



| Guía docente          |   |         |                    |                                 |
|-----------------------|---|---------|--------------------|---------------------------------|
| Datos Identificativos |   |         |                    | 2023/24                         |
| Asignatura (*)        | Recuperación de la información y web semántica  |         | Código             | 614502010                       |
| Titulación            | Mestrado Universitario en Enxeñaría Informática (plan 2012)   |         |                    |                                 |
| Descritores           |   |         |                    |                                 |
| Ciclo                 | Periodo   | Curso   | Tipo               | Créditos                        |
| Máster Oficial        | 1º cuatrimestre   | Primero | Obligatoria        | 6                               |
| Idioma                | Castellano  |         |                    |                                 |
| Modalidad docente     | Presencial  |         |                    |                                 |
| Prerrequisitos        |   |         |                    |                                 |
| Departamento          | Ciencias da Computación e Tecnoloxías da InformaciónComputación   |         |                    |                                 |
| Coordinador/a         | Parapar López, Javier   |         | Correo electrónico | javier.parapar@udc.es           |
| Profesorado           | Fernández Iglesias, Diego   |         | Correo electrónico | diego.fernandez@udc.es          |
|                       | Parapar López, Javier   |         |                    | javier.parapar@udc.es           |
|                       | Pérez Vila, Miguel Anxo   |         |                    | anxo.pvila@udc.es               |
|                       | Vázquez Naya, José Manuel   |         |                    | jose.manuel.vazquez.naya@udc.es |
| Web                   |   |         |                    |                                 |
| Descripción general   | <p>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 enfrenta 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.</p> |         |                    |                                 |

| Competencias del título |  |
|-------------------------|--|
| Código                  | Competencias del título  |
| 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 multidisciplinares) 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   |                         |   |                   |
|---|-------------------------|---|-------------------|
| Resultados de aprendizaje   | Competencias del título |   |                   |
| 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<br>CP8        |
| Conocer, comprender y analizar las plataformas software para la creación de estos sistemas.   | AP5                     |   | CP6<br>CP7<br>CP8 |
| Diseñar y construir nuevos sistemas o mejoras en sistemas existentes.   | AP5<br>AP12             | BP1<br>BP5<br>BP10<br>BP13<br>BP14<br>BP17<br>BM1<br>BM2<br>BM5 | CP6<br>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<br>BP5  | CP6<br>CP7        |
| Ser capaces de un correcto tratamiento de los aspectos éticos, de privacidad, confidencialidad y de seguridad de los estos sistemas.  |                         | BM3   | CP4<br>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. Especificaciones de freshness 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.<br>Recuperación de información paralela con arquitecturas MIMD y SIMD.<br>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 lóxgca y razonamiento. Representación de reglas semánticas. Motores de razonamiento.  |
| Aplicaciones de la Web Semántica                                      | Linked Data, FOAF, Dublin Core, WordNet. Anotación semántica. Buscadores semánticos. Servizos Web Semánticos.  |

| Planificación            |  |                    |  |               |
|--------------------------|--|--------------------|--|---------------|
| Metodologías / pruebas   | Competencias                               | Horas presenciales | Horas no presenciales / trabajo autónomo | Horas totales |
| Lecturas                 | A5 A12 B1 B5 B10<br>B13 B14 C4 C6 C7<br>C8 | 1                  | 15                                       | 16            |
| Prácticas de laboratorio | B10 B17 B21 B22<br>B23 B25                 | 20                 | 30                                       | 50            |
| Solución de problemas    | A5 A12 B1 B5 B13<br>B14 B17 B21 B22<br>B23 | 4                  | 12                                       | 16            |
| Prueba mixta             | A5 A12 B1 B5 B10<br>B13 B14 C4 C6 C7<br>C8 | 2                  | 18                                       | 20            |
| Sesión magistral         | A5 A12 B1 B5 B10<br>B13 C4 C6 C7 C8        | 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 | Seguimento do desenvolvemento das prácticas nas horas reservadas de laboratorio e atención ao estudante nos casos necesarios de problemas de particular dificultade |
| Solución de problemas    | Avaliarase o traballo individual do alumnado.<br>Promoveranse os valores de igualdade seguindo as recomendacións actuais.   |

| Evaluación               |  |  |              |
|--------------------------|--|--|--------------|
| Metodoloxías             | Competencias                               | Descrición   | Calificación |
| Prácticas de laboratorio | B10 B17 B21 B22<br>B23 B25                 | Seguimento de las prácticas y evaluación sobre el resultado alcanzado.   | 50           |
| Prueba mixta             | A5 A12 B1 B5 B10<br>B13 B14 C4 C6 C7<br>C8 | 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. Es obligatorio alcanzar un 40% de la calificación para superar la materia. | 50           |

| Observacións avaliación   |
|---|
| <p>p { margin-bottom: 0.25cm; line-height: 120%; }</p> <p>Para los alumnos a tiempo parcial el baremo de calificaciones y la evaluación continua son los mismos que para los otros alumnos.</p> |

| Fuentes de información |   |
|------------------------|---|
| <b>Básica</b>          | <ul style="list-style-type: none"> <li>- W.B. Croft, D. Metzler, T. Strohman. (2009). Search Engines. Information Retrieval in Practice. Pearson Education</li> <li>- C.D. Manning, P. Raghavan, H. Schütze. (2008). Introduction to Information Retrieval. Cambridge University Press</li> <li>- R. Baeza-Yates and B. Ribeiro-Neto. (2011). Modern Information Retrieval (second edition) . Addison Wesley/Pearson Education</li> <li>- F. Cacheda, J.M. Fernández, J. Huete (eds.) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. Ra-Ma</li> <li>- John Hebel, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, Mike Dean. (2009). Semantic Web Programming. Wiley</li> <li>- Bob DuCharme (2011). Learning SPARQL. O'Reilly</li> </ul> |
| <b>Complementaria</b>  |   |

| Recomendacións  |
|---|
| <b>Asignaturas que se recomenda haber cursado previamente</b>   |
| <b>Asignaturas que se recomenda cursar simultaneamente</b>  |
| <p>Análisis de sistemas de información/614502006</p> <p style="text-align: center;"><b>Asignaturas que continúan el temario</b></p> |



|                   |
|-------------------|
|                   |
| 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