UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Foundations of Computer Science

2021-1-F9101Q001

Aims

At the end of the course, the students will understand how query a database, and how to infer the implicit structure of a database from its tables.

Moreover, the students will be able to write and debug some simple programs in Python to manage and analyze datasets consisting of a few interconnected tables, such as those usually available at kaggle.com (CSV, TSV, JSON formats). This part will use the Jupyter Notebook.

Finally, the students will be able to query a relational database using SQL, both as a separate language as well as called from a Python program.

Knowledge and understanding

This course provides basic knowledge and understanding on:

- Data bases

- SQL
- Programming in Python
- Jupyter notebooks
- Managing tabular datasets with Pandas

Ability to apply knowledge and understanding

At the end of the course the students will be able to:

- Write SQL queries
- Write Python notebooks to manage datasets
- Write Python notebooks to analyze and clean datasets

Contents

Organizing raw datasets: file system, delimited files.

Introduction to data bases. Relational Data Model and SQL. Select ... From … Where.

Introduction to programming with Python. Explorative programmaing. Managing tabular data.

Introduction to testing and debugging.

Detailed program

1. Organizing raw datasets
2. files, directories, types of files
3. main command-line commands
4. delimiter-separated values
5. Introduction to data bases.
6. The Relational Data Model.
7. SQL: Select ... From … Where on a table.
8. Querying two or more tables.
9. Introduction to programming in Python.
10. Arrays, lists, dictionaries
11. Loops
12. Organization of a programma: functions
13. Modules and libraries
14. Explorative programming. Managing tabular data.
15. The Jupyter Notebook
16. Pandas
17. Introduction to testing and debugging.

Prerequisites

None

Teaching form

Lectures and exercises with a PC. During the emergency there will be some prerecorded video lectures: those asynchronous video lectures will be roughly 25% of the course. The remaining 75% of the course will be with synchronous videoconferences where some practical exercises will be attacked.

This course is taught in Italian.

Textbook and teaching resource

Downey, Pensare in Python (https://github.com/AllenDowney/ThinkPythonItalian)

Downey, Think Python (http://greenteapress.com/wp/think-python-2e/)

VanderPlas, Python Data Science Handbook (https://jakevdp.github.io/PythonDataScienceHandbook/

Allulli, Nanni. Fondamenti di Basi di Dati

Semester

First

Assessment method

The exam consists of 2 parts: a written exam (with open questions) on the Data bases topics, and a small group (max 3 people) project, with an oral discussion, on the Python topics.

The grading of the written exam is based on the correctness and the completeness of the answers. The written exam consists of writing 4 SQL queries, where each query corresponds to an open-ended question.

The grading of the project is based on the individual contribution that is displayed during the oral discussion and on the fitness of the project to perform the required analysis. The composition of the group is determined by prof. Della Vedova, with the goal of guaranteeing an adequate level of heterogeneity of the group. More precisely, each group will have a student with good programming skills and another student with good statistical skills.

The final grade is 1/3 of the written exam and 2/3 of the project.

There are no in-progress exams.

Office hours

Please reserve a meeting via email. During the emergency, office hours will be via videoconference.