



## Guía Docente

Datos Identificativos					2014/15
Asignatura (*)	Recuperación da Información		Código	614G01040	
Titulación					
Descritores					
Ciclo	Período	Curso	Tipo	Créditos	
Grao	2º cuatrimestre	Terceiro	Optativa	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	<a href="http://www.dc.fi.udc.es/~barreiro/IRdocen/IRcourse.html">http://www.dc.fi.udc.es/~barreiro/IRdocen/IRcourse.html</a>				
Descrición xeral	<p>Esta materia aborda a recuperación de información en repositorios de documentos textuais e a web. Estúdanse modelos, técnicas e algoritmos actuais que permiten o crawling, procesamento, indexación e procura en coleccións de textos do rango de gigabytes, ata os terabytes de información que se manexan na web. Nesta materia o estudante comprenderá a arquitectura dos motores de procura de internet usados polas grandes compañías de Search Engines (Google, Bing, Yahoo, Yandex, etc) e nas prácticas da mesma poderá desenvolver os módulos principais dun motor de procura. A Recuperación de Información e en particular na web expón extraordinarios retos debido ao volume e heteroxeneidade dos datos e fontes e ao amplo rango de intereses de usuarios privados e corporativos, por todo iso é un campo con amplas posibilidades de negocio e emprego en informática.</p>				

## Competencias da titulación

Código	Competencias da titulación

## Resultados da aprendizaxe

Competencias de materia (Resultados de aprendizaxe)	Competencias da titulación		
Coñecer, comprender e analizar os distintos modelos Recuperación de Información, as técnicas para a súa implementación eficiente e a metodoloxía de avaliación dos mesmos.	A17 A19	B1 B4 B5 B7	C2 C3 C6 C7 C8
Coñecer, comprender e analizar as plataformas software para a creación destes sistemas.	A17 A19	B1 B4 B5 B6	C2 C6 C7 C8
Planear e realizar a avaliación dos sistemas Recuperación de Información . Analizar os resultados da avaliación dos sistemas de RI para melloralos na súa eficacia e eficiencia.		B1 B4 B5 B6	C2 C6 C7 C8
Ser capaces dun correcto tratamento dos aspectos éticos, de privacidade, confidencialidade e de seguridade dos estes sistemas.		B1 B4 B5 B6	C2 C3 C6 C7 C8

## Contidos

Temas	Subtemas



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 estruturados, anchor text e análise de enlaces, internacionalización
Índices e procesado eficiente.	Índices Invertidos, compresión, construción, 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 avaliación. Métricas de eficacia e eficiencia. Training e test. Significancia estadística
Búsqueda distribuída e social.	eMeta-buscadores y búsqueda distribuída, blogs, redes sociais, 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	4	12	16
Proba mixta	2	14	16
Traballos tutelados	3	9	12
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 e complementar os coñecementos adquiridos. Temas: técnicas, aplicacións, sistemas industriais.
Prácticas de laboratorio	Prácticas de laboratorio sobre plataformas de desenvolvemento de amplo uso na industria, nas compañías de Search Engines e nos grupos de investigación (Lucene y Nutch)
Solución de problemas	Problemas e cuestións breves para asentir e profundizar nos contidos expostos nas sesións maxistrais.
Proba mixta	Prueba que versará sobre os contidos fundamentais da materia.
Traballos tutelados	Traballos e problemas realizados de forma autónoma polo estudante e tutelados polo I profesor
Sesión maxistral	O estudante asistirá ás explicacións dadas polo profesor sobre os distintos modelos, técnicas e algoritmos de Recuperación de Información. O profesor utilizará distintos niveis de abstracción-detalle e orientará ao estudante nas lecturas fundamentais e 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.

Avaliación		
Metodoloxías	Descrición	Cualificación
Prácticas de laboratorio	Seguimento das prácticas e avaliación sobre o resultado alcanzado. É obrigatorio alcanzar un 40% da cualificación para superar a materia.	30
Sesión maxistral	Os contidos das sesións maxistrais serán avaliados na proba mixta.	0
Solución de problemas	Asistencia ás actividades, participación e resultados na realización de problemas e traballos tutelados.	20
Proba mixta	Preguntas sobre os coñecementos adquiridos nas sesións maxistrais, actividades prácticas e de problemas e traballos. É obrigatorio alcanzar un 40% da cualificación para superar a materia	50

Observacións avaliación
Se non se obtén a puntuación mínima nas partes que o requiren, a nota máxima do alumno será 4'5 Para os alumnos a tempo parcial deben porse en contacto co profesor a primeira semana do cuadrimestre para a avaliación alternativa

Fontes de información	
<b>Bibliografía básica</b>	- W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education
<b>Bibliografía complementaria</b>	- 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

Recomendacións
<b>Materias que se recomenda ter cursado previamente</b>
<b>Materias que se recomenda cursar simultaneamente</b>
<b>Materias que continúan o temario</b>
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