Course type: seminar
Extent: two hours per week
Website: MOODLE

Outcomes: we will introduce current topics in data management research, focusing on challenges related to big data frameworks, Web data and the harvesting or processing of data with the crowd of Web users, namely crowdsourcing techniques.

Contents: the seminar is based on student presentations. Students are required to select one research paper and present it to their peers in class, in singles or couples (depending on the number of registered students). A single paper will be presented each week, over the ninety minutes of class.

The papers will involve the following topics:

**Big data management:** MapReduce, architectures, data exploration

**Web data management:** Semantic Web, Web-based knowledge bases, SPARQL, information extraction from semi-structured data

**Crowdsourcing:** data harvesting and analysis with the help of the crowd

data manipulation: data analysis, big data frameworks, Web data, harvesting data with the help of the crowd.
_requirements:

Prerequisites: first and second year courses, optionally database systems (89-851) or databases (89-281)

Assignments and other duties: presenting in class, in singles or couples (depending on the number of registered students). Attendance is mandatory.

Grade: 95% presentation, 95% active participation in discussion during peers’ presentations.

Bibliography: the list of recommended papers will be announced at the beginning of the semester, at the course Website.

Recommended reading:


