

## The Rocchio Algorithm

使用參數：

```
MAX_ITER = 20  
TOP_K = 8  
ALPHA = 0.5  
BETA = 0.5  
GAMMA = 0.5
```

照著公式實作而已，步驟大概是：

1. 先對 document 和 query 建立出 tfidf 的向量矩陣。
2. 再對 document 和 query 做第一次的 ranking，在這邊 similarity 計算是使用 cosine similarity。
3. 根據 ranking 的結果，找出前 K 個相似的 document 跟前 K 個不相似的 document，套用 The Rocchio Algorithm 的公式，計算出新的 query 向量矩陣並做 ranking。

$$\vec{q}' = \alpha \cdot \vec{q} + \beta \cdot \frac{1}{|R_q|} \cdot \left( \sum_{d_j \in R_q} \vec{d}_j \right) - \gamma \cdot \frac{1}{|\bar{R}_q|} \cdot \left( \sum_{d_{j'} \in \bar{R}_q} \vec{d}_{j'} \right)$$

4. 重複第 3 步 N 次之後，輸出 ranking 結果。