



Guía docente				
Datos Identificativos			2022/23	
Asignatura (*)	Comprensión del Lenguaje Natural		Código	614544008
Titulación	Máster Universitario en Intelixencia Artificial			
Descriptores				
Ciclo	Periodo	Curso	Tipo	Créditos
Máster Oficial	1º cuatrimestre	Primero	Obligatoria	6
Idioma	Inglés			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da Información			
Coordinador/a	Gómez Rodríguez, Carlos	Correo electrónico	carlos.gomez@udc.es	
Profesorado	Gómez Rodríguez, Carlos	Correo electrónico	carlos.gomez@udc.es	
	Vilares Calvo, David		david.vilares@udc.es	
	Vilares Ferro, Jesus		jesus.vilares@udc.es	
Web	campusvirtual.udc.es			
Descripción general	La asignatura introduce los conceptos y técnicas básicas asociadas al procesamiento del lenguaje natural, punto de partida para el diseño de entornos de explotación de información y de diálogo basados en el lenguaje humano, tanto a nivel léxico como sintáctico, semántico y pragmático.			
	El objetivo es introducir al estudiante en la complejidad inherente al análisis del lenguaje natural humano, fundamentalmente asociada a la ambigüedad y dependencias contextuales que presenta, y en el diseño de estructuras de datos y algoritmos que permitan su tratamiento práctico.			

Competencias del título	
Código	Competencias 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 del título		
Conocer, comprender y analizar la representación formal de diversos fenómenos léxicos, sintácticos y semánticos del lenguaje natural.	AM1	BM1 BM3 BM4 BM6 BM10	CM2 CM8
Conocer, comprender y saber usar las tecnologías, marcos y librerías para la construcción de sistemas de procesamiento del lenguaje natural.	AM1 AM2	BM3 BM4 BM6 BM7 BM10	CM2 CM3 CM7
Diseñar, implementar y saber usar algoritmos y estructuras de datos para tratar y dar soporte a los diversos fenómenos característicos del lenguaje natural.	AM1 AM2 AM3	BM1 BM3 BM4 BM6 BM7 BM10	CM2 CM3 CM7 CM8
Conocer, comprender y analizar las técnicas de procesamiento del lenguaje natural para el procesado y desambiguación a nivel léxico, sintáctico y semántico.	AM1 AM2 AM3	BM1 BM3 BM4 BM6 BM7 BM10	CM2 CM3 CM7 CM8
Conocer y comprender los problemas que plantea la ambigüedad e imprecisión en las fuentes de datos en lenguaje natural y técnicas para resolverlos.	AM1 AM2	BM1 BM3 BM4 BM6 BM7 BM10	CM2 CM3 CM7 CM8

Contenidos	
Tema	Subtema
Introducción.	Niveles de análisis. Ambigüedad y dependencias contextuales.
Análisis léxico.	Segmentación. Diccionarios y tesauros. Técnicas de etiquetación morfosintáctica.
Análisis sintáctico.	Gramáticas algebraicas. Gramáticas suavemente sensibles al contexto. Gramáticas de dependencias. Gramáticas probabilísticas.
Análisis semántico.	Semántica léxica. Dependencias semánticas. Grafos semánticos.

Planificación



Metodologías / pruebas	Competencias	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales
Sesión magistral	A2 A3 A4 B1 B3 B6 B7 B10 C2 C8	21	21	42
Prácticas de laboratorio	A2 A3 A4 B3 B4 B6 B7 B10 C2 C3 C7 C8	14	48	62
Solución de problemas	A2 A3 A4 B3 B4 B6 B7 B10 C2	7	25	32
Prueba objetiva	A2 A3 A4 B1 B6 B7 C2	3	9	12
Atención personalizada		2	0	2

(*) 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 alumno dispondrá de copias de las transparencias con anterioridad y el profesor promoverá una actitud activa, realizando preguntas que permitan aclarar aspectos concretos y dejando cuestiones abiertas para la reflexión del alumno.
Prácticas de laboratorio	Clases prácticas con uso de ordenador, que permiten al alumno 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.

Atención personalizada	
Metodologías	Descripción
Sesión magistral Prácticas de laboratorio Solución de problemas Prueba objetiva	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 alumnos 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 alumnos que presenten mayores dificultades en la tarea del aprendizaje y con un apoyo adicional a aquellos que presenten mayor desentendimiento 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 profesor las utilizará como una interacción que le permita extraer conclusiones respecto al grado de asimilación de la materia por parte de los alumnos.

Evaluación			
Metodologías	Competencias	Descripción	Calificación
Prácticas de laboratorio	A2 A3 A4 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.	40
Prueba objetiva	A2 A3 A4 B1 B6 B7 C2	Realización obligatoria. Se evaluará el dominio de los conocimientos teóricos y operativos de la asignatura.	60

Observaciones evaluación



Los alumnos deberán alcanzar al menos un 40% de la máxima nota cada parte (teoría, práctica) y en cualquier caso la suma de ambas 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 en alguna de las partes, el alumno tendrá una segunda oportunidad en la que solamente se le exigirá la entrega de dicha parte. No se guardarán calificaciones entre cursos académicos. 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? quien entregue todas las prácticas obligatorias o concurra a la prueba objetiva en el período oficial de evaluación. En el caso de realización fraudulenta de ejercicios o pruebas, se aplicará la Normativa de 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 de teoría supondrá el suspenso en las dos oportunidades del curso, con la calificación de 0,0 en ambos casos.

Fuentes de información

Básica	- Manning, C., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press - Goldberg, Y. (2017). Neural network methods for natural language processing. Synthesis lectures on human language technologies. Morgan Claypool - Jacob Eisenstein (2019). Introduction to Natural Language Processing. MIT Press - Jurafsky, D. & Martin, J. H. (2022). Speech and Language Processing (3rd ed. draft). Disponible en: https://web.stanford.edu/~jurafsky/slp3/ - Manning, C., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press
Complementaria	- Chollet, F. (2018). Keras: The python deep learning library. Astrophysics Source Code Library - Stuart Russell, Peter Norvig (2020). Artificial Intelligence: A Modern Approach, 4th Edition. Pearson - Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency Parsing. Synthesis lectures on human language technologies. Morgan Claypool Adicionalmente, manexaranse textos científicos dispoñibles nas bibliotecas dixitais da área, como o ACL Anthology ou ACM.

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Asignaturas que se recomienda cursar simultáneamente

Aprendizaje Automático I/614544012

Asignaturas que continúan el temario

Minería de Textos/614544011

Modelado del Lenguaje/614544009

Inteligencia Web y Tecnologías Semánticas/614544010

Otros comentarios

(*) 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