Information Retrieval

2025 - 2026

Is part of the next programmes:

- M0012004 Master of Computer Science: Software Engineering
- M0012005 Master of Computer Science: Data Science and Artificial Intelligence
- M0012006 Master of Computer Science: Computer Networks
- M0048004 Master of Computer Science: Software Engineering
- M0048005 Master of Computer Science: Data Science and Artificial Intelligence
- M0048006 Master of Computer Science: Computer Networks
- M0090004 Master of Teaching in Science and Technology: Computer Science
- U0001008 Courses open to exchange students in Sciences

Course Code:	2001WETINR
Study Domain:	Computer Science
Semester:	1E SEM
Contact Hours:	26

Credits:	6
Study Load (hours):	168
Contract Restrictions:	No contract restriction
Language of Instructions:	ENG
Lecturer(s):	T Toon Calders
Examperiod:	exam in the 1st semester

1. Prerequisites *

speaking and writing of:

- English

extra commentary:

This course will be taught in English only.

specific prerequisites for this course

Data structures: list, hash table, tree, map, graph.
Basic notions of probability: expected value, standard deviation, Bayes' theorem.
Programming skills: able to write small programs in a programming language like C++, Python, or Java.

2. Learning outcomes *

- Understand and use techniques for efficient keyword-based information retrieval, such as inverted indices, top-k query answering, resolving wildcard queries.

- Explain and use ranking techniques, including the vector-space model, the Boolean retrieval model, and language-based models.
- Understand the advantages and disadvantages of the different score models and be able to decide in which situation to use what model.
- Apply data preprocessing techniques to improve retrieval, such as stemming, lemmatization, stop-word removal, latent semantic indexing.
- Assess the importance of web-based resources based on link analysis such as pagerank and hubs-and-authorities.
- Efficiently implement high-dimensional search via latent-semantic indexing.
- Apply the retrieval techniques covered in the course to develop a retrieval engine that addresses an information retrieval need for a given data collection.

3. Course contents *

The course covers the foundations of current ranking techniques (which are often based on statistical models), index structures, and current implementation issues for the design of effective and scalable information-system architectures. The course covers a wide range of ranking principles, starting from the Boolean retrieval model, over to statistical ranking models such as TF-IDF, and on to probabilistic ranking techniques such as Okapi-BM25. The course will cover classical IR topics, including link-analysis methods such as PageRank, the user-specific personalization of queries, and relevance feedback.
The course topics are:

- Boolean retrieval
- Vector Space Model
- Language Based Model
- Boolean Independence Model
- Top-K querying and Index construction
- Index construction and compression
- Feedback-Expansion
- Evaluating IR
- Link analysis
- Dimensionality Reduction

- Neighbor search in high dimensional data
- Deep neural network approaches for information retrieval
- Retrieval Augmented Generation

4. International dimension *

- This course stimulates international and intercultural competences.
- Students use course materials in a foreign language.

5. Teaching method and planned learning activities

5.1 Used teaching methods *

Class contact teaching

- Lectures

Personal work

Assignments

- In group

Paper

- In group

Project

- In group

5.2 Planned learning activities and teaching methods

The course consist of 2h theory lectures per week, complemented with 3 assingments carried out in groups of 2 to 3 students. During the theory lectures the topics of the course are covered. The goal of the assignments is to deepen

knowledge about one or more of the course topics and will require gathering and processing additional information sources, and applying the learned retrieval techniques for a given set of documents. One of the assignments will include a presentation.

5.3 Facilities for working students *

Classroom activities

- no specific facilities

Individual work

- In group: individual alternative assignment possible

6. Assessment method and criteria *

6.1 Used assessment methods *

Examination

- Written examination without oral presentation
- - Closed book

Other assessment methods

- Project (no second assessment period)
- Presentation (no second assessment period)

6.2 Assessment criteria *

There will be a written exam and a project in small groups. Both components account for 50% of the grade. However, in order to obtain a passing grade, next to a passing grade in general, for both parts in isolation at least 40% of the passing mark needs to be obtained. There is hence only a limited form of compensation possible between the two components.

- **Closed book exam** (50% of grade): paper based exam regarding the thory of the course. A formularium will be offered to the students that can be used during

the exam.
- assignment(s) to be made in small groups (50% of grade)

Copied Code - Code created by Generative AI

For implementations supporting the project, students may copy code from the internet, but must do so explicit attribution of the source where it is copied from. Students may use generative AI tools for the code adopted in their project, but must do so with explicit acknowledgment, i.e. it must be clearly indicated which code was created by which tools. Students may be asked to elaborate on how they used generative AI.

Support from generative AI

Students may use generative AI tools as for writing reports, similar to initial search engines such as Google and for checking grammar and spelling. Students may be asked to elaborate on how they used generative AI.

7. Study material

7.1 Required reading *

- Book: “Introduction to Information Retrieval” by C. Manning et al. An electronic copy of this freely available book is made available on Blackboard.
- Course slides available
- Additional exercises and solutions
- Practice exams

All material will be offered in electronic form via Blackboard

7.2 Optional reading

The following study material can be studied voluntarily :

Optional material will be made available on Blackboard during the course.

8. Contact information *

toon.calders@uantwerpen.be

9. Tutoring

Students are expected to actively follow the lecture and exercises and to regularly and independently study the provided chapters of the background literature.