Agenda

1. Organization

2. Lab Project

Organization
Communication

- ❏ Slides, Announcements & Materials will be available at
  `temir.org/teaching/information-retrieval-ws23/`
  `information-retrieval-ws23.html`

- ❏ Communication channels are Discord and email
  - – Official announcements via Mail (check your student mails regularly!)
  - – Discord for Q&A and group communication

Organization
Lab Sessions

- Lab sessions throughout the semester

    - 13:15 - 14:00 timeslot → lab sessions / Q&A

    - 14:30 - 16:45 timeslot is intended as placeholder for project work
      no content from our side planned, but room can be used

- Not every week will have a lab session taking place (see course page)

- We are available for questions outside the scheduled times via Discord

    - Feel free to use the discussion board!

Lab Project
Overview

- ❑ Goal
  - Building and evaluating an information retrieval system for web search
  - Related work search, data handling, indexing, selection and implementation of suitable retrieval models, evaluation of search quality
  - Submission of a written report and software on the TIRA platform

Lab Project
Overview

❏ Goal

  – Building and evaluating an information retrieval system for web search
  – Related work search, data handling, indexing, selection and implementation of suitable retrieval models, evaluation of search quality
  – Submission of a written report and software on the TIRA platform

❏ Milestones

  – Lab is organized around 4 milestones throughout the semester
  – At each milestone, we expect each group to hand in a deliverable

Lab Project
LongEval - Task & Data

❑ Task: Web Search

  – Special focus on robustness of systems to changes in queries over time

  – Evolving document and query sets to reflect changes

❑ Data: LongEval web search collection

  – Provided by a commercial search engine (Qwant)

  – Corpus of pages, queries, user interaction logs

  – TIRA as an evaluation platform & leaderboard

❑ Additional evaluation data will be collected in the lab itself

Lab Project

Create topics for the supplied search task and data collection.

- ❑ A topic is a description of a users' information need
    - – A text entered into the IR system as query
    - – A description of the underlying information need
    - – A narrative describing what is relevant to the query

ID	1
Text	retrieval system improving effectiveness
Description	What papers focus on improving the effectiveness of a retrieval system?
Narrative	Relevant papers include research on what makes a retrieval system effective and what improves the effectiveness of a retrieval system. Papers that focus on improving something else or improving the effectiveness of a system that is not a retrieval system are not relevant.

- ❑ Due Date: 23.10.2023
- ❑ Deliverable: Valid topic file (XML, see course page for example)

Lab Project
Milestone II: Relevance Assessments

Assess the relevance of documents retrieved for your topic

- ❑ Given your topics, we will supply a set of retrieved documents
    - – Retrieval is done by pooling several baseline systems
    - – For each topic and system, documents are retrieved and pooled
- ❑ Annotate these documents w.r.t. their relevance to the topic

Query ID	1
Text	retrieval system improving effectiveness
Description	What papers focus on improving the effectiveness of a retrieval system?
Narrative	Relevant papers include research on what makes a retrieval system [...]
Document ID	2005.ipm_journal-ir0anthology0volumeA41A1.7
Document	In this paper we will present a language-independent probabilistic model [...]
Relevancy	1

- ❑ Due Date: 06.11.2023
- ❑ Deliverable: Completed batch on the annotation platform

Lab Project
Milestone III: IR System

Build and evaluate your own IR system using
your topics and relevance assessments.

❑ Implement your IR system

– Training data will be supplied; compute resources available
– Final system should be deployed to the TIRA platform

❑ Evaluate your IR system

– The previously annotated topics are used for testing
– Testing is carried out using the TIRA platform

❑ Shortly reflect on the assignment in a written report

❑ Due Date: 27.11.2023
❑ Deliverable: Short reflection (approx. half page), TIRA submission

Lab Project
Milestone IV: Research Assignment

Derive and test a research question with your IR system.

- Come up with a hypothesis that can be tested with your system

- Conduct statistical analysis to address the hypothesis
  - Hyp. should separate two configurations (modified c., control c.)
  - Hyp. should motivate a distribution over both groups to test against

- Summarize your system and findings in a short report

"System effectiveness will be higher when boosting
the rank of documents under 1000 characters."

- Due Date: 08.01.2024
- Deliverable: Report (max. 2 pages)

Questionnaire

Raise your hand if you...

- ❏ ... have used Python before?

- ❏ ... have preexisting knowledge in ML?

- ❏ ... have worked on data analysis before?

- ❏ ... have done data annotation before?

- ❏ ... have done scientific writing before?

- ❏ ... have worked in a cluster environment before?

- ❏ ... have used Docker before?

Project Groups

❑ You can work in groups of up to 4 people

❑ Already know whom you want to work with?
 – Yes! → see below
 – No! → raise your hand to be assigned to a group

❑ Send us a mail with names of all group members until next week!

❑ Each group will then receive:
 – A unique group name
 – A Discord channel
 – A TIRA account

What makes a good topic?