WM HW2 report

資工三 b05902058 陳竣宇

執行方式

```
python3 vsm.py $1 $2 $3 $4 $5
```

- $1: path for inverted file
- $2: path for url to content
- $3: path for QS_1.csv
- $4: path for template.csv
- $5: path for output csv file

1. 使用的方法

- Vector space model
  - 只使用2, 3, 4個字的term
  - Okapi weighting based document score
  - parameters
    - k_a = 1000
    - k_1 = 2
    - b = 0.75
- Rocchio Relevance Feedback
  - $\alpha = 1$
  - $\beta = 0.2$

2. 嘗試過的實驗

- 在計算新聞字數時不考慮換行字元
  - 結果和原本相同
- 增加unigram、5-gram的term
  - 結果和原本相同
- 拿掉4-gram
  - 結果降為0.2329866

3. 調整過的參數

- k_a = 1000, k_1 = 2, b = 0.75
  - score = 0.2372526
- k_a = 1000, k_1 = 1.2, b = 0.75
  - score = 0.2325742
- k_a = 1000, k_1 = 2, b = 1
  - score = 0.2342754
- k_a = 666, k_1 = 2, b = 0.75
  - score = 0.2372526