

Obtenció i Emmagatzematge de Dades

Codi: 104851

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503852 Estadística Aplicada	FB	1	2

Professor/a de contacte

Nom: Marc Vallribera Ros

Correu electrònic: marc.vallribera@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: català (cat)

Grup íntegre en anglès: No

Grup íntegre en català: Sí

Grup íntegre en espanyol: No

Equip docent

Albert Grau Perise

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.

Competències

- Avaluar de manera crítica i amb criteris de qualitat el treball realitzat.
- Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
- Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- Seleccionar les fonts i tècniques d'adquisició i gestió de dades adequades per a fer-ne un tractament estadístic.
- Utilitzar correctament un ampli espectre del programari i llenguatges de programació estadístiques, escollint el més apropiat per a cada anàlisi i ser capaç d'adaptar-lo a noves necessitats.

Resultats d'aprenentatge

1. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
2. Entendre els algorismes informàtics utilitzats per gestionar una base de dades i el llenguatge SQL.
3. Fer operacions bàsiques relacionades amb la depuració de la informació.
4. Gestionar una base de dades.
5. Identificar correctament els tipus de dades i de mesures.
6. Identificar els avantatges i els inconvenients d'internet com a font important d'informació en estadística.
7. Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
8. Que els estudiants sàpiguin aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
9. Utilitzar les fonts d'informació adequades per a cada tipus d'estudi aplicat.

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

Metodologia

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.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de teoria	26	1,04	1, 2, 3, 4, 5, 6, 8, 9
Tipus: Supervisades			
Entregues Avaluació Continuada	9	0,36	1, 2, 4, 5, 9
Pràctiques	36	1,44	1, 2, 3, 4, 5, 8, 9
Tipus: Autònomes			
Estudi	15	0,6	1, 2, 3, 4, 5, 6, 8, 9
Lectura de llibres	20	0,8	2, 5, 6, 9
Preparació de les pràctiques	10	0,4	1, 2, 3, 4, 5, 8, 9
Problemes proposats	23	0,92	5, 6, 8, 9

Avaluació

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 es realitzarà l'últim dia de classe. L'examen de recuperació es realitzarà el dia reservat a aquesta assignatura dins el calendari d'exàmens.

Pràctiques (30%): Es farà una entrega avaluada 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.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen final	40%	3	0,12	1, 2, 3, 4, 5, 6, 7, 8
Lliurament de les pràctiques	30%	5	0,2	1, 2, 3, 4, 5, 6, 7, 8, 9
Lliurament exercicis	30%	3	0,12	1, 2, 3, 4, 5, 7, 8

Bibliografia

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

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

https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991000598339706709

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.