需求：

1.找一個新聞媒體(蘋果日報)

2.用一個主題，例如：新冠肺炎

3.爬文下來會呈現文字

4.文字雲(作者名字)、情緒分析等等

研究動機：

2019年，新冠病毒爆發，疫情初期，台灣部分學校、公司採遠距上課、辦公，所幸疫情控制得宜，遠距政策只持續了幾個月。但今年疫情再度爆發，教育部宣布停課不停學，部分公司也改彈性上班、遠距上班，遠程會議系統再度被台灣人所重視；且從疫情爆發初期，全球遠程會議系統的需求逐漸提高，資安風險也成了個系統廠商需要考慮的問題。

但資訊安全的概念並非人人都熟知，在不夠了解的情況下，記者不應隨意撰寫文章來誤導民眾，因此本研究想找出是否有較專業撰寫資訊安全知識的記者，並且其是否運用了較正面的措辭來撰寫。

實驗步驟：

**News\_crawlers.py**

將搜尋結果爬蟲下來

1.將所有要import的library引入

2.利用Chrome開發的chromedriver.exe開啟自動監控，開啟的網頁為聯合新聞網中"資訊安全"的搜尋結果

3.新冠病毒於2019年10~11月開始爆發，因此我們將取2019.10月後的新聞，

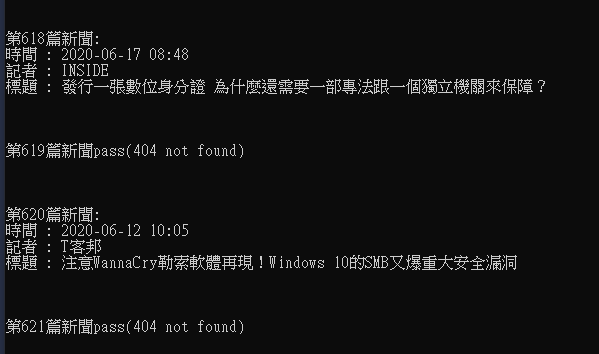
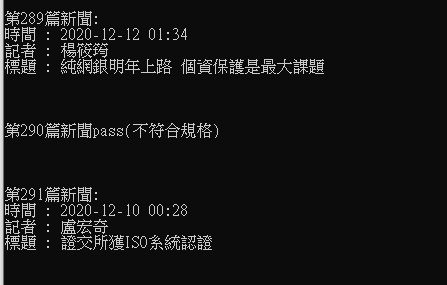
但因此網頁為動態網頁，並不會將搜尋出的800多篇新聞全部展出，因此需要利用driver.execute\_script()將滾輪往下滑

4.所有的目標搜尋結果都展示出來後，將所有符合要求的連結存進href陣列裡

5.儲存所有href後，逐一爬每一個網頁的新聞內容，並記錄下[時間, 作者, 標題, 內容, 網址]

每一筆爬取的結果顯示如圖(僅顯示時間、作者、標題)

而部分不符合我們要的規格的新聞會被pass掉(通常是需要會員登入才可以看的新聞)



6.最後共爬下了701篇新聞

**News\_crawlers\_filter.py**

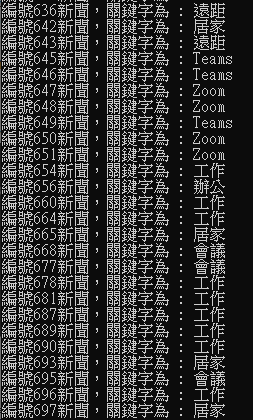
將網站爬蟲後的csv進行篩選，找出哪些是針對遠端會議軟體的新聞

1.將所有要import的library引入

2.News\_crawlers.csv讀入

3.定義關鍵字，若內文不包含任一關鍵字，即不採用

4.利用自定義的key\_word搜尋新聞內容，篩選出符合我們要求的新聞



5.將篩選出的結果存至News\_crawlers\_filter.csv

6.共篩選出299篇

**News\_crawlers\_sensitive\_analysis\_wordcloud.py**

將篩選出的新聞進行文字雲的分析，找出較常撰寫此類新聞的記者

1.將所有要import的library引入

2.News\_crawlers.csv讀入

3.將記者名字分成三種類型串在一起，{"正面記者(positive\_text\_author)", "負面記者(negative\_text\_author)”, “所有記者(all\_author)”}以方便丟入文字雲的function

4.進行文字的清洗，因有的文章是轉載自其他新聞網，因此把這些影響結果的字去掉

5.輸出文字雲，因為是人名的關係，所以不加上結巴分割字元的功能。並且加上mask、edge的參數，將輸出造型進行變換

最終輸出結果：

(1)所有記者(wordcloudAuthor.jpg)



(2)正面記者(wordcloudPositiveAuthor.jpg)



(3)負面記者(wordcloudNegativeAuthor.jpg)

