

## Task Sheet 1

### IR Basics and System Modeling

Work on the practical task is mandatory in order to be eligible to take part in the exam!  
The tasks are designed to be done in groups of two. Therefore you should find a partner before starting.

#### Task Description

1. You will find the Test module 'Practical Task 1 - Questions' on the Moodle page of the class. Open it and answer all questions.
2. Create a fundamental design of an implementation for an IR system. Use a UML class diagram for this and describe it briefly ( $\geq 1$  page) in writing. The system should have at least the following basic functionalities:
  - Creation of a document collection from a given text file (by splitting the file into individual documents by a given criteria).
  - Elimination of stop words and base/stem form reduction of terms in documents.
  - Searching the document collection using a single search term or several search terms (linked by Boolean junctors) and outputting the search results.
  - User interface for starting the individual functions (decomposition, stop word elimination, stem form reduction, search, ...).
  - Support of different retrieval models for the search (Boolean model, vector space model, ...).
  - Support of different implementations of a retrieval model (e.g. linear search, inverted list and signature method for Boolean model).
  - Ability to output a document textually after different processing steps (original document; after stop word elimination; after basic/root form reduction; title/filename only).
  - Storage of the documents (incl. the individual processing steps) on the hard disk and loading this data from the hard disk after restarting the program.
  - Automatic determination of the key figures Recall and Precision for predefined test queries.

Save your diagram in PDF format and upload the file to the module 'Practical Task 1 - Diagram upload' Moodle page.

**Fulfill both of these tasks until the deadline (May 05th, 2025, 11:59pm)!**

Please also refer to the introduction slides shown in the first exercise for organizational issues.

## Known Pitfalls

- Not using UML notation and/or violating the UML specifications. For reference, see e.g. [https://en.wikipedia.org/wiki/Class\\_diagram](https://en.wikipedia.org/wiki/Class_diagram) or <http://agilemodeling.com/artifacts/classDiagram.htm>
- Just creating the diagram without explanations or just a text without a diagram. If parts of the task are missing (even small parts), we consider the submission failed.
- Submitting in other formats than PDF. We do not want to convert your solution into something readable for us so please stick to the requirement.
- Unclean or unreadable drawings, usage of ASCII art or similarly obscure methods. Please prepare a clean and proper diagram, for example with MS Visio, Inkscape, Dia, draw.io, etc.