Guia docente				
Datos Identificativos				2024/25
Asignatura (*)	Lenguaje Natural y Minería de Textos		Código	614G02043
Titulación	Grao en Ciencia e Enxeñaría de Datos			
Descriptores				
Ciclo	Periodo	Curso	Tipo	Créditos
Grado	2º cuatrimestre	Cuarto	Optativa	6
Idioma	Castellano			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da Información			
Coordinador/a	Vilares Calvo, David	Correo electrónico	david.vilares@udc.es	
Profesorado	Gómez Rodríguez, Carlos Muñoz Ortiz, Alberto Vilares Calvo, David	Correo electrónico	carlos.gomez@udc.es alberto.munoz.ortiz@udc.es david.vilares@udc.es	
Web	campusvirtual.udc.es			
Descripción general	El procesamiento del lenguaje natural (PLN) es el área de la inteligencia artificial que se encarga del estudio y desarrollo de modelos computacionales que sean capaces de procesar y comprender las particularidades del lenguaje natural con la misma eficacia con la que lo hacemos los humanos. En esta asignatura se introducirá al alumnado en fundamentos básicos y técnicas de aprendizaje automático asociadas al PLN, que son utilizados como punto de partida para el desarrollo de numerosas tecnologías de lenguaje y de la explotación automática de textos. El alumnado se familiarizará con algoritmos y técnicas para representar en forma de árbol y grafos la información latente presente en textos escritos, con técnicas para la representación de palabras de manera que capturen eficientemente su significado, con la implementación de modelos mediante técnicas de aprendizaje automático para su aplicación a diversos problemas relacionados con el PLN que son de utilidad para usuarios no especializados, así como con las técnicas existentes para que las tecnologías del lenguaje puedan ser aplicadas a multitud de idiomas, incluso para aquellos para los que existe una limitada cantidad de recursos disponibles, ya sean anotados o no.			

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A28	CE28 - Comprensión y dominio de los fundamentos y técnicas para el procesado de datos escritos, tanto en lenguaje formal como en lenguaje natural.
B2	CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
B3	CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética
B4	CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado
B7	CG2 - Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables.
B8	CG3 - Ser capaz de mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo.
B9	CG4 - Capacidad para abordar con éxito todas las etapas de un proyecto de análisis de datos: exploración previa de los datos, preprocesado, análisis, visualización y comunicación de resultados.
B10	CG5 - Ser capaz de trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones.

Resultados de aprendizaje			
Resultados de aprendizaje	Competencias / Resultados del título		
Conocer, comprender y analizar las técnicas de procesamiento del lenguaje natural para el procesado y desambiguación a nivel sintáctico y semántico.	A28	B2 B3 B4 B7 B8 B9 B10	
Saber usar las técnicas y métodos del procesamiento del lenguaje natural para resolver problemas reales de minería de textos	A28	B2 B3 B4 B7 B8 B9 B10	
Conocer y comprender los problemas que plantea el multilingüismo en las fuentes de datos y técnicas para resolverlos	A28	B2 B8 B9 B10	
Conocer y analizar paradigmas emergentes de computación con el potencial de mejorar el paralelismo en la minería de textos	A28	B2 B4 B7 B8	

Contenidos	
Tema	Subtema
Análisis sintáctico de constituyentes para minería de textos	Sintaxis de constituyentes Análisis de constituyentes estadístico con programación dinámica Análisis de constituyentes shift-reduce con redes neuronales Análisis de constituyentes discontinuos Análisis de constituyentes secuencia a secuencia
Análisis sintáctico de dependencias para minería de textos	Sintaxis de dependencias Criterios de anotación y dependencias universales Análisis de dependencias basado en transiciones Análisis de dependencias basado en grafos No proyectividad
Semántica	Análisis de dependencias semánticas Vectores densos mediante SVD Vectores densos mediante predicción de palabras: skip-gram y CBOW Propiedades de los vectores densos Clustering de Brown
Computación con sentidos de las palabras	Sentidos de las palabras Relaciones entre sentidos Bases de datos de relaciones léxicas Desambiguación del sentido de las palabras
Aplicaciones prácticas de la minería de textos	-

Procesamiento multilingüe del lenguaje	Procesamiento de idiomas de morfología rica Procesamiento de idiomas no segmentados Procesamiento de idiomas con pocos recursos Procesamiento translingüe
Tecnologías emergentes	-

Planificación				
Metodologías / pruebas	Competencias / Resultados	Horas lectivas (presenciales y virtuales)	Horas trabajo autónomo	Horas totales
Sesión magistral	A28 B3 B8 B9	21	21	42
Prácticas de laboratorio	A28 B2 B4 B7 B9 B10	10	30	40
Solución de problemas	A28 B2 B3 B8	8	30	38
Prueba objetiva	A28 B2 B3 B8	2	10	12
Trabajos tutelados	A28 B3 B4 B7 B10	1	16	17
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 alumnado.
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 objetiva	Se evaluará el dominio de los conocimientos teóricos y operativos de la materia.
Trabajos tutelados	Trabajo en el que el alumnado consultarán fuentes de información actualizada para familiarizarse con aspectos de investigación del campo

Atención personalizada	
Metodologías	Descripción
Prueba objetiva Sesión magistral Prácticas de laboratorio Trabajos tutelados Solución de problemas	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 que presenten mayor desenvoltura y deseen ampliar conocimientos. En los trabajos tutelados se proporcionará una atención personalizada al alumnado para orientarlos en su trabajo autónomo. 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**

Prueba objetiva	A28 B2 B3 B8	Realización obligatoria. Se evaluará el dominio de los conocimientos teóricos y operativos de la asignatura.	50
Prácticas de laboratorio	A28 B2 B4 B7 B9 B10	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.	40
Trabajos tutelados	A28 B3 B4 B7 B10	Se evaluará el dominio del alumnado para comprender y asimilar trabajos de investigación.	10

Observaciones evaluación
Los/as alumnos/as deberán alcanzar al menos un 40% de la máxima nota de las partes de teoría y práctica, y en cualquier caso la suma de tres partes debe superar un 5 para superar la asignatura. Si no se cumple alguno de los requisitos anteriores, la nota de la convocatoria se establecerá de acuerdo a la menor nota obtenida. En caso de no alcanzar el mínimo bien en teoría o práctica, el/la alumno/a tendrá una segunda oportunidad en la que solamente se le exigirá la entrega de dicha parte. Los trabajos tutelados se consideran como evaluación continua y no serán entregables en la segunda oportunidad. 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. Tendrá la condición de ?Presentado/a? quien concurra a la prueba objetiva en el período oficial de evaluación. Todos los aspectos relacionados con dispensa académica, dedicación al estudio, permanencia y fraude académico se regirán de acuerdo a la normativa académica vigente de la UDC.

Fuentes de información	
Básica	- Manning, Chris., & Schutze, Hinrich (1999). Foundations of statistical natural language processing. MIT Press - Jacob Eisenstein (2019). Introduction to Natural Language Processing. MIT Press - Jurafsky, Dan. & Martin, J. H. (2022). Speech and Language Processing (3rd ed. draft). Disponible en: https://web.stanford.edu/~jurafsky/slp3/ - Kochmar, Ekaterina (2022). Getting Started with Natural Language Processing. Oreilly - Bengfort, Benjamin & Bilbro, Rebeca & Tony Ojeda (2018). Applied Text Analysis with Python. Oreilly
Complementária	- Chollet, François. (2018). Keras: The python deep learning library. Astrophysics Source Code Library - Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - Kübler, Sandra., McDonald, Ronald., & Nivre, Joakim (2009). Dependency Parsing. Synthesis lectures on human language technologies. Morgan Claypool - Goldberg, Yoav (2017). Neural network methods for natural language processing. Synthesis lectures on human language technologies. Morgan Claypool

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Procesamiento de Lenguaje Escrito/614G02029 Aprendizaje Automático III/614G02026 Aprendizaje Automático I/614G02019 Aprendizaje Automático II/614G02021
Asignaturas que se recomienda cursar simultáneamente
Asignaturas que continúan el temario
Otros comentarios

Se fomentará el desarrollo de una ciudadanía crítica, abierta y respetuosa con la diversidad en nuestra sociedad, destacando la igualdad de derechos del alumnado sin discriminación por cuestiones de género o condición sexual. Se empleará un lenguaje inclusivo en el material y en el desarrollo de las sesiones. Se trabajará para identificar y modificar prejuicios y actitudes sexistas e influir en el entorno para modificarlos y fomentar valores de respeto e igualdad. Se facilitará la plena integración del alumnado que, por razones físicas, sensoriales, psíquicas o socioculturales, experimenten dificultades para un acceso adecuado, igualitario y provechoso a la vida universitaria.

(*) 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