UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Foundations of Computer Science

2324-1-FDS01Q009

Obiettivi

Al termine del corso lo studente sarà in grado di interrogare e di inferire la struttura implicita della base di dati a partire dalle tabelle che la compongono.

Inoltre lo studente saprà scrivere e correggere gli errori di semplici programmi in Python per la gestione e l'analisi di dataset formati da poche tabelle interconnesse, come quelle disponibili a kaggle.com (formati CSV, TSV, JSON). In particolare, le analisi verranno fatte utilizzando il Jupyter Notebook.

Infine, lo studente potrà interrogare database relazionali usando SQL, sia come linguaggio separato, sia come chiamata in un programma Python.

Conoscenza e comprensione

Questo insegnamento fornisce le conoscenze basilari e capacità di comprensione relativamente a:

- Database.
- SQL
- Programmazione in Python
- Notebook Jupyter
- Gestione di dati tabellari con Pandas

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- Scrivere interrogazioni in SQL
- Scrivere notebook in Python per gestire dataset
- Scrivere notebook in Python per analizzare e pulire dataset

Contenuti sintetici

Organizzazione di dati grezzi: file system, file con delimitatori.

Introduzione ai database. Il modello relazionale dei dati e linguaggio SQL. Select ... From … Where.

Introduzione alla programmazione in Python. Programmazione esplorativa. Gestione dati in formato tabellare.

Cenni di test e debugging.

Programma esteso

1. Organizzazione di dati grezzi
2. file, directory, tipi di file
3. principali comandi da terminale
4. file con delimitatori.
5. Introduzione ai database.
6. Il modello relazionale dei dati
7. Il linguaggio SQL: Select ... From … Where su una tabella.
8. Interrogazioni su più tabelle.
9. Introduzione alla programmazione in Python.
10. Array, liste, dizionari
11. Cicli
12. Organizzazione di un programma: funzioni
13. Librerie
14. Programmazione esplorativa. Gestione dati in formato tabellare.
15. Jupyter Notebook
16. Pandas
17. Cenni di test e debugging.

Prerequisiti

Conoscenza di un linguaggio di programmazione

Modalità didattica

Lezioni e esercizi al computer.
Le lezioni non vengono registrate nè trasmesse in streaming.

Materiale didattico

- Downey Think Python 2e: How To Think Like a Computer Scientist
- Downey Think Stats 2e
- Downey Elements of Data Science. This book is a bit too basic.
- VanderPlas, Python Data Science Handbook
- Elmasri, Navathe, Fundamentals of Database Systems

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

L'esame consiste di due parti: uno scritto su SQL e un progetto relativo alla parte su Python.

L'esame scritto prevede quattro domande a risposta aperta per la parte di Basi di Dati, dove ogni domanda è una query SQL. Nello scritto viene valutata la completezza e la correttezza delle risposte.

Il progetto in Python è svolto in piccoli gruppi (max 3 persone), e prevede una discussione del progetto dove viene valutata la capacità individuale di realizzare le analisi richieste e di giustificare le decisioni progettuali. Sia lo scritto che la presentazione del progetto devono essere completati entro Settembre 2024.

Il voto finale è ottenuto per 1/4 dalla prova scritta e per 3/4 dal progetto.

Non sono previste prove in itinere. Si ricorda che è necessario essere iscritti alle prove d'esame tramite segreterie online. Non verranno fatte eccezioni al riguardo.

Orario di ricevimento

Il ricevimento è online e su appuntamento a https://www.unimib.it/gianluca-della-vedova

Sustainable Development Goals