



Course Specification Document

Title	Information Retrieval Systems
--------------	-------------------------------

Credits	2.5 ECTS
----------------	----------

Aims	This course aims to provide the student with various knowledge related to the concepts and basics of information retrieval and text and web mining, enabling him to use it on real data sets to obtain new knowledge, useful information and conclusions.
-------------	---

Intended learning outcomes

On successful completion of this course, the student will be able to:

- Understand the fundamentals of information retrieval.
- Identify the components of search engines and their operation.
- Understand clustering and classification methods.
- Familiarize himself with programming tools used to build information retrieval systems.
- Utilize programming tools in constructing information retrieval systems.

Syllabus

- **Fundamentals of information retrieval and logical search model:** Non-structured data, inverted indices, queries.
- **Building inverted indices:** Document analysis, tokenization, linguistic processing.
- **Sentence Retrieval:** Binary indices, positional indices, permutation indices, part-of-speech tagging, n-gram word sequences, similarity queries
- **Spelling error correction:** Query correction, document correction, index correction, isolated word correction, context-based correction.
- **Search results ranking:** TF-IDF metric, vector space model, similarity metrics, other criteria, results ranking acceleration.
- **Evaluation of information retrieval systems and search engines:** Evaluation criteria, precision, recall, relevance, major search engine metrics, result improvement, result presentation.
- **Web search:** Web searching, search engine operation, document privacy, user evaluation, commercial advertising models.
- **Web crawlers:** Basic operations, operational complexities, Mercator crawling ...