

Obtenció i Emmagatzematge de Dades

Codi: 104851
Crèdits: 6

2024/2025

Titulació	Tipus	Curs
2503852 Estadística Aplicada	FB	1

Professor/a de contacte

Nom: Marc Vallribera Ros

Correu electrònic: marc.vallribera@uab.cat

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

Coneixements sobre operacions lògiques.
Coneixements sobre conjunts i relacions entre conjunts.
Coneixements bàsics de Python.

Objectius

En aquesta assignatura s'introdueixen els conceptes bàsics de Bases de Dades (BD) necessaris tant a nivell de dissenyador de BD com d'usuari, així com els mecanismes d'obtenció d'informació d'internet a través de "WebScraping", i emmagatzematge d'aquesta informació en les BD.

Resultats d'aprenentatge

1. CM03 (Competència) Depurar la informació, tenint-ne en compte les implicacions ètiques, per a emmagatzemar-la en suport informàtic per a l'explotació i ús ulteriors.
2. CM03 (Competència) Depurar la informació, tenint-ne en compte les implicacions ètiques, per a emmagatzemar-la en suport informàtic per a l'explotació i ús ulteriors.
3. KM06 (Coneixement) Reconèixer els conceptes bàsics de les bases de dades, tant pel que fa a l'usuari com al dissenyador, tenint cura de les implicacions ètiques vinculades a la informació recollida.
4. KM06 (Coneixement) Reconèixer els conceptes bàsics de les bases de dades, tant pel que fa a l'usuari com al dissenyador, tenint cura de les implicacions ètiques vinculades a la informació recollida.
5. SM05 (Habilitat) Obtenir informació a través de webscraping per a emmagatzemar-la en bases de dades adequades.
6. SM06 (Habilitat) Explotar bases de dades de qualsevol volum.

Continguts

1. Introducció a les bases de dades
 1. Sistemes d'informació i bases de dades

2. Bases de dades
 1. Conceptes
 2. Característiques
 3. Evolució de les Bases de Dades
3. Definició i característiques d'un SGBD
4. Arquitectura dels SGBD
5. Principals SGBD
6. Fases de desenvolupament d'aplicacions de BD
7. Etapes del disseny d'una BD
 1. Disseny conceptual
 2. Disseny lògic
 3. Disseny d'aplicacions
2. El model Entitat - Relació (E-R)
 1. El diagrama E-R
 2. Entitats, atributs, interrelacions
 3. Atributs de les interrelacions
 4. Dependència d'existència i participació
3. El Model Relacional
 1. Concepte de relació o taula, atributs, tuples, dominis, claus primàries i foranies
 2. Restriccions de domini, integritat de claus i referència
 3. Transformació del model E/R al relacional
 4. Operadors d'àlgebra relacional
4. Implementació de les Bases de Dades
 1. Structured Query Language (SQL)
 2. Tractament de les dades
 3. Consulta de les dades
 4. Gestió de la Base de Dades amb SQL
 5. Treballant amb bases de dades SQLite
5. Nocions de HTML i expressions regulars
 1. Estructura del codi d'una pàgina web
 2. Etiquetes i atributs HTML i CSS
 3. Cerca de textos amb expressions regulars
 4. Caràcters especials, conjunts, agrupaments i repeticions
6. Obtenció i emmagatzemament de dades de pàgines web
 1. Introducció a les eines de Web Scraping
 2. Programació d'eines de Web Scraping en Python
 3. Cerca i obtenció d'informació amb expressions regulars
 4. Cerca i obtenció d'informació amb BeautifulSoup
 5. Emmagatzemament en Bases de Dades
 6. Exportació de resultats en fitxers de valors separats per comes

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de teoria	26	1,04	
Tipus: Supervisades			
Entregues d'exercicis individuals	9	0,36	
Pràctiques	36	1,44	
Tipus: Autònomes			

Estudi	15	0,6
Lectura de llibres	20	0,8
Preparació de les pràctiques	10	0,4
Problemes proposats	23	0,92

Teoria

Les classes de teoria s'imparteixen mitjançant classes magistrals amb transparències. Aquestes transparències són accessibles, i els alumnes les poden obtenir des de la pàgina web de l'assignatura del Campus Virtual.

Exercicis

Es realitzaran 2 entregues (individuals) per tal que l'alumne pugui demostrar que està adquirint els coneixements que s'expliquen a classe. L'entrega es realitzarà a través del Moodle que hi ha al Campus Virtual.

Problemes proposats

Durant el transcurs del curs s'entregarà un llistat de problemes, referent als temes més pràctics de l'assignatura, per tal que l'alumne pugui adquirir i/o afermar el seus coneixements sobre les diferents etapes en el disseny, implementació i explotació de les bases de dades.

Preparació de les pràctiques

L'alumne haurà d'haver llegit i preparat les pràctiques per tal de poder-les realitzar dins l'horari de pràctiques establert i fora d'aquest.

Pràctiques

L'objectiu de les sessions de laboratori és donar una ampla visió de les bases de dades, des de la gestió i creació fins a la connexió amb un aplicatiu que permeti consultar i modificar la base de dades. L'alumne haurà d'assolir competències en creació, gestió i manipulació de bases de dades, així com l'obtenció de informació d'internet, i l'emmagatzematge d'aquesta a la Base de Dades.

Al llarg d'aquestes sessions de laboratori, el professor anirà supervisant i guiant cada grup d'alumnes durant el procés de realització de la pràctica.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen final	40%	3	0,12	CM03, KM06, SM05
Lliurament de les pràctiques	30%	5	0,2	CM03, KM06, SM05, SM06
Lliurament exercicis	30%	3	0,12	CM03, KM06, SM05, SM06

Un 70% de la nota de l'assignatura es basarà en la nota de pràctiques i un examen final, que es podrà recuperar. El 30% restant s'avaluarà a través de les entregues de l'avaluació continuada. Totes les notes que s'indiquen a continuació són sobre 10.

La nota final serà: $\text{Nota Final} = 0.4 * \text{Nota examen} + 0.3 * \text{Nota pràctiques} + 0.3 * \text{Nota exercicis}$

Per aprovar l'assignatura caldrà aprovar les pràctiques i l'examen per separat.

Examen (40%): L'examen principal de l'assignatura i l'examen de recuperació es realitzaran el dia reservat a aquesta assignatura dins el calendari d'exàmens.

Pràctiques (30%): Es faran dues entregues avaluades de les pràctiques, resultat de les diferents sessions.

A final de curs es podran recuperar les pràctiques, fent una entrega especial. La nota màxima que es podrà obtenir a la recuperació de les pràctiques serà un 5.

Exercicis (30%): La nota d'exercicis s'obtindrà a partir dels problemes que s'aniran entregant durant el transcurs de l'assignatura. La manera concreta i els dies que es faran les entregues dels problemes s'avisarà amb suficient antelació al Campus Virtual de l'assignatura. La nota dels exercicis no és recuperable.

Avaluació Única

En cas d'acollir-se a la Avaluació Única, a final de curs es demanarà fer una entrega d'un exercici similar als realitzats a les pràctiques, a més de fer un examen amb algunes preguntes addicionals. En aquest cas la nota serà del 70% de l'examen final i 30% de l'entrega de l'exercici.

S'aplicarà el mateix sistema de recuperació que per l'avaluació continuada.

La revisió de la qualificació final segueix el mateix procediment que per a l'avaluació continuada.

Bibliografia

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

Programari

- [Visual Studio Code - Code Editing. Redefined](#): Es farà servir per crear els programes en Python per fer Web Scraping i exportar les dades a una Base de Dades o a un fitxer.
- [Welcome to Python.org](#): Llenguatge de programació utilitzat per al Web Scraping.
- [SQLiteStudio](#): Entorn de gestió de Bases de Dades en SQLite.

Llista d'idiomes

Nom	Grup	Idioma	Semestre	Torn
(PLAB) Pràctiques de laboratori	1	Català	segon quadrimestre	tarda
(PLAB) Pràctiques de laboratori	2	Català	segon quadrimestre	tarda
(SEM) Seminaris	1	Català	segon quadrimestre	tarda
(SEM) Seminaris	2	Català	segon quadrimestre	tarda
(TE) Teoria	1	Català	segon quadrimestre	tarda