# Information Retrieval and Applications- HW1

# M10915045-施信宏

# 此次作業讓我們實作了tf-idf，然而再延伸至vector space model，第一版的程式完成後，上傳kaggle只有0.02的準確率。之後將vector space改成全部queries內出現過且不同的單字，之後準確率提升至0.15左右。最後回頭看投影片，發現在term-weight定義時，若term的tf = 0的情況，則該項的tf-idf應為0，加入此條件後再執行，準確率則來到了0.745左右。

# 藉由此次作業，也更理解了前期做文本搜尋的方法，idf定義此單字是否重要的概念也很有趣，然而也想知道現今的NLP界的巨人-bert在計算相似度上面是否有更精確的值，是否準確率也來的更高，本來有打算有此方法，但怕破壞了其他人衝分數的體驗，就此作罷。