

社群媒體分析

期中報告

【Ptt八卦版:台海危機】

十四組

N104020006 曾怡霽

N104020012 蔡京叡

N104020013 吳映儒

N104020030 林欣民

大綱

- 研究動機與分析目的
- 資料前處理
- 情緒分析
- 不同版文字雲分析
- 結論





研究動機與分析目的

研究動機與分析目的

隨著兩岸關係越來越緊張，不僅台灣人民關注自身國家的未來，眾多與台灣有關聯的國家也隨時都在關注台海危機。其中兩岸關係較有起伏的兩個事件點，分別是2022年8月的裴洛西訪台與2022年4月蔡英文訪美與麥卡錫會面。

使用PTT八卦版作為分析的主要資料來源，並且探討民眾對台海危機的看法，以及裴洛西訪台、蔡英文訪美與麥卡錫會面民眾的情緒。





資料前處理

Ptt爬蟲、替換字串、中文斷詞、清除停用字

資料集基本介紹

資料來源--

主要:ptt的八卦版

其他:ptt的其他版如政黑板、股票版、房屋版

時間:2022/08/01 -2023/4/20

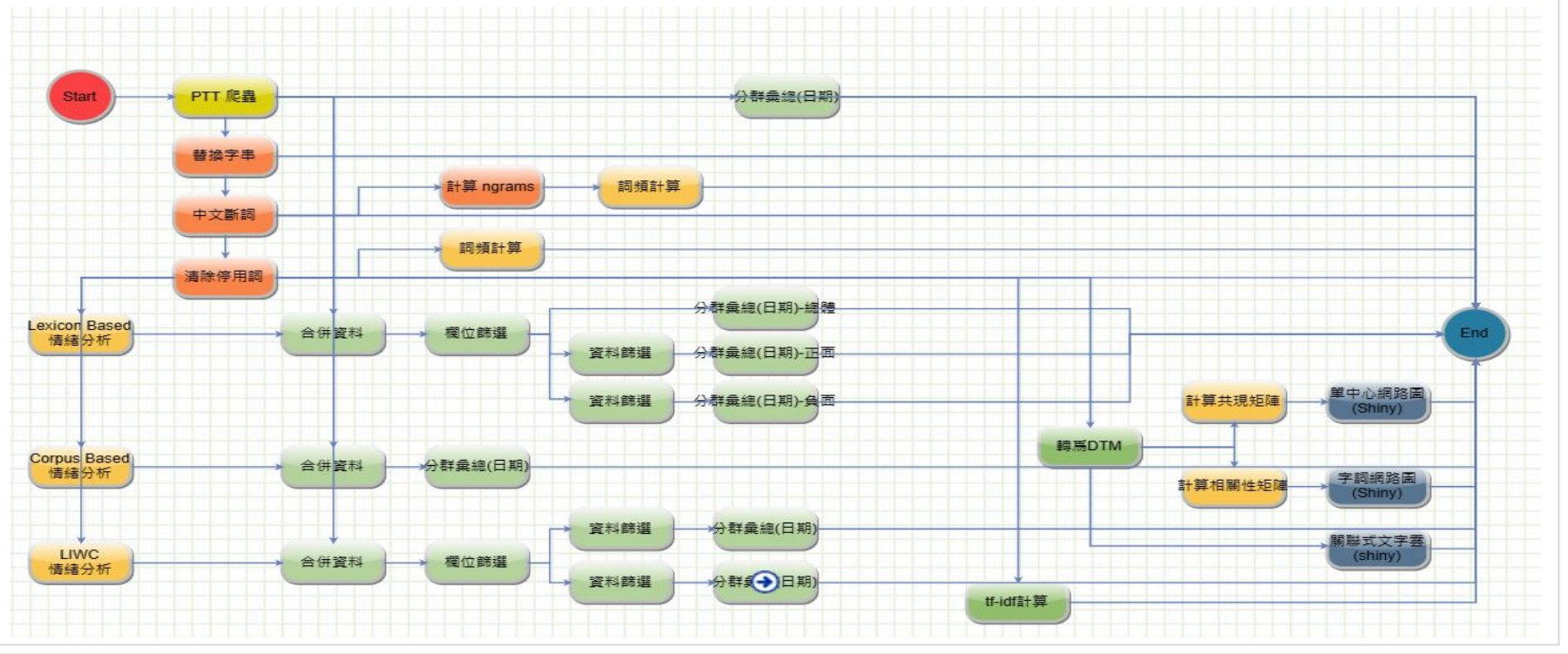
關鍵字:台海危機、裴洛西、麥卡錫

資料筆數:5132筆



分析資料

使用 workflow 平台 TarFlow 工作流程:台海(原版)



資料前處理 替換字串

先去除換行符號、網址、html標籤、底線、多餘的空格

替換字串 (7)

參數設定

Input - 58

任務結果

選擇處理欄位 *

artContent

替換字串設定 ①

```
((http|ftp|https):/){([a-zA-Z0-9\._-]+\.[a-zA-Z]{2,6})|([0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3})}(:[0-9]{1,4})*  
/[a-zA-Z0-9\&%_\.\/~]*?>>  
Sent from JPTT on my \w+>>  
\w+:\{2\}[\d\w-]+\.[\d\w-]+*(?:\{:\[^{s/}\})*>>  
\b(?:https?://www\.)|S+|b>>
```

選擇替換規則檔案 ①

-----請選擇-----

儲存更改

資料前處理 中文斷詞、清除停用字

進行中文斷詞、清除可能干擾分析結果的停用字，並運用「計算ngrams」以及「詞頻計算」功能，反覆觀察，進行斷詞及停用字處理。

清除停用詞 (9)

參數設定

Input - 8

任務結果

語言 *	使用預設停止詞
Chinese	是
是否清除單字元 ⓘ	是否轉為小寫英文
是	是
清除英文字母 *	清除數字 *
是	是
清除換行符號 *	清除特殊標點符號 *
是	是
清除html tag *	自定義停止詞
是	表示 認為 內文 指出 記者

儲存更改

本次分析內容自訂義的停用字為表示、認為、內文、指出、記者、連結、網址、轉載、標題、附註、媒體、報導、來源、一下、一直、想法、心得、新聞、報導、現在、完整、今天、目前、包括、這種等。

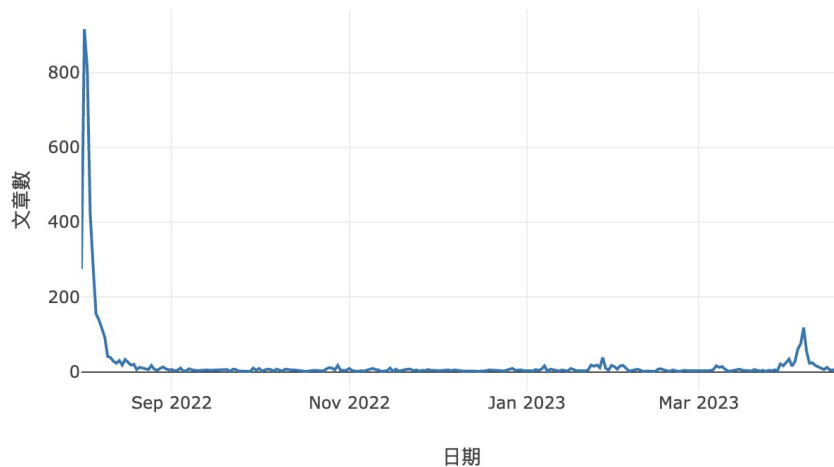


情緒分析

Lexicon Based 情緒分析、LIWC 情緒分析、Corpus Based 情緒分析

每日文章趨勢

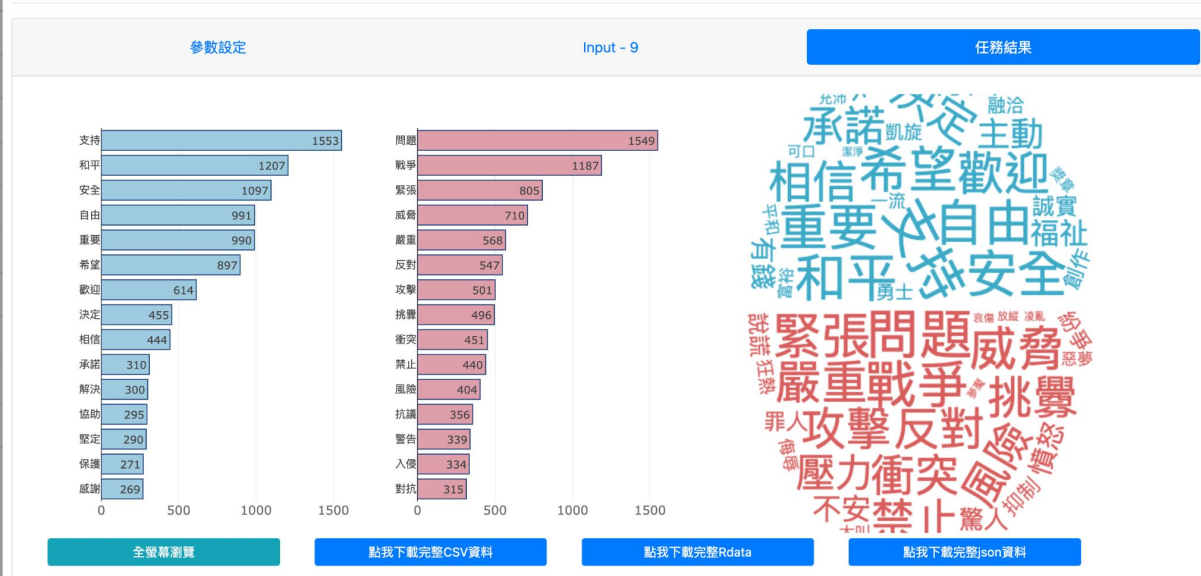
每日文章統計



可以看到2022年8月的文章數比較多，因為裴洛西訪台的因素；而2023年4月則是因為蔡英文訪美會麥卡錫，而且可以觀察到2022年8月文章數遠大於2023年4月，可推測當時的討論熱度較為激烈的原因，推測因為中國大陸2022年8月時對台灣施壓力道較強所致。

Lexicon Based 情緒分析

Lexicon Based 情緒分析 (24)



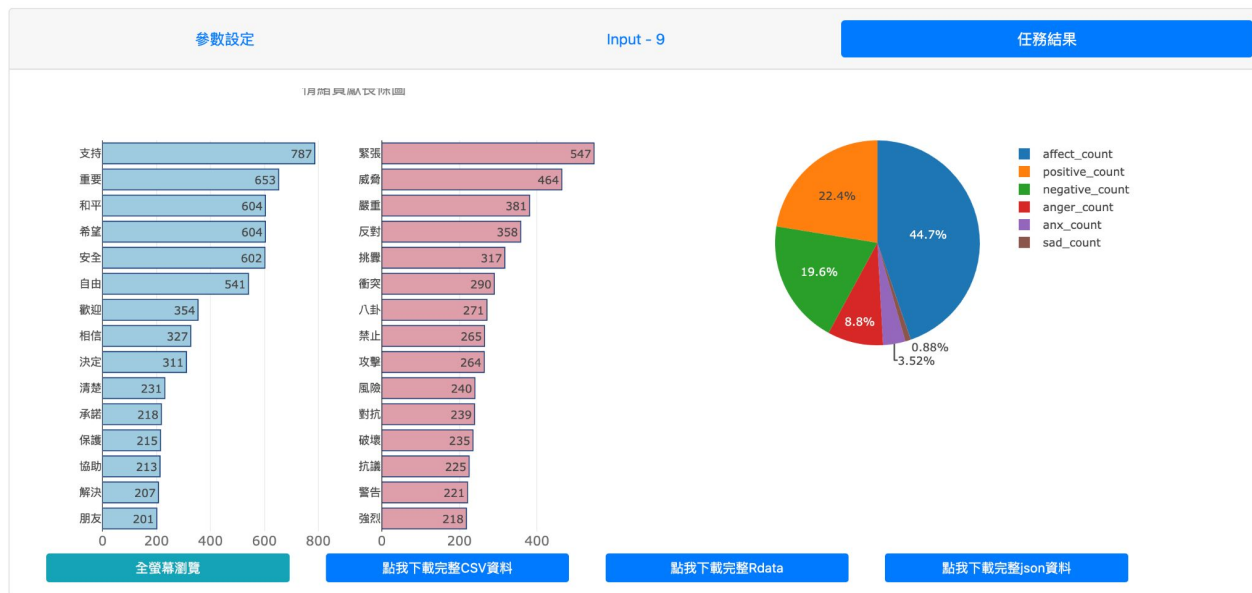
排除問題與戰爭後得到-

負面詞彙：緊張、威脅、嚴重

正面詞彙：有支持、和平與安全

LIWC 情緒分析

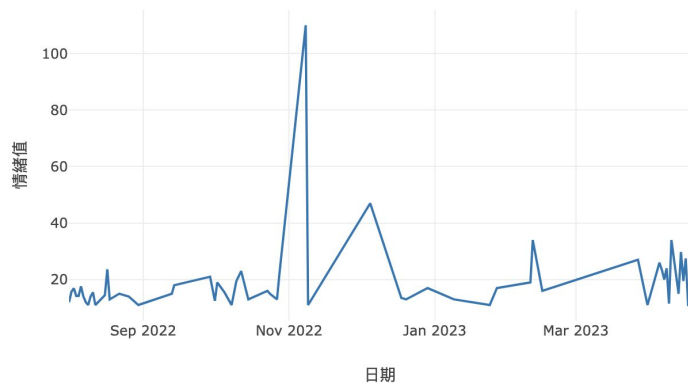
LIWC 情緒分析 (114)



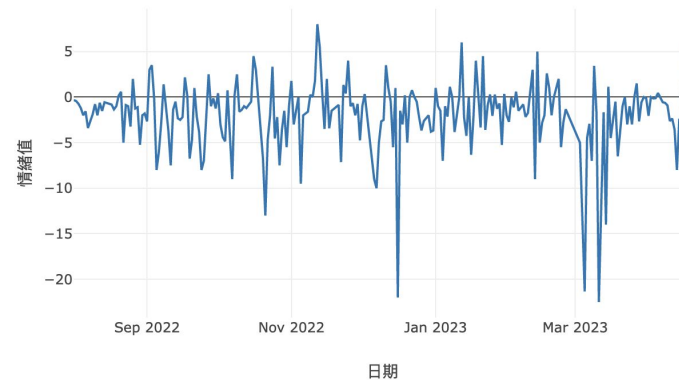
正面詞彙22.4%與負面詞彙19.6%
比例沒太大的懸殊

正負情緒平均分數Lexicon Based

正面情緒平均值分析圖



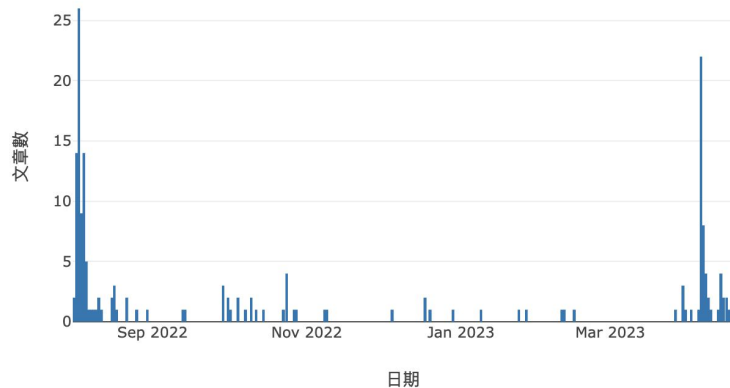
負面情緒平均值分析圖



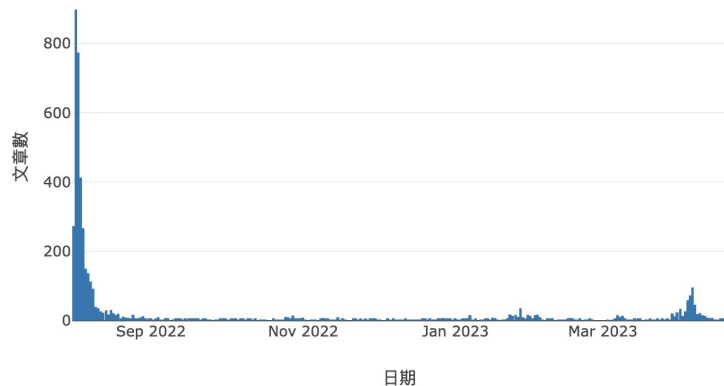
2022年11月的正面情緒平均值有個比較高峰，推測是因為習近平與拜登的會議在2022年11月進行。

正負面情緒文章統計

正面文章數長條圖



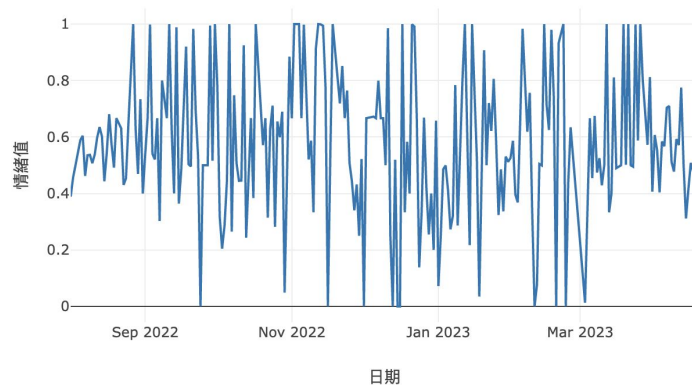
負面文章數長條圖



由正負面情緒文章數量(情緒分數 >10 為正面、情緒分數 <10 為負面)可以觀察到，負面文章討論度遠大於正面文章討論度。

Corpus Based 情緒平均值

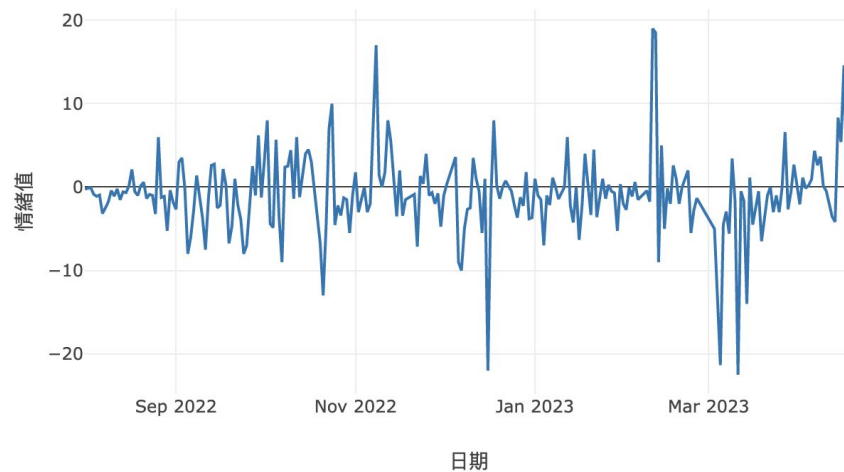
總體情緒平均值分析圖(Corpus Based)



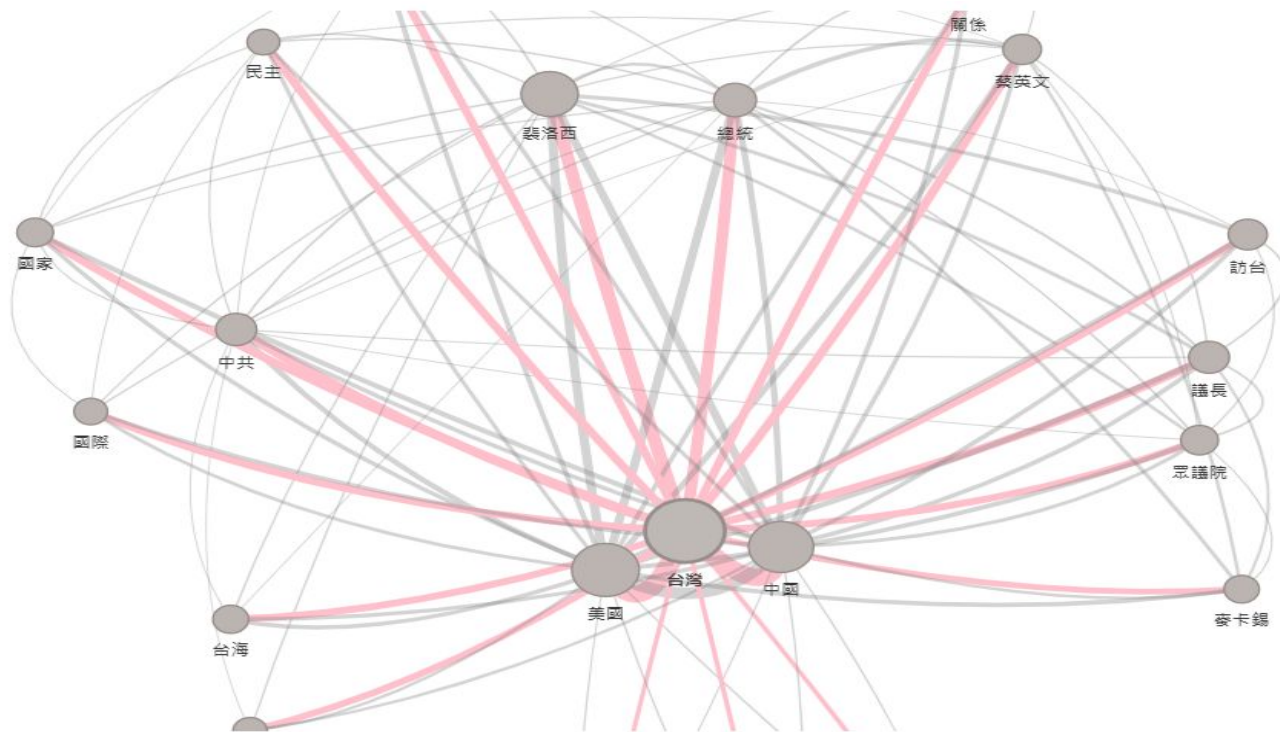
文章的情緒分數平均值正負皆差不多。

情緒平均分數Lexicon Based

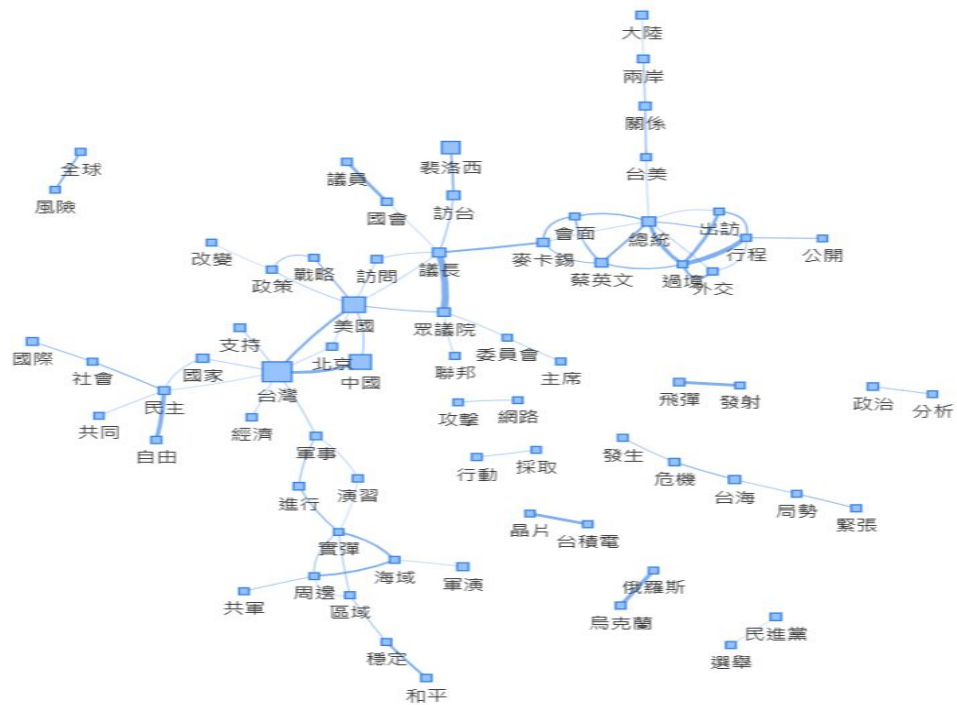
總體情緒平均值分析圖



單中心網路圖



字詞網路圖



關聯式文字雲shiny



透過轉為DTM後所產生的單中心網路圖、字詞網路圖與關聯式文字雲中，可發現美國、中國與台灣的關聯程度很高，可見台灣與中國之間的危機，美國也扮演相當重要的角色。

詞頻計算文字雲



左邊為使用八卦版、政黑板、股版所爬蟲出的文字雲結果。以台灣、中國與美國居多。



不同版文字雲分析

使用與此主題相關的PTT版分析

Gossiping 八卦版：

分析主題為兩岸關係、台海危機，由於八卦版對於政治、國際情勢均有一定的討論度，因此選擇他。

Home_sale房屋版：

若兩岸真的發生戰爭，推測台灣房市、房價一定會受到影響，因此在此版應也有相關討論度。

Stock 股票版：

戰爭對於經濟來說有相當大的影響，因此股票版也是能獲得相當數量的討論。

各版的文字雲

Gossiping 八卦版



Home_sale房屋版



Stock 股票版



各版的文字雲關鍵字分析

在**八卦版**基本上還是以國際情勢本身為討論熱度最高，出現了**裴洛西**、**中國**、**台灣**、**美國**等字。

而**房屋版**則出現了**房價**、**房市**、**危機**等關鍵字，可以看到戰爭與房價的確是大家相當關注的主題。

最後在**股票版**可以看到**台積電**佔據了很大的版面，可以推測網友們對於台積電與台海危機間的關係，是有相當的關注度。



結論

資料分析結果的發現與討論

- 1.在情緒分析上來看，可以發現有關台海危機的文章，情緒分數有很大的起伏，而討論度也會隨著重大事件的發生，EX:裴洛西訪台、蔡英文訪美有較高的相關文章數量。
- 2.而從不同的PTT版文字雲中也可以看出，不同主題的看板所關注的重點、熱度，即使是使用相同的主題，所關注的重點亦或是討論度皆可能有所不同。
- 3.而原本預期有關戰爭、台海危機的主題情緒分數可能會偏向負面，但實際分析後發現，其實正面與負面的情緒分數其實相差不遠，因此可以推斷台灣人民對於兩岸情勢發展還是抱有著相當程度的正面想法。

Thank You!

