Information Retrieval Galápagos Tortoise
WiSe 23/24 FSU Jena

Report - Milestone 3 November 29, 2023

We took this milestone as an opportunity to explore some of the features of information retrieval provided by `pyterrier` [2]. Unfortunately, our knowledge in the field is still very limited, hence, we were not always able to assess the relevance of the measures taken. `Pyterrier`'s experiments provided some assistance, but with quite some evaluation criteria at hand, it was hard to decide which one we should strive to maximize.
Below we briefly outline our retrieval pipeline.

1. Initial retrieval with `PL2`.
   We wanted to expore some alternatives to retrieval with `BM25`.
2. Tune hyperparameter c of `PL2`.
   We tested several values for `PL2`'s c parameter and visualized the mean average precision (`map`) depending on the choice of c. Setting $c = 1.4$ yielded the best result.
3. Second retrieval with `BM25`.
   With `BM25` we introduced a second ranking function.
4. Experiment with query expansion for `BM25`.
   `Pyterrier` provides several query expansion models. We tested Bose-Einstein statistics (`Bo1`) and Kullback-Leibler divergence (`KL`). The first one yielded better results on our training data.
5. Combine multiple retrieval systems linearly.
   We combined our two retrieval models (`PL2` and `BM25` with `Bo1` query expansion) linearly, weighting `BM25`'s score twice as much, as it seemed to be the stronger model. Yet, we hoped to benefit from their respective strengths by combining them linearly.
6. Run experiments to evaluate performance.
   The combination of `PL2` and `BM25` with `Bo1` query expansion performed best.
7. Rerank the results with neural reranking using pyterrier_t5[1]. For this, the following steps where necessary:
   - Split the documents into snippets, since the neural reranker has a maximum token size of 512 token. For this we used a sliding window approach.
   - Apply the reranking to the snippets.
   - Use an aggregation approach to get a score for a single document based on the scores of the snippets. Here we used the maximum score of any snippet of a given document.

Results

References

[1] Sean MacAvaney. pyterrier_t5. `https://github.com/terrierteam/pyterrier_t5`.

[2] Craig Macdonald and Nicola Tonellotto. Declarative experimentation ininformation retrieval using pyterrier. In *Proceedings of ICTIR 2020*, 2020.