



Guía Docente				
Datos Identificativos				2012/13
Asignatura (*)	Recuperación da información e web semántica	Código	614502010	
Titulación	Mestrado Universitario en Enxeñaría Informática (plan 2012)			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Primeiro	Obrigatoria	6
Idioma	Castelán			
Prerrequisitos				
Departamento	ComputaciónTecnoloxías da Información e as Comunicaci3ns			
Coordinaci3n	Barreiro Garcia, lvoro	Correo electr3nico	alvaro.barreiro@udc.es	
Profesorado	Barreiro Garcia, lvoro Blanco Gonzalez, Roi Cacheda Seijo, Fidel Vzquez Naya, Jos Manuel	Correo electr3nico	alvaro.barreiro@udc.es roi.blanco@udc.es fidel.cacheda@udc.es jose.manuel.vazquez.naya@udc.es	
Web				
Descrici3n xeral	<p>Los modelos, tcnicas y algoritmos de recuperaci3n de informaci3n estudiados en esta materia permitirn a los estudiantes comprender la arquitectura de los Search Engines para el web. Adems los contenidos prcticos de la misma les capacitarn para construir sus propios buscadores para trabajar sobre repositorios de documento o la web. Adems durante los ltimos aos ha habido un inters creciente en idear una web semntica a partir de meta-datos y anotaciones. Una web basada en documentos xml y tags, meta-datos y esquemas, sin duda facilitara los enormes retos a los que se enfrenta la recuperaci3n de informaci3n web. En esta asignatura se abordan tambin los modelos, tcnicas y algoritmos de mayor impacto desarrollados en los ltimos aos con el objetivo de materializar una web semntica. La Recuperaci3n de Informaci3n en grandes colecciones de documentos y en la web plantea enormes retos (volumen de datos, datos distribuidos, alto porcentaje de datos voltiles, datos no estructurados y redundantes, heterogeneidad, calidad de los datos y confianza) y la Web Semntica parte ya del gran reto de la extracci3n de informaci3n cuando los meta-datos no son expuestos pblicamente y plantea nuevos retos como los del matching de ontol3gias, resoluci3n de entidades o una dificultad mayor en cuanto a la heterogeneidad y calidad de los datos y a la indexaci3n y bsqueda semntica. Por todo ello la Recuperaci3n de Informaci3n y la Web semntica constituyen uno de los campos de mejores salidas profesionales en informtica con oportunidades de negocio y empleo no s3lo en las grandes compaas de Search Engines sino tambin en muchas pequeas y medianas compaas.</p>			

Competencias da titulaci3n	
C3digo	Competencias da titulaci3n
A5	Capacidade de comprender e saber aplicar o funcionamento e organizaci3n da internet, as tecnoloxias e protocolos de redes de nova xeraci3n, os modelos de compoentes, s3ftware intermediario e servizos.
A9	Capacidade para desear e avaliar sistemas operativos e servidores, e aplicaci3ns e sistemas baseados en computaci3n distribuida.
B1	Capacidade de resoluci3n de problemas.
B3	Capacidade de anlise e sntese.
B5	Habilidades de xesti3n da informaci3n.
B7	Preocupaci3n pola calidade.
B9	Capacidade para xerar novas ideas (creatividade).
C2	Dominar a expresi3n e a compresi3n de forma oral e escrita dun idioma estranxeiro.
C3	Utilizar as ferramentas bsicas das tecnoloxias da informaci3n e as comunicaci3ns (TIC) necesarias para o exercicio da sa profesi3n e para a aprendizaxe ao longo da sa vida.
C5	Entender a importancia da cultura emprendedora e coecer os medios ao alcance das persoas emprendedoras.
C6	Valorar criticamente o coecemento, a tecnoloxia e a informaci3n dispoible para resolver os problemas cos que deben afrontarse.
C7	Asumir como profesional e cidadn a importancia da aprendizaxe ao longo da vida.
C8	Valorar a importancia que ten a investigaci3n, a innovaci3n e o desenvolvemento tecnol3xico no avance socioecon3mico 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 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

Contidos

Temas	Subtemas
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 non presenciais / traballo autónomo	Horas totais
Lecturas	1	15	16
Prácticas de laboratorio	20	30	50
Solución de problemas	4	12	16
Proba mixta	2	18	20
Sesión maxistral	16	32	48
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
Prácticas de laboratorio	Prácticas de laboratorio sobre plataformas de desenvolvemento de uso comercial (Lucene, Terrier, Nutch, Jena, Protege, Pellet)
Solución de problemas	Problemas e cuestións breves para asentarse e profundizar nos contidos expostos nas sesións magistrais.
Proba mixta	Proba que versará sobre os contidos fundamentais da materia.
Sesión maxistral	El estudante asistirá a las explicacións dadas por el profesor sobre los distintos modelos, técnicas e algoritmos de Recuperación de Información e Web Semántica. El profesor utilizará distintos niveles de abstracción-detalle e orientará al estudante en las lecturas fundamentais e complementarias.

Atención personalizada	
Metodoloxías	Descrición
Prácticas de laboratorio Solución de problemas	Seguimento do desenvolvemento das prácticas nas horas reservadas de laboratorio e atención ao estudante nos casos necesarios de problemas de particular dificultade

Avaliación		
Metodoloxías	Descrición	Cualificación
Prácticas de laboratorio	Seguimento das prácticas e avaliación sobre o resultado alcanzado.	50
Proba mixta	Cuestións sobre os coñecementos adquiridos. Cuestións que impliquen razoamento en base a los coñecementos adquiridos para resolver problemas prácticos de interese real en recuperación de información e web semántica.	50

Observacións avaliación

Fontes de información



Bibliografía básica	<ul style="list-style-type: none">- Bob DuCharme (2011). Learning SPARQL. O'Reilly- C.D. Manning, P. Raghavan, H. Schutze. (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
Bibliografía complementaria	

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