



Guía docente				
Datos Identificativos				2012/13
Asignatura (*)	Recuperación de la información y web semántica	Código	614502010	
Titulación	Mestrado Universitario en Enxeñaría Informática (plan 2012)			
Descriptorios				
Ciclo	Periodo	Curso	Tipo	Créditos
Máster Oficial	1º cuatrimestre	Primero	Obligatoria	6
Idioma	Castellano			
Prerrequisitos				
Departamento	ComputaciónTecnoloxías da Información e as Comunicaciós			
Coordinador/a	Barreiro Garcia, Álvaro	Correo electrónico	alvaro.barreiro@udc.es	
Profesorado	Barreiro Garcia, Álvaro Blanco Gonzalez, Roi Cacheda Seijo, Fidel Vázquez Naya, José Manuel	Correo electrónico	alvaro.barreiro@udc.es roi.blanco@udc.es fidel.cacheda@udc.es jose.manuel.vazquez.naya@udc.es	
Web				
Descripción general	<p>Los modelos, técnicas y algoritmos de recuperación de información estudiados en esta materia permitirán a los estudiantes comprender la arquitectura de los Search Engines para el web. Además los contenidos prácticos de la misma les capacitarán para construir sus propios buscadores para trabajar sobre repositorios de documento o la web. Además durante los últimos años ha habido un interés creciente en idear una web semántica a partir de meta-datos y anotaciones. Una web basada en documentos xml y tags, meta-datos y esquemas, sin duda facilitaría los enormes retos a los que se enfrenta la recuperación de información web. En esta asignatura se abordan también los modelos, técnicas y algoritmos de mayor impacto desarrollados en los últimos años con el objetivo de materializar una web semántica. La Recuperación de Información en grandes colecciones de documentos y en la web plantea enormes retos (volumen de datos, datos distribuidos, alto porcentaje de datos volátiles, datos no estructurados y redundantes, heterogeneidad, calidad de los datos y confianza) y la Web Semántica parte ya del gran reto de la extracción de información cuando los meta-datos no son expuestos públicamente y plantea nuevos retos como los del matching de ontologías, resolución de entidades o una dificultad mayor en cuanto a la heterogeneidad y calidad de los datos y a la indexación y búsqueda semántica. Por todo ello la Recuperación de Información y la Web semántica constituyen uno de los campos de mejores salidas profesionales en informática con oportunidades de negocio y empleo no sólo en las grandes compañías de Search Engines sino también en muchas pequeñas y medianas compañías.</p>			

Competencias de la titulación	
Código	Competencias de la titulación
A5	Capacidad de comprender y saber aplicar el funcionamiento y organización de Internet, las tecnologías y protocolos de redes de nueva generación, los modelos de componentes, software intermediario y servicios.
A9	Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.
B1	Capacidad de resolución de problemas.
B3	Capacidad de análisis y síntesis.
B5	Habilidades de gestión de la información.
B7	Preocupación por la calidad.
B9	Capacidad para generar nuevas ideas (creatividad).
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.
C5	Entender la importancia de la cultura emprendedora y conocer los medios al alcance de las personas emprendedoras.
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

Competencias de materia (Resultados de aprendizaje)	Competencias de la titulación		
	AP5	BP3	CP2 CP6 CP8
Conocer, comprender y analizar los distintos modelos Recuperación de Información y Web Semántica, las técnicas para su implementación eficiente y la metodología de evaluación de los mismos.	AP5	BP3	CP2 CP6 CP8
Conocer, comprender y analizar las plataformas software para la creación de estos sistemas.	AP5	BP3	CP2 CP3 CP6 CP7 CP8
Diseñar y construir nuevos sistemas o mejoras en sistemas existentes.	AP5 AP9	BP1 BP3 BP5 BP9	CP3 CP5 CP6 CP7
Planear y realizar la evaluación de los sistemas de Recuperación de Información y Web Semántica . Analizar los resultados de la evaluación de los sistemas para mejorarlos en su eficacia y eficiencia.	AP5 AP9	BP1 BP5 BP7	CP3 CP5 CP6 CP7
Ser capaces de un correcto tratamiento de los aspectos éticos, de privacidad, confidencialidad y de seguridad de los estos sistemas.			CP6

## Contenidos

Tema	Subtema
Introducción.	Recuperación de Información y Arquitectura de los Search Engines
Recopilación de información.	Crawling y feeds.
Procesamiento de textos y páginas web.	Preprocesamiento de textos y parsing. Anchor text y análisis de enlaces en la web, internacionalización.
Indices y ranking.	Construcción y compresión de índices. Procesado eficiente de consultas.
Formulación de consultas y presentación de resultados.	Formulación y reformulación de consultas. Snippets. Visualización de resultados.
Modelos de recuperación de información.	Booleano, vectorial, probabilístico, language models.
Evaluación.	Evaluación de sistemas de Recuperación de Información. Iniciativas de evaluación. Métricas de eficacia y eficiencia. Diseño de la evaluación: training, test y significancia estadística. Evaluación crowdsourcing.
Minería de textos.	Clasificación y clustering de documentos.
Búsqueda distribuida y social	Meta-buscadores y búsqueda distribuida, blogs, micro-blogs y redes sociales.
Sistemas de recomendación.	Filtrado colaborativo. Modelos y algoritmos para recomendación. Sistemas de recomendación
Introducción a la Web Semántica	La Web Semántica. Ontologías: definición, tipos y ejemplos.
Descripción y consulta de recursos	Lenguajes XML, RDF y RDF Schema. Lenguaje de consultas SPARQL. Lenguaje OWL. Herramientas de desarrollo de ontologías. Librerías para el manejo de ontologías. Repositorios RDF.
Razonamiento y reglas	Fundamentos de lógica y razonamiento. Representación de reglas semánticas. Motores de razonamiento.
Aplicaciones de Web Semántica	Linked Data, FOAF, Dublin Core, WordNet. Anotación semántica. Buscadores semánticos. Servicios Web Semánticos.

## Planificación



Metodoloxías / probas	Horas presenciais	Horas no presenciais / traballo autónomo	Horas totais
Lecturas	1	15	16
Prácticas de laboratorio	20	30	50
Solución de problemas	4	12	16
Prueba mixta	2	18	20
Sesión magistral	16	32	48
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
Prácticas de laboratorio	Prácticas de laboratorio sobre plataformas de desarrollo de uso comercial (Lucene, Terrier, Nutch, Jena, Protege, Pellet)
Solución de problemas	Problemas y cuestiones breves para asentar y profundizar en los contenidos expuestos en las sesiones magistrales.
Prueba mixta	Prueba que versará sobre los contenidos fundamentales de la materia.
Sesión magistral	El estudiante asistirá a las explicaciones dadas por el profesor sobre los distintos modelos, técnicas y algoritmos de Recuperación de Información y Web Semántica. 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	Seguimiento del desarrollo de las prácticas en las horas reservadas de laboratorio y atención al estudiante en los casos necesarios de problemas de particular dificultad
Solución de problemas	

Evaluación		
Metodoloxías	Descrición	Calificación
Prácticas de laboratorio	Seguimiento de las prácticas y evaluación sobre el resultado alcanzado.	50
Prueba 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 y web semántica.	50

Observaciones evaluación

Fuentes de información



<b>Básica</b>	<ul style="list-style-type: none"><li>- Bob DuCharme (2011). Learning SPARQL. O'Reilly</li><li>- C.D. Manning, P. Raghavan, H. Schütze. (2008). Introduction to Information Retrieval. Cambridge University Press</li><li>- R. Baeza-Yates and B. Ribeiro-Neto. (2011). Modern Information Retrieval (second edition) . Addison Wesley/Pearson Education</li><li>- F. Cacheda, J.M. Fernández, J. Huete (eds.) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. Ra-Ma</li><li>- W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education</li><li>- John Hebel, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, Mike Dean. (2009). Semantic Web Programming. Wiley</li></ul>
<b>Complementaria</b>	

#### 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