

IR assignment 2 results

Caleb Wentworth

October 2024

1 Analysis

Both my TF & BM25 models were moderately successful for approximately 10% of queries. I believe most issue came from the pre-processing of questions and answers, since I did not apply stemming to the tokens. Non-English, and fused URL tokens were also in the dataset, increasing the size of the corpus and affecting the IDF values.

My BM25 model's best result was for query 115844: "How do I politely refuse to lend my pen?" this query was extremely suspect, with little opportunity for preprocessing outside of html tag removal. It resulted in a p@5 of 0.6

My BM25 models worst result was query 16: "Is there a good website to plan a trip via trains in Europe?" this query had numerous grammatical mistakes (improper spacing of punctuation) and multiple urls in a short document. This lead to a short query with rare tokens.

My TF-IDF model had a good result for query 45054: "Why do we fasten seatbelts for turbulence?" This query had no urls, and lots of repeated terms in a short query, and no strange non english tokens. It resulted in a p@5 of 0.6

My TF-IDF model had a bad result for query 148734: "Finding big cacti between Phoenix, Las Vegas, and Los Angeles" This query lots of URLs, it was also fairly short, thus the nonsense tokens were fairly high percentage wise.

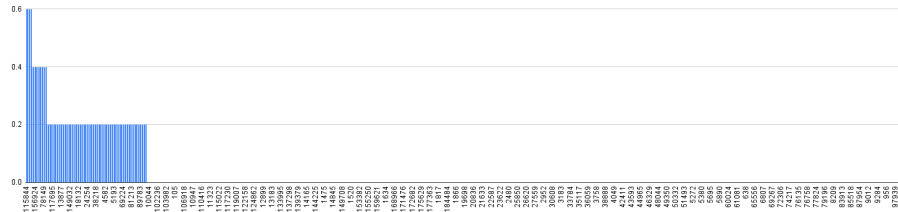


Figure 1: BM25 P@5

