Information Retrieval Lab

Welcome to week 5 of the Information Retrieval class 2023

Agenda

1. Organization

2. Project Info: Milestone 2

Organization

Deadline Milestone 1 is THURSDAY, 04.05.23! (23:59 h) (Note: Yes, it has changed. You should have received an email about that. If not, notify us!)

Project Info: Milestone 2 – Methods

Overall Goals

- You will receive a ranked list of documents (run file) from us based on your datasets that you submitted in Milestone 1
- Create relevance assessments for the documents from Milestone 1
- Create a baseline information retrieval system that produces a run file using the topics you have created to assess the relevance of your information retrieval system
- Evaluate your baseline information retrieval system

Project Info: Milestone 2 – Methods

Sidenote

- What are relevance assessments?
- What is a **baseline** information retrieval system?
- What is a run file?

Project Info: Milestone 2 – Methods

Sidenote

- What are relevance assessments?
  Relevance assessments are (mostly human) judgements, that indicate if a retrieved document is relevant to a specific search query. Binary relevance assessments only judge a document in terms of *relevant* or *irrelevant*.
- What is a **baseline** information retrieval system?
- What is a run file?

Project Info: Milestone 2 – Methods

Sidenote

- What are relevance assessments?
  Relevance assessments are (human) judgements, that indicate if a retrieved document is relevant to a specific search query. Binary relevance assessments only judge a document in terms of *relevant* or *irrelevant*.
- What is a **baseline** information retrieval system?
  A baseline information retrieval system is a simple and widely spread retrieval method applied to a dataset. It retrieves documents using known methods (e.g. BM25).
- What is a run file?

Project Info: Milestone 2 – Methods

Sidenote

- What are relevance assessments?
  Relevance assessments are (mostly human) judgements, that indicate if a retrieved document is relevant to a specific search query. Binary relevance assessments only judge a document in terms of *relevant* or *irrelevant*.
- What is a **baseline** information retrieval system?
  A baseline information retrieval system is a simple and widely spread retrieval method applied to a dataset. It retrieves documents using known methods (e.g. BM25).
- What is a run file?
  Your TIRA submission from milestone 1 produced a run file on tira. This file contains a ranked list of documents for specific topics/queries.

Project Info: Milestone 2 – Methods

What to do

- Create binary relevance assessments (qrels) of the output from Milestone 1 (run file). The qrels file must follow the standard TREC style:

```
qid 0 docno relevance
```

  `qid`: query number, `0`: literal 0, `docno`: id of a document in your collection, and `relevance`: how relevant is docno for qid.
  Include your relevance assessments into the dataset you created in Milestone 1 and register it into ir_datasets as before.

- Create a baseline retrieval system that will produce a run file in the same format as the output of Milestone 1, that is standard TREC run file format:

```
qid Q0 docno rank score tag
```

  `qid`: query number, `Q0`: literal Q0, `docno`: id of a retrieved document, `rank`: position (1 to max 1000) in the ranked list, `score`: score computed by your retrieval system, and `tag`: identifying name of the retrieval system.

- Evaluate the effectiveness of this baseline information retrieval system using the relevance assessments you created in the first step of Milestone 2.

- Persist your run in TIRA by following this tutorial:

https://github.com/tira-io/ir-experiment-platform/blob/main/tira-ir-starters/pyterrier/full-rank-pipeline.ipynb

Project Info: Milestone 2 – Methods

What to hand in

- A docker image with an updated version of your dataset containing qrels registered in ir_datasets and upload it to TIRA as before in Milestone 1
- Upload a second docker image containing a jupyter notebook with the implementation of your baseline retrieval system that produces a run file to TIRA

Project Info: Milestone 2 – Methods

What to hand in

DEADLINE FOR MILESTONE 2 is the 06.06.23