R12725049徐尚淵 作業二

1. 執行環境：Jupyter Notebook
2. 一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

   自動產生的描述程式語言：Python (版本3.11.4)
3. 執行方式：

一張含有 文字, 字型, 白色, 標誌 的圖片

自動產生的描述直接使用Jupyter Notebook GUI Run code

1. 作業處理邏輯說明
   1. 先做與作業一相同的前處理 (簡化截圖)

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

* 1. 計算每個文字在Document當中出現的頻率，利用term\_info\_List儲存之後可能會用到的資訊。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. 將Dictionary輸出並命名為dictionary.txt

一張含有 文字, 螢幕擷取畫面, 陳列, 字型 的圖片

自動產生的描述

* 1. 計算每個文件的tf-idf vector 因term\_document\_frequency在前面做過了因此此處可以直接讀取

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述 一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

* 1. 將每個文件的 tf-idf 儲存

一張含有 文字, 電子產品, 螢幕擷取畫面, 陳列 的圖片

自動產生的描述

* 1. 讀取文件並將其儲存為df的形式以便日後處理資料的Function

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

* 1. 分別計算其unit vector的Function

一張含有 文字, 字型, 行, 螢幕擷取畫面 的圖片

自動產生的描述

* 1. 讀取檔案並計算兩者的cosine similarity

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述