Agenda

1. Organization

2. Project Info

3. Project groups

Organization
Communication

❑ Slides, Announcements & Materials will be available at
`temir.org/teaching/information-retrieval-ws20/`
`information-retrieval-ws20.html`

❑ Online sessions will take place on BigBlueButton:
`https://conf.informatik.uni-leipzig.de/`

❑ Communication channels are Discord and email
  – Official announcements via Mail (check your student mails regularly!)
  – Discord for Q&A and group communication [1]

---

[1]server *"irlecture"*, see mail for access code

Organization
Lab Sessions

❑ Lab contents:

– building and evaluating an information system for a specific domain

– related work search, data handling, indexing, selection and implementation of suitable retrieval models, evaluation of search quality

– submission of a written report and documented source code

❑ Two classes in addition to the lecture:

– Introduction to Scientific Writing

– Introduction to Docker and Tira

❑ Semi-regular Q&A sessions throughout the semester

❑ We are available via Discord and BigBlueButton during the lab time

Lab Project

❏ Projects organized as part of a Shared Task

- IR problem that is publicly posted and research groups can compete on
- you will work with the same resources as leading researchers in the field
- Conference participation is possible with the lab results

❏ Different projects to choose from:

- Touché - Conversational Argument Retrieval
- Touché - Comparative Argument Retrieval
- ChatonIR (Scientific Document Retrieval, 2 groups max)

Lab Project

> *"Build a search engine that retrieves strong arguments for and against
> a given controversial issue."*

❑ Objective: system for conversational argument retrieval

❑ Deliverables: build an argument retrieval system, deploy it on the Tira platform
and hand in a written summary of your work

Lab Project

❑ What should an conversational argument retrieval system do?

   – Support users who search for arguments to be used in conversations

   – Retrieve "strong" pro/con arguments on a given topic

❑ Example Topic:

Query	*Is climate change real?*
Information Need	*User requires arguments supporting the claim that climate change is in fact real. Relevant arguments will support the given stance that climate change is real or attack a hoax side's argument.*

Lab Project
Touché Task 1 – Resources

❏ Data

– Source dataset consisting of more than 300'000 arguments
[Ajjour et. al., 2019]

– Training & evaluation datasets for argument quality and argument relevance

– Query dataset

❏ Material

– list of related papers as starting point for your own literature research

❏ Evaluation platform

– deploy your solution on [tira.io] for automated and reproducible evaluation

❏ Example

– the [args.me] search engine and its source code

Lab Project

> *"Given a comparative question, retrieve and rank documents that help to answer the comparative question."*

❑ Objective: system for comparative argument retrieval

❑ Deliverables: build an argument retrieval system, deploy it on the Tira platform and hand in a written summary of your work

Lab Project
Touché Task 2 – Details

❑ What should an comparative argument retrieval system do?

  – Support users facing some choice problem

  – Retrieve Documents arguing for/against certain aspects of a given topic

❑ Example Topic:

Query	*Which is better, laptop or desktop?*
Information Need	*Find arguments that show in what personal situation what kind of machine is preferable. This can range from situations like frequent traveling where a mobile device is to be favored to situations where a stationary desktop PC is preferable.*

Lab Project
Touché Task 2 – Resources

❑ Data

    – Source dataset access via Chatnoir API

    – Query dataset & Relevance jugdements

❑ Material

    – list of related papers as starting point for your own literature research

❑ Evaluation platform

    – deploy your solution on [tira.io] for automated and reproducible evaluation

Lab Project
ChatonIR – Overview

> *"Build a search engine that retrieves relevant information on IR and NLP-related topics from a collection of scientific papers."*

- **Objective**:
  - index a corpus of scientific publications and their meta information
  - conceptualize an idea for a document retrieval system (classic search engine, conversational search, interactive search, ...)
  - implement your chosen approach and evaluate the solution
  - develop a process/tool that will help researches identify relevant publications in the IR/NLP domain

- **Deliverables**: conceptualize and build a document retrieval system, deploy it as a demo and hand in a written summary of your work

Lab Project
Milestones

- **Literature Research**
  Find existing research relevant to the task.

- **Data Analysis**
  Take a closer look at the data, use descriptive statistics, identify interesting patterns.

- **Technology Stack**
  Decide upon the software libraries you are going to use.

- **Vertical Prototype**
  Build a working prototype with a basic retrieval model.

- **Refined Prototype**
  Build a prototype that uses an advanced/refined retrieval model.

- **Deployment on Tira**
  Build a containerized version of your software and deploy it on the Tira platform.

- **Evaluation**
  Evaluate the results of your retrieval models.

- **Documentation**
  Write a README, including deployment instructions.

- **Report**
  Write the final report.

Project Groups

❑ you can work in groups of up to 4 people

❑ there will be a dedicated channel on the Discord server for finding groups

❑ if you assemble a group yourself, write a mail listing all group members (name & mail address) to us

❑ we will assign groups for everyone else

❑ each group will receive:
  – a unique group name
  – a Discord channel
  – a Tira account and VM