



1. Identificación

1.1. De la asignatura

Curso Académico	2024/2025
Titulación	GRADO EN CIENCIA E INGENIERÍA DE DATOS
Nombre de la asignatura	RECUPERACIÓN DE INFORMACIÓN
Código	6594
Curso	TERCERO
Carácter	OBLIGATORIA
Número de grupos	1
Créditos ECTS	6.0
Estimación del volumen de trabajo	150.0
Organización temporal	1º Cuatrimestre
Idiomas en que se imparte	Español

1.2. Del profesorado: Equipo docente

FERNANDEZ MAIMO, LORENZO

Docente: **GRUPO 1**

Coordinación de los grupos: **GRUPO 1**

Coordinador de la asignatura

Categoría

PROFESORES TITULARES DE UNIVERSIDAD

Área

ARQUITECTURA Y TECNOLOGÍA DE COMPUTADORES

Departamento

INGENIERÍA Y TECNOLOGÍA DE COMPUTADORES

Correo electrónico / Página web / Tutoría electrónica

lfmaimo@um.es <http://ditec.um.es/personal/8> Tutoría electrónica: **Sí**

Teléfono, horario y lugar de atención al alumnado

Duración: C1 **Día:** Martes **Horario:** 15:30-17:00 **Lugar:** 868884651, Facultad de Informática B1.3.033

Observaciones:
Despacho 3.35. Avisar previamente por correo electrónico

Duración: C2 **Día:** Miércoles **Horario:** 16:00-17:30 **Lugar:** 868884651, Facultad de Informática B1.3.033

Observaciones:
Despacho 3.35. Avisar previamente por correo electrónico

Duración: C2 **Día:** Miércoles **Horario:** 11:00-12:30 **Lugar:** 868884651, Facultad de Informática B1.3.033

Observaciones:
Despacho 3.35. Avisar previamente por correo electrónico

Duración: C1 **Día:** Miércoles **Horario:** 11:00-12:30 **Lugar:** 868884651, Facultad de Informática B1.3.033

Observaciones:
Despacho 3.35. Avisar previamente por correo electrónico

QUILES PEREZ, MARIO

Docente: GRUPO 1

Coordinación de los grupos:

Categoría

INVESTIGADOR/A PREDOCTORAL (SÉNECA)

Área

ARQUITECTURA Y TECNOLOGÍA DE COMPUTADORES

Departamento

No consta

Correo electrónico / Página web / Tutoría electrónica

mqp@um.es Tutoría electrónica: No

Teléfono, horario y lugar de atención al alumnado

2. Presentación

IntrodLa tecnología de recuperación de la información empezó hace años con los primeros esfuerzos de digitalización en grandes organizaciones y gobiernos, pero actualmente todos dependemos de poder acceder a dicha información, ya que prácticamente todo está ya digitalizado, y la mayoría de dicha información no saldrá ya del mundo digital. La web lo cambió todo, haciendo que todo el mundo pueda acceder a la información, permitiendo que cualquier pueda participar compartiendo su propia información y haciendo que podamos influir en la información que publican los demás.

Recuperar información cuando se encuentra distribuida por la web en muy diversos formatos supone un reto importante. El primer reto es cómo localizar la información que necesitamos, y el siguiente es cómo acceder a ella. En esta asignatura nos centraremos en recuperar documentos de la web, entendiendo por documentos cualquier formato que admita términos y enlaces, donde sea necesario poder hacer consultas sobre la información que contiene. Veremos las bases de los sistemas buscadores como Google y los métodos eficientes para indexar la información. Todo esto nos servirá como primer paso para la asignatura del segundo cuatrimestre de procesamiento del lenguaje natural.

3. Condiciones de acceso a la asignatura

3.1. Incompatibilidades

No constan

3.2. Requisitos

No constan

3.3. Recomendaciones

No hay exigencias de conocimientos previos, pero es altamente recomendable saber programar en Python, tanto en imperativo como en orientación a objeto, así como haber superado las asignaturas de Álgebra Lineal, Estructuras de Datos, Programación para la Ciencia de Datos y Bases de Datos II.

4. Competencias

4.1. Competencias básicas

- CB1: Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio
- CB2: Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
- CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

4.2. Competencias de la titulación

- CG2: Conocer y aplicar los fundamentos de ingeniería y tecnologías informáticas actuales para diseñar e implementar nuevas aplicaciones de análisis de datos.
- CG3: Capacidad para desarrollar experimentos y para implementar sistemas, infraestructuras, procesos y herramientas con el fin de soportar la manipulación de los datos durante todo el ciclo de vida de estos.

- CG5: Conocer, desarrollar e implementar estrategias de gestión de datos con el fin de realizar su recolección, almacenamiento, preservación y disponibilidad para posteriores procesamientos.
- CE3: Extraer información de fuentes tipos y formatos de datos semiestructurados y estructurados de interés para la ciencia de datos, incluyendo texto, imágenes, y conjuntos de datos de libre disposición, y hacerlos disponibles para análisis y usos posteriores.
- CE14: Desarrollar e implementar modelos de datos relevantes, definir metadatos usando estándares y prácticas comunes para distintos tipos de fuentes de datos en una variedad de dominios científicos e industriales.
- CE15: Procesar y analizar datos procedentes de fuentes, tipos y formatos de datos semi estructurados y estructurados de interés para Ciencia de datos.
- CE17: Asegurar la calidad de los datos, su accesibilidad, interoperabilidad, cumplimiento de estándares y publicación (data curation).

4.3. Competencias transversales y de materia

No constan

5. Contenidos

5.1. Teoría

Tema 1: Introducción

Tema 2: Modelos básicos de recuperación de información

- Modelo booleano de recuperación de información
- Índices invertidos. Conceptos básicos y funcionamiento
- Consultas en el modelo booleano. Procesamiento mediante un índice invertido. Optimización de consultas
- Modelo booleano extendido: consultas de frases. Índices posicionales.
- Construcción de índices.
- Compresión de índices: Variable-byte codes.
- Evaluación de índices

Tema 3: Modelos de espacio vectorial

- Scoring
- Term-weighting
- TF-IDF
- Distancia del coseno
- BM25
- Problemas: versiones mínimas de python.

- Evaluación NDGC
- Aprendizaje pesos

Tema 4: Análisis de enlaces

- Técnicas de ranking basadas en análisis matricial
- PageRank
- Análisis de grafos para redes sociales

Tema 5: Metadatos y semántica

- APIs de acceso datos
- CSV, XML, JSON, parquet, y otros
- OpenData
- Web Scraping/crawling

5.2. Prácticas

▪ Práctica 1: Práctica 1

Ficheros invertidos / compresión / búsquedas booleanas.

Relacionado con:

- Tema 2: Modelos básicos de recuperación de información

▪ Práctica 2: Práctica 2

Modelos espacio vectorial: TF-IDF / BM25.

Relacionado con:

- Tema 3: Modelos de espacio vectorial

▪ Práctica 3: Práctica 3

PageRank

Relacionado con:

- Tema 4: Análisis de enlaces

▪ Práctica 4: Práctica 4

Web-crawling + scraping. Enlaces HTML. Fuentes OpenData (formatos parquet, XML, JSON y otros)

Relacionado con:

- Tema 5: Metadatos y semántica

6. Actividades Formativas

Actividad Formativa	Metodología	Horas	Presencialidad
AF1: Exposición teórica: Exposición de los contenidos teóricos de la asignatura por parte del profesor. También se contemplan las sesiones informativas sobre el desarrollo del trabajo de fin de grado o prácticas externas.		20.0	100.0
AF2: Seminarios y actividades de aula: Exposición, análisis y debate dentro del contexto de aplicaciones específicas de contenidos teóricos, así como planteamiento y resolución de ejercicios y casos prácticos en el aula, tanto al grupo completo como en grupos reducidos. También se contemplan conferencias, debates y seminarios temáticos.		5.0	100.0
AF3: Prácticas de laboratorio: Ejercicios y resolución de problemas, aprendizaje orientado a proyectos, estudio de casos, exposición y discusión de trabajos, simulaciones y/o prácticas con ordenadores, generalmente desarrolladas en grupos reducidos.		28.0	100.0
AF4: Trabajo autónomo del alumno: Estudio y preparación de contenidos teórico-prácticos, trabajo individual consistente en lecturas, búsquedas de información, sistematización de contenidos, elaboración de informes o estudio para la elaboración de casos entre otras actividades.		90.0	0.0
AF5: Tutorías formativas y resolución de dudas: Asistencia individualizada -tutorías individuales- o en grupo -tutorías colectivas- a los estudiantes por parte del profesor.		3.0	100.0
AF6: Evaluación: Pruebas individuales, ya sean escritas, orales o con medios informáticos, donde el estudiante demostrará los conocimientos teóricos y prácticos adquiridos durante las actividades formativas asociadas a la enseñanza de la materia.		4.0	100.0
	Totales	150,00	

7. Horario de la asignatura

<https://www.um.es/web/estudios/grados/ciencia-ingenieria-datos/2024-25#horarios>

8. Sistemas de Evaluación

Identificador	Denominación del instrumento de evaluación	Criterios de Valoración	Ponderación
SE1	Exámenes individuales: Pruebas objetivas, de desarrollo, de respuesta corta, de ejecución de tareas o de escala de actitudes, realizadas por los estudiantes para mostrar los conocimientos teóricos y prácticos adquiridos ya sea de forma oral, escrita o utilizando un ordenador.	El examen será teórico-práctico, entendiéndose que puede incluir toda la materia vista en teoría y en prácticas.	50.0

SE3	Procedimientos de observación del trabajo del estudiante: Registros de participación, de realización de actividades, cumplimiento de plazos, participación en foros, informes de seguimiento del trabajo fin de grado y registros sobre el desarrollo de las prácticas externas.		0.0
SE4	Informe técnico. En este instrumento incluimos los resultados de actividades prácticas, o de laboratorio, junto con sus memorias descriptivas y posibles resúmenes del estado del arte sobre temas concretos. La opción de realizar entrevistas personales o presentaciones de los trabajos realizados también entran en esta categoría.	Existe la posibilidad de que se evalúe partes de las prácticas en el laboratorio dentro del horario de clase como forma de probar que se ha seguido la parte práctica de la asignatura.	50.0

9. Fechas de exámenes

<https://www.um.es/web/estudios/grados/ciencia-ingenieria-datos/2024-25#examenes>

10. Resultados del Aprendizaje

No constan

11. Bibliografía

Bibliografía básica

- [Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. \(versión online: <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>\)](#)
- [Versión online del libro "Introduction to Information Retrieval", Cambridge University Press. 2008.](#)
- [R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval \(2nd ed.\). Addison Wesley, 2010](#)
- [Russell, Matthew , Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Site. 3ª edición. O' Reilly , 2019](#)

Bibliografía complementaria

No constan

12. Observaciones

Cada parte, teoría y práctica, se consideran aprobadas si se obtiene una calificación de 5 o más en una escala de 0 a 10. La nota final será la media ponderada de la teoría y las prácticas en el caso de haber aprobado ambas, o la nota del menor de los suspensos en el caso de haber suspendido alguna de las dos. En cualquier otro caso, será NP.

Esta asignatura no se encuentra vinculada de forma directa con los Objetivos de Desarrollo Sostenible.

NECESIDADES EDUCATIVAS ESPECIALES

Aquellos estudiantes con discapacidad o necesidades educativas especiales podrán dirigirse al Servicio de Atención a la Diversidad y Voluntariado (ADYV - <https://www.um.es/adyv>) para recibir orientación sobre un mejor aprovechamiento de su proceso formativo y, en su caso, la adopción de medidas de equiparación y de mejora para la inclusión, en virtud de la Resolución Rectoral R-358/2016. El tratamiento de la información sobre este alumnado, en cumplimiento con la LOPD, es de estricta confidencialidad.

REGLAMENTO DE EVALUACIÓN DE ESTUDIANTES

El artículo 8.6 del Reglamento de Evaluación de Estudiantes (REVA) prevé que "salvo en el caso de actividades definidas como obligatorias en la guía docente, si el o la estudiante no puede seguir el proceso de evaluación continua por circunstancias sobrevenidas debidamente justificadas, tendrá derecho a realizar una prueba global".

Se recuerda asimismo que el artículo 22.1 del Reglamento de Evaluación de Estudiantes (REVA) estipula que "el o la estudiante que se valga de conductas fraudulentas, incluida la indebida atribución de identidad o autoría, o esté en posesión de medios o instrumentos que faciliten dichas conductas, obtendrá la calificación de cero en el procedimiento de evaluación y, en su caso, podrá ser objeto de sanción, previa apertura de expediente disciplinario".