

MÓDULO III: Procesamiento y representación de contenidos digitales

MATERIA:

Sistemas de indización y recuperación de la información digital

1.- Datos de la Asignatura

Código	306085	Plan	2022	ECTS	6
Carácter	Obligatoria	Curso	1	Periodicidad	Semestral S1
Área	Lenguajes y Sistemas Informáticos				
Departamento	Informática y Automática				
Plataforma Virtual	Plataforma:	STUDIUM / DIAWEB			
	URL de Acceso:	http://studium.usal.es / http://mastersid.usal.es			

Datos del profesorado

Profesor Coordinador	L. Carlos García Figuerola	Grupo / s	
Departamento	Informática y Automática		
Área	Lenguajes y Sistemas Informáticos		
Centro	Facultad de Traducción y Documentación		
Despacho	2		
Horario de tutorías			
URL Web	https://studium.usal.es		
E-mail	fique@usal.es	Teléfono	ext 3099

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

Bloque formativo al que pertenece la materia

Módulo 3: Procesamiento y representación de contenidos digitales

Papel de la asignatura dentro del Bloque formativo y del Plan de Estudios.

Las funciones y procesos vinculados al tratamiento técnico, la representación, la recuperación y la organización automática de información digital centran el desarrollo formativo del tercer módulo. La parte correspondiente a la organización automática digital es la correspondiente a esta asignatura.

Perfil profesional.

- Recuperación clásica: Modelos de recuperación de información, categorización y clasificación automática, extracción de entidades, resumen automático de documentos, procesamiento de lenguaje natural, técnicas de aprendizaje automático.
- Recuperación semántica
- Web Semántico
- Organización automática de la información
- Extracción automática de información

3.- Recomendaciones previas

4.- Objetivos de la asignatura

Objetivos

- Conocer el alcance y los objetivos que persiguen los sistemas de organización automática de información.
- Conocer las herramientas y métodos básicos utilizados en el desarrollo de sistemas de organización automática de la información.
- Conocer la importancia de la organización automática de la información en el desarrollo de sistemas y servicios de información.
- Comprender el ámbito de la organización automática de la información dentro de los perfiles científicos y profesionales, así como los principios teóricos y prácticos de la Ciencia de Datos que la sustentan.
- Conocer los mecanismos para representar de la manera más eficiente la información contenida en los documentos electrónicos, con el objetivo de optimizar el procesamiento automático de información.
- Conocer la importancia en la evolución de los sistemas de tratamiento automatizados de la información hacia modelos avanzados.
- Conocer los algoritmos y estructuras básicas para representar y procesar la información.
- Conocer los criterios de evaluación de la recuperación de información.
- Conocer y aplicar las técnicas básicas de Ciencia de Datos y aprendizaje automático (Big Data) aplicables a la documentación digital.

5.- Contenidos

1. Introducción
1. Concepto de documento digital
2. Documentos dinámicos vs. documentos estáticos
3. Documentos colaborativos
2. Técnicas básicas de búsqueda
1. Operadores relacionales y operadores booleanos
2. Búsqueda de subcadenas
3. Expresiones regulares
3. Modelos de representación (vectores, pesos, similitudes)
4. Aprendizaje Automático. Modelos heurísticos, modelos estadísticos y modelos mixtos.
5. Aprendizaje supervisado. Aplicación a la categorización de documentos. Técnicas de entrenamiento y construcción de clasificadores. Métricas de evaluación.
6. Otras aplicaciones del aprendizaje supervisado en el ámbito documental. Detección de idiomas, detección de entidades, extracción de información
7. Aprendizaje no supervisado. Clasificación Automática de Documentos. Modelos planos y modelos jerárquicos. Granularidad y técnicas de corte. Etiquetado automático y web semántica.
8. Análisis automático de contenidos digitales.

6.- Competencias a adquirir

Básicas / Generales:

CB6: 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.

CB7. Los estudiantes sabrán 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

CB8. Los estudiantes serán 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.

CB9. Los estudiantes sabrán comunicar sus conclusiones –y los conocimientos y razones últimas que las sustentan- a públicos especializados y no especializados de un modo claro y sin ambigüedades. CB10. Los estudiantes poseerán las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Específicas:

CE2. Capacidad para seleccionar y aplicar las técnicas adecuadas, en función de las necesidades de los usuarios, en la selección, organización, acceso y recuperación, conservación, difusión e intercambio de la información digital.

CE3. Desempeño en el uso de las tecnologías de la información que se emplean en las unidades y servicios de información digital.

CE5. Capacidad para utilizar las herramientas informáticas para la implantación, desarrollo y explotación de sistemas de información digital.

CE6. Capacidad para analizar, asesorar y formar a productores, usuarios y clientes de servicios de información digital.

7.- Metodologías docentes

Clases de teoría con apoyo de material audiovisual. Se trata de clases en las que se presentan los contenidos básicos de la materia. La clase comenzará con una breve introducción de los contenidos que se pretenden transmitir en la clase. El desarrollo de la clase se llevará a cabo con medios audiovisuales, textos, transparencias, etc., que permitan un adecuado nivel de motivación e interés en los alumnos. Se facilitarán a los alumnos en la página web de la asignatura previamente a su exposición. Estas transparencias son una guía para el estudio, pero no son sustitutas de la bibliografía recomendada

Clases prácticas. Las clases prácticas presenciales estarán dedicadas a la resolución colaborativa de problemas de representación y análisis de la información, para lo cual se utilizarán varias de las herramientas existentes.

Trabajo obligatorio. Al alumno se le darán unas pautas para la realización de trabajos. Serán propuestos varios tipos de trabajos: algunos serán de revisión bibliográfica de temas actuales, otros de aplicación práctica de aspectos no vistos en los talleres de prácticas, y algunos otros serán trabajos de iniciación a la investigación.

Presentación de los trabajos. Defensa pública de los trabajos, la cual estimula la interacción con el resto de los alumnos de la asignatura.

Tutorías. El alumnado tiene a su disposición horas de tutorías en las que puede consultar cualquier duda relacionada con los contenidos, organización y planificación de la asignatura. Las tutorías pueden ser individualizadas, pero se admiten tutorías grupales para resolver problemas relacionados con las actividades realizadas en grupo.

8.- Previsión de distribución de las metodologías docentes

Actividad Formativa		Horas Presenciales		Horas de Trabajo personal		Porcentaje. Presencial
Sesiones magistrales		20		30		40 %
Prácticas en aula		20		30		40 %
Preparación de trabajos		10		20		33,3 %
Seminarios		10		10		50 %
Total Horas	150	Total Horas Presenciales	60	Total Horas Trabajo Autón.	90	40 %

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.MorganKaufmann 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.**Revistas**

- Information Processing & Management.
- Journal of Documentation.
- Journal of the American Society for Information Science and Technology.
- Library & Information Science Research.
- Scientometrics.

10.- Evaluación**Consideraciones Generales**

Los alumnos deberán asistir regularmente a las actividades presenciales y poner interés en el desarrollo de la materia. También en el desarrollo del trabajo obligatorio

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.