



Guía docente				
Datos Identificativos			2014/15	
Asignatura (*)	Recuperación de la información y web semántica		Código	614502010
Titulación	Mestrado Universitario en Enxeñaría Informática (plan 2012)			
Descriptor				
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	Correo electrónico	alvaro.barreiro@udc.es	
	Cacheda Seijo, Fidel		fidel.cacheda@udc.es	
	Parapar López, Javier		javier.parapar@udc.es	
	Vázquez Naya, José Manuel		jose.manuel.vazquez.naya@udc.es	
Web				
Descripción general	Os modelos, técnicas e algoritmos de recuperación de información estudados nesta materia permitirán aos estudantes comprender a arquitectura dos Search Engines para a web. Ademais os contidos prácticos da mesma capacitaranlles para construír os seus propios buscadores para traballar sobre repositorios de documento ou a web. Ademais durante os últimos anos houbo un interese crecente en idear unha web semántica a partir de meta-datos e anotacións. Unha web baseada en documentos xml e tags, meta-datos e esquemas, sen dúbida facilitaría os enormes retos aos que se enfronta a recuperación de información web. Nesta materia abórdanse tamén os modelos, técnicas e algoritmos de maior impacto desenvolvidos nos últimos anos co obxectivo de materializar unha web semántica. A Recuperación de Información en grandes coleccións de documentos e na web expón enormes retos (volumen de datos, datos distribuídos, alta porcentaxe de datos volátiles, datos non estruturados e redundantes, heteroxeneidade, calidade dos datos e confianza) e a Web Semántica parte xa do gran reto da extracción de información cando os meta-datos non son expostos publicamente e expón novos retos como os do matching de ontoloxías, resolución de entidades ou unha dificultade maior en canto á heteroxeneidade e calidade dos datos e á indexación e procura semántica. Por todo iso a Recuperación de Información e a Web semántica constitúen un dos campos de mellores saídas profesionais en informática con oportunidades de negocio e emprego non só nas grandes compañías de Search Engines senón tamén en moitas pequenas e medianas compañías.			

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.
A12	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.
B1	Capacidad de resolución de problemas.
B5	Habilidades de gestión de la información.
B10	Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática
B13	Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática
B14	Capacidad para la elaboración, planificación estratégica, dirección, coordinación y gestión técnica y económica de proyectos en todos los ámbitos de la Ingeniería en Informática siguiendo criterios de calidad y medioambientales
B17	Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinarios, siendo capaces de integrar estos conocimientos
B21	Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
B22	Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinarios) relacionados con su área de estudio



B23	Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
B25	Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
C4	Desarrollarse para el ejercicio de una ciudadanía abierta, culta, crítica, comprometida, democrática y solidaria, capaz de analizar la realidad, diagnosticar problemas, formular e implantar soluciones basadas en el conocimiento y orientadas al bien común.
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	
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		CP6 CP8
Conocer, comprender y analizar las plataformas software para la creación de estos sistemas.	AP5		CP6 CP7 CP8
Diseñar y construir nuevos sistemas o mejoras en sistemas existentes.	AP5 AP12	BP1 BP5 BP10 BP13 BP14 BP17 BM1 BM2 BM5	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	BP1 BP5	CP6 CP7
Ser capaces de un correcto tratamiento de los aspectos éticos, de privacidad, confidencialidad y de seguridad de los estos sistemas.		BM3	CP4 CP6

Contenidos	
Tema	Subtema
Introducción.	Recuperación de Información y los retos de la Web
Recuperación de Información en la Web.	Topología de la web: el grafo web. Arquitecturas de Search Engines. Ránking basado en contenido y análisis de enlaces. Learning to rank. Web spam. Gestión de datos web: identificadores de documentos, metadatos, duplicados. Interfaces para búsqueda y navegación.
Web Crawling.	Tipos de crawlers. Arquitectura. Esigenciasas de fresness y politeness. Algoritmos de crawling. Evaluación
Indexación.	Construcción y compresión de índices invertidos. Procesado de consultas.
Recuperación de Información en la Web con paralelismo y distribución.	Particionamiento y selección de colecciones. Particionamiento de índices. Recuperación de información paralela con arquitecturas MIMD y SIMD. Recuperación de Información basada en cluster. Recuperación de información distribuida y federada.



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 lo manejo de ontologías. Repositorios RDF.
Razonamiento y reglas	Fundamentos de lóxico y razonamiento. Representación de reglas semánticas. Motores de razonamiento.
Aplicaciones de la Web Semántica	Linked Data, FOAF, Dublin Core, WordNet. Anotación semántica. Buscadores semánticos. Servicios Web Semánticos.

Planificación			
Metodologías / pruebas	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales
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			

Metodologías	
Metodologías	Descripción
Lecturas	Lecturas para consolidar y complementar los conocimientos adquiridos
Prácticas de laboratorio	Prácticas de laboratorio sobre plataformas de desarrollo de uso comercial (Lucene, Terrier, Apache Solr, Nutch, Jena, Protege, Pellet)
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.
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	
Metodologías	Descripción
Prácticas de laboratorio Solución de problemas	Seguimiento do desenvolvemento das prácticas nas horas reservadas de laboratorio e atención ao estudante nos casos necesarios de problemas de particular dificultade

Evaluación		
Metodologías	Descripción	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

Aquellos estudiantes con matrícula a tiempo parcial o cualquier circunstancia justificada que impida la asistencia a las clases, deberán contactar con los docentes para determinar alternativas al seguimiento y evaluación de la materia.

Fuentes de información

Básica	- Bob DuCharme (2011). Learning SPARQL. O'Reilly - 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 (eds.) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. Ra-Ma - W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education - John Hebel, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, Mike Dean. (2009). Semantic Web Programming. Wiley
Complementaria	

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