

Obtención y Almacenamiento de Datos

Código: 104851
Créditos ECTS: 6

2024/2025

Titulación	Tipo	Curso
2503852 Estadística Aplicada	FB	1

Contacto

Nombre: Marc Vallribera Ros

Correo electrónico: marc.vallribera@uab.cat

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

Conocimientos sobre operaciones lógicas.
Conocimientos sobre conjuntos y relaciones entre conjuntos.
Conocimientos básicos de Python.

Objetivos y contextualización

En esta asignatura se introducen los conceptos básicos de Bases de Datos (BD) necesarios tanto a nivel de diseñador de BD como de usuario, así como los mecanismos de obtención de información de internet a través de Web Scrapping y el almacenamiento de esta información en las BD.

Resultados de aprendizaje

1. CM03 (Competencia) Depurar la información, teniendo en cuenta sus implicaciones éticas, para almacenarla en soporte informático para la ulterior explotación y uso.
2. CM03 (Competencia) Depurar la información, teniendo en cuenta sus implicaciones éticas, para almacenarla en soporte informático para la ulterior explotación y uso.
3. KM06 (Conocimiento) Reconocer los conceptos básicos de bases de datos, tanto a nivel de usuario como de diseñador, cuidando las implicaciones éticas vinculadas con la información recogida.
4. KM06 (Conocimiento) Reconocer los conceptos básicos de bases de datos, tanto a nivel de usuario como de diseñador, cuidando las implicaciones éticas vinculadas con la información recogida.
5. SM05 (Habilidad) Obtener información a través de webscrapping para almacenarla en bases de datos adecuadas.
6. SM06 (Habilidad) Explotar bases de datos de cualquier volumen.

Contenido

1. Introducción a las bases de datos
 1. Sistemas de información y bases de datos

2. Bases de datos
 1. Conceptos
 2. Características
 3. Evolución de las bases de datos
3. Definición y características de un SGBD
4. Arquitectura de los SGBD
5. Principales SGBD
6. Fases de desarrollo de aplicaciones de BD
7. Etapas de diseño de BD
 1. Diseño conceptual
 2. Diseño lógico
 3. Diseño de aplicaciones
2. El modelo Entidad-Relación (E-R)
 1. El diagrama de E-R
 2. Entidades, atributos, interrelaciones
 3. Atributos de interrelaciones
 4. Dependencia de la existencia y de participación
3. El modelo relacional
 1. Concepto de relación o tabla, atributos, tuplas, dominios, claves principales y externas
 2. Restricciones de dominio, integridad de claves y referencia
 3. Transformación del modelo E/R a relacional
 4. Operadores de álgebra relacional
4. Implementación de la base de datos
 1. Lenguaje de consulta estructurado (SQL)
 2. Procesamiento de datos
 3. Consulta de datos
 4. Gestión de la Base de Datos con SQL
 5. Trabajando con Bases de Datos SQLite
5. Nociones de HTML y expresiones regulares
 1. Estructura y código de una página web
 2. Etiquetas y atributos HTML y CSS
 3. Búsqueda de textos con expresiones regulares
 4. Caracteres especiales, conjuntos, grupos y repeticiones
6. Recopilación y almacenamiento de datos de páginas web
 1. Introducción a las herramientas de Web Scraping
 2. Programación de herramientas de Web Scraping en Python
 3. Búsqueda y obtención de información con expresiones regulares
 4. Búsqueda y obtención de información con BeautifulSoup
 5. Almacenamiento en bases de datos
 6. Exportación de resultados en ficheros de valores separados por comas

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases de teoría	26	1,04	
Tipo: Supervisadas			
Entregas de ejercicios individuales	9	0,36	
Prácticas	36	1,44	
Tipo: Autónomas			

Estudio	15	0,6
Lectura de libros	20	0,8
Preparación de las practicas	10	0,4
Problemas propuestos	23	0,92

Teoría

Se imparten a través de clases magistrales con transparencias. Estas transparencias son accesibles, y los alumnos pueden obtenerlas en la página web de la asignatura del Campus Virtual.

Ejercicios

Se harán 2 entregas (individuales) para que el estudiante pueda demostrar que está adquiriendo el conocimiento que se explica en clase. La entrega se realizará a través de Moodle que se encuentra en el Campus Virtual.

Problemas propuestos

durante el curso se proporcionarán una lista de problemas, sobre los temas más prácticos de la asignatura, para que el alumno pueda adquirir y/o consolidar su conocimiento de las diferentes etapas del diseño, implementación y explotación de las bases de datos.

Preparación de la práctica

El estudiante debe haber leído y preparado las prácticas para poder hacerlas dentro del horario establecido de prácticas y fuera de él.

Prácticas

El objetivo de las sesiones de laboratorio es dar una visión amplia de las bases de datos, desde la gestión y creación hasta la conexión con una aplicación que permite consultar y modificar la base de datos. Los estudiantes tendrán que adquirir competencias en la creación, gestión y manipulación de bases de datos, así como en la obtención de información de Internet, y en el almacenamiento de los datos en la base de datos. A lo largo de estas sesiones de laboratorio, el profesor supervisará y guiará a cada grupo de estudiantes durante el proceso de práctica.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de las prácticas	30%	5	0,2	CM03, KM06, SM05, SM06
Entrega ejercicios	30%	3	0,12	CM03, KM06, SM05, SM06
Examen final	40%	3	0,12	CM03, KM06, SM05

El 70% de la calificación del curso se basará en la calificación de las prácticas y un examen final, que se puede recuperar. El 30% restante se evaluará mediante entregas de la evaluación continua. Todas las notas enumeradas a continuación son sobre 10.

La calificación final será: $\text{Nota final} = 0.4 * \text{Nota examen} + 0.3 * \text{nota de prácticas} + 0.3 * \text{Nota ejercicios}$

Para aprobar el curso es necesario pasar las prácticas y el examen por separado.

Examen (40%): El examen principal del curso y el examen de recuperación se llevarán a cabo el día reservado para esta asignatura dentro del calendario de exámenes.

Prácticas (30%): Se hará una entrega evaluada de las prácticas, resultado de las diferentes sesiones.

Al final del curso, las prácticas se pueden recuperar, con una entrega especial. La nota máxima que se puede obtener en la recuperación de las prácticas será de 5.

Ejercicios (30%): La nota de ejercicios se obtendrá de los problemas que se entregarán durante el curso. La forma específica y los días que se realizarán las entregas de los problemas serán notificados con previo aviso al Campus Virtual del sujeto. La nota de ejercicios no es recuperable.

Evaluación Única

En caso de acogerse a la Evaluación Única, a final de curso se pedirá realizar una entrega de un ejercicio similar a los realizados en las prácticas, además de realizar un examen con algunas preguntas adicionales. En este caso, la nota será del 70% del examen final y 30% de la entrega del ejercicio.

Se aplicará el mismo sistema de recuperación que por la evaluación continua.

La revisión de la calificación final sigue el mismo procedimiento que para la evaluación continua.

Bibliografía

A. Silberschatz, H.F. Korth, S. Sudarshan (2006), *Fundamentos de Bases de Datos*, McGraw-Hill

Ian Mackie (2020), *A Beginners Guide to Python 3 Programming*

https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991010431569006709

Michael Heydt (2018), *Python web scraping cookbook: over 90 proven recipes to get you scraping with Python, microservices, Docker, and AWS*

https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991009832849706709

Ryan Mitchell (2018), *Web scraping with Python: collecting data from the modern web*

https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991009832809706709

Software

- [Visual Studio Code - Code Editing. Redefined](#): Se utilizará para crear los programas en Python para hacer Web scraping y exportar los datos a una base de datos o en un archivo.
- [Welcome to Python.org](#): Lenguaje de programación utilizado para el Web scraping.
- [SQLiteStudio](#): Entorno de gestión de Bases de Datos en SQLite.

Lista de idiomas

Nombre	Grupo	Idioma	Semestre	Turno
(PLAB) Prácticas de laboratorio	1	Catalán	segundo cuatrimestre	tarde
(PLAB) Prácticas de laboratorio	2	Catalán	segundo cuatrimestre	tarde
(SEM) Seminarios	1	Catalán	segundo cuatrimestre	tarde
(SEM) Seminarios	2	Catalán	segundo cuatrimestre	tarde
(TE) Teoría	1	Catalán	segundo cuatrimestre	tarde