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. Probabilistic Information Retrieval
5. Query Expansion
6. Inverted Index
7. Distributed Retrieval
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. Mining Social Graphs
6. Classification Methodology
7. Document Classification
8. Recommender Systems

Knowledge Bases
1. Semi-structured data
2. Semantic Web
3. RDF Resource Description Framework
4. Semantic Web Resources
5. Information Extraction
6. Taxonomy Induction
7. Ontology Mapping

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 + programming exercises (Python)

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