

Agenda

1. Organization
2. Project Info

Organization

Communication

- Slides, Announcements & Materials will be available at
`temir.org/teaching/information-retrieval-ss22/information-retrieval-ss22.html`

- Main Communication channel is email
 - Official announcements via Mail (check your student mails regularly!)

- Lab Project group organization via moodle:
`https://moodle2.uni-leipzig.de/course/view.php?id=38544`
 - Optional Discord for group communication ¹

¹server “*webis-lectures*”, mail us for access code

Exam

- Exam at the end of the term
- Lab Project (optional, for bonus points in exam), Deadline: **12th of July**
- Grade: passed exam grade + bonus (0 / 0.5 / 1.0) for lab project (not passed / passed / passed well)
 - E.g. exam grade is 2.3, but lab project is passed well, then the final grade is $2.3 - 1.0 = 1.3$
 - E.g. exam grade is 1.0, but lab project is not passed, then the final grade will stay $1.0 - 0 = 1.0$
- Biweekly programming exercises to build a basic search engine

Organization

Lab Sessions

- Exercise contents:
 - Building and evaluating an basic information retrieval system
 - Data handling, indexing, implementation of suitable retrieval models, evaluation of search quality
 - Biweekly exercises throughout the semester

- Project contents:
 - Building and evaluating an information retrieval system for a specific domain
 - Related work search, data handling, indexing, selection and implementation of suitable retrieval models, evaluation of search quality
 - Write a research paper
 - How to do literature research
 - Introduction to scientific writing
 - Submission of a written report and documented source code

Lab Project

Milestones

- ❑ **Task Definition [~ 1 week]**
Find a domain and get corresponding dataset for your domain-specific retrieval system.
- ❑ **Literature Research [~ 1 week]**
Find existing research relevant to the task (similar domains, specific retrieval systems,...).
- ❑ **Data Analysis [~ 1 week]**
Take a closer look at the data, use descriptive statistics, identify interesting patterns.
- ❑ **Technology Stack [~ 1 week]**
Decide upon the software libraries you are going to use.
- ❑ **Vertical Prototype [~ 2 weeks]**
Build a working prototype with a basic retrieval model.
- ❑ **Refined Prototype [~ 2 weeks]**
Build a prototype that uses an advanced/refined retrieval model.
- ❑ **Evaluation [~ 3 weeks]**
Evaluate the results of your retrieval models.
- ❑ **Documentation [~ 1 week]**
Write a README, including deployment instructions.
- ❑ **Report [~ 1 week]**
Write the paper.

Project Groups

- You can work in groups of 4 people
- Each group will receive:
 - A unique group name
 - A Discord channel (optional, request as group per email)

Project Hand-Ins

- At the end of the term you hand in
 - A written report in the form of a scientific paper in [ECIR Demonstration paper style](#) (4 pages + references) [[Latex packages/Word templates](#)]
 - Your code + README with instructions and full history in a git repository
 - Deadline: **12th of July 2022**

Project Datasets

- Dataset sources(selection):
 - Huggingface
 - Tensorflow Datasets
 - Allen Institute
 - Kaggle

- You can choose dataset from other sources as well.

- Datasets should contain more than 10000 documents!