



Guía Docente				
Datos Identificativos				2013/14
Asignatura (*)	Recuperación da Información	Código	614G01040	
Titulación	Grao en Enxeñaría Informática			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Grao	2º cuatrimestre	Terceiro	Obrigatoria	6
Idioma	Castelán			
Prerrequisitos				
Departamento	Computación			
Coordinación	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Profesorado	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Web	http://www.dc.fi.udc.es/~barreiro/IRdocen/IRcourse.html			
Descrición xeral	<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 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>			

Competencias da titulación	
Código	Competencias da titulación
A17	Coñecemento e aplicación das características, funcionalidades e estrutura dos sistemas distribuídos, as redes de computadores e internet, e deseñar e implementar aplicacións baseadas nelas.
A19	Coñecemento e aplicación das ferramentas necesarias para o almacenamento, procesamento e acceso aos sistemas de información, incluídos os baseados en web.
B1	Capacidade de resolución de problemas
B4	Capacidade para organizar e planificar
B5	Habilidades de xestión da información
B6	Toma de decisións
B7	Preocupación pola calidade
C2	Dominar a expresión e a comprensión de forma oral e escrita dun idioma estranxeiro.
C3	Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.
C6	Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.
C7	Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.
C8	Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.

Resultados da aprendizaxe			
Competencias de materia (Resultados de aprendizaxe)			Competencias da titulación
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	B1	C2
	A19	B4	C3
		B5	C6
		B7	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.		B1 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

Contidos	
Temas	Subtemas
Introducción	Recuperación de Información y Search Engines. Arquitectura de un Search Engine. Grandes retos.
Recopilación de información.	Crawling y feeds.
Procesamiento de texto.	Preprocesamiento. Parsing, documentos estructurados, anchor text y análisis de enlaces, internacionalización
Indices y procesado eficiente.	Indices Invertidos, compresión, construcción, procesado eficiente de consultas sobre índices invertidos
Formulación de consultas y presentación de resultados	Transformación de consultas, relevance feedback, pseudo-feedback, snippets y visualización de resultados
Modelos de recuperación de información.	Booleano, espacio vectorial, probabilístico, BM25, Language Models y Relevance Models.
Evaluación de sistemas de Recuperación de Información.	Colecciones e iniciativas de evaluación. Métricas de eficacia y eficiencia. Training y test. Significancia estadística
Búsqueda distribuida y social.	Meta-buscadores y búsqueda distribuida, blogs, redes sociales, sistemas de recomendación.

Planificación			
Metodoloxías / probas	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Lecturas	2	12	14
Prácticas de laboratorio	14	21	35
Solución de problemas	7	21	28
Proba mixta	2	14	16
Sesión maxistral	19	38	57
Atención personalizada	0		0

*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

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 amplio uso en la 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 en los contenidos expuestos en las sesiones magistrales.
Proba mixta	Prueba que versará sobre los contenidos fundamentales de la materia.
Sesión maxistral	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
Lecturas	Lecturas: Las lecturas complementarias pueden requerir atención personalizada.
Prácticas de laboratorio	Problemas: Algunos problemas de mayor dificultad pueden requerir atención personalizada.
Solución de problemas	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.

Avaliación

Metodoloxías	Descrición	Cualificación
Prácticas de laboratorio	Seguimiento de las prácticas y evaluación sobre el resultado alcanzado.	15
Sesión maxistral	A participación activa nestas sesión será avaliada con un 5% e o contido das sesións maxistrais será obxecto do 70% avaliada na proba mixta	5
Solución de problemas	Asistencia a las actividades, participación y resultados en la realización de problemas.	10
Proba mixta	Cuestiones sobre los conocimientos adquiridos. Cuestiones que impliquen razonamiento en base a los conocimientos adquiridos para resolver problemas prácticos de interés real en recuperación de información.	70

Observacións avaliación

--

Fontes de información

Bibliografía básica	- W.B. Croft, D. Metzler, T. Strohmán. (2009). Search Engines. Information Retrieval in Practice. Pearson Education
Bibliografía 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. Cacheda, J.M. Fernández, J. Huete (editores) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. . Ra-Ma

Recomendacións

Materias que se recomienda ter cursado previamente
Materias que se recomienda cursar simultaneamente
Materias que continúan o temario
Observacións



(*)A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías