Information Retrieval: Homework 4
Retrieval evaluation

Introduction

In this assignment, you will delve into the implementation of fundamental retrieval evaluation metrics. These metrics will empower you to critically evaluate and compare the results obtained from different search engines, fostering a deeper insight into the intricacies of information retrieval techniques.

You will be provided with a `main.py` file with the `TODO`'s and cached search results in JSON format of four different search engines.

- Bing
- DuckDuckGo
- Google
- Yahoo

There will be two directories for two different search queries :

query1_cache: Modern Information Retrieval

query2_cache: information retrieval evaluation

The JSON format should be familiar to you since these are formatted the same way as in the first assignment.

How to run the provided code

The code was tested using `python` version `3.10` and you can run it using multiple tools:

Docker

You have been provided with a `Dockerfile` **and** a `docker-compose.yml`. The command to run :

```
docker compose up --build
```

`--build` ensures that if any changes happened the images gets rebuilt, you should not need this all the time but always include it to make sure.

Without Docker

You can also run the code without Docker.

virtualenv

First make sure you have **virtualenv** installed then run:

```
virtualenv venv
source venv/bin/activate
pip install -r requirements.txt
python main.py
```

without virtualenv

```
pip install -r requirements.txt
python main.py
```

Important Note

It is probably a good idea to redirect your output to a `results.txt` file, you can do so by `python main.py > results.txt` which is also what the `Dockerfile` does. Additionally, the graphs will be saved to a directory called `plots`.

Finally, the reference plots will be available on themis, you can find them here.

Additionally, we will be putting some results on themis, so you can submit your code to test whether it works. Say themis has an issue, you can always and is encouraged to look at the expected out and compare that to your output. Remember that passing on themis is **not** required.

Description

Implement fundamental retrieval evaluation metrics. Use these metrics to evaluate the retrieved results of the 4 search algorithms (this also includes the baseline/reference algorithm).

Precision and Recall:

- Implement a method to compute the precision and recall.
- Evaluate the ranking results produced by the 4 search algorithms

Precision vs Recall:

- Plot the precision figure for the 11 standard recall values.
- Evaluate the ranking results produced by the 4 search algorithms.
- Pay attention to the boundary cases (refer to the example queries $q1$ and $q2$ from the lecture)

Single valued summaries:

- Implement the metric Precision at rank 5 (P@5), Precision at rank 10 (P@10), and the F-measure.
- Evaluate the ranking results produced by the 4 search algorithms

Deliverables:

- Code + Documentation
- A brief report comparing the evaluation of the three search algorithms. For each metric briefly describe the metric, the scores/plots obtained on the three search techniques, and a comparative evaluation of these scores.