

# SISTEMAS DE INDIZACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN DIGITAL

## 1.- Datos de la Asignatura

Código	304134	Plan	2013	ECTS	6
Carácter	obligatoria	Curso	20/21	Periodicidad	Semestral
Área	Lenguajes y Sistemas Informáticos				
Departamento	Informática y Automática				
Plataforma Virtual	Plataforma:	Studium			
	URL de Acceso:	<a href="http://studium.usal.es">http://studium.usal.es</a>			

## Datos del profesorado

Profesor	Carlos G. Figuerola	Grupo / s	
Departamento	Informática y Automática		
Área	Lenguajes y Sistemas Informáticos		
Centro	Facultad de Traducción y Documentación		
Despacho	2		
Horario de tutorías	Pendiente de conocer el horario de las asignaturas		
URL Web			
E-mail	figue@usal.es	Teléfono	

Repetir análogamente para otros profesores implicados en la docencia

## 2.- Sentido de la materia en el plan de estudios

<b>Bloque formativo al que pertenece la materia</b>
MÓDULO 3. Procesamiento y representación de contenidos digitales
<b>Papel de la asignatura dentro del Bloque formativo y del Plan de Estudios.</b>
Las funciones y procesos vinculados al tratamiento técnico, la representación y la recuperación de información digital centran el desarrollo formativo del tercer módulo. La parte correspondiente a la recuperación digital es la correspondiente a esta asignatura.
<b>Perfil profesional.</b>
<ul style="list-style-type: none"> <li>Recuperación clásica: Modelos de recuperación de información, categorización y clasificación automática, indización en recuperación de información, expansión de consultas, normalización terminológica (lematización), extracción de entidades para la recuperación de información, algoritmos de búsqueda, ordenación y árboles, resumen automático de documentos, recuperación de pasajes de texto, recuperación en grandes volúmenes de información (Terabytes), escalabilidad de</li> </ul>

- las colecciones de evaluación.
- Recuperación semántica
- Web Semántico
- Organización automática de la información
- Extracción automática de información

### 3.- Recomendaciones previas

### 4.- Objetivos de la asignatura

- Conocer el alcance y los objetivos que persiguen los sistemas de recuperación de información.
- Conocer las herramientas y métodos básicos utilizados en el desarrollo de sistemas de recuperación de información.
- Conocer la importancia de la recuperación de información en el desarrollo de sistemas y servicios de información.
- Comprender el ámbito de la recuperación de información dentro de los perfiles científicos y profesionales.
- Conocer los mecanismos para representar de la manera más eficiente la información contenida en los documentos electrónicos, con el objetivo de optimizar la recuperación de información.
- Conocer la importancia en la evolución de los sistemas de recuperación de información hacia modelos avanzados.
- Conocer los algoritmos y estructuras básicas para recuperar información.
- Conocer los criterios de evaluación de la recuperación de información.
- Conocer y aplicar las técnicas de recuperación de información para encontrar información en el enorme fondo documental que es Internet.

### 5.- Contenidos

1. Documentos digitales.  
Características básicas. Implicaciones
2. Técnicas básicas de recuperación y búsqueda
3. Modelos de representación (vectores, pesos, similitudes)
4. Categorización Automática de Documentos, machine learning
5. Etiquetado automático y web semántica. Detección de entidades
6. Organización automática: clustering

### 6.- Competencias a adquirir

#### Específicas.

Tecnologías de la información: informática, tecnologías de la información: telecomunicaciones, técnicas de producción y edición, elaboración y difusión de la información.

#### Transversales.

Capacidad de análisis y síntesis, capacidad de organización y planificación, conocimientos de informática en el ámbito de estudio, capacidad de gestión de la información, resolución de problemas, toma de decisiones, razonamiento crítico, creatividad, iniciativa y espíritu emprendedor.

### 7.- Metodologías docentes

#### Actividades presenciales:

- Clases teóricas: sesiones expositivas del profesor
- Clases prácticas: ejercicios, supuestos prácticos, etc.
- Tutorías por grupos: seguimiento de las prácticas
- Tutorías individuales:
- Defensa del trabajo: valoración individual de lo aprendido y realizado por el alumno

#### Actividades no presenciales

- Preparación previa de las clases teóricas
- Preparación de las clases prácticas:
- Elaboración del dossier de prácticas:
- Preparación de la documentación del trabajo y defensa:

## 8.- Previsión de distribución de las metodologías docentes

	Horas dirigidas por el profesor		Horas de trabajo autónomo	HORAS TOTALES	
	Horas presenciales.	Horas no presenciales.			
Sesiones magistrales	20	20	20	60	
Prácticas	- En aula				
	- En el laboratorio				
	- En aula de informática	20	20	20	60
	- De campo				
	- De visualización (visu)				
Seminarios					
Exposiciones y debates	10	10	10	30	
Tutorías					
Actividades de seguimiento online					
Preparación de trabajos					
Otras actividades (detallar)					
Exámenes					
<b>TOTAL</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>150</b>	

## 9.- Recursos

### Libros de consulta para el alumno

Baeza-Yates, R.; Ribeiro-Neto, B. (eds.). Modern Information Retrieval. New York: ACM Press, 1999

Cacheda Seijo, F; Fernández Luna, J.M. Huetae Guadix, J.F. eds. (2011): Recuperación de Información. Un enfoque práctico y multidisciplinar.

Ra-Ma, Madrid, 2011

Ceri, S; Bozzon,A.; Brambilla, M. et al.(2013). Web Information Retrieval, Springer, Heildeberg 2013

Chakrabarti, S.(2002). Mining the Web. Discovering Knowledge from Hypertext Data.Morgan-Kaufmann Publishers

Manning, C.D.; Raghavan, P. and Schütze, H. (2008). Introduction to Information Retrieval, Cambridge University Press, 2008. <https://nlp.stanford.edu/IR-book/>

Otras referencias bibliográficas, electrónicas o cualquier otro tipo de recurso.

Se facilitarán oportunamente durante el curso

## 10.- Evaluación

Las pruebas de evaluación que se diseñen deben evaluar si se han adquirido las competencias descritas, por ello, es recomendable que al describir las pruebas se indiquen las competencias y resultados de aprendizaje que se evalúan.

### Consideraciones Generales

La asignatura se plantea con el sistema de evaluación continua. A lo largo del curso se realizarán una serie de actividades de carácter obligatorio que deberán ser superadas.

#### Criterios de evaluación

- comprensión de los principios que rigen el funcionamiento de los Sistemas de Recuperación Automatizados
- comprensión de las fases y operaciones que conforman el proceso de la Recuperación Automatizada
- demostrar que se es capaz de configurar adecuadamente un Sistema de Recuperación
- demostrar que se es capaz de gestionar adecuadamente un Sistema de Recuperación Automatizado
- demostrar que se es capaz de detectar puntos fuertes y puntos débiles en diferentes Sistemas de Recuperación
- demostrar que se es capaz de documentarse y resolver problemas técnicos sencillos en los Sistemas de Recuperación

#### Instrumentos de evaluación

Participación activa del alumno en las actividades docentes  
Realización satisfactoria de los trabajos propuestos  
Entrega en los plazos previstos de los informes y otros trabajos encomendados  
Superación de una prueba escrita (examen)

#### Recomendaciones para la evaluación.

Con carácter general, se recomienda:

- asistir activamente a las sesiones presenciales de la asignatura.
- seguir las instrucciones para la elaboración y presentación de las tareas.
- cumplir los plazos marcados para la entrega de tareas.

#### Recomendaciones para la recuperación.

A la vista del trabajo desarrollado por cada alumno, se indicará de forma personalizada las actividades a realizar para la recuperación.