CS-423  Distributed information systems

Aberer Karl

Cursus | Sem. | Type
--- | --- | ---
Biocomputing minor | H | Opt.
Civil & Environmental Engineering | | Opt.
Communication systems minor | H | Opt.
Computer science | MA1, MA3 | Opt.
Cybersecurity | MA1, MA3 | Opt.
Data Science | MA1, MA3 | Opt.
Digital Humanities | MA1, MA3 | Opt.
Electrical and Electronical Engineering | MA1, MA3 | Opt.
Energy Management and Sustainability | MA1, MA3 | Opt.
Environmental Sciences and Engineering | MA1, MA3 | Opt.
SC master EPFL | MA1, MA3 | Obl.

Summary
This course introduces the key concepts and algorithms from the areas of information retrieval, data mining and knowledge bases, which constitute the foundations of today's Web-based distributed information systems.

Content

Information Retrieval
1. Information Retrieval - Introduction
2. Text-Based Information Retrieval
3. Vector Space Retrieval
4. Inverted Files
5. Distributed Retrieval
6. Probabilistic Information Retrieval
7. Query Expansion
8. Latent Semantic Indexing
9. Word Embeddings
10. Link-Based Ranking

Data Mining
1. Data Mining – Introduction
2. Association Rule Mining
3. Clustering
4. Classification
5. Classification Methodology
6. Document Classification
7. Recommender Systems
8. Mining Social Graphs

Knowledge Bases
1. Semi-structured data
2. Semantic Web
3. RDF Resource Description Framework
4. Semantic Web Resources
5. Keyphrase extraction
6. Named entity recognition
7. Information extraction
8. Taxonomy Induction
9. Entity Disambiguation
10. Label Propagation
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 collaborative information management models, like crowd-sourcing, recommender systems, social networks
• Apply knowledge 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

Teaching methods
Ex cathedra + programming exercises (Python)

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