UNIVERSIDADE DA CORUÑA

Guía Docente			
Datos Identificativos			**2016/17**
Asignatura (*)	Recuperación da información e web semántica	**Código**	614502010
Titulación	Mestrado Universitario en Enxeñaría Informática (plan 2012)		

Descriptores				
Ciclo	**Período**	**Curso**	**Tipo**	**Créditos**
Mestrado Oficial	1º cuadrimestre	Primeiro	Obrigatoria	6

Idioma	Castelán		
Modalidade docente	Presencial		
Prerrequisitos			
Departamento	ComputaciónTecnoloxías da Información e as Comunicacións		
Coordinación	Barreiro Garcia, Álvaro	**Correo electrónico**	alvaro.barreiro@udc.es
Profesorado	Barreiro Garcia, Álvaro Fernández Iglesias, Diego Parapar López, Javier Vázquez Naya, José Manuel	**Correo electrónico**	alvaro.barreiro@udc.es diego.fernandez@udc.es javier.parapar@udc.es jose.manuel.vazquez.naya@udc.es
Web			
Descrición 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 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 (volume 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 ontologías, 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 do título	
Código	**Competencias do título**
A5	Capacidade de comprender e saber aplicar o funcionamento e organización da internet, as tecnoloxías e protocolos de redes de nova xeración, os modelos de compoñentes, sóftware intermediario e servizos.
A12	Capacidade para aplicar métodos matemáticos, estatísticos e de intelixencia artificial para modelar, deseñar e desenvolver aplicacións, servizos, sistemas intelixentes e sistemas baseados no coñecemento.
B1	Capacidade de resolución de problemas.
B5	Habilidades de xestión da información.
B10	Capacidade para proxectar, calcular e deseñar produtos, procesos e instalacións en todos os ámbitos da enxeñaría informática
B13	Capacidade para o modelado matemático, cálculo e simulación en centros tecnolóxicos e de enxeñaría de empresa, particularmente en tarefas de investigación, desenvolvemento e innovación en todos os ámbitos relacionados coa Enxeñaría en Informática
B14	Capacidade para a elaboración, planificación estratéxica, dirección, coordinación e xestión técnica e económica de proxectos en todos os ámbitos da Enxeñaría en Informática seguindo criterios de calidade e ambientais
B17	Capacidade para a aplicación dos coñecementos adquiridos e de resolver problemas en contornas novas ou pouco coñecidos dentro de contextos máis amplos e multidisciplinares, sendo capaces de integrar estes coñecementos
B21	Posuír e comprender coñecementos que acheguen unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación

B22	Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en contornas novas ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B23	Que os estudantes sexan capaces de integrar coñecementos e enfrontarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos
B25	Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun modo que haberá de ser en gran medida autodirixido ou autónomo
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.
C7	Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.
C8	Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade

Resultados da aprendizaxe			
Resultados de aprendizaxe	Competencias do título		
Coñecer, comprender e analizar os distintos modelos Recuperación de Información e Web Semántica, as técnicas para a súa implementación eficiente e a metodoloxía de avaliación dos mesmos.	AP5		CP6 CP8
Coñecer, comprender e analizar as plataformas software para a creación 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. Esixenciias de fresness e politeness. Algoritmos de crawling. Avaliación.
Indexación.	Construcción e compresión de indices invertidos. Procesado de consultas.
Recuperación de Información na Web con paralelismo e distribución.	Particionamiento e selección de colecciones. Particionamiento de índices. Recuperación de información paralela con arquitecturas MIMD e SIMD. Recuperación de Información basada en cluster. Recuperación de información distribuida 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.
Descripción e consulta de recursos	Linguaxes XML, RDF e RDF Schema. Linguaxe de consultas SPARQL. Linguaxe OWL. Ferramentas de desenvolvemento de ontologías. Librarías para o manexo de ontologías. Repositorios RDF.
Razonamiento e regras	Fundamentos de lóxica e razoamiento. Representación de regras semánticas. Motores de razoaamiento.
Aplicacións da Web Semántica	Linked Data, FOAF, Dublin Core, WordNet. Anotación semántica. Buscadores semánticos. Servizos Web Semánticos.

Planificación				
Metodoloxías / probas	**Competencias**	**Horas presenciais**	**Horas non presenciais / traballo autónomo**	**Horas totais**
Lecturas	A5 A12 B1 B5 B10 B13 B14 C4 C6 C7 C8	1	15	16
Prácticas de laboratorio	B10 B17 B21 B22 B23 B25	20	30	50
Solución de problemas	A5 A12 B1 B5 B13 B14 B17 B21 B22 B23	4	12	16
Proba mixta	A5 A12 B1 B5 B10 B13 B14 C4 C6 C7 C8	2	18	20
Sesión maxistral	A5 A12 B1 B5 B10 B13 C4 C6 C7 C8	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

UNIVERSIDADE DA CORUÑA

Avaliación			
Metodoloxías	**Competencias**	**Descrición**	**Cualificación**
Prácticas de laboratorio	B10 B17 B21 B22 B23 B25	Seguimento das prácticas e avaliación sobre o resultado alcanzado.	50
Proba mixta	A5 A12 B1 B5 B10 B13 B14 C4 C6 C7 C8	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	- W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education - 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 - John Hebeler, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, Mike Dean. (2009). Semantic Web Programming. Wiley - Bob DuCharme (2011). Learning SPARQL. O'Reilly
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