EPFL

CS-423   **Distributed information systems**

Aberer Karl

Cursus	Sem.	Type
Biocomputing minor	E	Opt.
Communication systems minor	E	Opt.
Computer science minor	E	Opt.
Computer science	MA2	Obl.
Data Science	MA2	Opt.
Digital Humanities	MA2	Opt.
Electrical and Electronical Engineering	MA2, MA4	Opt.
Energy Management and Sustainability	MA2, MA4	Opt.
Environmental Sciences and Engineering	MA2, MA4	Opt.
SC master EPFL	MA2, MA4	Obl.

Language of teaching	English
Credits	4
Session	Summer
Semester	Spring
Exam	Written
Workload	120h
Weeks	14
Hours	**3 weekly**
Courses	2 weekly
Exercises	1 weekly
Number of positions	

Summary

This course introduces in detail several key technologies underlying today's distributed information systems, including Web data management, information retrieval and data mining.

Content

Web Information Management: Semi-structured data - graph data model, web ontologies, schema integration

Information Search: Web search - vector space retrieval, inverted files, advanced retrieval models, word embeddings, web search

Big Data Analytics: Data mining - associations rules, clustering, classification, model selection; Crowd-sourcing; Recommender systems - collaborative filtering and content-based recommendation

Learning Prerequisites

Recommended courses

Introduction to Database Systems

Learning Outcomes

By the end of the course, the student must be able to:

• Characterize the main tasks performed by information systems, namely data, information and knowledge management

• Apply semi-structured data models, their representation through Web standards and algorithms for storing and processing semi-structured data

• Apply fundamental models and techniques of text retrieval and their use in Web search engines

• Apply main categories of data mining techniques, local rules, predictive and descriptive models, and master representative algorithms for each of the categories

• Apply collaborative information management models, like crowd-sourcing, recommender systems, social networks

Teaching methods

Ex cathedra + exercises

Assessment methods

25% Continuous evaluations with bonus system during the semester
75% Final written exam (180 min) during exam session

Supervision

Office hours	Yes
Assistants	Yes
Forum	Yes

Resources

Websites

- http://lsir.epfl.ch/teaching/current-courses/

Moodle Link

- http://moodle.epfl.ch/course/view.php?id=4051