



Guía docente				
Datos Identificativos				2021/22
Asignatura (*)	Recuperación de la Información	Código	614G01040	
Titulación	Grao en Enxeñaría Informática			
Descritores				
Ciclo	Periodo	Curso	Tipo	Créditos
Grado	2º cuatrimestre	Tercero	Optativa	6
Idioma	Castellano			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da InformaciónComputación			
Coordinador/a	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Profesorado	Barreiro Garcia, Álvaro Landín Piñeiro, Alfonso	Correo electrónico	alvaro.barreiro@udc.es alfonso.landin@udc.es	
Web	http://www.dc.fi.udc.es/~barreiro/IRdocen/IRcourse.html			
Descripción general	<p>Esta materia aborda la recuperación de información en repositorios de documentos textuales y el web. Se estudian modelos, técnicas y algoritmos actuales que permiten el crawling, procesamiento, indexación y búsqueda en colecciones de textos del rango de gigabytes, hasta los terabytes de información que se manejan en el web. En esta materia el estudiante comprenderá la arquitectura de los motores de búsqueda de internet usados por las grandes compañías de Search Engines (Google, Bing, Yahoo, Yandex, Baidu, etc) y en las prácticas de la misma podrá desarrollar los módulos principales de un motor de búsqueda. La Recuperación de Información y en particular en el web plantea extraordinarios retos debido al volumen y heterogeneidad de los datos y fuentes y al amplio rango de intereses de usuarios privados y corporativos, por todo ello es un campo con amplias posibilidades de negocio y empleo en informática.</p>			
Plan de contingencia	<p>1. Modificacións nos contidos Sin modificacións.</p> <p>2. Metodoloxías *Metodoloxías docentes que se manteñen Todas *Metodoloxías docentes que se modifican Ninguna</p> <p>3. Mecanismos de atención personalizada ao alumnado Mail, teams e moodle.</p> <p>4. Modificacións na avaliación Ningunha *Observacións de avaliación: El examen final en las dos oportunidades será presencial si las condiciones lo permiten</p> <p>5. Modificacións da bibliografía ou webgrafía Sin modificacións</p>			

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A17	Conocimiento y aplicación de las características, funcionalidades y estructura de los sistemas distribuidos, las redes de computadores e internet, y diseñar e implementar aplicaciones basadas en ellas.
A19	Conocimiento y aplicación de las herramientas necesarias para el almacenamiento, procesamiento y acceso a los sistemas de información, incluidos los basados en web.



A45	Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
B1	Capacidad de resolución de problemas
B3	Capacidad de análisis y síntesis
B4	Capacidad para organizar y planificar
B5	Habilidades de gestión de la información
B6	Toma de decisiones
B7	Preocupación por la calidad
C2	Dominar la expresión y la comprensión de forma oral y escrita de un idioma extranjero.
C3	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.
C6	Valorar críticamente el conocimiento, la tecnología y la información disponible para resolver los problemas con los que deben enfrentarse.
C7	Asumir como profesional y ciudadano la importancia del aprendizaje a lo largo de la vida.
C8	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		
Conocer, comprender y analizar los distintos modelos Recuperación de Información, las técnicas para su implementación eficiente y la metodología de evaluación de los mismos.	A17 A19 A45	B1 B4 B5 B7	C2 C3 C6 C7 C8
Conocer, comprender y analizar las plataformas software para la creación de estos sistemas.	A17 A19	B1 B4 B5 B6	C2 C6 C7 C8
Planear y realizar la evaluación de los sistemas Recuperación de Información . Analizar los resultados de la evaluación de los sistemas de RI para mejorarlos en su eficacia y eficiencia.	A45	B1 B3 B4 B5 B6	C2 C6 C7 C8
Ser capaces de un correcto tratamiento de los aspectos éticos, de privacidad, confidencialidad y de seguridad de los estos sistemas.		B1 B4 B5 B6	C2 C3 C6 C7 C8

Contenidos	
Tema	Subtema
Introducción	Recuperación de Información e Search Engines. Arquitectura dun Search Engine. Grandes retos.
Recopilación de información.	Crawling e feeds.
Procesamento de texto.	Preprocesamento. Parsing, documentos estructurados, anchor text e análisis de enlaces, internacionalización
Indices e procesado eficiente.	Indices Invertidos, compresión, construcción, procesado eficiente de consultas sobre índices invertidos
Formulación de consultas e presentación de resultados	Transformación de consultas, relevance feedback, pseudo-feedback, snippets e visualización de resultados



Modelos de recuperación de información.	Booleano, espacio vectorial, probabilístico, BM25, Language Models e Relevance Models.
Evaluación de sistemas de Recuperación de Información.	Datasets e iniciativas de evaluación. Métricas de eficacia e eficiencia. Training e test. Significancia estadística
Búsqueda distribuida e social.	eMeta-buscaadores y búsqueda distribuida, blogs, redes sociais, sistemas de recomendación.

Planificación				
Metodoloxías / pruebas	Competencias / Resultados	Horas lectivas (presenciales y virtuales)	Horas traballo autónomo	Horas totales
Lecturas	A17 A19 A45 B3 B5 B7 C2 C6 C7 C8	2	12	14
Prácticas de laboratorio	A17 A19 A45 B1 B4 B5 B6 B7 C3	14	21	35
Solución de problemas	A17 A19 A45 B1 B5 B6 C6 C7 C8	4	12	16
Prueba mixta	A17 A19 A45 B1 C2 C6 C7 C8	2	14	16
Trabajos tutelados	A17 A19 A45 B7 C2 C3 C6 C7 C8	3	9	12
Sesión magistral	A17 A19 A45 B7 C2 C6 C7 C8	19	38	57
Atención personalizada		0		0

(*) Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos

Metodoloxías	
Metodoloxías	Descrición
Lecturas	Lecturas para consolidar y complementar los conocimientos adquiridos. Temas: técnicas, aplicaciones, sistemas industriales.
Prácticas de laboratorio	Prácticas de laboratorio sobre plataformas de desarrollo de amplo uso na industria, en las compañías de Search Engines y en los grupos de investigación (Lucene y Nutch)
Solución de problemas	Problemas y cuestiones breves para asentar y profundizar los contenidos expuestos en las sesiones magistrales.
Prueba mixta	Prueba que versará sobre los contenidos fundamentales de la materia.
Trabajos tutelados	Trabajos y problemas realizados de forma autónoma por el estudiante y tutelados por el profesor
Sesión magistral	El estudiante asistirá a las explicaciones dadas por el profesor sobre los distintos modelos, técnicas y algoritmos de Recuperación de Información. El profesor utilizará distintos niveles de abstracción-detalle y orientará al estudiante en las lecturas fundamentales y complementarias.

Atención personalizada	
Metodoloxías	Descrición
Prácticas de laboratorio	Lecturas: Las lecturas complementarias pueden requerir atención personalizada.
Solución de problemas	Problemas: Algunos problemas de mayor dificultad pueden requerir atención personalizada. Prácticas laboratorio: Además de evaluar el resultado de la práctica conforme a los requisitos exigidos se hace un seguimiento del desarrollo de las mismas. Debe respetarse la autonomía del estudiante para que adquiera mayor destreza con las plataformas software empleados pero el profesor podrá resolver ciertas dificultades que puedan bloquear al estudiante un tiempo excesivo dada la planificación de la asignatura.



Evaluación			
Metodologías	Competencias / Resultados	Descripción	Calificación
Prácticas de laboratorio	A17 A19 A45 B1 B4 B5 B6 B7 C3	Seguimiento, defensa y evaluación de los resultados de las prácticas realizados en las horas de las clases prácticas de laboratorio.	30
Sesión magistral	A17 A19 A45 B7 C2 C6 C7 C8	Los contenidos de las sesiones magistrales serán evaluados en la prueba mixta.	0
Solución de problemas	A17 A19 A45 B1 B5 B6 C6 C7 C8	Resultados obtenidos en la realización de problemas y/o trabajos y/o cuestiones y/o cuestionarios realizados en las clases y controles llevados a cabo.	20
Prueba mixta	A17 A19 A45 B1 C2 C6 C7 C8	Preguntas sobre los conocimientos adquiridos en las sesiones magistrales, actividades prácticas y de problemas, trabajos, cuestiones y cuestionarios. Es obligatorio alcanzar un 40% de la calificación para superar la materia.	50

Observaciones evaluación

Si no se obtiene la puntuación mínima en la parte que lo requieren, la nota máxima del alumno será 4'5. Para los alumnos a tiempo parcial el baremo de calificaciones y la evaluación continua son los mismos que para los otros alumnos.

Fuentes de información

Básica	- W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education
Complementaria	- C.D. Manning, P. Raghavan, H. Schütze. (2008). Introduction to Information Retrieval. Cambridge University Press - R. Baeza-Yates and B. Ribeiro-Neto (2011). Modern Information Retrieval (second edition). Addison Wesley/Pearson Education - F. Casheda, J.M. Fernández, J. Huete (editores) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. . Ra-Ma

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Asignaturas que se recomienda cursar simultáneamente

Asignaturas que continúan el temario

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