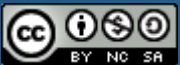


Information Retrieval Project (IE 691)

Dr. Pedro Ortiz Suarez
Prof. Dr. Simone Paolo Ponzetto

Data and Web Science Group
Fakultät für Wirtschaftsinformatik und Wirtschaftsmathematik
Universität Mannheim



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

Team project

2

- Tentative schedule
 - Topics published: Oct 5
 - Topics selected and confirmed: Oct 12
- **Project coaching:**
 - Two (**optional**) sessions: Nov 2, Nov 30
 - We check the progress of your projects
 - Help you resolve dilemmas and problems you might be facing
- **Project presentations: Dec 7**
 - Present what you did: methods/models, implementation, evaluation
 - 10-15 minutes per team
 - **All team members should present and clearly state what their contribution was**
 - **We will ask questions to all team members**

Team Project

3

- **Purpose:** „hands-on” experience **implementing** and **evaluating** information retrieval model(s) and performing IR tasks
 - Best way to understand something is to (try to) implement it
- Other goals:
 - Experiencing **teamwork**
 - Learning how to **coherently** and **concisely report** on the results of the work
 - Exercising how to **clearly present** results of your work

Topics

4

1. TREC 2022 Fair Ranking Track: Provide a fair exposure to articles that are associated to particular protected attributes.
 2. LiLAS @ CLEF2021: Ad-hoc retrieval of scientific documents and Research dataset recommendation
 3. ARQMath-3 @ CLEF2022: Effective question answering systems for math
- You can propose your own topic as well
 - Has to undoubtedly be an **IR topic** and of **similar scope/difficulty** as the five above

Topic 1: TREC 2022 Fair Ranking Track

5

- Participants will be provided with a corpus of documents and a set of queries in the form of a short list of search terms that represent a **WikiProject**. Each document in the corpus is relevant to zero to many WikiProjects and associated with potentially many **fairness categories**.
- There are **two tasks** in the 2022 Fair Ranking Track. In each of the tasks, for a given query, participants are to produce document rankings that are:
 1. Relevant to a particular WikiProject.
 2. Provide a **fair exposure** to articles that are associated to particular protected attributes.
- Task 1: WikiProject Coordinators
- Task 2: Wikipedia Editors

Topic 1: TREC 2022 Fair Ranking Track

6

- Task 1: WikiProject Coordinators
 - Produce a ranked [list of articles needing work](#) that editors can then consult when looking for work to do.
 - Relevance to a [WikiProject topic](#).
 - [Fairness](#) with respect to the exposure of different fairness categories
- Task 2: Wikipedia Editors
 - Focused on individual Wikipedia editors. Different editors may receive different rankings for the same query.
 - [Relevance to a WikiProject](#).
 - Work needed on the article ([articles needing more work](#) preferred).
 - [Fairness](#) with respect to the exposure of different fairness categories
- Datasets:
 - <https://data.boisestate.edu/library/Ekstrand/TRECFairRanking/>
 - https://fair-trec.github.io/docs/Fair_Ranking_2022_Participant_Instructions.pdf

Topic 2: LiLAS @ CLEF2021

7

- The Living Labs for Academic Search (LiLAS) lab aims to strengthen the concept of user-centric living labs for the **domain of academic search**.
- Finding the most relevant publications to a query remains a challenge in scholarly Information Retrieval systems, even more in **multi-lingual** and **cross-domain** environments.
- **Task:** Given a query and a 100-document candidate list, participants will **rank documents** from most to least relevant with regards to the query
- Participants are asked to define and implement their ranking approach for a **multi-lingual candidate documents** list.
- A good ranking should present the most relevant documents regarding a query on top of the result set. Multiple languages can be used to pose the query (e.g. English, German, French). **Regardless of the language** used on the query, the **retrieval can include candidate documents in other languages**.
- <https://clef-lilas.github.io/tasks/>

Topic 3: ARQMath-3 @ CLEF2022

8

- Effective **question answering systems for math** would be valuable for math Community Question Answering
- The ARQMath lab was established to support research into **retrieval models that incorporate mathematical notation**.
- **Task 1: Answer Retrieval:**
 - Participants are presented with a question posted on MSE after 2018, and are asked to return a ranked list of up to 1,000 answers from prior years (2010–2018).
- **Task 2: Formula Retrieval:**
 - Given a single **query formula**, **similar formulae** in a collection were to be returned.
 - However in ARQMath, **relevance is decided by context** and systems **receive credit** for finding **visually distinct formulae**.
- <https://www.cs.rit.edu/~dpri/ARQMath/>

Organization

9

- Form groups of **5 students**
 - Students who cannot find a group should
 - post it on ILIAS forum
 - email us **if cannot find** by **Wednesday, Oct 12**
 - **All students must contribute** to the project
 - Ideally, **equally** (We **will ask** about individual contributions)
- Each group is allowed to pick a topic **they like the most**
 - Theoretically, all groups **could** pick the **same topic**
 - **More competition**, we will directly **compare your results** when evaluating
- Topic selection and team forming
 - **Deadline: Wednesday, Oct 12 (23:59)**
 - **Post under ILIAS project submission:**
 - Student names and IDs (Matrikelnummer)
 - Selected topic
 - **We will form teams from students without a team**

Participation

10

- **Important:**
 - If you're thinking about **dropping this course**, now would be the **last time** to do it
 - If so, **please inform us as soon as possible**
 - If you couldn't find fellow students for to form a team, send us an individual email by **October 12** stating that you don't have a team
 - Dropping at a later point is **not allowed**:
 - (1) not fair to your colleagues in the team, who will have to do your part of the work too
 - (2) for motivated students on the waiting list, whose spot in the course you'll have taken and wasted