

Information Retrieval

Exercise – Winter term 2026/2027

`klara.gutekunst@uni-kassel.de`

Agenda

1. Organization
2. Lab Projects
3. Next Steps

Organization

Communication

- ❑ Slides, Announcements & Materials on the course website [\[temir.org\]](https://temir.org)
- ❑ Questions and communication via Discord and email [\[discord.gg\]](https://discord.gg/temir)
 - Category “Group Channels WiSe 2026”, channel “uk-ir-ws26”
 - Please use your full name as the server nickname



Organization

Lab / Exercise

- ❑ Joint collaborative exercises with Friedrich Schiller University Jena
- ❑ Lab project as group work (3 people per group)
- ❑ Lab/ theoretical exercise sessions throughout the semester
 - **Tuesday, 14:00 – 15:30, Room 1114**
 - Lab
 - Each group shortly presents their current progress
 - Discuss open questions and next steps
 - Exercise
 - Exercises for lecture contents
 - Find (additional) notebooks/ tutorials on [teaching-ir-with-shared-tasks](#)
- ❑ We are also available for questions outside the scheduled times via [Discord](#)
 - Feel free to use the [discussion board](#) to discuss across universities!

Lab Project

Overview

- ❑ Goal
 - Build & evaluate an information retrieval system for web search
 - Search for related work, learn data handling, indexing, select & implement suitable retrieval models, evaluate search effectiveness
 - Identify (and solve) retrieval errors
 - Submit written report and software on the [TIRA](#) platform

Lab Project

Overview

□ Goal

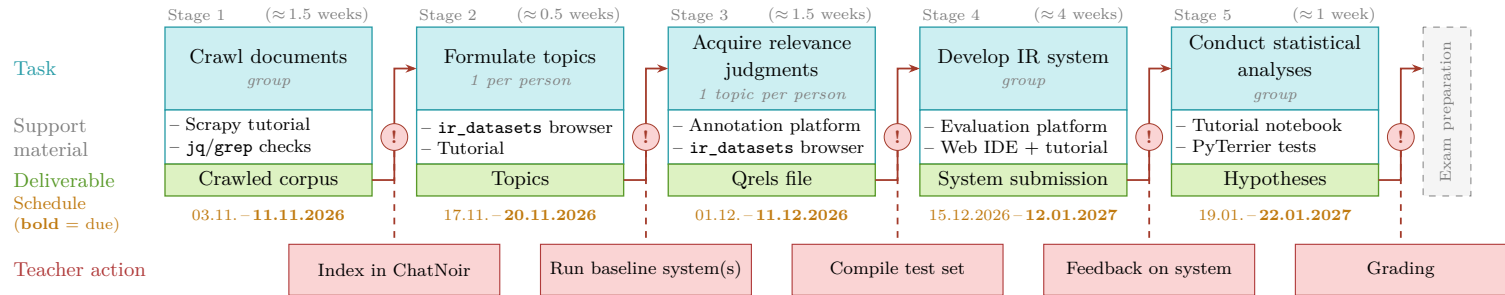
- Build & evaluate an information retrieval system for web search
- Search for related work, learn data handling, indexing, select & implement suitable retrieval models, evaluate search effectiveness
- Identify (and solve) retrieval errors
- Submit written report and software on the [TIRA](#) platform

□ Resources

- [TIRA](#) as an evaluation platform & leaderboard [[website](#), [paper](#)]
- `ir_datasets` browser [[website](#)]
- Experimentation with [PyTerrier](#) [[paper](#)]
- Tutorials [[IR](#), [crawling](#)]

Lab Project

Overview



Overview of the course stages.

- ❑ Lab organized in 5 stages throughout the semester
- ❑ Each stage is finished by a deliverable
→ Groups hand in their intermediate work
- ❑ Deliverables are exchanged with other universities

Lab Project

Task & Data

- ❑ Task: Web search
 - Aka: What you search everyday...
 - Very broad set of topics
- ❑ Crawled data
 - [Scrapy tutorial](#)
 - Collect data with upcoming tasks in mind
- ❑ Further evaluation data will be collected in the lab itself

Lab Project

Stage 1: Crawl data

- ❑ TODO
- ❑ **Due Date:** Wednesday, 11.11.2026, 23:59
- ❑ **Deliverable:** data corpus submitted [online](#)

Lab Project

Stage 2: Formulate topics

Create topics for the supplied search task and data collection.

- ❑ A topic is a description of a user's information need [[lecture 2025: Topics](#)]
 - A *text* that could be entered into the search system as a query
 - A precise *description* of the underlying information need
 - A *narrative* describing what is relevant to the topic (and what not)
- ❑ **Due Date:** Friday, 20.11.2026, 23:59
- ❑ **Deliverable:** valid topics submitted online

Topic 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.

Lab Project

Stage 3: Acquire relevance judgments

Judge the relevance of documents retrieved for your topic.

- ❑ Given your topics, we will retrieve documents
 - Top-10 for each topic and baseline system [[lecture 2024: Pooling](#)]
 - Some manually added documents
- ❑ Annotate these documents w.r.t. relevance to the topic (read the *narrative*)
- ❑ **Due Date:** Friday, 11.12.2026, 23:59
- ❑ **Deliverable:** Relevance judgments finished on [Doccano](#)

Topic ID	1
Query	retrieval system improving effectiveness
Topic desc.	What papers focus on improving the effectiveness of a retrieval system?
Topic narrative	Relevant papers include research on what makes a retrieval system [...]
Document ID	2005.ipm_journal-ir0anthology0volumeA41A1.7
Document text	In this paper we will present a language-independent probabilistic [...]
Relevance	1 (relevant)

Lab Project

Stage 4: Develop information retrieval 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 deployed to the **TIRA** platform
- ❑ Evaluate your IR system
 - Topics from **Open Web Index** and **TREC** are used for validation
 - Tests carried out on the **TIRA** platform
- ❑ Summarize your system and findings in a preliminary report
- ❑ **Due Date:** Tuesday, 12.01.2027, 23:59
- ❑ **Deliverable:** Preliminary report (max. 2 pages), **TIRA** submission

Lab Project

Stage 5: Conduct statistical analyses

Derive and test hypotheses with your previously constructed IR system.

- ❑ Look at types of retrieval errors
- ❑ Observe your system's behavior compared to baselines
- ❑ Formulate 2 hypotheses
- ❑ Hypotheses must be testable (more on that later)
- ❑ **Due Date:** Friday, 22.01.2027, 23:59
- ❑ **Deliverable:** Report (max. 2 pages)

Questionnaire

Raise your hand if you...

- ☐ ... have used Python before?
- ☐ ... have previous knowledge in ML?
- ☐ ... have worked on data analysis before?
- ☐ ... have done data annotation before?
- ☐ ... have done scientific writing before?
- ☐ ... have done statistical analyses before?
- ☐ ... have used Docker before?
- ☐ ... would like to participate in an international shared task?

Next Steps

Project Groups

- ❑ Form groups of 2-3 people
- ❑ Each group will receive a **TIRA** account
- ❑ Each group should have a unique group name (be creative!)
- ❑ **Email me** the names of all group members, the name of your group, and Discord handles until Friday, 06.11.2026, 23:59

Next Steps

Assignment

- ❑ Exercise sheet on the website [\[temir.org\]](https://temir.org)

- ❑ Crawl & submit your own crawled data [\[form\]](#)
 - One crawl per team member
 - Group submission
 - **Deliverable:** crawled data submitted [online](#)
 - **Due Date:** Wednesday, 11.11.2026, 23:59