

## Obtención y Almacenamiento de Datos

Código: 104851  
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503852 Estadística Aplicada	FB	1	2

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

### Contacto

Nombre: Marc Vallribera Ros  
Correo electrónico: Marc.Vallribera@uab.cat

### Uso de idiomas

Lengua vehicular mayoritaria: catalán (cat)  
Algún grupo íntegramente en inglés: No  
Algún grupo íntegramente en catalán: Sí  
Algún grupo íntegramente en español: No

### Prerequisitos

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 Scraping y el almacenamiento de esta información en las BD.

### Competencias

- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- 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.
- 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.
- Seleccionar las fuentes y técnicas de adquisición y gestión de datos adecuadas para su tratamiento estadístico.
- Utilizar correctamente un amplio espectro del software y lenguajes de programación estadísticos, escogiendo el más apropiado para cada análisis y ser capaz de adaptarlo a nuevas necesidades.

### Resultados de aprendizaje

1. Entender los algoritmos informáticos utilizados para gestionar una base de datos y el lenguaje SQL.
2. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
3. Gestionar una base de datos.

4. Identificar correctamente los tipos de datos y de medidas.
5. Identificar las ventajas y los inconvenientes de internet como fuente importante de información en estadística.
6. 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.
7. 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.
8. Realizar operaciones básicas relacionadas con depuración de la información.
9. Utilizar las fuentes de información adecuadas para cada tipo de estudio aplicado.

## 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

## Metodología

### 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 3 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.

## Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases de teoría	26	1,04	2, 1, 8, 3, 4, 5, 7, 9
Tipo: Supervisadas			
Entregas Evaluación Continuada	9	0,36	2, 1, 3, 4, 9
Prácticas	36	1,44	2, 1, 8, 3, 4, 7, 9
Tipo: Autónomas			
Estudio	15	0,6	2, 1, 8, 3, 4, 5, 7, 9
Lectura de libros	20	0,8	1, 4, 5, 9
Preparación de las prácticas	10	0,4	2, 1, 8, 3, 4, 7, 9
Problemas propuestos	23	0,92	4, 5, 7, 9

## Evaluación

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:  $\text{Nota final} = 0.4 * \text{Nota examen} + 0.3 \text{ nota prácticas} + 0.3 * \text{Nota continuada}$

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

Examen (40%): El examen principal del curso se llevará a cabo el último día de clase. El examen de recuperación se llevará 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.

## Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de las prácticas	30%	5	0,2	2, 1, 8, 3, 4, 5, 6, 7, 9
Entrega ejercicios	30%	3	0,12	2, 1, 8, 3, 4, 6, 7
Examen final	40%	3	0,12	2, 1, 8, 3, 4, 5, 6, 7

## Bibliografía

A. Silberschatz, H.F. Korth, S. Sudarshan (2006), *Fundamentos de Bases de Datos*, McGraw-Hill  
Ryan Mitchell (2018), *Web Scraping with Python*, O'Reilly Media, Inc, USA [ISBN 1491985577]  
Michael Heydt (2018) *Python Web Scraping Cookbook*, Packt Publishing [ISBN 1787285219]  
Ian Mackie (2020) *A Beginners Guide to Python 3 Programming*

[UAB Biblioteques - Cercador -- Python web scraping cookbook : over 90 proven recipes to get you scraping with Python, microservices, Docker, and AWS / Michael Heydt](#)

[UAB Biblioteques - Cercador -- Web scraping with Python : collecting data from the modern web / Ryan Mitchell](#)

## 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.