



Guía Docente				
Datos Identificativos			2020/21	
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			
Modalidade docente	Híbrida			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da InformaciónComputación			
Coordinación	Barreiro Garcia, Álvaro		Correo electrónico	alvaro.barreiro@udc.es
Profesorado	Barreiro Garcia, Álvaro		Correo electrónico	alvaro.barreiro@udc.es
	Fernández Iglesias, Diego			diego.fernandez@udc.es
	Parapar López, Javier			javier.parapar@udc.es
	Vázquez Naya, José Manuel			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 capacitaránlles 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.			



Plan de continxencia	1. Modificacións nos contidos
	Sin cambios
	2. Metodoloxías
	*Metodoloxías docentes que se manteñen
	Todas
	*Metodoloxías docentes que se modifican
	Ningunha
	3. Mecanismos de atención personalizada ao alumnado
	Esta asignatura esta catalogada como híbrida porque as tutorías serán por mail e Teams dende inicio do curso. O resto das actividades e probas serán presenciais e cambiarían a facerse por Teams, moodle e mail no caso de continxencia.
	4. Modificacións na avaliación
	Ningunha
	Sin cambios
	*Observacións de avaliación:
	5. Modificacións da bibliografía ou webgrafía
	Sin cambios

Competencias / Resultados do título	
Código	Competencias / Resultados do título

Resultados da aprendizaxe			
Resultados de aprendizaxe		Competencias / Resultados 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 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. Librerías 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	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas 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

Avaliación

Metodoloxías	Competencias / Resultados	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

p { margin-bottom: 0.25cm; line-height: 120%; }

Para os alumnos a
tempo parcial o baremo de cualificación e a avaliación continua son
os mesmos que para os outros alumnos.

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 Hebel, 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