WM HW1 report

資工三 b05902058 陳竣宇

1. VSM

- Okapi weighting based document score
- parameters
  - k_a = 1000
  - k_1 = 2
  - b = 0.75

2. Rocchio Relevance Feedback

- pick the top 40 as relevant documents, others as irrelevant documents
- parameters
  - α = 0.95
  - β = 0.05
  - γ = 0.01

3. Results of Experiments

- MAP value under different parameters of VSM
  - k_a = 1000, k_1 = 2
    - score = 0.80247
  - k_a = 1000, k_1 = 1.2
    - score = 0.79991
- With Feedback vs. without Feedback
  - k_a = 1000, k_1 = 1.2
  - with Feedback
    - score = 0.79673
  - Without Feedback
    - score = 0.79991
- Other experiments
  - 在取query term的時候只取concepts，結果與取所有tag(title, question, narrative, concepts)的結果相似
  - 因為query file裡面會出現 是不相關的 等字眼，於是嘗試把第一個句號之後的文字捨棄，結果仍與不捨棄相差不多

4. Learned from the work

- how to read a `.xml` file
- how to make a Chinese vector space model
- how to use Rocchio Feedback algorithm