



Guía Docente				
Datos Identificativos				2012/13
Asignatura (*)	Modelos e Técnicas Avanzadas de Recuperación da Información	Código	614451119	
Titulación	Mestrado Universitario en Enxeñaría de Sistemas Informáticos			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Primeiro	Optativa	3
Idioma				
Prerrequisitos				
Departamento	Computación			
Coordinación	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Profesorado	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Web	www.dc.fi.udc.es/~barreiro/mtari.html			
Descrición xeral				

Competencias da titulación	
Código	Competencias da titulación
A1	Análise estatístico.
A2	Arquitectura de computadores.
A3	Arquitectura de redes.
A4	Bases de datos.
A5	Capacidade para entender e avaliar especificacións internas e externas.
A6	Cifrado e protección de datos.
A7	Dirección, planificación e xestión de proxectos.
A8	Deseño e arquitectura de Sistemas de Información.
A10	Xestión do cambio e do coñecemento.
A11	Enxeñaría do software.
A12	Integración de sistemas.
B1	Capacidade de análise e síntese.
B2	Capacidade de organización e planificación de proxectos informáticos.
B3	Capacidade de xestión da información.
B4	Capacidade de resolución de problemas.
B6	Traballo en equipo.
B8	Razoamento crítico.
B9	Compromiso ético.
B10	Aprendizaxe autónoma.
B12	Creatividade.
B15	Motivación pola calidade.
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.
C4	Desenvolverse para o exercicio dunha cidadanía aberta, culta, crítica, comprometida, democrática e solidaria, capaz de analizar a realidade, diagnosticar problemas, formular e implantar solucións baseadas no coñecemento e orientadas ao ben común.
C6	Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.
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 de Recuperación de Información, las técnicas para su implementación eficiente y la metodología de evaluación de los mismos.	AP8 AP12	BP1 BP3 BP8 BP10	CM3
Conocer, comprender y analizar las plataformas software para la creación de sistemas de RI	AP8 AP12	BP1 BP10	CM3
Diseñar y construir nuevos sistemas de RI o mejorar los sistemas existentes	AP7 AP8 AP11 AP12	BP1 BP2 BP3 BP4 BP6 BP8 BP10 BP12 BP15	CM3 CM6 CM8
Planear y realizar la evaluación de los sistemas de RI. Analizar los resultados de la evaluación de los sistemas de RI para mejorarlos en su eficacia y eficiencia	AP1 AP5 AP10 AP11	BP1 BP3 BP4 BP8 BP10 BP12 BP15	CM3 CM6
Ser capaces de un correcto tratamiento de los aspectos éticos, de privacidad y de seguridad de los sistemas de recuperación de información.	AP2 AP3 AP4 AP6 AP8 AP10	BP1 BP2 BP3 BP4 BP8 BP9 BP10 BP12 BP15	CM3 CM4 CM8

Contidos	
Temas	Subtemas
Introducción a RI (1 horas)	Introducción a los modelos, técnicas, evaluación, sistemas de RI.
Evaluación de sistemas de RI (2 horas)	Tareas y métricas. Colecciones de referencia. TREC. Significancia estadística.
Modelo booleano de RI (1 hora)	Representación de documentos, consultas y medidas de similaridad. Aplicaciones tradicionales exitosas. Implementaciones eficientes.
Modelos de espacio vectorial de RI (2 horas)	Representación de documentos, consultas y medidas de similaridad. Esquemas de pesado. Normalización. Implementación eficiente.
Modelo clásico probabilístico de RI (2 horas)	Probability Ranking Principle Derivación del modelo clásico probabilístico. Otros modelos probabilísticos: 2-Poisson, Okapi, Redes de Inferencia. Implementación eficiente.



Modelos estadísticos de lenguaje para RI (2 horas)	Modelos de lenguaje. Suavización. Aprendizaje y estimación de parámetros. Modelos de lenguaje basados en relevancia. Implementación eficiente.
Modelo Latent Semantic Indexing de RI (1 horas)	Reducción de dimensionalidad basada en SVD. Derivación del modelo LSI. Cuestiones sobre la escalabilidad del modelo y nuevas aproximaciones: LSI eficiente, LPI, etc.
Realimentación de relevancia en RI (2 horas)	Realimentación de relevancia bajo el modelo vectorial (Rocchio) y probabilístico. Local Context Analysis (LCA) y expansión de consultas.
Compresión y construcción de índices. Procesamiento de queries. (3 horas)	Algoritmos de construcción de índices. Algoritmos de compresión de índices: compresión de listas, compresión de frecuencias, compresión del léxico. Procesamiento eficiente de consultas.
WebIR: modelos de retrieval, eficiencia, escalabilidad, problemas propios y oportunidades en RI en el web. (3 horas)	Modelos de retrieval para el web. Análisis de links. Page Rank y HITS. Implementación de search engines. Oportunidades de RI en el web.
IR distribuida (3 horas)	IR paralela y distribuida. Modelos de IR distribuida: selección de recursos, enrutado de consultas, fusión de resultados. Aplicaciones novedosas en IR distribuida.
NLP para IR (2 horas)	Preprocesado. Parsing. Stemming. Extracción de información.

### Planificación

Metodoloxías / probas	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Sesión maxistral	16	16	32
Prácticas de laboratorio	15	0	15
Seminario	5	5	10
Traballos tutelados	0	7	7
Proba obxectiva	2	0	2
Lecturas	0	7	7
Atención personalizada	2	0	2

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

### Metodoloxías

Metodoloxías	Descrición
Sesión maxistral	O alumnos asistirá e atenderá as explicación dadas sobre os profesor dos modelos, técnicas e Sistemas de Recuperación de Información.
Prácticas de laboratorio	Indexación, recuperación e evaluación de eficacia e rendimento con unha colección de documentos pública sobre unha plataforma educativa e de investigación en RI (Terrier o Lemur).



Seminario	Alguns temas trataranse como seminarios poñendo máis énfasis na contextualización, resultados e implicación e deixando aspectos formales e técnicos muy precisos solo para os alumnos que manifesten interese especial por esos temas.
Traballos tutelados	Identificación de unha nova aplicación exitosa das técnicas de Recuperación de Información. Diseño da arquitectura software de unha implementación posible para esa aplicación.
Proba obxectiva	Examen dos contidos fundamentais expostos nas clases maxistrais e seminarios.
Lecturas	Lecturas para consolidación dos coñecementos expostos en clases maxistrais e seminarios.

### Atención personalizada

Metodoloxías	Descrición
Traballos tutelados Lecturas	A propia natureza do traballo, deseño de unha solución novedosa de unha aplicación usando técnicas de RI, implica revisar o deseño do estudante.  A consolidación dos coñecementos expostos en clases maxistrais e seminarios pode requirir atención personalizada.

### Avaliación

Metodoloxías	Descrición	Cualificación
Prácticas de laboratorio	Seguimento na aula de prácticas da correcta realización das prácticas encomendadas	0
Traballos tutelados	Corrección, adecuación e viabilidade técnica do deseño proposto para unha nova aplicación das técnicas de RI.	0
Proba obxectiva	Cuestións curtas sobre coñecemos adquiridos.  Cuestión que impliquen razoamento sobre a base dos coñecementos adquiridos para resolver problemas novos.	100
Outros		

### Observacións avaliación

Esta materia non se impartirá presencialmente no curso 2012/2012 e non ten recursos docente asignado, razón pola que o profesor so pode comprometerse a realizar a proba obxectiva.

### Fontes de información

<b>Bibliografía básica</b>	<ul style="list-style-type: none"> <li>- R. Baeza-Yates and B. Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley, May 1999.</li> <li>- H. Witten, A. Moffat, and T. C. Bell (1999). Managing Gigabytes: Compressing and Indexing Documents and Images, 2nd edition. Morgan Kaufmann</li> <li>- C. J. Van Rijsbergen (1979). Information Retrieval (2nd Edition). Butterworths, London</li> </ul>
<b>Bibliografía complementaria</b>	<ul style="list-style-type: none"> <li>- W.B. Croft and J. Lafferty (2003). Language Modeling for Information Retrieval. Kluwer Academic Publishers</li> <li>- A. Moffat and A. Turpin (2002). Compression and Coding Algorithms. Kluwer Academic Publishers</li> <li>- R. K. Belew. (2001). Finding Out About.. Cambridge Press</li> <li>- C. D. Manning and H. Schütze. (1999). Foundations of Statistical Natural Language Processing. MIT Press</li> <li>- David A. Grossman and Ophir Frieder. (1998). Information Retrieval: Algorithms and Heuristics. Kluwer Academic Publishers,</li> <li>- E. Voorhees and D.K. Harman. (2005). TREC: experiment and evaluation in information retrieval.. MIT Press</li> </ul>

### Recomendacións

**Materias que se recomenda ter cursado previamente**

**Materias que se recomenda 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