



Guia docente				
Datos Identificativos				2022/23
Asignatura (*)	Procesamiento de Lenguaje Escrito		Código	614G02029
Titulación	Grao en Ciencia e Enxeñaría de Datos			
Descriptorios				
Ciclo	Periodo	Curso	Tipo	Créditos
Grado	2º cuatrimestre	Tercero	Obligatoria	6
Idioma	CastellanoGallego			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da Información			
Coordinador/a	Dafonte Vazquez, Jose Carlos		Correo electrónico	carlos.dafonte@udc.es
Profesorado	Alonso Pardo, Miguel angel		Correo electrónico	miguel.alonso@udc.es
	Dafonte Vazquez, Jose Carlos			carlos.dafonte@udc.es
	Vilares Calvo, David			david.vilares@udc.es
Web	https://campusvirtual.udc.gal			
Descripción general	La asignatura introduce nociones necesarias para procesar información, tanto en lenguajes formales usados para datos estructurados como en lenguaje humano. La asignatura incluye tanto los fundamentos básicos del procesamiento de lenguaje estructurado para el diseño de reconocedores: análisis léxico, sintáctico y semántico; como del procesamiento del lenguaje natural: modelado de lenguaje, análisis morfológico y sintáctico.			

Competencias del título	
Código	Competencias 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.
C1	CT1 - 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.
C4	CT4 - 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	A28	B3 B4 B7 B8 B10	C4
Conocer, comprender y saber usar las tecnologías, marcos y librerías para la construcción de sistemas de procesamiento del lenguaje	A28	B2 B3 B4 B7 B8 B9 B10	C1 C4
Diseñar, implementar y saber usar algoritmos y estructuras de datos para tratar y dar soporte a los diversos fenómenos característicos del lenguaje	A28	B2 B3 B4 B7 B8 B9 B10	C1 C4

Contenidos	
Tema	Subtema
Tema I. Fundamentos de lenguaje estructurado	1.1. Conceptos básicos: Gramáticas y lenguajes 1.2. Fases y componentes de un procesador del lenguaje 1.3. Entornos de aplicación de la Teoría de Lenguajes
Tema II. Análisis léxico	2.1. Autómatas finitos, expresiones regulares y gramáticas regulares 2.2. Conversión entre diferentes representaciones 2.3. Implementación de autómatas y expresiones regulares
Tema III. Análisis sintáctico	3.1. Lenguajes de contexto libre y Gramáticas independientes del contexto 3.2. Reconocedores descendentes recursivos y LL(k) 3.3. Reconocedores LR(k)
Tema IV. Análisis semántico y diseño de acciones semánticas	4.1. Control de atributos 4.2. Manejo de errores 4.3. Generación dirigida por la sintaxis
Tema V. Introducción al procesamiento del lenguaje natural	5.1. Niveles de análisis 5.2. Ambigüedad
Tema VI. Modelado del lenguaje	6.1. Modelos del lenguaje basados en n-gramas 6.2. Evaluación de modelos del lenguaje 6.3. Técnicas de suavizado 6.4. Modelos del lenguaje basados en redes neuronales
Tema VII. Análisis morfológico del lenguaje natural	7.1. Fundamentos de análisis morfológico 7.2. Análisis morfológico estadístico 7.3. Análisis morfológico con redes neuronales
Tema VIII. Fundamentos de análisis sintáctico del lenguaje natural	8.1. Fundamentos de análisis sintáctico de constituyentes 8.2. Fundamentos de análisis sintáctico de dependencias

Planificación				
Metodologías / pruebas	Competencias	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales



Sesión magistral	A28 B4 B8 B9 C4	21	27.3	48.3
Prueba objetiva	A28 B2 B3 B7 B9	3	18	21
Prácticas de laboratorio	A28 B2 B3 B4 B7 B8 B9 B10 C1	21	35.7	56.7
Atención personalizada		24	0	24
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	Descripción
Sesión magistral	En las clases de teoría, los profesores desarrollarán los contenidos teóricos y algunos ejercicios complementarios
Prueba objetiva	Prueba en la que se evaluarán los conocimientos adquiridos
Prácticas de laboratorio	Permiten al alumno aprender de forma eficaz a través de actividades de carácter aplicado, en este caso mediante la realización de proyectos en forma de prácticas. En su enunciado se detallará si es necesaria la entrega y su formato.

Atención personalizada	
Metodologías	Descripción
Prácticas de laboratorio	El desarrollo de las clases se realizará de acuerdo con el progreso de los estudiantes en las capacidades de comprensión y asimilación de los contenidos enseñados, combinando el progreso general de la clase con una atención específica a los estudiantes que presentan grandes dificultades en la tarea de aprendizaje y con un apoyo adicional a aquellos que presentan mayor facilidad y quieren ampliar conocimientos.
Sesión magistral	Dado el carácter personalizado de las tutorías, estas no deben dedicarse a ampliar el contenido con nuevos conceptos, sino a aclarar los conceptos ya comentados. El profesor también debe utilizarlos como una interacción que le permita sacar conclusiones sobre el grado de asimilación de la asignatura por parte de los alumnos.

Evaluación			
Metodologías	Competencias	Descripción	Calificación
Prácticas de laboratorio	A28 B2 B3 B4 B7 B8 B9 B10 C1	Realización de ejercicios de diverso tipo, asociados en buena parte a los contenidos de las clases magistrales y bajo las condiciones establecidas en el enunciado de cada práctica.	40
Prueba objetiva	A28 B2 B3 B7 B9	Se realizará un examen al final del cuatrimestre y su correspondiente segunda oportunidad. Este examen se centrará principalmente en los aspectos teóricos y ejercicios asociados. Sin embargo, podrá preguntarse alguna cuestión relacionada con las prácticas de laboratorio realizadas.	60

Observaciones evaluación
El examen computa un 60% de la nota y el 40% restante se reparte entre las prácticas y cualquier otra actividad evaluable realizada a lo largo del curso. Será necesario conseguir un mínimo de 3 (sobre 6 puntos, en el examen) y un 2 (sobre 4 puntos, en el conjunto de las prácticas y demás actividades). En caso de no conseguir alguno de los mínimos, la máxima nota que se podrá obtener es un 4.5. Si las prácticas u otras actividades se llevan a cabo en grupos, todos los miembros del grupo serán responsables solidariamente por el trabajo realizado y entregado así como sus posibles implicaciones. En la segunda oportunidad se mantendrán los resultados obtenidos en cada sección (examen y prácticas) de la primera oportunidad. En el caso de estudiantes a tiempo parcial, la falta de asistencia justificada a las clases de prácticas y TGR no será penalizada. El estudiante que no concurra a la prueba objetiva en el período oficial de evaluación tendrá la condición de ?No presentado? (NP). La realización fraudulenta de las pruebas o actividades de evaluación, una vez comprobada, implicará directamente la calificación de suspenso '0' en la materia en la oportunidad correspondiente.

Fuentes de información



Básica	- Aho, A.V.; Lam, M.; Sethi, R.; Ullman, J.D. (2008). Compiladores: Principios, técnicas y herramientas. Addison-Wesley - Hopcroft, J.E.; Motwani, R.; Ullman, J. D. (2002). Introducción a la teoría de autómatas, lenguajes y computación. Addison-Wesley - Garrido, A.; Iñesta, J.M.; Moreno, F.; Pérez, J.A. (2004). Diseño de compiladores. Publicaciones de la Universidad de Alicante - Daniel Jurafsky; James H. Martin (2022). Speech and Language Processing, Draft 3rd Edition. https://web.stanford.edu/~jurafsky/slp3/ - Christopher D. Manning; Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing. MIT Press (ISBN 78-0262133609)
Complementaria	- Sudkamp, T.A. (1994). Languages and machines. Addison-Wesley - Aho, A.V.; Ullman, J.D. (1972). The theory of parsing, translation and compiling. Prentice-Hall - Loudon, D.K. (2004). Construcción de compiladores. Principios y Práctica. Paraninfo Thomson Learning - Yoav Goldberg (2017). Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers (ISBN 978-1627052986) - Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022). Natural Language Processing with Transformers. O'Reilly Media (ISBN 9781098103248) - Yue Zhang and Zhiyang Teng (2021). Natural Language Processing: A Machine Learning Perspective. Cambridge University Press (ISBN 9781108420211)

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Álgebra Lineal/614G02001

Fundamentos de Programación II/614G02009

Fundamentos de Programación I/614G02004

Inferencia Estadística/614G02007

Probabilidad y Estadística Básica/614G02003

Asignaturas que se recomienda cursar simultáneamente

Asignaturas que continúan el temario

Lenguaje Natural y Minería de Textos/614G02043

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