

Agenda

1. Organization
2. Project Info
3. Project groups

Organization

Communication

- ❑ Slides, Announcements & Materials will be available at `temir.org/teaching/information-retrieval-ws21/information-retrieval-ws21.html`
- ❑ Communication channels are Discord and email
 - Official announcements via Mail (check your student mails regularly!)
 - Discord for Q&A and group communication ¹

¹server “*irlecture*”, see mail for access code

Exam

- ❑ Lab Project will be your exam
- ❑ Final Talk: 20%, Paper (+ additional resources): 80%
- ❑ Deadline Paper: 28th of February 2022

Organization

Lab Sessions

- ❑ Lab 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
 - Submission of a written report and documented source code

- ❑ Classes in addition to the lecture:
 - How to do literature research
 - How to do oral presentations
 - Introduction to Scientific Writing
 - Introduction to Docker and Tira

- ❑ Q&A sessions throughout the semester

- ❑ We are available via Discord

Lab Project

- ❑ Projects organized as part of a Shared Task
 - Shared Task = {input dataset, scientific problem, evaluation measures}
 - Scientific problem (IR) 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 Task 1 - Argument Retrieval for Controversial Questions
 - Touché Task 3 - Image Retrieval for Arguments

Lab Project

Touché Task 1 – Overview

“Build a search engine that retrieves argumentative sentences for a given controversial topic.”

- ❑ **Objective:** System for retrieval of argumentative sentences from a collection of arguments on controversial topics
- ❑ **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 1 – Details

- ❑ What should an argument retrieval system do?
 - Support users to make informed decisions by retrieving relevant arguments for a controversial topic with a specific stance (pro or con)
 - In this task instead of retrieving complete arguments, we will retrieve a pair of argumentative sentences
 - An argumentative sentence can be: a claim, a premise, or a conclusion
 - Sentences in this pair may come from two different arguments
- ❑ Example Topic:

Query *Should teachers get tenure?*

Description *A user has heard that some countries do give teachers tenure and others don't. Interested in the reasoning for or against tenure, the user searches for positive and negative arguments. [...]*

Narrative *Highly relevant statements clearly focus on tenure for teachers in schools or universities. Relevant statements consider tenure more generally, not specifically for teachers, or [...]*

Lab Project

Touché Task 1 – Resources

- ❑ Data
 - Args.me corpus with more than 300,000 arguments, each split into its constituent sentences [Ajjour et al., 2019]
 - Document-level relevance judgments and quality judgments
 - Query dataset (Topics)
- ❑ 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

Touché Task 3 – Overview

“Given a stance on some controversial topic and a collection of argumentative documents with images, the task is to retrieve and rank images that show support or opposition to that stance.”

- ❑ **Objective:** System for retrieving images for arguments (e.g., based on the web pages the images are on)
- ❑ **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 3 – Details

What should an image retrieval system for arguments do?

- Support users in getting an overview of opinions on controversial topics
- Retrieve images (from web pages) showing support or opposition for a given topic

This task uses the same topics as Task 1

👍 cartoons covers photos posters memes protesters quotes statistics

The grid contains 18 images with the following captions:

- Can Nuclear power save us? – NUES web.northeastern.edu
- Californians for Green Nuclear Power cgnp.org
- Anti-nuclear energy protesters take ... alamy.com
- Business Cartoon | TOONPOOL toonpool.com
- Nuclear Energy Grows: Global Poll ... globescan.com
- marketers oppose clean energy subsidies ... redgreenandblue.org
- Anti-Nuclear Opposition | comparenuclear comparenuclear.wordpress.com
- No to nuclear: Japan wants reactors ... aljazeera.com
- Orano U.S. on | Nuclear energy, ... pinterest.dk
- zero emissions must use nuclear energy policyoptions.ipp.org
- www.laka.org | anti-nuclea... laka.org
- 26 Famous Quotes About Nuclear Energy getintonuclear.com
- Life-cycle greenhouse gas emissions by electricity generation method. Method of Generation. letstalkscience.ca
- 3 Reasons Why Nuclear Energy Is Awe... youtube.com
- Anti-nuclear | IF WORL... ifworlddesignguide.com
- Nuclear Power Is Not L... thenewpress.com
- One million people against nu... facebook.com

Lab Project

Touché Task 3 – Resources

- ❑ Data
 - Documents and images retrieved by Google for the topics
 - Relevance annotations for ~1000 images [Kiesel et al., 2021]
 - Query dataset (Topics)
- ❑ 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

Milestones

- ❑ **Literature Research [~ 2 weeks]**
Find existing research relevant to the task.
- ❑ **Data Analysis [~ 2 weeks]**
Take a closer look at the data, use descriptive statistics, identify interesting patterns.
- ❑ **Technology Stack [~ 2 weeks]**
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.
- ❑ **Deployment on Tira [~ 2 weeks]**
Build a containerized version of your software and deploy it on the Tira platform.
- ❑ **Evaluation [~ 3 weeks]**
Evaluate the results of your retrieval models.
- ❑ **Documentation [~ 2 weeks]**
Write a README, including deployment instructions.
- ❑ **Report [~ 3 weeks]**
Write the final report.

Project Groups

- ❑ You can work in groups of up to 4 people
- ❑ Each group will receive:
 - A unique group name
 - A Discord channel
 - A Tira account and VM