

Report Assignment 2

Posting List

The posting list was in order of terms and within those terms docID

Top-K methods

I only used one top-k method, index elimination

Threshold and Query terms %

The threshold of df of 100 : (ie $\log(N/100)$)

Query term % of

If terms greater than 10 than terms match% 10%

If terms less than 10 any matching terms

TD-IDF weighting scheme

For both documents and queries I used the standard weighting scheme of $TF \cdot IDF$

($TF = 1 + \log(F)$ and $IDF = \log(N/DF)$)

Where terms share the same IDF for all