	A2 A3 B3 B4 B6 B7 B10 C2 C8	6	15	21
Prueba de respuesta múltiple	A2 A3 B1 B6 B7 B10 C2	0	1	1

Prueba objetiva	A2 A3 B1 B6 B7 B10 C2 C3	2	8	10
Atención personalizada		1	0	1
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	**Descripción**
Sesión magistral	Clases teóricas, en las que se expone el contenido de cada tema. El alumnado dispondrá de copias de las transparencias con anterioridad y el profesorado promoverá una actitud activa, realizando preguntas que permitan aclarar aspectos concretos y dejando cuestiones abiertas para la reflexión del estudiantado.
Prácticas de laboratorio	Clases prácticas con uso de ordenador, que permiten al alumnado familiarizarse desde un punto de vista práctico con las cuestiones expuestas en las clases teóricas.
Solución de problemas	Aprendizaje basado en problemas, seminarios, estudio de casos y proyectos.
Prueba de respuesta múltiple	Breves cuestionarios que se habilitarán después de algunas sesiones teóricas para comprobar la asimilación de contenidos.
Prueba objetiva	Se evaluará el dominio de los conocimientos teóricos y operativos de la materia.

Atención personalizada	
Metodologías	**Descripción**
Prácticas de laboratorio Solución de problemas Prueba objetiva Sesión magistral	El desarrollo, tanto de las clases magistrales como de las de resolución de problemas y los laboratorios de prácticas, se realizará atendiendo al progreso de los/as alumnos/as en las capacidades de comprensión y asimilación de los contenidos impartidos. El avance general de la clase se compaginará con una atención específica a aquellos/as alumnos/as que presenten mayores dificultades en la tarea del aprendizaje y con un apoyo adicional a aquellos/as que presenten mayor desenvoltura y deseen ampliar conocimientos. En lo que respecta a las tutorías individuales, dado su carácter personalizado, no deben dedicarse a extender los contenidos con nuevos conceptos, sino a aclarar los conceptos ya expuestos. El profesorado las utilizará como una interacción que le permita extraer conclusiones respecto al grado de asimilación de la materia por parte del estudiantado.

Evaluación			
Metodologías	**Competencias / Resultados**	**Descripción**	**Calificación**
Prácticas de laboratorio	A2 A3 B3 B4 B6 B7 B10 C2 C3 C7 C8	Las entregas de las prácticas 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.	50
Prueba objetiva	A2 A3 B1 B6 B7 B10 C2 C3	Realización obligatoria. Se evaluará el dominio de los conocimientos teóricos y operativos de la asignatura.	45
Prueba de respuesta múltiple	A2 A3 B1 B6 B7 B10 C2	Pequeños cuestionarios de evaluación continua que se propondrán al final de unas sesiones teóricas y donde se preguntará de forma sencilla sobre algunos de los conceptos explicados en esa sesión. Se avisará con antelación.	5

Observaciones evaluación

Cada alumno/a 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/la alumno/a 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/as? 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 del estudiantado 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.

Fuentes de información	
Básica	Jurafsky, Daniel & James H. Martin (2021). ?N-gram Language Models.? Speech and Language Processing, Capítulo 3. https://web.stanford.edu/~jurafsky/slp3/Jurafsky, Daniel & James H. Martin (2021). ?Vector Semantics and Embeddings.? Speech and Language Processing, Capítulo 6. https://web.stanford.edu/~jurafsky/slp3/Jurafsky, Daniel & James H. Martin (2021). ?Neural Networks and Neural Language Models.? Speech and Language Processing, Capítulo 7. https://web.stanford.edu/~jurafsky/slp3/Jurafsky, Daniel & James H. Martin (2021). ?Sequence Labeling for Parts of Speech and Named Entities.? Speech and Language Processing, Capítulo 8. https://web.stanford.edu/~jurafsky/slp3/Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2018). ?BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171?4186, Minneapolis, Minnesota. Association for Computational Linguistics.Erk, Katrin (2012). "Vector space models of word meaning and phrase meaning: A survey." Language and Linguistics Compass 6.10: 635-653.
Complementária	Baroni, Marco, Raffaella Bernardi & Roberto Zamparelli (2014). ?Frege in space: A program for compositional distributional semantics.? Linguistic Issues in Language Technologies 9(6): 5-110.Baroni, Marco, Georgiana Dinu & Germán Kruszewski (2014). ?Don?t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors.? In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) , pp. 238?247, Baltimore, Maryland. Association for Computational Linguistics.Church, Kenneth Ward, Zeyu Chen & Yanjun Ma (2021). ?Emerging trends: A gentle introduction to fine-tuning.? Natural Language Engineering, 27: 763?778.Hirschberg, Julia & Christopher D. Manning (2015). "Advances in natural language processing." Science 349.6245: 261-266.Linzen, Tal (2016). "Issues in evaluating semantic spaces using word analogies." In Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, pp. 13?18, Berlin, Germany. Association for Computational Linguistics.Mikolov, Tomas, Wen-tau Yih & Geoffrey Zweig (2013). "Linguistic Regularities in Continuous Space Word Representations." In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 746?751, Atlanta, Georgia. Association for Computational Linguistics.Taher Pilehvar, Mohammad & Jose Camacho-Collados (2021). Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Morgan & Claypool (Synthesis Lectures on Human Language Technologies, volume 47).

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Comprensión del Lenguaje Natural/614544008 Aprendizaje Automático I/614544012
Asignaturas que se recomienda cursar simultáneamente

Aprendizaje Profundo/614544013 Aprendizaje Automático II/614544014
Asignaturas que continúan el temario
Minería de Textos/614544011 Inteligencia Web y Tecnologías Semánticas/614544010
Otros comentarios
Tal y como recogen las distintas normativas docentes universitarias, esta materia incorpora la perspectiva de género. Se fomentará la participación del alumnado en clase, y se trabajará para identificar y modificar prejuicios y actitudes sexistas e incidir en el entorno para modificarlos y promover valores de respeto e igualdad. Deberán detectarse las situaciones de discriminación por razón de género y se propondrán actuaciones y medidas para corregirlas.

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías