

IR Programming homework 2 report

B04902025 資工四 施博瀚

How to run my program:

```
python3 main.py -h
```

Report:

作業用的方法是 sample code + Okapi/BM25

用 TD.csv 的資料計算後五筆 MAP 當作參考，但隨著實驗進行發現常常並不能代表整體表現。

總上傳次數：30

Highest score：0.2107125

(加上答案後 0.3551797)

Basic	0.1558846
-------	-----------

Ignore query and IDF ($k = 1.2$, $b = 0.75$) :

Okapi without Doc Length Norm.	0.1641908
Okapi with Doc Length Norm.	0.1543037
Using cosine sim.	0.0543396
$k = 0.1$ (Best performance on my MAP)	0.1706159

幾乎就是錯誤的嘗試，令人意外的是 Cosine 慘不忍睹

Keep IDF (Okapi, $k = 1.2$, $b = 0.75$)

Using $dtf * idf$ as score	0.1901624
Using $\log(idf)$ as score (Original $idf = N/df$)	0.2060402
Using $qtf * dtf * idf$ as score	0.2091199
$k = 1.3$	0.2094794
$k = 1.4$	0.2098871
$k = 1.5$	0.2098071

k = 1.6	0.2102734
k = 1.7	0.2107125

k = 1.2 時最後五筆的 MAP 表現是最好的，然而我實驗性地增加 k 卻得到更好的結果，顯示自己測出的 MAP 不能代表整體。