

Grado en Ingeniería Informática

DATOS DE LA ASIGNATURA

Nombre:

Motores de Búsqueda

Denominación en inglés:

Search engines

Código:

606010304

Carácter:

Optativo

Horas:

	Totales	Presenciales	No presenciales
Trabajo estimado:	150	60	90

Créditos:

Grupos reducidos				
Grupos grandes	Aula estándar	Laboratorio	Prácticas de campo	Aula de informática
3	0	0	0	3

Departamentos:

Tecnologías de la Información

Áreas de Conocimiento:

Lenguaje y Sistemas Informáticos

Curso:

4º - Cuarto

Cuatrimestre:

Primer cuatrimestre

DATOS DE LOS PROFESORES

Nombre:	E-Mail:	Teléfono:	Despacho:
De la Villa Cordero, Manuel	manuel.villa@dti.uhu.es	959217639	Edif. Torreumbria, Dcho. 41
*Maña López, Manuel J.	manuel.mana@dti.uhu.es	959217389	Edif. Torreumbria, Dcho. 47
Domínguez Olmedo, Juan Luis	juan.dominguez@dti.uhu.es	959217371	TUPB-61

*Profesor coordinador de la asignatura

Consultar los horarios de la asignatura

1. Descripción de contenidos**1.1. Breve descripción (en castellano):**

Modelos fundamentales para la representación de información no estructurada. Evaluación de la efectividad de sistemas de recuperación de información. Búsqueda y posicionamiento en la Web. Recuperación de información textual, multimedia y geográfica.

1.2. Breve descripción (en inglés):

Fundamental models in non-structured data retrieval. Effectiveness evaluation of information retrieval systems. Searching the Web. Search engine optimization. Full-text, multimedia and geographical information retrieval.

2. Situación de la asignatura**2.1. Contexto dentro de la titulación:**

Los sistemas inteligentes son productos con un gran valor añadido que están generando gran cantidad de proyectos y puestos de trabajo. El interés creciente en la investigación y desarrollo de herramientas que ayuden a los usuarios a organizar, buscar y comprender la información de tipo textual, multimedia y geográfica viene avalado por un mercado de herramientas software que está en continua expansión. Por otra parte, según Oracle, el 90% de la información que manejan las organizaciones no es estructurada; se trata fundamentalmente de texto, objetos multimedia (vídeo, fotografías, música y otros tipos de audio) o información geográfica.

Sin embargo, tradicionalmente, en los planes de estudio de Informática se incide fundamentalmente en la parte de información estructurada (el 10% restante) en asignaturas de estructuras de datos, bases de datos e ingeniería del software. Por tanto, la incorporación de esta asignatura al plan de estudios del Grado en Ingeniería Informática de la Universidad de Huelva viene a cubrir ese hueco significativo que supone el 90% de la información de las organizaciones y permitirá a los alumnos integrarse, con ciertas garantías, en este emergente sector.

2.2. Recomendaciones:

Se exponen, a continuación, algunas de las competencias que deberían poseer los alumnos antes de comenzar la asignatura:

- Tener destreza en el lenguaje de programación Java.
- Poseer destreza para buscar información útil en la Red.
- Poseer conocimientos suficientes para leer documentos escritos en inglés.
- Saber manejar fuentes bibliográficas.
- Tener capacidad de lectura comprensiva.
- Saber expresarse correctamente de forma oral y escrita.

3. Objetivos (Expresados como resultados del aprendizaje):

El objetivo general es proporcionar al alumno los conocimientos fundamentales, teóricos y prácticos, de los sistemas de gestión y recuperación de información no estructurada (textual, multimedia y geográfica).

Este objetivo general que se desea alcanzar con el estudio de la materia propuesta para esta asignatura se puede refinar en los siguientes objetivos específicos:

- Comprender la dificultad de representar y recuperar información no estructurada.
- Comprender los modelos fundamentales para la representación de la información no estructurada.
- Aprender las técnicas más importantes de evaluación de la efectividad.
- Comprender la complejidad del problema de la búsqueda en la Web, así como conocer en detalle las herramientas que existen en la actualidad para abordar dicho problema.
- Conocer las técnicas para conseguir un buen posicionamiento en la Web.
- Ser capaz de implementar un sistema básico de recuperación de información.

4. Competencias a adquirir por los estudiantes**4.1. Competencias específicas:****4.2. Competencias básicas, generales o transversales:**

- **CB5:** Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía
- **G01:** Capacidad de organización y planificación así como capacidad de gestión de la Información.
- **G04:** Capacidad para tomar decisiones basadas en criterios objetivos (datos experimentales, científicos o de simulación disponibles) así como capacidad de argumentar y justificar lógicamente dichas decisiones, sabiendo aceptar otros puntos de vista
- **G06:** Capacidad para el aprendizaje autónomo así como iniciativa y espíritu emprendedor
- **G08:** Capacidad para adaptarse a las tecnologías y a los futuros entornos actualizando las competencias profesionales.
- **T01:** Uso y dominio de una segunda lengua
- **T02:** Conocimiento y perfeccionamiento en el ámbito de las TIC's

5. Actividades Formativas y Metodologías Docentes

5.1. Actividades formativas:

- Sesiones de Teoría sobre los contenidos del Programa.
- Sesiones de Resolución de Problemas.
- Sesiones Prácticas en Laboratorios Especializados o en Aulas de Informática.
- Actividades Académicamente Dirigidas por el Profesorado: seminarios, conferencias, desarrollo de trabajos, debates, tutorías colectivas, actividades de evaluación y autoevaluación.

5.2. Metodologías docentes:

- Clase Magistral Participativa.
- Desarrollo de Prácticas en Laboratorios Especializados o Aulas de Informática en grupos reducidos.
- Resolución de Problemas y Ejercicios Prácticos.
- Tutorías Individuales o Colectivas. Interacción directa profesorado-estudiantes.
- Planteamiento, Realización, Tutorización y Presentación de Trabajos.
- Evaluaciones y Exámenes.

5.3. Desarrollo y justificación:

En las **clases teóricas** se presentarán los conceptos de manera clara y concisa utilizando para ello las herramientas docentes más adecuadas al alcance del profesor. Para cada tema se proporcionará al alumno material de lectura que servirá de complemento a cada sesión teórica. En ocasiones, el alumno deberá trabajar ciertos contenidos de forma personal, con ayuda del material proporcionado, estimulando, de esta forma, el aprendizaje autónomo.

Las **prácticas de laboratorio** se enfocarán, fundamentalmente, al desarrollo de ejercicios prácticos de pequeña/mediana envergadura que permitirán al alumno afianzar los conocimientos adquiridos en teoría. Además se presentarán al alumno APIs muy utilizados para el análisis de texto y la construcción de motores de búsqueda que permitirán al alumno construir sistemas de recuperación con poco esfuerzo.

Las **actividades académicas dirigidas** tienen como objetivo profundizar en aspectos más avanzados y concretos en recuperación de información. Se llevarán a cabo varios tipos de actividades:

- *Resolución de problemas.* Se solicitará al alumno que resuelva y entregue determinados problemas.
- *Trabajos en grupo.* Se proporcionará una lista de temas objeto de trabajo que estarán relacionados con aspectos actuales en recuperación de información no estructurada y que no se abordan con profundidad durante el curso. El trabajo se ajustará a las directrices que se marquen para su desarrollo. Los alumnos, en equipos, elegirán uno de los temas, lo llevarán a cabo y lo expondrán en el aula al resto de los compañeros.

6. Temario desarrollado:

1. Recuperación de información. Tipos especiales de datos

Motivación. Características del texto y de la información multimedia. La investigación en acceso a la información textual en la UHU.

2. Documentos. Lenguajes y formatos

Documentos. Metadatos. Formatos de documentos. Lenguajes de marcado.

3. Análisis del texto

Introducción. Análisis léxico. Eliminación de palabras vacías. Reducción a la raíz. Análisis morfológico. Análisis sintáctico. Reconocimiento de entidades nombradas. Índice invertido: construcción y búsqueda. Tesoros y bases de datos léxicas.

4. Modelos para la recuperación de información textual

Descripción de la tarea. Características generales de los modelos clásicos. Modelo booleano. Modelo del espacio vectorial.

5. Evaluación de los sistemas de recuperación de información

Tipos de evaluación. Evaluación de la efectividad. Precisión y cobertura. Utilización de un único valor. Otras medidas de evaluación. Recursos para la evaluación

6. Búsqueda en la Web

Introducción. Características de la Web. Arquitectura de un buscador. Robot. Indexador. Motor de búsqueda. Interfaz de usuario. Otros mecanismos de búsqueda.

7. Posicionamiento en la Web

Introducción. Posicionamiento publicitario y posicionamiento orgánico. Estadísticas sobre buscadores. Técnicas para conseguir un buen posicionamiento: contenido y enlaces.

8. Recuperación de información multimedia

Conceptos fundamentales. Recuperación de imagen basada en contenido. Recuperación de audio y música. Recuperación de video.

9. Recuperación de información geográfica

Conceptos fundamentales. Aplicaciones de la recuperación de información geográfica. Arquitectura de un sistema de recuperación de información geográfica. Recursos de información geográfica. Procesamiento de lenguaje natural aplicado a la información geográfica.

7. Bibliografía

7.1. Bibliografía básica:

R. Baeza-Yates y B. Ribeiro-Neto. Modern Information Retrieval: The Concepts and Technology behind Search. Second Edition. Pearson, 2011.

Christopher D. Manning, Prabhakar Raghavan y Hinrich Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.

7.2. Bibliografía complementaria:

David A. Grossman y Ophir Frieder. Information Retrieval: Algorithms and Heuristics. Springer, 2004.

P. Jackson y I. Moulinier. Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization. John Benjamins Publishing Company, 2002.

I.H. Witten, A. Moffat y T.C. Bell. Managing Gigabytes: Compressing and Indexing Documents and Images. Second Edition. Morgan Kaufmann Publishers, 1999.

8. Sistemas y criterios de evaluación.

8.1. Sistemas de evaluación:

- Examen de teoría/problemas
- Defensa de Prácticas
- Defensa de Trabajos e Informes Escritos

8.2. Criterios de evaluación y calificación:

La asignatura se califica utilizando diferentes instrumentos, seleccionados según la competencia que se desea evaluar:

I1. Examen objetivo. 40% de la nota. Mediante el examen teórico se evalúa la adquisición de la competencia CB5.

I2. Trabajos de laboratorio. 40% de la nota. Mediante la presentación y defensa de las prácticas de laboratorio se evalúa la adquisición de las competencias CB5, G01, G04, G06, G08 y T02.

I3. Trabajos/problemas individuales y en grupo. 20% de la nota. Mediante la presentación y defensa de las prácticas de laboratorio se evalúa la adquisición de las competencias CB5, G01, G06, T01 y T02.

La asistencia a clase es obligatoria. Si algún alumno no pudiera hacerlo por motivos justificados debe comunicárselo a los profesores de la asignatura lo antes posible. En ese caso, se diseñará un plan de evaluación y calificación alternativo y adaptado a las necesidades del alumno.

Para aprobar la asignatura es necesario aprobar por separado I1 e I2 y sumar 5 puntos entre las 3 calificaciones.

9. Organización docente semanal orientativa:

	Semanas	Grupos Grandes	Grupos Reducidos Aula Estándar	Grupos Reducidos Aula de Informática	Grupos Reducidos Laboratorio	Grupos Reducidos prácticas de campo	Pruebas y/o actividades evaluables	Contenido desarrollado
#1	2	0	2	0	0		Presentación de la asignatura y Tema 1	
#2	2	0	2	0	0		Temas 1 y 2.	
#3	2	0	2	0	0		Temas 2 y 3.	
#4	2	0	2	0	0		Tema 3.	
#5	2	0	2	0	0		Temas 3 y 4.	
#6	2	0	2	0	0		Tema 4.	
#7	2	0	2	0	0		Tema 5.	
#8	2	0	2	0	0		Temas 5 y 6.	
#9	2	0	2	0	0		Tema 6.	
#10	2	0	2	0	0	Entrega trabajo en grupo.	Temas 6 y 7.	
#11	2	0	2	0	0		Tema 7 y Presentación de trabajos.	
#12	2	0	2	0	0		Presentación de trabajos.	
#13	2	0	2	0	0		Tema 8.	
#14	2	0	2	0	0		Tema 8.	
#15	2	0	2	0	0		Tema 9.	
	30	0	30	0	0			