



# ESCUELA POLITECNICA SUPERIOR

## Guía Docente

Curso 2010-2011

### Máster

## Tecnologías Informáticas Avanzadas

### DATOS DE LA ASIGNATURA

<b>Nombre:</b>		
Recuperación de la Información		
<b>Denominación en inglés<sup>1</sup>:</b>		
Information Retrieval		
<b>Tipo:</b>	<b>Cuatrimestre:</b>	<b>Idioma:</b>
<input checked="" type="checkbox"/> Obligatoria <input type="checkbox"/> Optativa	1º Cuatrimestre	Español
<b>Créditos E.C.T.S.:</b>		
Totales:	Teóricos:	Prácticos:
4,0	2,3	1,7
<b>Web de la asignatura:</b>		
<a href="http://www.uhu.es/sevirtual/">http://www.uhu.es/sevirtual/</a>		

<sup>1</sup>Para su inclusión en el Complemento Europeo al Título

### DATOS DE LOS PROFESORES

Nombre:	e-mail:	Teléfono:	Despacho:
Manuel Maña López	manuel.mana@dti.uhu.es	959 217389	50
Jacinto Mata Vázquez	mata@uhu.es	959 217652	26

## DATOS ESPECÍFICOS DE LA ASIGNATURA

### 1.1. Descriptores de la asignatura:

Modelos fundamentales para la representación de información textual. Evaluación de la efectividad de sistemas de recuperación de información. Búsqueda en la Web.

### 1.2. Descriptores de la asignatura (en inglés)<sup>2</sup>:

Fundamental models in information retrieval. Effectiveness evaluation of information retrieval systems. Searching the Web.

<sup>2</sup>Para su inclusión en el Complemento Europeo al Título

## 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 textual viene avalado por un mercado de herramientas software que en 2001 alcanzó los 423,3 millones de dólares y se esperaba que en 2006 alcanzase la cifra de los 1.300 millones de dólares. La incorporación de asignaturas de estas áreas al Máster, como *Recuperación de la Información*, *Minería de Textos* o *Integración de Información y Aplicaciones*, permitirá a los alumnos integrarse, con ciertas garantías, en este emergente sector, más allá de la formación básica que recibirán en el Grado.

En esta asignatura se tratan las técnicas fundamentales en recuperación de información, que servirán como base para las asignaturas de 2º cuatrimestre con las que está relacionada (*Minería de Textos* e *Integración de Información y Aplicaciones*).

### 2.2. Recomendaciones:

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

- Tener destreza en lenguajes de programación orientados a objetos de propósito general (Java, C++, ...)
- Poseer destreza para buscar información útil en la Red
- Poseer conocimientos básicos de inglés
- Saber manejar fuentes bibliográficas
- Tener capacidad de lectura comprensiva
- Saber expresarse correctamente de forma oral y escrita

3. Competencias a adquirir por los estudiantes.			
3.1. Competencias transversales o genéricas.			
<input type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Analizar y resolver problemas técnicos complejos
<input checked="" type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Dominar las prácticas y los estándares de la investigación en Tecnologías de la Información.
<input checked="" type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Comunicarse, oralmente y por escrito, de manera eficiente y eficaz con diversas audiencias.
<input checked="" type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Trabajar en equipos.
<input checked="" type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Perseguir la calidad en el trabajo desarrollado.
<input type="checkbox"/> Alto	<input checked="" type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Aprender de forma autónoma.
<input type="checkbox"/> Alto	<input type="checkbox"/> Medio	<input type="checkbox"/> Bajo	Tener iniciativa, espíritu emprendedor y capacidad de generar nuevas ideas.
3.2. Competencias específicas.			
3.2.1. Competencias cognitivas (saber):			
<ul style="list-style-type: none"> <li>▪ Comprender la dificultad de representar y recuperar documentos de texto</li> <li>▪ Comprender los modelos fundamentales para la representación de la información textual</li> <li>▪ Conocer las técnicas más importantes de evaluación de la efectividad</li> <li>▪ 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</li> <li>▪ Conocer los estándares de investigación en el área de recuperación de información</li> </ul>			
3.2.2. Competencias procedimentales e instrumentales (saber hacer):			
<ul style="list-style-type: none"> <li>▪ Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información</li> <li>▪ Ser capaz de entender un texto científico</li> </ul>			
3.2.3. Competencias actitudinales (ser):			
<ul style="list-style-type: none"> <li>▪ Saber comunicarse de forma efectiva en español, tanto de forma oral como escrita</li> <li>▪ Saber colaborar con otros compañeros para resolver problemas complejos</li> <li>▪ Saber proponer soluciones alternativas a una dada</li> <li>▪ Preocuparse por la calidad del diseño de un sistema de recuperación de información</li> </ul>			

#### 4. Objetivos:

El objetivo general es proporcionar al alumno los conocimientos fundamentales, teóricos y prácticos, de los sistemas de recuperación de información.

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 documentos de texto
- Comprender los modelos fundamentales para la representación de la información textual
- 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
- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información

#### 5. Metodología (en horas de trabajo del estudiante):

	Primer Cuatrimestre	Segundo Cuatrimestre
	Presenciales	
Clases de teoría	20,0	0,0
Clases de problemas	0,0	0,0
Clases prácticas	11,0	0,0
Actividades académicas dirigidas	6,0	0,0
Exámenes	0,0	0,0
	No presenciales	
Estudio de clases teóricas (factor de trabajo: 1,00)	20,0	0,0
Estudio de clases de problemas y prácticas (factor de trabajo: 3,00)	33,0	0,0
Preparación de actividades académicamente dirigidas y otras actividades	10,0	0,0
<b>Total:</b>	<b>100,0</b>	<b>0,0</b>
<b>Trabajo total del estudiante: 100,0 horas.</b>		
<b>Horas presenciales: 37,0</b>	<b>Horas no presenciales: 63,0</b>	

#### 6. Técnicas docentes.

##### 6.1. Técnicas docentes utilizadas:

- Sesiones académicas de teoría
- Sesiones académicas de problemas
- Sesiones prácticas en laboratorio
- Seminarios técnicos
- Actividades tutorizadas: trabajos en grupos reducidos
- Tutorías individuales y en grupo
- Apoyo virtual

## 6.2. 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 clases teóricas se desarrollarán durante 20 h. en sesiones de 1 ó 2 horas. Para el estudio de las sesiones teóricas, las lecturas previas a cada tema y la preparación de los contenidos adicionales, se estima que el alumno necesitará, de forma aproximada, 20 h. de estudio.

Las **prácticas de laboratorio** se enfocarán, fundamentalmente, al diseño e implementación de un *proyecto de curso* que deberá realizarse en grupos, fomentando así el aprendizaje cooperativo. Algunas de las sesiones se dedicarán también a la realización de ejercicios de mucha menor entidad que el proyecto. Se inculcará, además, en los alumnos, el sentido ético que debe primar en los estudios universitarios, previniendo así comportamientos fraudulentos como la copia de prácticas.

Por este concepto, se impartirán un total de 11 h. en aulas de laboratorio. Se estima que el alumno necesitará unas 33 h. de dedicación para el estudio de los contenidos prácticos.

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, como por ejemplo:

- **Comentarios de artículos científicos.** Se solicitará al alumno que lea varios artículos científicos relacionados con la materia y se realizarán debates en el aula sobre los mismos.
- **Evaluación de motores de búsqueda en la Web.** La actividad consiste en evaluar un conjunto de buscadores utilizando las métricas de evaluación que se explican en clases de teoría. Esta actividad se realiza en grupo.

Para estas actividades se dedicarán 6 horas en sesiones de 1 h, cada dos semanas aproximadamente. El esfuerzo del alumno se estima en 10 horas de trabajo individual.

## 7. Bloques temáticos:

A continuación se presentan los bloques temáticos junto con las competencias específicas que se desarrollarán en cada uno de ellos.

### **Bloque 1 Motivación y conceptos básicos**

- Comprender la dificultad de representar y recuperar documentos de texto
- Conocer los estándares de investigación en el área de recuperación de información

### **Bloque 2 Análisis del texto**

- Comprender los modelos fundamentales para la representación de la información textual
- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información

### **Bloque 3 Modelos de representación del texto**

- Comprender los modelos fundamentales para la representación de la información textual
- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información

### **Bloque 4 Evaluación de sistemas de recuperación de información**

- Conocer las técnicas más importantes de evaluación de la efectividad
- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información

### **Bloque 5 Operaciones sobre la consulta**

- Comprender los modelos fundamentales para la representación de la información textual
- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información

### **Bloque 6 Búsquedas en la Web**

- 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

### **Bloque 7 Temas avanzados en RI**

- Saber comunicarse de forma efectiva en español, tanto de forma oral como escrita
- Saber colaborar con otros compañeros para resolver problemas complejos
- Ser capaz de entender un texto científico
- Conocer los estándares de investigación en el área de recuperación de información

### **Bloque 8 Prácticas**

- Ser capaz de implementar, ejecutar y evaluar un sistema básico de recuperación de información
- Saber colaborar con otros compañeros para resolver problemas complejos
- Preocuparse por la calidad del diseño de un sistema de recuperación de información
- Saber proponer soluciones alternativas a una dada

## 8. Temario desarrollado:

### **Tema 1 Motivación y conceptos básicos (2)**

- 1.1 La Era de la Información
- 1.2 Recuperación de información y acceso a la información
- 1.3 Recuperación de información vs. recuperación de datos
- 1.4 El proceso de recuperación
- 1.5 Un paso más allá: minería de texto
- 1.6 Procesamiento de lenguaje natural
- 1.7 La investigación en acceso a la información en la UHU
- 1.8 Recursos, conferencias y publicaciones relevantes

### **Tema 2 Análisis del texto (3)**

- 2.1 Introducción
- 2.2 Análisis léxico
- 2.3 Eliminación de palabras vacías
- 2.4 Reducción a la raíz
- 2.5 Selección de términos índice
- 2.6 Índice invertido: construcción y búsqueda
- 2.7 Tesauros y bases de datos léxicas
- 2.8 Otras técnicas de análisis de texto

### **Tema 3 Modelos de representación del texto (4)**

- 3.1 Descripción de la tarea
- 3.2 Modelos clásicos
  - 3.2.1 Booleano
  - 3.2.2 Espacio vectorial
  - 3.2.3 Probabilístico
- 3.3 Modelos alternativos

### **Tema 4 Evaluación de sistemas de recuperación de información (3)**

- 4.1 Tipos de evaluación
- 4.2 Evaluación de la efectividad
  - 4.2.1 Precisión y cobertura
  - 4.2.2 Utilización de un único valor
  - 4.2.3 Otras medidas
- 4.3 Recursos para la evaluación

### **Tema 5 Operaciones sobre la consulta (2)**

- 5.1 Introducción
- 5.2 Realimentación de relevancia
- 5.3 Análisis local
- 5.4 Análisis global

### **Tema 6 Búsquedas en la Web (4)**

- 6.1 Introducción
- 6.2 Características de la Web
- 6.3 Buscadores
  - 6.3.1 Robot
  - 6.3.2 Indexador
  - 6.3.3 Motor de búsqueda
  - 6.3.4 Interfaz de usuario
  - 6.3.5 Estadísticas
- 6.4 Otros mecanismos de búsqueda

## 9. Bibliografía.

### 9.1. Bibliografía general:

Modern Information Retrieval  
Autores: R. Baeza-Yates y B. Ribeiro-Neto  
Editorial: ACM Press - Addison Wesley  
Año: 1999

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

### 9.2. Bibliografía específica:

Information Retrieval: Algorithms and Heuristics  
Autores: David A. Grossman y Ophir Frieder  
Editorial: Springer  
Año: 2004

Information Retrieval: Data Structures & Algorithms  
Autores: W.B. Frakes y R. Baeza-Yates (eds.)  
Editorial: Prentice-Hall International  
Año: 1992

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

Advances in Information Retrieval: Recent Research from de Center for Intelligent Information Retrieval  
Autor: W. B. Croft  
Editorial: Morgan Kaufmann Publishers  
Año: 1997

Readings in Information Retrieval  
Autores: K. Sparck Jones y P. Willett (eds.)  
Editorial: Kluwer Academic Publishers  
Año: 2000

Managing Gigabytes: Compressing and Indexing Documents and Images, 2ª edición  
Autores: I.H. Witten, A. Moffat y T.C. Bell  
Editorial: Morgan Kaufmann Publishers  
Año: 1999

## 10. Técnicas de evaluación.

### 10.1. Técnicas de evaluación utilizadas:

- Examen teórico-práctico
- Trabajos desarrollados durante el curso
- Participación activa en las sesiones académicas
- Controles periódicos de adquisición de conocimientos
- Examen práctico en aula de informática
- Otras: Proyecto de curso
- Otras: Especificar

**10.2. Criterios de evaluación y calificación:**

<b>Aspecto</b>	<b>Criterio</b>	<b>Instrumento</b>	<b>Peso</b>
Conocimientos teórico-prácticos	Nivel de consecución de los objetivos teórico-prácticos	Entrega y defensa del proyecto en grupo	70%
Actividades individuales	Participación activa en el aula Calidad de la documentación presentada	Observaciones del profesor Resúmenes de las lecturas	10%
Actividades en grupo	Capacidad para trabajar en grupo Nivel de comprensión de las métricas de evaluación	Observaciones del profesor Resultados de la evaluación de los motores de búsqueda	20%

Para aprobar la asignatura, el alumno deberá sumar 5 puntos entre todos los aspectos evaluables. Además, deberá realizar y defender el trabajo en grupo.

**11. Mecanismos de control y seguimiento:**

Los establecidos por la propia Universidad y los definidos en el Máster.