



Guía Docente				
Datos Identificativos			2014/15	
Asignatura (*)	Recuperación da información e web semántica		Código	614502010
Titulación				
Descriptor				
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, Álvaro		Correo electr3nico	alvaro.barreiro@udc.es
Profesorado	Barreiro Garcia, Álvaro		Correo electr3nico	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				
Descrici3n xeral	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 anotaci3ns. Unha web baseada en documentos xml e tags, meta-datos e esquemas, sen dúbida facilitaríaa os enormes retos aos que se enfronta a recuperación de informaci3n web. Nesta materia ab3rdanse tamén os modelos, técnicas e algoritmos de maior impacto desenvolvidos nos últimos anos co obxectivo de materializar unha web semántica. A Recuperaci3n de Informaci3n en grandes coleccións de documentos e na web exp3n 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 extracci3n de informaci3n cando os meta-datos non son expostos publicamente e exp3n novos retos como os do matching de ontoloxías, resoluci3n de entidades ou unha dificultade maior en canto á heteroxeneidade e calidade dos datos e á indexaci3n e procura semántica. Por todo iso a Recuperaci3n de Informaci3n 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 sen3n tamén en moitas pequenas e medianas compañías.			

Competencias da titulaci3n	
C3digo	Competencias da titulaci3n

Resultados da aprendizaxe			
Competencias de materia (Resultados de aprendizaxe)		Competencias da titulaci3n	
Coecer, comprender e analizar os distintos modelos Recuperaci3n de Informaci3n e Web Semntica, as tcnicas para a sa implementaci3n eficiente e a metodoloxa de avaliaci3n dos mesmos.		AP5	CP6 CP8
Coecer, comprender e analizar as plataformas software para a creaci3n destes sistemas.		AP5	CP6 CP7 CP8



Deseñar e construír novos sistemas ou melloras en sistemas existentes.	AP5 AP12	BP1 BP5 BP10 BP13 BP14 BP17 BM1 BM2 BM5	CP6 CP7
Planear e realizar a avaliación dos sistemas de Recuperación de Información e Web Semántica . Analizar os resultados da avaliación dos sistemas para melloralos na súa eficacia e eficiencia.	AP5	BP1 BP5	CP6 CP7
Ser capaces dun correcto tratamento dos aspectos éticos, de privacidade, confidencialidade e de seguridade dos estes sistemas.		BM3	CP4 CP6

Contidos	
Temas	Subtemas
Introducción.	Recuperación de Información e o retos da Web
Recuperación de Información na Web.	Topoloxía da web: o grafo web. Arquitecturas de Search Engines. Ranking baseado en contido e análise de ligazóns. Learning to rank. Web spam. Xestión de datos web: identificadores de documentos, metadatos, duplicados. Interfaces para procura e navegación.
Web Crawling.	Tipos de crawlers. Arquitectura. Esixencias de fresness e politeness. Algoritmos de crawling. Avaliación.
Indexación.	Construcción e compresión de índices invertidos. Procesado de consultas.
Recuperación de Información na Web con paralelismo e distribución.	Particionamiento e selección de coleccións. Particionamiento de índices. Recuperación de información paralela con arquitecturas MIMD e SIMD. Recuperación de Información baseada en cluster. Recuperación de información distribuída e federada.
Sistemas de recomendación.	Filtrado colaborativo. Modelos e algoritmos para recomendación. Sistemas de recomendación
Introducción a Web Semántica	A Web Semántica. Ontoloxías: definición, tipos e exemplos.
Descrición e consulta de recursos	Linguaxes XML, RDF e RDF Schema. Linguaxe de consultas SPARQL. Linguaxe OWL. Ferramentas de desenvolvemento de ontoloxías. Librarías para o manexo de ontoloxías. Repositorios RDF.
Razonamiento e regras	Fundamentos de lóxica e razoamento. Representación de regras semánticas. Motores de razoamento.
Aplicacións da Web Semántica	Linked Data, FOAF, Dublin Core, WordNet. Anotación semántica. Buscadores semánticos. Servizos 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, Apache Solr, Nutch, Jena, Protege, Pellet)
Solución de problemas	Problemas e cuestións breves para asentar e profundizar os contidos expostos nas sesións maxistrais.
Proba mixta	Prueba que versará sobre os contidos fundamentais da materia.
Sesión maxistral	O estudante asistirá ás explicacións dadas polo profesor sobre os distintos modelos, técnicas e algoritmos de Recuperación de Información e Web Semántica. O profesor utilizará distintos niveis de abstracción-detalle e orientará ao estudante nas 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 aos coñecementos adquiridos para resolver problemas prácticos de interese real en recuperación de información e web semántica.	50

Observacións avaliación

Aqueles estudantes con matrícula a tempo parcial ou calquer circunstancia xustificada que impida a asistencia as clases, deberán contactar cos docentes para determinar alternativas ao seguimento e avaliación da materia.

Fontes de información

Bibliografía 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
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