# CAB431- Search Engine Technology

## Task 1 - Index the collection

The Terrier search engine architecture was used for indexing which happen to use Termpipelines stopword with the Porter Stemmer for just English terms as default, I thought this was a good choice because Porter Stemming is usually one of the better algorisms to use in information interval.

A picture containing text

Description generated with high confidence



A close up of a sign

Description generated with high confidence

A close up of a keyboard

Description generated with high confidence

# Task 2-Execute baseline retrieval runs and evaluation

## The results

Results of baseline runs with all query documents put together

|  |  |  |
| --- | --- | --- |
| Model | TF-IDF | BM25 |
| Map | 0.5400 | 0.3017 |
| Gmap | 0.1603 | 0.5455 |
| P.10 | 0.5424 | 0.5400 |
| Ndcg10 | 0.5505 | 0.5455 |



A picture containing device

Description generated with high confidence



A close up of a sign

Description generated with high confidence

The graphs didn’t turn out well

A close up of a device

Description generated with high confidence

A group of people in a room

Description generated with high confidence

The results show not a great difference between both models but at 120 query’s there is a small variation between BM25 and TF\_IDF but similar result overall.

# Task 3

I tried to do the glove task but couldn’t get it to work, got word to vector to work and some similarity between words and top k using cosine similarity.

A screenshot of a social media post

Description generated with very high confidence

# Task 4

I was trying to load in the files with query id and the vectors and for each term in the query I would compare it to each word in the vector files using cosine similarity. I could not complete this task because of formatting issues and couldn’t get trec\_eval to run. The literature I was going off is Query Expansion Using Word Embeddings by Saar Kuzi, Anna shtok and Oren Kurland and was trying to attempt the fusion example.

A close up of a logo

Description generated with very high confidence

Trying to get the terms

A screenshot of a cell phone

Description generated with very high confidence

Couldn’t get it to work

A close up of a person

Description generated with high confidence

Top k if it worked

# Task 5

Didn’t complete this task because couldn’t complete task 4.