一、專題摘要 **(解釋實作與說明需要解決的問題，限 300~500 字。)**

1. 期末專題主題 : 爬蟲 PTT 政黑板
2. 期末專題基本目標 : 利用 request package 抓取 PTT 政黑板所有的文章，將每一篇文章擷取其中的標題、作者、IP、文章內容、網友留言等資料。再將每篇文章的文章內容與網友留言，透過 TFIDF 的統計方方式計算出高頻率的關鍵詞，增加詞彙給 Jieba 做切詞分析，並將自訂的 stop word 移除。將每篇文章的文章內容與網友留言做切詞後，發現蔡英文和柯文哲是大家最常提到的政治人物。於是，爬蟲後的分析切詞資料目標為，針對蔡英文和柯文哲，分析出，在 PTT 政黑板中，他們與哪些事情是常常被一起討論的？被討論的內容是好的還是不好的？

二、實作方法介紹 **(介紹使用的程式碼、模組，並附上實作過程與結果的截圖，需圖文並茂。)**

1. 使用的程式碼介紹

   a. 使用模組:

   b. 爬取 PTT 政黑板所有文章:

   c. 整理出每篇文章的資料

   d. 將每篇文章的文章內容與網友留言，透過 TFIDF 的統計方方式計算出高頻率的關鍵詞，增加至詞彙庫

   e. 利用 Jieba 作分析，並利用 snownlp 針對文章內容作情緒分析

   d. 整體分析，文字雲分析

三、成果展示 **(介紹成果的特點為何，並撰寫心得。)**

收集了選舉後，與疫情爆發時期，網友對於政治討論議題的比較。並分別針對蔡英文跟柯文哲相關文章加以分析報告。可以發現選舉後，大家還是會圍繞選舉字眼討論，例如:蟑螂、中共等。就此發現 PTT 鄉民有一常用語為政治蟑螂，是網友用來諷稱政黨買來帶文章風向的網軍。不管是柯文哲或是蔡英文，他們都難以逃脫跟蟑螂牽扯的命運。

大概在選舉過後的兩個禮拜後，武漢肺炎話題開始蔓衍全球，網友也開始討論了政治人物對於處理疫情的態度，於是從文字雲開始出現了，病毒、疫情等字眼，反映了大家開始慢慢把政治討論重點轉移到了疫情。

1/12– 1/26

ALL Data

常見用詞：

1. 柯	4. 支持
2. 立委	5. 蟑螂
3. 政治	6. 中共

情緒分析	
N	0.845
P	0.155

蔡英文

常見用詞：

1. 柯	4. 蟑螂
2. 中共	5. 合作
3. 支持	6. 立委

情緒分析	
N	0.85
P	0.15

柯文哲

常見用詞：

1. 蟑螂	4. 政治
2. 立委	5. 希望
3. 合作	6. 台北

情緒分析	
N	0.87
P	0.13

1/27 – 1/28

ALL Data

常見用詞：

1. 柯文哲
2. 肺炎
3. 蟑螂
4. 病毒
5. 疫情
6. 口罩

情緒分析	
N	0.813
P	0.187

蔡英文

常見用詞：

1. 柯
2. 肺炎
3. 蟑螂
4. 病毒
5. 疫情
6. 口罩

情緒分析	
N	0.814
P	0.185

柯文哲

常見用詞：

1. 蟑螂
2. 中共
3. 政治
4. 台北
5. 政府
6. 肺炎

情緒分析	
N	0.82
P	0.18

四、結論 (總結本次專題的問題與結果)

很開心這次參加爬蟲的活動，原本就懂一點點前端，透過這次的活動，能將前端訊息運用到更廣的地方，獲益良多。更因為這次活動，了解到多線程與非同步程式的運作，讓我發想了在這次活動結束後，想要在這部分再多著墨些。唯一可惜的地方，就是因為時間的關係最後期末專題未能應用上。但是初接觸到了，爬完蟲後，可運用蒐集的資料，再作文字雲、情緒分析。這也是我期末專題最令人開心的事。最後，還因為這次的分析，讓我這個對政治比較冷漠的人，認識的 PTT 鄉民用語:政治蟑螂☺。(還要謝謝這次陪跑訓練的教練們，總是不厭其煩的回答各種問題)

五、期末專題作者資訊 (請附上作者資訊)

1. 個人 Github 連結: https://github.com/stella0320/1st-PyCrawlerMarathon-
2. 個人在百日馬拉松顯示名稱: Jia Xin Chen