Information Retrieval Lab

Welcome to week 10 of the Information Retrieval class 2023

Agenda

1. Organization

2. Project Info: Milestone 3

Organization

SHARKI SURVEY

Please fill in the survey for the SharKI project about this lab

- Deutsch:
  https://umfrage.uni-leipzig.de/index.php/869486?lang=de
- English:
  https://umfrage.uni-leipzig.de/index.php/869486?lang=en

Organization

Deadline Milestone 2 is TODAY, 06.06.23! (23:59 h)

Project Info: Milestone 3 – Methods II

Overall Goals

- You will produce a more effective retrieval system than that of Milestone 2. Your system should be effective for topics it has not seen.
- We will provide a leaderboard for you to compete to develop the most effective retrieval system.
  - We encourage you to submit multiple runs/approaches!

Project Info: Milestone 3 – Methods II

Sidenote - Your new approach

- How does your new approach defer from a baseline system?
  - A baseline system is typically a general approach to retrieval.
  - Your approach should implement a whole retrieval pipeline, that combines multiple (standard) retrieval systems.
  - Your approach should be targeted to your search domain and dataset.
- You need to develop something of your own that is more effective than the baseline system.

Project Info: Milestone 2 – Methods II

What to do

- Develop a new retrieval approach that is more effective than your baseline you deployed in Milestone 2.
- Do not make changes to your dataset!
- There will be two test datasets in tira to run your systems on: 1) a version of your dataset, and 2) the tutors dataset for multi field retrieval. These test datasets contain all topics and will be the basis for the leaderboard.
- There are three example notebooks on github: `https://github.com/mam10eks/ir-lab-sose-2023/tree/main/milestone-03`
  - These notebooks implement three different ideas on how a custom approach could look like.
  - The notebooks are only examples, please upload your own retrieval pipelines to tira, you can find more in the pyterrier documentation: `https://pyterrier.readthedocs.io/en/latest/pipeline_examples.html` (Note: You do not have enough data for learning to rank)

What to do

- You can reuse those notebooks in git, but overwrite the cells with the retrieval pipeline with your custom approach.
- When implementing a retrieval pipeline, we expect you to document the code cells with your approach with an explanation of why you chose it and what it does. Do not copy-paste explanations from the pyterrier notbook but understand the methods and write it down in your own words.
- Don’t forget the reflection cell at the end of EACH notebook (we encourage you to submit multiple methodically different approaches)

Project Info: Milestone 3 – Methods

What to hand in

- Minimum: Upload at least one docker image containing a jupyter notebook with the implementation of your new retrieval approach that produces a run file to TIRA (EXACTLY as you did for Milestone 2 but with a new approach).
- Better: Implement methodically different retrieval approaches. The more different your approaches are, the better!
- Don’t forget the explanation cells explaining your intent for each retrieval approach and how it works and the reflection cell at the bottom of each notebook.

Project Info: Milestone 3 – Methods

Grading

- Your grade for the module will be the exam grade (of a passed exam) + a bonus
- Bonus: Groups that submitted ALL three milestones in time will be applicable for the bonus.
  - Best Bonus: Exam grade - 1.0 grade points (e.g. Exam grade == 2.3, final grade == 1.3)
  - Normal Bonus: Exam grade - 1/3 grade points (e.g. exam grade == 2.3, final grade == 2.0)
  - No Bonus: Exam grade - 0 grade points (e.g. exam grade == 2.3, final grade == 2.3)
- No bonus will go to groups that did not submit all milestones
- Normal bonus will go to all groups that submitted all notebooks with what was required
- Best bonus will go to groups that submitted all notebooks and did more than what was required

Project Info: Milestone 3 – Methods II

When to hand in

DEADLINE FOR MILESTONE 3 is the 27.06.23

Project Info: Milestone 3 – Methods II

When to meet again

See you in TWO weeks, on the 20th of June