



國立中山大學資訊管理系碩士班

社群媒體分析期末專案報告 與食安議題有關之政府聲量討論

食安小組成員：

N094020030 陳詠琳

M121020012 涂宥安

N114320027 吳政翰

N094220008 龔敖摩

指導教授：黃三益 教授

助教：蔡易航、蔡睿澤、張宸瑜、呂育真

中華民國 113 年 6月11日

目錄

一、動機和分析目的.....	1
二、資料集的描述.....	2
三、資料分析的過程.....	5
四、視覺化的分析結果與解釋.....	21
五、結論.....	28

一、動機和分析目的

在當今社會，食品安全一直是一個備受關注的議題。人們對於他們所食用的食品是否安全及可靠的關注程度日益增加。社群媒體平台如PTT、Dcard和Mobile01是民眾討論各種話題的重要場所之一，也包括對於食品安全的關注。因此，我們的分析目的是利用爬蟲程式從這些平台上抓取相關文章，尤其是在Dcard（美食、閒聊及時事板）、PTT（美食、八卦及政黑板）及Mobile01（閒聊、旅遊美食及時事板）等板面上討論食品安全議題的文章。我們希望透過分析這些文章，了解以下幾個方面：

- (1). 食安議題討論的熱門時期：我們想要找出食品安全議題在社群媒體上討論較熱門的時段，這有助於我們了解民眾對於食品安全的關注趨勢。
- (2). 共現頻率較高的關鍵詞：透過分析文章中出現頻率較高的關鍵詞，我們可以了解民眾對於食品安全議題的主要關注點和討論重點。
- (3). 熱門討論時段的相關主題：根據討論熱門時段的高頻詞分析，可以找出該時段最常被討論的主題，並透過主題模型找出更多相關的關鍵字。
- (4). 對政府處理食安事件的態度：透過情感分析，我們希望了解一般民眾對於政府在處理食品安全事件時的態度，是正面的信任或負面的質疑。

綜合以上分析目的，我們將能夠更深入地了解民眾對於食品安全議題的關注程度之變化、頻繁出現的焦點關鍵字、情感表達以及對政府處理相關事件的態度，進而提供相應的建議和改善方案。

二、資料集的描述

(1) 資料來源：

本次分析的食安議題資料來源出自於Dcard、PPT及Mobile01可能與食安比較有關的討論版來蒐集所需的資料。

(2) 資料取得：文字探勘工作流程設計平台

(3) 資料範圍：

針對Dcard（美食、閒聊及時事版）、PPT（美食、八卦及政黑版）及Mobile01（閒聊、旅遊美食及時事版）討論區內容透過系統爬蟲搜尋113年1月1日起至5月31日期間與食安關鍵字有關的議題來獲得資料。再透過下載其CSV檔後，利用Python讀取資料後，再設定中央政府、民進黨、執政及衛福部等關鍵字擷取與之有關的內容為其資料範圍。

≡ Dcard 爬蟲 (58) 多數有更新點，建議重新執行

≡ PTT 爬蟲 (57) 多數有更新點，建議重新執行

參數設定

任務結果

選擇看板

pc(電腦)

scoder(海龜)

sport(運動)

talk(閒聊)

travelfood(旅遊美食)

trading(證券)

video(影音)

搜尋起始日期

2024/01/01

搜尋結束日期

2024/05/31

搜尋關鍵字

食安

排除關鍵字

以神戶區區、e-g
富士動物園
標子
—

GET/POST

(4) 資料內容：

因資料都是文字敘述內容，經過爬蟲後再分別利用欄位篩選將各版一致性的欄位名稱擷取出來。

欄位篩選 (63) 參數有做更動，建議重新執行

步驟設定

Input - 58

任務結果

選擇要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id
dataSource
artUrl
artDate
artTitle
artContent
boardID

儲存設定

欄位篩選 (67)

參數設定

input - 57

任務結果

選擇要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

artId

artTitle

artDate

artPoster

artCategory

artContent

儲存設定

欄位篩選 (64) 參數有變更，建議重新執行

參數設定

Input - 59

任務結果

選擇要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

dataSource

artId

artDate

artTitle

artContent

commentCount

儲存設定

再利用垂直合併的方式把三版的內容合併成一個CSV檔。這裡要注意的是系統的垂直合併僅適合連結兩個爬蟲元件所獲取的資料來整合，所以要把Dcard及PPT的資料垂直合併後再使用一次垂直合併來整合Mobile01的資料，最後再把整合後的CSV檔下載下來。

垂直合併 (73)

數據設定

Input - 63

Input - 67

任務結果

選擇要保留的交集欄位(按住ctrl(Windows)或command(MAC)可以複選)

artDate

artTitle

artContent

儲存設定

垂直合併 (76)

數據設定

Input - 64

Input - 73

任務結果

選擇要保留的交集欄位(按住ctrl(Windows)或command(MAC)可以複選)

artContent

artDate

artTitle

儲存設定

由於我們想要探討的議題是分析討論版的網民對政府維護食安的努力是保持著正面抑或是負面的情緒反應，但爬蟲僅能針對可能出現的所有關鍵字來搜尋資料，無法做到先針對食安關鍵字來搜尋取得的資料，再利用這些資料挑出與中央政府、民進黨、執政及衛福部有關的內容來分析，故這裡額外使用Python讀取剛剛下載的CSV檔來擷取我們要的內容。

```

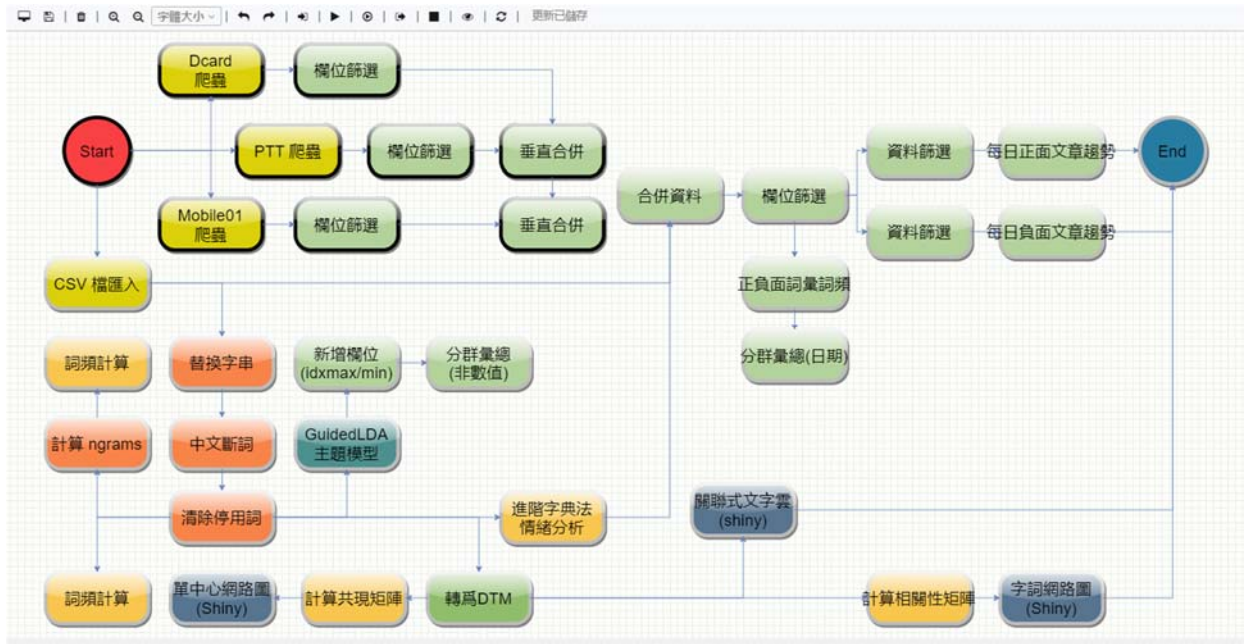
main.py x
1 import pandas as pd
2 import re
3 # 讀取CSV檔案
4 df = pd.read_csv('foodsafety.csv')
5 # 定義多個特定關鍵字
6 keywords = ['中央政府', '民進黨', '執政', '衛福部'] # 請替換為你的特定關鍵字列表
7 # 使用正規表達式匹配第四欄，並保留符合任何一個關鍵字的列
8 pattern = '|'.join(keywords)
9 df = df[df['artContent'].str.contains(pattern, na=False, flags=re.IGNORECASE)]
10 # 將結果存回CSV檔案
11 df.to_csv('foodsafetyfinal.csv', index=False, encoding='utf-8-sig')

```

三、資料分析的過程

(一) 工作流程名稱：期末專題報告-食安

(二) 工作流程設計架構：



(三) 流程概述：

由於前一段已經針對資料如何取得的步驟詳細說明了，這裡就不再贅述，以下僅就資料清洗及如何使用分析工具的過程說明流程：

1. CSV檔上傳系統後匯入
2. 替換字串
3. 中文斷詞
4. 清除停用詞
5. 計算ngrams(含詞頻計算)
6. 詞頻計算
7. 轉為DTM
8. 計算共現矩陣並完成單中心網路圖
9. 計算相關性矩陣並完成字詞網路圖
10. 關聯式文字雲
11. GuidedLDA主題模型分類出4大主題
12. 進階字典法情緒分析
13. 分析結果與原始資料合併後經欄位篩選及重構後完成正負情緒統計
14. 欄位篩選後分別經資料篩選完成每日正負面文章趨勢統計

(四) 流程講解執行細節：

1. CSV檔上傳系統後匯入：

用Python經過關鍵字篩選取得的資料轉成命名為foodsafetyfinal的CSV檔上傳到工作流程系統。



然後再利用CSV檔匯入元件把已上傳系統的檔案匯入。

≡ CSV 檔匯入 (4)



2. 替換字串：

(1) \n>> ,

(2) \n\n>> 。

(3) Sent from JPTT on my .*>>

(4) ((http|ftp|https):\/\/)[^\s]+(\/(?:[^\s]))*>>

≡ 替換字串 (7)



3. 中文斷詞：

初步先不定義任何需要斷詞的內容，直接依系統內建斷詞辭典來斷詞。

三 中文斷詞 (9)

後續再依計算ngrams及詞頻計算所呈現的文字雲找出兩個常連在一起的詞彙來定義斷詞內容。

三 中文斷詞 (9)

4. 清除停用詞：

初步先不定義停止詞直接依下圖設定內容來清除停用詞。

三 清除停用詞 (11)

後續再依詞頻計算的文字雲反覆找出較無意義的詞彙來定義停止詞。

參數設定	Input - 9	任務結果
<p>語言 *</p> <div>Chinese</div>	<p>使用預設停止詞</p> <div>是</div>	
<p>是否清除單字元 ⓘ</p> <div>是</div>	<p>是否轉為小寫英文</p> <div>是</div>	
<p>清除英文字母 *</p> <div>否</div>	<p>清除數字 *</p> <div>是</div>	
<p>清除換行符號 *</p> <div>是</div>	<p>清除特殊標點符號 *</p> <div>是</div>	
<p>清除html tag *</p> <div>是</div>	<p>自定義停止詞</p> <div>新聞內容 心得想法 完整新聞 新聞網址 附註心得</div>	
<div>儲存更改</div>		

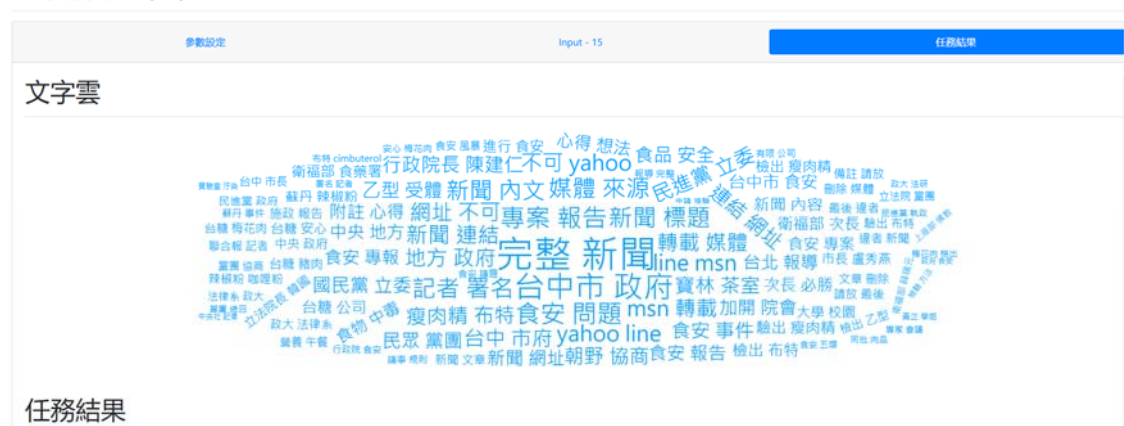
n設定2，藉此找出兩個最可能一起出現的詞彙。

參數設定		Input - 11	任務結果
忽略包含停止詞的組合 *		n 大小 *	
<input type="text" value="否"/>		<input type="text" value="2"/>	
儲存設定			

三 詞頻計算 (17)

參數設定	Input - 15	任務結果
<div>自定義停止詞 以換行符號區隔，e.g. 你好 不要 ...</div>	<div>最多篩選詞彙數量 ① 200</div>	
<div>儲存更改</div>		

三 詞頻計算 (17)



6. 詞頻計算：

在連接清除停用詞的詞頻計算元件裡設定篩選的詞彙數量為200。

≡ 詞頻計算 (13)

參數設定

Input - 11

任務結果

自定義停止詞

以換行符號區隔, e.g.
你好
不要
...

最多篩選詞彙數量 1

200

儲存更改

再依任務結果的文字雲找出較無意義的詞彙，反覆修改定義停止詞內容。

≡ 詞頻計算 (13)

[illegible]

完成定義停止詞後，詞頻計算最終呈現的文字雲如下圖。

詞頻計算 (13)

任務結果

7. 轉為DTM：

在清除停用詞後方連接轉為DTM元件，設定篩選詞彙數量為200。

≡ 轉為DTM (19)

參數設定	Input - 11	任務結果
保留詞彙 ① 以換行符號區隔，e.g. 國立中山大學 西子灣 壽山...	最多篩選詞彙數量 ① 200	
儲存英文		

任務結果會把原本的內容轉為DTM格式如下圖。

≡ 轉為DTM (19)

參數設定

Input - 11

任務結果

system_id	ppm	一下	上午	上 映	下午	下 架	不用	中國	中央	乙 型 雙 體	代 表	佛 教	個 案	健 康	備 詢	全 台	公 司	公 告	公 布	公 開	共 同	加 開 院 會	包 括	北 市	協 商	廢 因	廢 料	召 開	台 中	台 中 市	中 市 政 府	
1	0.0	0.0	0.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	3.0	0.0	0.0	3.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	5.0	1.0	37.0	7.0	3.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	9.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
8	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	4.0	0.0	0.0	3.0	1.0	0.0	0.0	0.0
9	0.0	1.0	1.0	0.0	0.0	0.0	0.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	2.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	2.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	0.0	0.0	4.0	1.0	0.0	3.0	0.0

Showing 1 to 10 of 100 entries

Previous

1

2

3

4

5

...

10

Next

全部搜尋

點以下載完整CSV資料

點以下載完整Rdata

點以下載完整JSON資料

8. 計算共現矩陣並完成單中心網路圖：
連接計算共現矩陣完成計算如下圖。

三 計算共現矩陣 (29)


查詢狀態

Input - 19

任務結果

200

字數



任務結果

Show10▼entries

Search

index	食安	檢驗	台糖	輸出	瘦肉 精	西布 特羅	民進 黨	立委	國內	美國	校園	老師	律師	上場	電影	大學	高中	中央	食農 部	衛生 局	衛福 部	農業 部	ppm	王必 勝	台中 市	台南	公布	台中 市政府	檢驗	寶樹 堂	輸出
食安	0.0	4642.0	2719.0	2599.0	2689.0	2667.0	2042.0	738.0	468.0	271.0	271.0	124.0	147.0	136.0	128.0	228.0	138.0	2333.0	1688.0	1200.0	2056.0	798.0	907.0	988.0	2239.0	2121.0	1913.0	1652.0	1552.0	1681.0	1642.0
檢驗	4642.0	0.0	3515.0	3106.0	2429.0	2879.0	249.0	203.0	382.0	63.0	36.0	0.0	2.0	0.0	80.0	7.0	1600.0	2077.0	837.0	1425.0	1324.0	1529.0	1188.0	1826.0	1646.0	1678.0	1347.0	1654.0	1748.0	1664.0	
台糖	2719.0	3515.0	0.0	2721.0	2721.0	2434.0	361.0	157.0	206.0	26.0	5.0	0.0	0.0	0.0	20.0	7.0	1067.0	1366.0	730.0	813.0	1240.0	759.0	891.0	1487.0	1025.0	1013.0	1224.0	1432.0	1187.0	1226.0	
輸出	2599.0	3106.0	2721.0	0.0	2108.0	2499.0	242.0	117.0	330.0	10.0	27.0	0.0	0.0	0.0	21.0	12.0	900.0	1481.0	658.0	953.0	1273.0	811.0	847.0	1315.0	979.0	993.0	965.0	1161.0	1006.0	1060.0	
瘦肉 精	2689.0	2429.0	2721.0	2108.0	0.0	1977.0	535.0	261.0	149.0	47.0	11.0	1.0	0.0	0.0	36.0	13.0	1034.0	1186.0	507.0	535.0	696.0	771.0	872.0	1231.0	1142.0	1048.0	1187.0	1147.0	911.0	1186.0	
西布 特羅	2667.0	2879.0	2434.0	2499.0	1977.0	0.0	305.0	146.0	326.0	7.0	1.0	0.0	0.0	0.0	0.0	30.0	3.0	950.0	1270.0	473.0	777.0	940.0	574.0	683.0	1178.0	994.0	976.0	1291.0	1158.0	1563.0	1288.0
民進 黨	2042.0	249.0	361.0	242.0	535.0	305.0	0.0	2057.0	1578.0	1695.0	2087.0	2062.0	3260.0	3232.0	3019.0	2941.0	2607.0	346.0	126.0	121.0	221.0	75.0	39.0	93.0	388.0	220.0	177.0	186.0	75.0	157.0	248.0
立委	738.0	203.0	157.0	117.0	261.0	146.0	2057.0	0.0	854.0	948.0	1106.0	1111.0	1759.0	1741.0	1627.0	1666.0	1398.0	272.0	113.0	55.0	182.0	22.0	56.0	83.0	155.0	120.0	127.0	101.0	40.0	67.0	117.0

<

全部顯示

數據下載次數: 11/資料

數據下載次數: 1/Data

數據下載次數: 100/資料

>

再計算共現矩陣後方連接單中心網路圖元件，並依下圖系統內建設定執行。

單中心網路圖 (Shiny) (25)

參數設定 Input: 29 任務結果

選擇關鍵字檔案 0
-----請選擇-----

DTM *
19

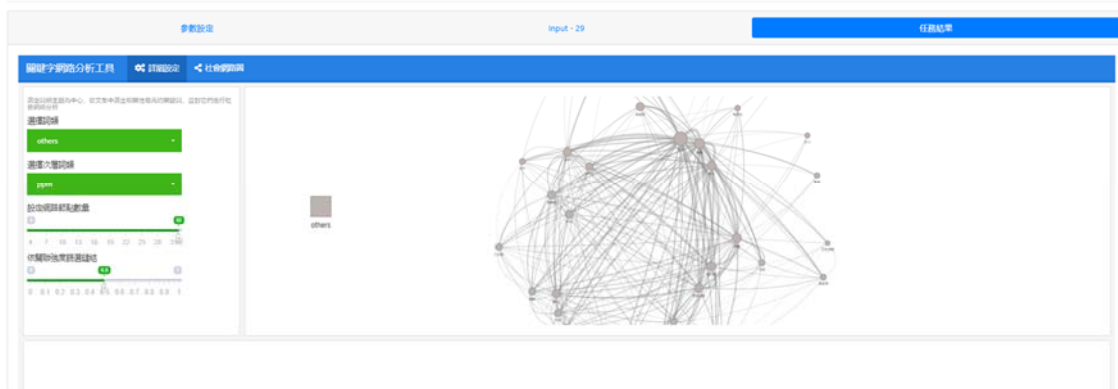
原始檔案元件 *
?

共線/相關性矩陣 *
29

儲存設定

觀看任務結果，系統已完成單中心網路圖繪製(有關圖文解釋詳見第四段內容)，後續再依實際需要設定節點數量及關聯強度鏈結，修改單中心網路圖呈現的內容。

單中心網路圖 (Shiny) (25)



9. 計算相關性矩陣並完成字詞網路圖：

在轉為DTM元件後方再連接計算相關性矩陣元件，跑出任務結果如下圖。

計算相關性矩陣 (23)

參數設定 Input: 19 任務結果

統計資訊

200
字數

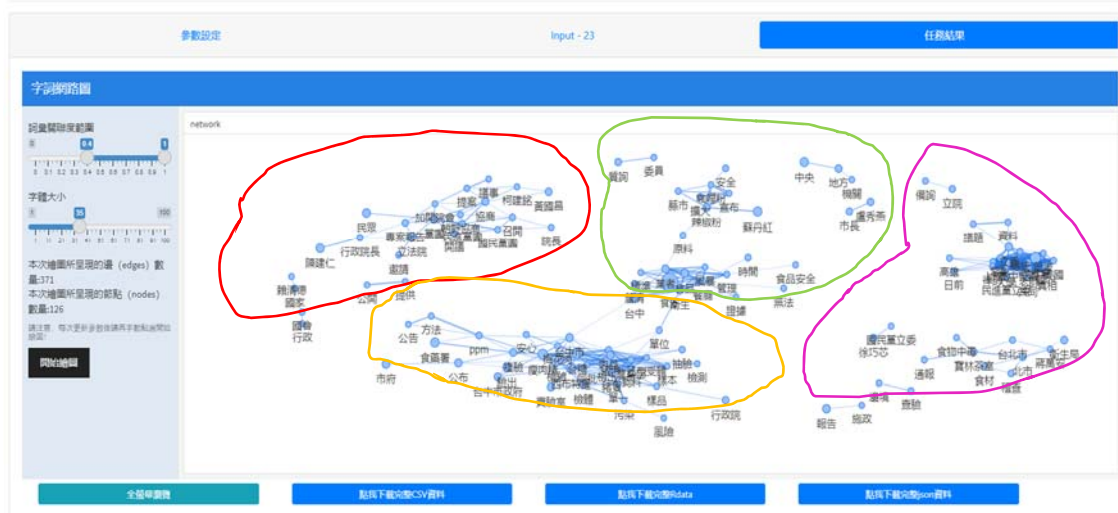
任務結果

Show 10 entries

Index	國內	校園	老師	大學	高中	律師	上地	電影	真相	風暴	民進黨	總統	美國	民進黨立委	立委	共匪	唐國	日報	市長	油價	工作	柯文哲	國民黨團	陳鴻
國內	1.000000	0.878929	0.864621	0.875610	0.891244	0.893948	0.891997	0.894332	0.640900	0.687767	0.630881	0.492450	0.671047	0.419979	0.464231	0.461830	0.368383	0.398270	0.280661	0.381732	0.246077	0.225361	-0.027205	-0.032876
校園	0.878929	1.000000	0.934997	0.944754	0.965585	0.963343	0.962131	0.963693	0.692728	0.756628	0.703494	0.535763	0.728679	0.465649	0.316356	0.438424	0.330594	0.398996	0.305171	0.397213	0.301176	0.256129	-0.014743	-0.023127
老師	0.864621	0.934997	1.000000	0.954041	0.970059	0.969064	0.969958	0.970244	0.695315	0.733372	0.718972	0.536496	0.724361	0.468364	0.323296	0.424798	0.327744	0.394588	0.308435	0.395895	0.296974	0.278747	-0.020188	-0.014221
大學	0.875610	0.944754	0.954041	1.000000	0.978281	0.979739	0.980180	0.980655	0.704492	0.741209	0.722782	0.554470	0.739402	0.468340	0.354286	0.431945	0.328636	0.420463	0.313387	0.404349	0.310365	0.289674	-0.018780	-0.021749
高中	0.891244	0.965585	0.970059	0.978281	1.000000	0.996573	0.996684	0.997381	0.717074	0.759353	0.735528	0.554819	0.746325	0.477496	0.332121	0.439944	0.335714	0.413642	0.324134	0.410282	0.295226	0.271523	-0.018164	-0.018165
律師	0.893948	0.963343	0.969064	0.979739	0.996573	1.000000	0.997211	0.996510	0.718754	0.756417	0.732569	0.556358	0.747578	0.479937	0.333357	0.440580	0.336276	0.413867	0.325361	0.405579	0.295742	0.268830	-0.017824	-0.017825
上地	0.891997	0.962131	0.969958	0.980180	0.996684	0.997211	1.000000	0.999181	0.718384	0.758767	0.733341	0.555419	0.749030	0.478352	0.332904	0.441719	0.337356	0.414355	0.322775	0.405001	0.294739	0.270033	-0.017086	-0.017087
電影	0.640900	0.687767	0.630881	0.671047	0.630881	0.671047	0.630881	1.000000	0.720129	0.737687	0.733385	0.556967	0.748715	0.481248	0.533349	0.444141	0.337317	0.416668	0.322709	0.405015	0.296696	0.270030	-0.015776	-0.016434

全部結果 顯示下載連結(10個) 顯示下載連結(10個) 顯示下載連結(10個) 顯示下載連結(10個)

然後再計算相關性矩陣元件後方連接字詞網路圖元件後執行，跑出任務結果後設定詞彙關聯度範圍為1到0.4、字體大小35，按下開始繪圖呈現如下圖(有關圖文解釋詳見第四段內容)，可以看出概略分成四大區塊(主題)。



10. 關聯式文字雲：

在轉為DTM後方再連接關聯式文字雲元件，並依下圖設定後執行。

三 關聯式文字雲 (shiny) (27)

參數設定	Input - 19	任務結果
<div>分群數 *</div> <div><input type="text" value="20"/></div>	<div>文字雲顯示字數 *</div> <div><input type="text" value="500"/></div>	
<div>迭代次數(最少250次)</div> <div><input type="text" value="1000"/></div>	<div>距離計算公式 *</div> <div><input type="text" value="euclidean"/></div>	
<div>聚合演算法 *</div> <div><input type="text" value="weighted"/></div>		
<div>儲存更改</div>		

觀看任務結果，完成關聯式文字雲的繪製如下圖（有關圖文解釋詳見第四段內容）。

三 關聯式文字雲 (shiny) (27)



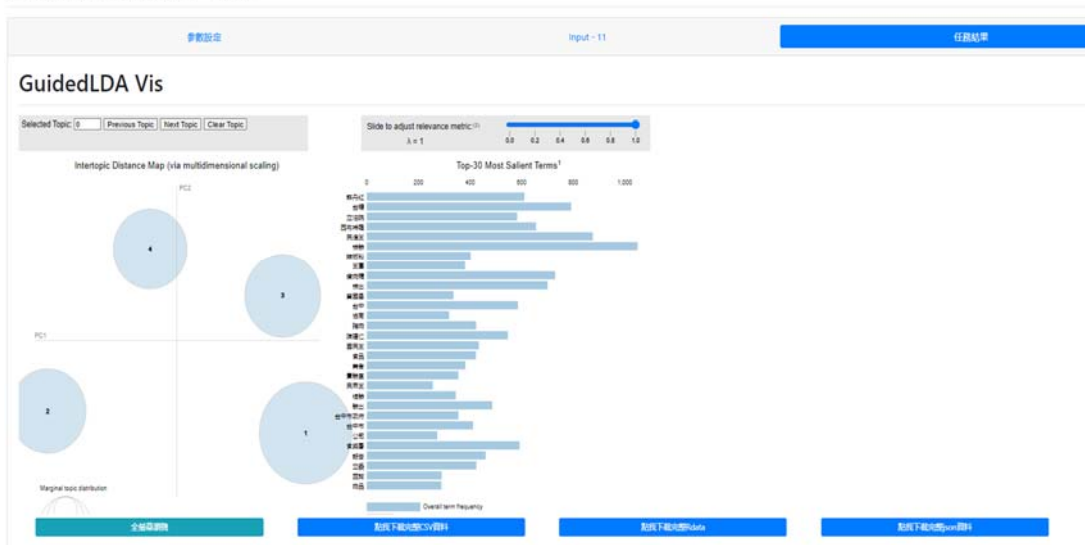
11. GuidedLDA主題模型分類出4大主題：

在清除停用詞後方接上GuidedLDA主題模型元件，依下圖設定迭代次數50、主題數為4、主題保留關鍵字數量20、詞彙頻率下限20、詞彙頻率上限0.5、alpha及Beta使用預設值，主題種子字則依關聯式文字雲及字詞網路圖分別設定各主題的10個種子字後執行。

GuidedLDA 主題模型 (21)

觀看執行結果如下圖，可以看出剛剛的設定已很清楚地把4大主題明顯區分出來(有關圖文解釋詳見第四段內容)。

GuidedLDA 主題模型 (21)



後續在接上新增欄位(idxmax/min)元件，依下圖設定匯總函數max、計算欄位選定0-3共4個主題，並新增欄位名稱為class後執行。

新增欄位 (idxmax/min) (36)

參數設定

Input - 21

任務結果

匯總函數

max

新增的欄位名稱

class

計算欄位(按住ctrl(Windows)或command(MAC)可以拖選)

system_id

0

1

2

3

儲存設定

完成任務結果如下圖所示。

新增欄位 (idxmax/min) (36)

參數設定

Input - 21

任務結果

任務結果

Show 10 entries

Search:

system_id	0	1	2	3	class
1	0.000096	0.154713	0.263794	0.581408	3
2	0.993007	0.002331	0.002331	0.002331	0
3	0.678922	0.004902	0.004902	0.311275	0
4	0.996896	0.001035	0.001035	0.001035	0
5	0.299682	0.001590	0.438792	0.259936	2
6	0.000543	0.998371	0.000543	0.000543	1
7	0.149085	0.014332	0.001076	0.835307	3
8	0.000386	0.067914	0.140266	0.791434	3
9	0.000659	0.000659	0.000659	0.998022	3
10	0.012204	0.000239	0.000239	0.987318	3

Showing 1 to 10 of 100 entries

Previous

1

2

3

4

5

...

10

Next

在新增欄位(idxmax/min)後方接上分群彙總(非數值)元件，並依下圖設定。

分群彙總 (非數值) (38)

參數設定

Input - 36

任務結果

使用...欄位進行分群(按住ctrl(Windows)或command(MAC)可以拖選)

system_id

0

1

2

3

class

計算欄位(按住ctrl(Windows)或command(MAC)可以拖選)

system_id

0

1

2

3

class

匯總函數

count

unique

min

max

first

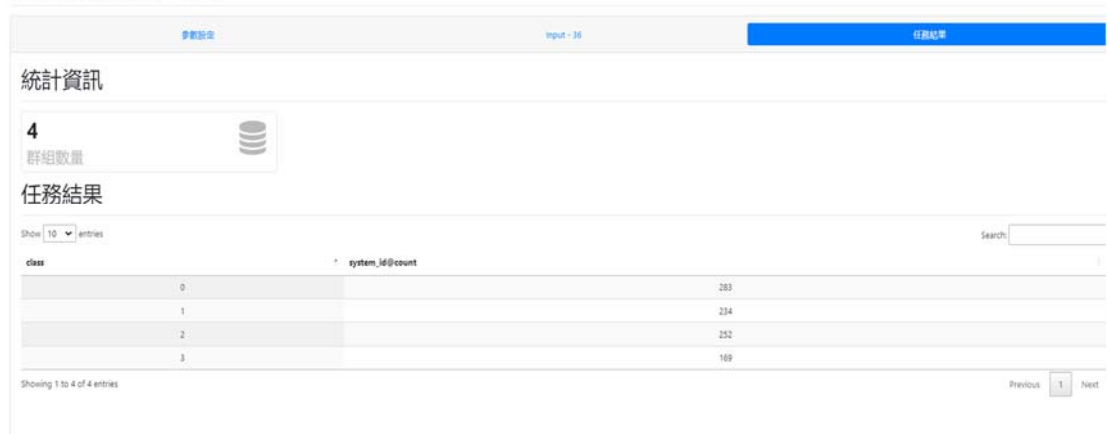
last

sum

儲存設定

觀看任務結果，完成計算4大主題中每個主題的文章數量。

分群總覽 (非數值) (38)



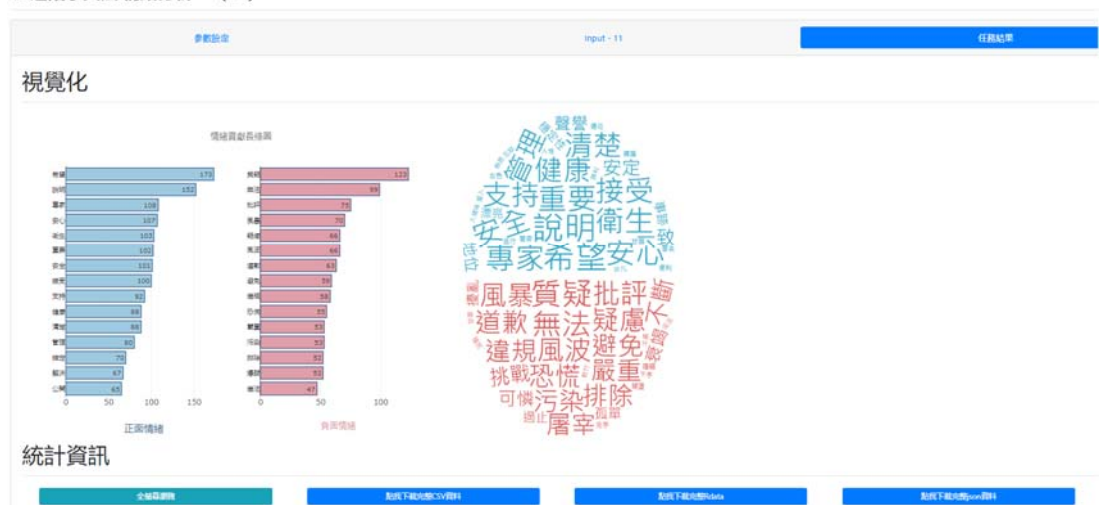
12. 進階字典法情緒分析：

在清除停用詞後方連接進階字典法情緒分析元件，並依下圖選取字典NTUSD，自行定義正面及負面詞彙，並設定要使用否定詞及加強詞。

進階字典法 情緒分析 (34)

觀看任務結果如下圖(有關圖文解釋詳見第四段內容)。

進階字典法 情緒分析 (34)



13. 分析結果與原始資料合併後經欄位篩選及重構後完成正負情緒統計

把CSV檔匯入元件及進階字典法情緒分析元件連接合併資料，利用兩筆資料的system_id來合併內容。

合併資料 (40)

觀看任務結果，已完成資料合併如下圖。

合併資料 (40)

system_id	artDate	artContent	artTitle	positive_count	positive_words	negative_count	negative_words	booster_count	booster_words	negator_count	negator_words	sentiment_value
1	2024/5/15 09:24	66中兩軍和牛境內保護。2024年4月被檢獲，共獲新鮮66增110萬元。當晚被檢獲。遭到這批食安局事件，該公司應到化處理以下問題：一、檢獲一批批。檢中經警參考。該批我們繼續尋下至。Y66中市食安處增處10家。	你總覺得不覺得西。萬。食安處。檢10次化。技巧。和牛。境內保護。食安處。警事件等	90	[技巧, 培養, 預防, 發展, 合作, 新鮮, 優秀, 訓練, 請求, 陪伴, 了解, 安心, 完整, 警備, 管理, 治療, 預防, 主動, 衛生, 認真, 最佳, 有效, 新裝, 實際, 掌握...]	55	[混亂, 腐爛, 違反, 避免, 危險, 霉壞, 不足, 不法, 開除, 疑慮, 無法, 導航, 計劃, 警備, 警發, 擔心, 對於, 申訴, 憤怒, 責任, 治罪, 無能, 否認, 徵求, 嚴重...]	0	[]	0	[]	35.00000
2	2024/5/14 02:20	真可怕跟威士忌+那! 韓國咖啡室咖啡社 囉「學小鞋」 被檢獲市審判所成立審判之庭為事有保護。警發起韓國檢區增處市審判所政警防範的防範。警發之在TVS55防範節目 (TVS55防範) 中通過。韓國被檢獲到所...	跟威士忌+那! 韓國咖啡室咖啡社 囉「學小鞋」 被檢獲市審判所成立審判之庭為事有保護。警發起韓國檢區增處市審判所政警防範的防範。警發之在TVS55防範節目 (TVS55防範) 中通過。韓國被檢獲到所...	3	[期望, 實際, 確實]	8	[可怕, 防範, 被檢, 逃竄, 沒想到, 警發, 不對勁, 認真]	0	[]	0	[]	-5.00000
3	2024/5/17 11:16	警報通知: 國際貿易發展委員會的決議 已全案統一口徑, 反對日本核食進口, 所謂「開放進口」這正是要金銀打博狗。通關的利權, 小案企業萬能, 反對日本核食的時候, 增獲保知。查與訪少量而位候選人, 位提到增獲保知...	政通暢的: 保友黨為何。拒絕爭議。[審判之庭] ?	1	[金案]	5	[反對, 麻煩, 慘狀]	0	[]	0	[]	-2.00000
4	2024/2/28 00:06	柯P連任黨主席會。林進春: 沒興趣跟實數說記者陳炳輝。/ 台北經濟 和學見114小時VOCnews4已新聞 民視黨主席柯文政致函後。續前在立法院黨團開會。更被質疑...	柯P連任黨主席會。林進春: 沒興趣跟實數說記者陳炳輝。/ 台北經濟 和學見114小時VOCnews4已新聞 民視黨主席柯文政致函後。續前在立法院黨團開會。更被質疑...	8	[確實, 容易, 解決, 真正, 理解, 自主, 整理, 務理]	7	[難免, 質疑, 困難, 放任, 擔憂, 負累, 忘記]	0	[]	0	[]	1.00000

在合併資料後方連接欄位篩選元件，並依下圖設定要保留的欄位system_id、artDate、artTitle、positive_count、booster_count、negative_count、negator_count及sentiment_value後執行。

欄位篩選 (43)

觀看任務結果，完成欄位篩選如下圖。

欄位篩選 (43)

任務結果

Show 10 entries

system_id	artDate	artTitle	positive_count	negative_count	booster_count	negator_count	sentiment_value
1	2024/5/15 03:24	你等開氣血不啱就訂罰萬，食安局擬10大化驗技巧，和中環內餐廳、食安區舉辦餐事件	90	55	0	0	35.00000
2	2024/3/4 02:20	跟喝威士忌+茶！韓國咖啡室新位牌「學小龍」被點齊內幕	3	8	0	0	-5.00000
3	2024/1/17 11:16	歡迎轉封：保潔為何唔過淨潔「停止進口日本核食」？	1	3	0	0	-2.00000
4	2024/2/28 00:06	柯P連派生黨會，林進春：這陣台灣實數受	8	7	0	0	1.00000
5	2024/3/1 17:11	有証有人不被技術和說，是不係無能無能出台灣電？	13	9	0	0	4.00000
6	2024/2/9 02:38	台灣國內國中來電指出！正有李繼博有台中市府：應舉辦食安講座	4	14	0	0	-12.00000
7	2024/2/21 06:59	高雄轉報和出動升紅，進口零件北南黃金，食安問題最大就是中國菜	1	4	0	0	-3.00000
8	2024/3/8 00:16	蘇丹紅黑草再擴大 與白胡椒(黑)中網	13	8	0	0	5.00000
9	2024/3/8 00:55	維多維進口蘇丹紅黑草 九華公司登記同一地址	3	4	0	0	1.00000
10	2024/3/8 16:02	不丹蘇丹紅黑草轉報，他國粉，白胡椒，等等調味品「全球進入黑名單」	14	15	0	0	-1.00000

Showing 1 to 10 of 100 entries

在欄位篩選後方接上欄位重構(Melt)元件，並設定下圖設定，把要轉換的positive_count及negative欄位，轉成新的欄位名稱melt_sentiment_count(其值設定為melt_sentiment_count_value)，並保留artDate欄位後執行。

正負面詞彙詞頻 (45)

要轉換的欄位(按住ctrl(Windows)或command(MAC)可以複選)

- artDate
- artTitle
- positive_count
- negative_count
- booster_count
- negator_count
- sentiment_value

轉換的欄位新的欄位名稱

melt_sentiment_count

要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選)

- artDate
- artTitle
- positive_count
- negative_count
- booster_count
- negator_count
- sentiment_value

值的欄位名稱

melt_sentiment_count_value

儲存更改

觀看任務結果，完成每日正負面詞彙詞頻統計如下圖。

正負面詞彙詞頻 (45)

任務結果

Show 10 entries

system_id	artDate	melt_sentiment_count	melt_sentiment_count_value
1	2024/5/15 03:24	positive_count	90
2	2024/3/4 02:20	positive_count	3
3	2024/1/17 11:16	positive_count	1
4	2024/2/28 00:06	positive_count	8
5	2024/3/1 17:11	positive_count	13
6	2024/2/5 02:38	positive_count	4
7	2024/2/21 06:59	positive_count	1
8	2024/3/8 00:16	positive_count	13
9	2024/3/8 00:55	positive_count	5
10	2024/3/8 16:02	positive_count	14

Showing 1 to 10 of 100 entries

全部欄位 點以下欄位CSV資料 點以下欄位Data 點以下欄位Data資料

在欄位重構(Melt)後方接上分群彙總(日期)元件，並依下圖完成參數設定後執行。

分群彙總(日期) (47)

參數設定

Input - 45

任務結果

選擇日期欄位 *

artDate

選擇日期類型 *

日

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

artDate

melt_sentiment_count

melt_sentiment_count_value

日期格式 *

%Y-%m-%d %H:%M:%S%z

匯總函數 *

sum

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) *

system_id

artDate

melt_sentiment_count

melt_sentiment_count_value

儲存更改

觀看任務結果，完成每日正負面情緒值統計如下圖。

分群彙總(日期) (47)

參數設定

Input - 45

任務結果

統計資訊

226

群組數量

任務結果

Show 10 entries

Search:

melt_sentiment_count	artDate	melt_sentiment_count_value
negative_count	2024-01-01	8
negative_count	2024-01-02	67
negative_count	2024-01-03	65
negative_count	2024-01-04	78
negative_count	2024-01-05	66
negative_count	2024-01-07	72

全部單選

點選下載CSV資料

點選下載JSONData

點選下載JSON資料

- 欄位篩選後分別經資料篩選完成每日正負面文章趨勢統計：
在欄位篩選後方分別接上兩個資料篩選元件，其中1個條件式設定sentiment_value>0，另1個設定sentiment_value<0。

資料篩選 (49)

參數設定

Input - 43

任務結果

條件式 *

sentiment_value>0

儲存更改

資料篩選 (51)

參數設定

Input - 43

任務結果

條件式 *

sentiment_value<0

儲存更改

分別執行後，觀看執行結果，分別篩選出正面文章及負面文章內容如下圖。

資料篩選 (49)

資料篩選

Input - 43

任務結果

統計資訊

505
移除數量

433
保留數量

任務結果

Show 10 entries

Search

system_id	artDate	artTitle	positive_count	negative_count	booster_count	negator_count	sentiment_value
1	2024/5/15 09:24	你發現餐廳不想被訂百萬，食安危機10大化解技巧，和串連內保，食安風暴新聞事件簿	90	55	0	0	35.000
4	2024/2/28 00:06	相連民衆生面會，林俊聖：這舞台確實難受	8	7	0	0	1.000
5	2024/3/1 17:11	有國籍人士被控刑罰，應不應對刑罰出外轉電？	13	9	0	0	4.000
8	2024/3/8 00:16	蘇丹紅風暴再擴大，黑白場域+毒品中藥	13	8	0	0	5.000
9	2024/3/8 00:55	檢察官進口蘇丹紅藥劑，九家公同登記同一地址	5	4	0	0	1.000
11	2024/3/14 12:16	國民黨律師質疑林俊聖食安會經被蘇丹紅打臉，廖偉明一神話已次發話	10	5	0	0	5.000
15	2024/3/17 09:56	六位選到的貴賓來！讓事得到少數黨關注，吳思遠算計：不應要時間+香	4	1	0	0	3.000
19	2024/4/8 12:34	你（妳）的人生需要被洗腦？	17	11	0	0	6.000
20	2024/4/12 13:12	陳政府口說無憑了！卓榮泰宣布「五打七案」讓入打這智慧科技局	17	3	0	0	14.000

全部篩選

點此下載內容CSV資料

點此下載內容xlsx

點此下載內容json資料

資料篩選 (51)

資料篩選

Input - 43

任務結果

577
移除數量

361
保留數量

任務結果

Show 10 entries

Search

system_id	artDate	artTitle	positive_count	negative_count	booster_count	negator_count	sentiment_value
2	2024/3/4 02:20	張曉斌土生+茶：韓國咖啡賣到地庫，像「穿小鞋」被防衛內幕	3	8	0	0	-5.00000
3	2024/1/17 11:16	政選檢討：侯友宜為何拒絕承認「停止進口日本核食」？	1	3	0	0	-2.00000
6	2024/2/3 02:38	核轉疫內閣中央等推出！王金平繼續與台中市府：應嚴密食安措施	4	14	0	0	-10.00000
7	2024/2/21 06:59	高雄轉輸輸出蘇丹紅，進口商在野北供職金絲，食安問題最大就是中國菜	1	4	0	0	-3.00000
10	2024/3/8 16:02	不只蘇丹紅毒轉輸，咖哩粉、白胡椒、等等調味品「全被捲入影響區大	14	15	0	0	-1.00000
12	2024/3/4 05:59	黃金寶轉輸輸出到當發刊雜誌，陳光內穿牆：真想「交換」？	5	14	1	0	-6.70700
13	2024/3/16 10:47	公營不分青紅皂白修理路障！黃瑞明：公營管理別真，假裝中立的評述業績	9	10	0	0	-1.00000
14	2024/3/15 05:00	最嚴政權交棒，蔡政府官員歸罪，潘漢、羅社	15	23	0	0	-6.00000
16	2024/3/17 09:17	民進黨的「韓國咖啡地位」好不了	16	24	1	0	-7.96335
17	2024/4/2 13:19	半部黨能怎麼輸出？請市府私下拜託林大豐阮	9	14	0	0	-5.00000

Showing 1 to 10 of 100 entries

Previous

1

2

3

4

5

...

10

N

全部篩選

點此下載內容CSV資料

點此下載內容xlsx

點此下載內容json資料

在不同正面文章及負面文章資料篩選後方分別接上分群彙總(日期)元件，分別依下圖完成設定。

每日正面文章趨勢 (53)

資料篩選

Input - 49

任務結果

選擇日期欄位

artDate

選擇日期類型

日

計算欄位(按住ctrl(Windows)或command(MAC)可以複選)

system_id
artDate
artTitle
positive_count
negative_count
booster_count
negator_count

日期格式

%Y-%m-%d %H:%M:%S%z

匯總函數

count

保留欄位(按住ctrl(Windows)或command(MAC)可以複選)

artDate
artTitle
positive_count
negative_count
booster_count
negator_count
sentiment_value

儲存設定

每日負面文章趨勢 (55)

參數設定

Input - 51

任務結果

選擇日期欄位 *
artDate

選擇日期類型 *
日

計算欄位(按住ctrl(Windows)或command(MAC)可以拖選) *
system_id
artDate
artTitle
positive_count
negative_count
booster_count
negator_count

日期格式 * ❶
%Y-%m-%d %H:%M:%S%z

匯總函數 * ❷
count

保留欄位(按住ctrl(Windows)或command(MAC)可以拖選) *
artDate
artTitle
positive_count
negative_count
booster_count
negator_count
sentiment_value

儲存更改

分別觀看任務結果，完成每日正面及負面文章趨勢如下圖。

每日正面文章趨勢 (53)

參數設定

Input - 49

任務結果

279
群組數量

任務結果

Show 10 entries

Search:

artDate	sentiment_value	system_id
2024-01-02	2.00000	1
2024-01-08	5.00000	1
2024-01-09	2.00000	1
2024-01-09	7.00000	1
2024-01-09	15.00000	1
2024-01-20	5.00000	1
2024-01-31	1.00000	1
2024-01-31	17.00000	1
total: 85.000	75.00000	8

全部群組欄

點此下載完整CSV資料

點此下載完整Data

點此下載完整json資料

每日負面文章趨勢 (55)

參數設定

Input - 51

任務結果

統計資訊

244
群組數量

任務結果

Show 10 entries

Search:

artDate	sentiment_value	system_id
2024-01-01	-2.00000	1
2024-01-02	-19.00000	1
2024-01-03	-18.00000	1
2024-01-04	-19.00000	1
2024-01-04	-6.00000	1
2024-01-05	-19.00000	1
2024-01-07	-21.00000	1

全部群組欄

點此下載完整CSV資料

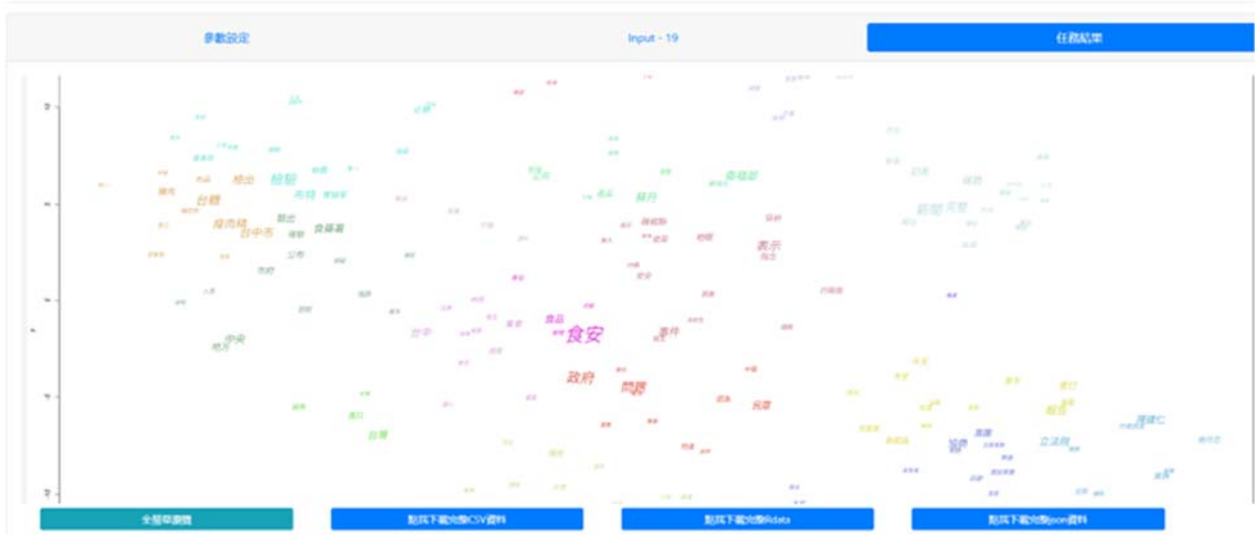
點此下載完整Data

點此下載完整json資料

四、視覺化的分析結果與解釋

經過前處理後首先化出文字雲，以直觀方式初步了解篩選出的字詞是否為有意義的字詞，轉為DTM後，繪製出關聯式文字雲

三 關聯式文字雲 (shiny) (27)

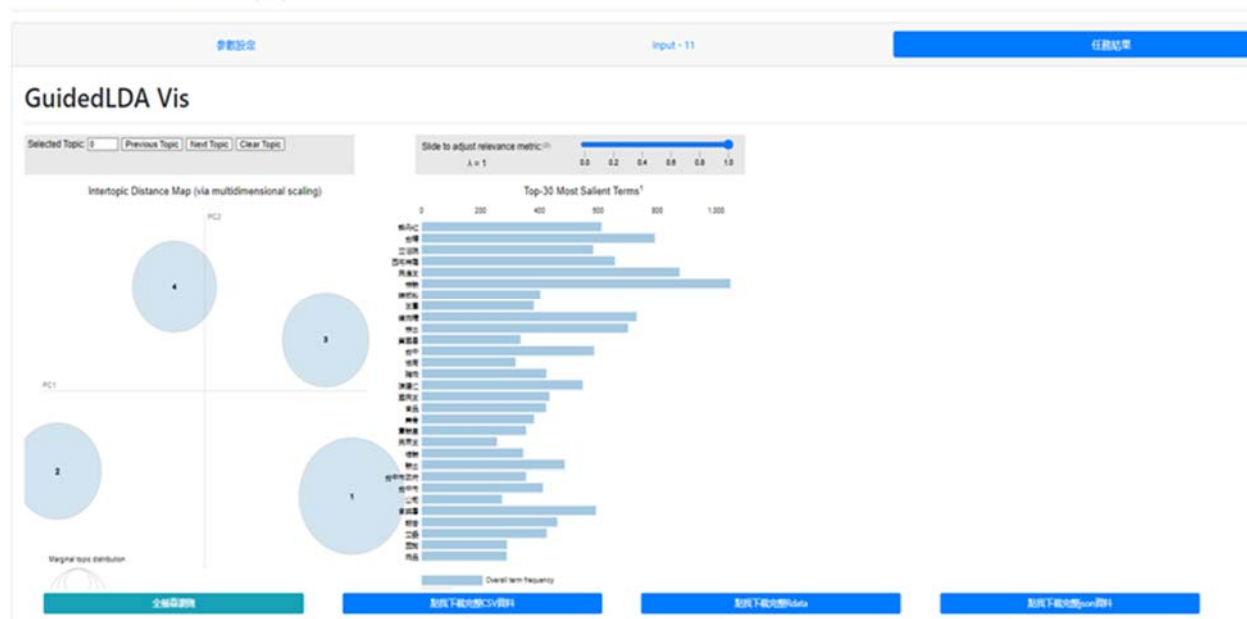


依顏色及分區可看出食安，瘦肉精、辣椒粉與民進黨較顯眼且分散，接者以計算共現矩陣繪製出單中心網路圖了解字詞強度。另計算相關性矩陣，依關聯性繪製出字詞網路圖，

三 字詞網路圖 (Shiny) (32)

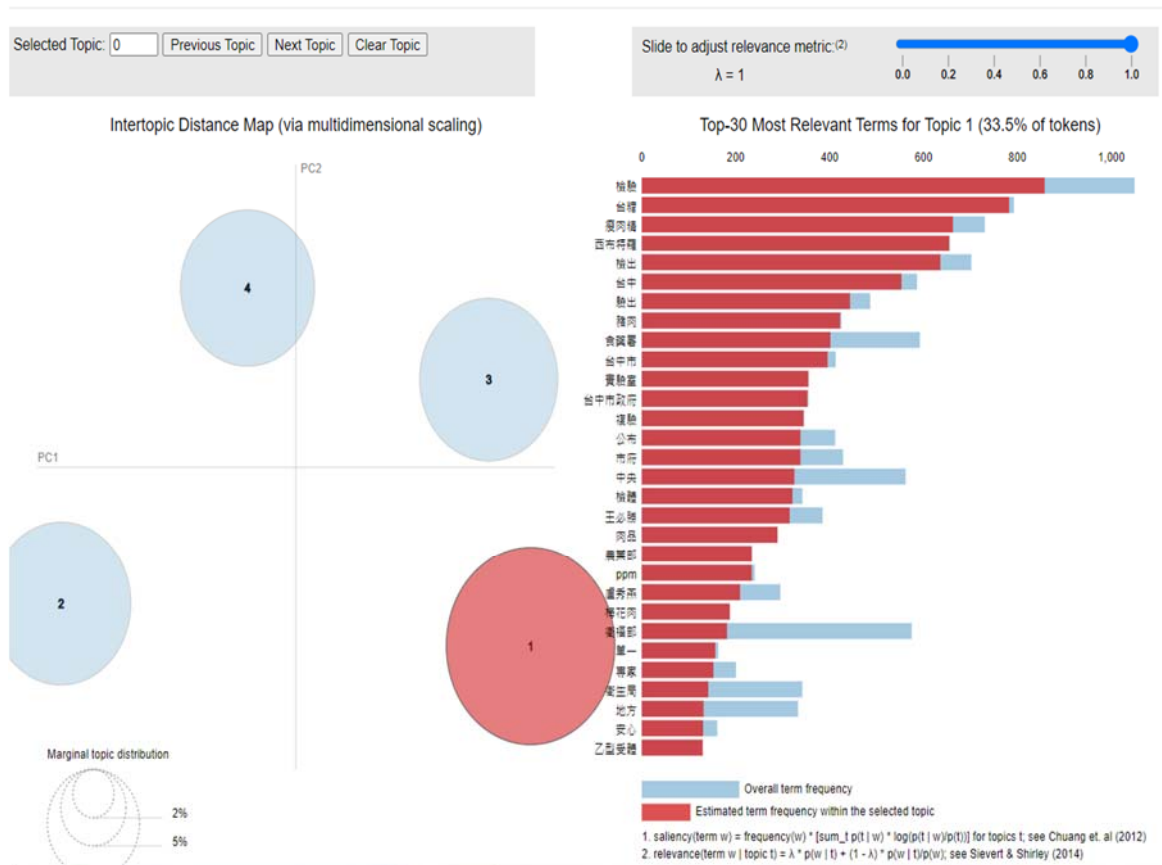


可知在0.4-1關聯性下可分為四群，後將四群主題加入種子分別為，辣椒粉，瘦肉精，辣椒粉，食安報告字導入GuidedLDA主題模型找出各主題關鍵字

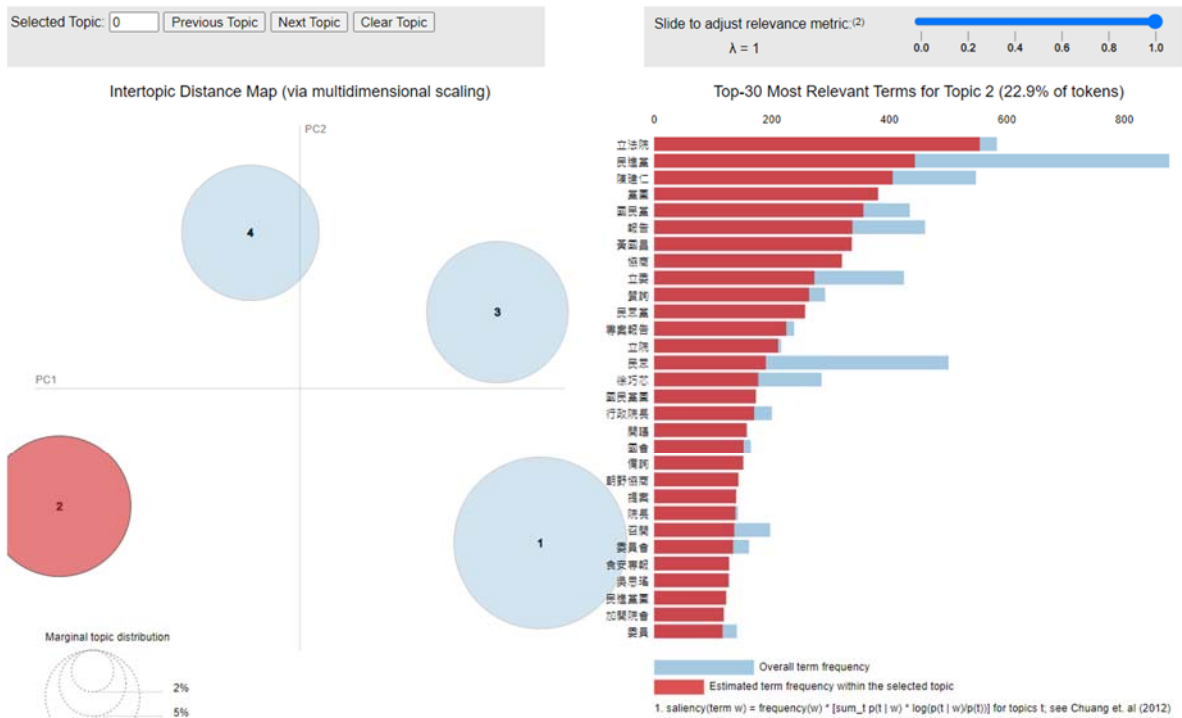


可看出4主題分布無重疊，分類出的各主題關鍵字與文本重疊度高，表示分類出的辨識度高。

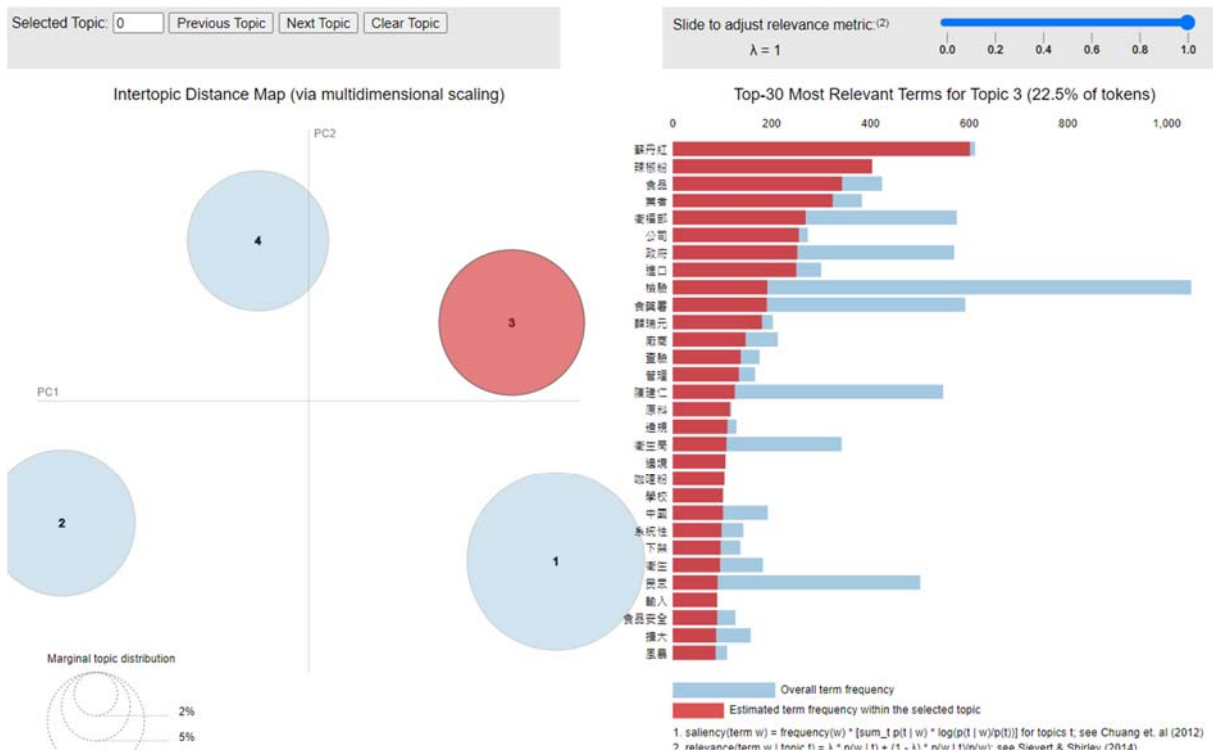
● 瘦肉精



GuidedLDA Vis

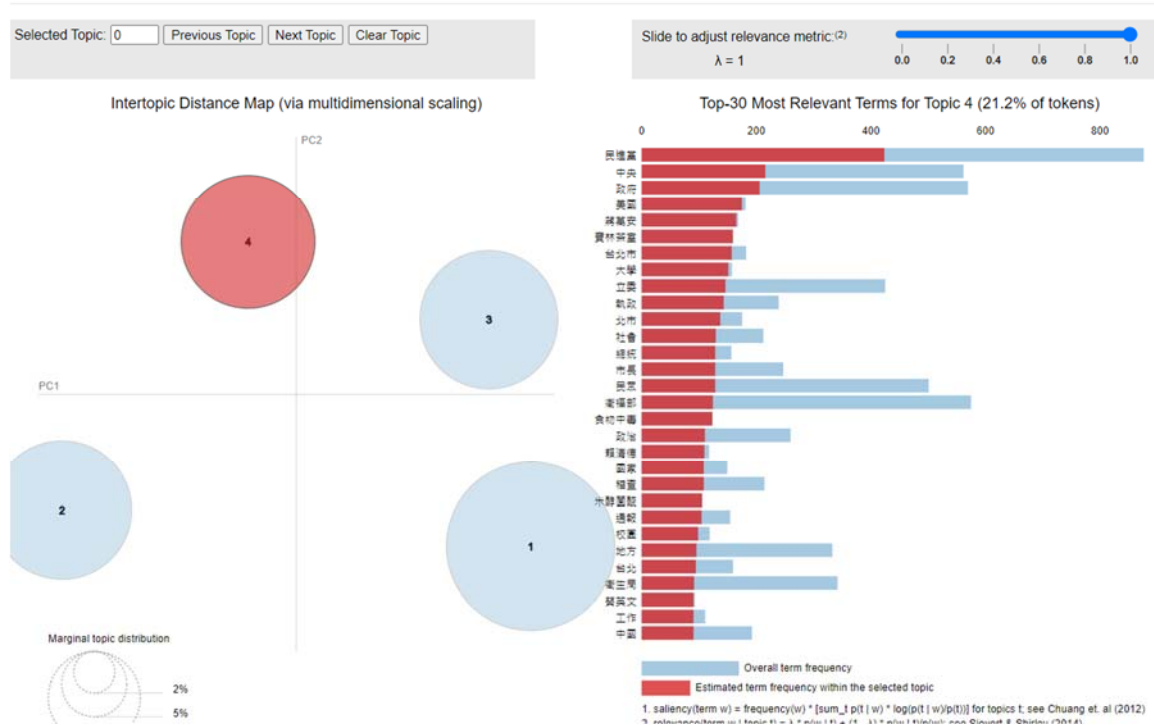


GuidedLDA Vis



● 食物中毒(寶林茶室)

GuidedLDA Vis



● 情緒分析：

以情緒分析字典法可畫出文本正負面情緒字文字雲及長條圖，可列出主要正面及負面詞，比較情緒貢獻長條圖數值，可知正面態度數值多於負面。

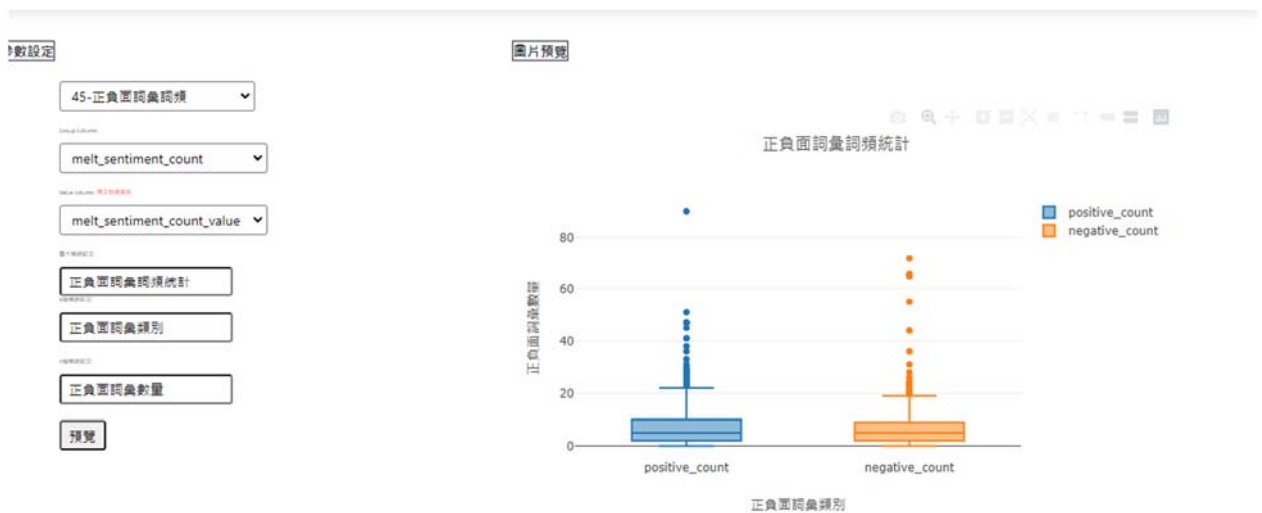
進階字典法 情緒分析 (34)



● 爬蟲文本中各主題數目表



● 統計文本正負面詞頻比較



圖示無顯著差異

● 依日期畫出每日正面文章統計圖

參數設定

圖片預覽

53-每日正面文章趨勢

日期

artDate

系統ID

system_id

聚合方式

sum

顯示數量

160

圖表標題

每日正面文章統計

日期

日期

正面情緒篇數

正面情緒篇數

預覽

每日正面文章統計



● 依日期畫出每日負面文章統計圖

參數設定

圖片預覽

55-每日負面文章趨勢

日期

artDate

系統ID

system_id

聚合方式

sum

顯示數量

160

圖表標題

每日負面文章統計

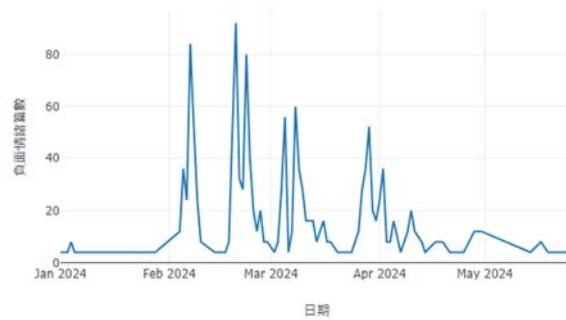
日期

負面情緒篇數

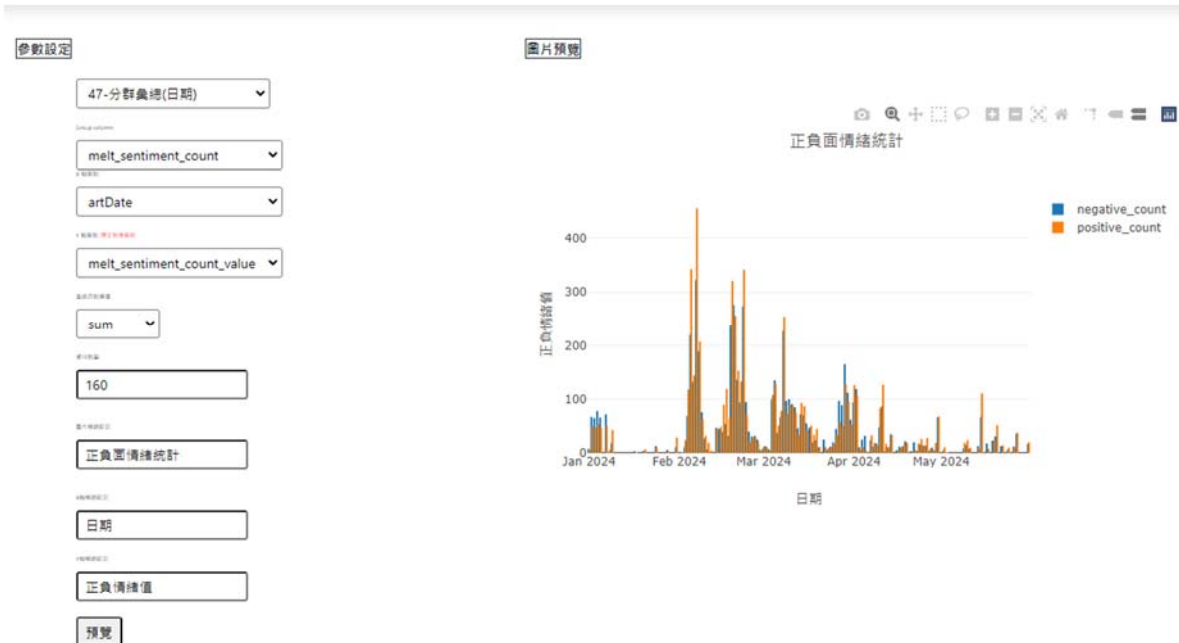
負面情緒篇數

預覽

每日負面文章統計

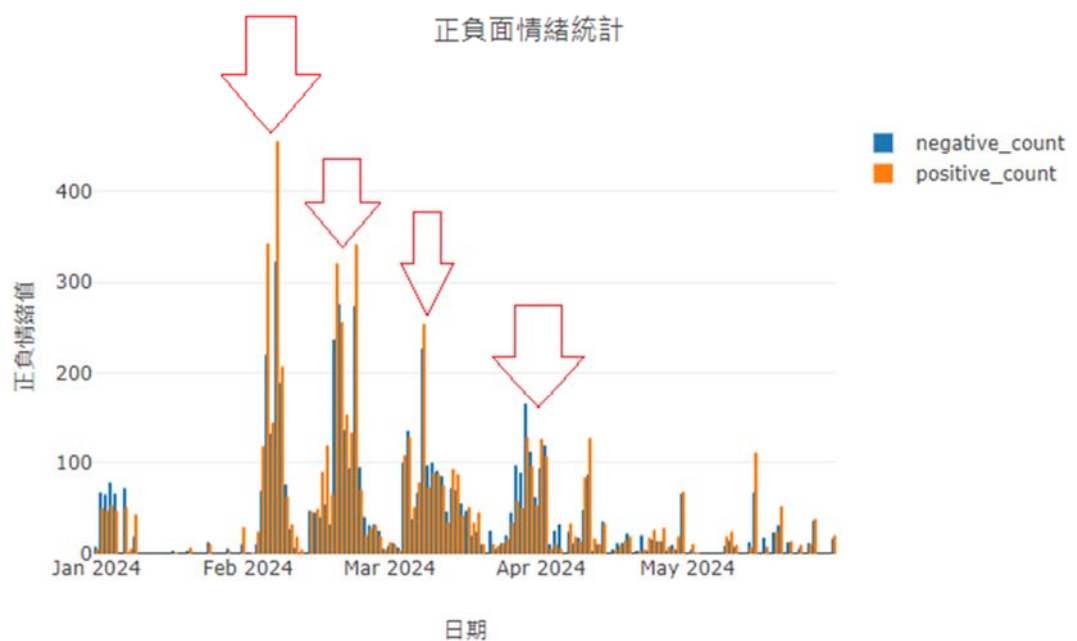


● 合併兩圖以長條圖比較



● 可看出有幾個時間點有高峰值

圖表預覽



可發現討論度高有4個時間，比對該時間點食安相關新聞

- 2024年2月台中瘦肉精事件
- 2024年3月辣椒粉含蘇丹紅事件
- 2024年3月政府食安五環報告
- 2024年4月寶林茶室食物中毒事件

五、結論

在這個課題當中我們使用了情緒分析，來判斷對於食安議題是否對於政府的舉措與方法有所不妥，而有明顯且激烈的負面情緒。

以下是過程與總結

- (1). 資料來源:我們從Dcard的美食版、閒聊版、時事版以及PPT的美食版、八卦版、政黑板與Mobile01的閒聊版、旅遊美食版、以及時事版的平台版面上(由TarFlow內建資料來源取得)進行資料的收集。
這些平台具有評論的公信力，並且有足夠的資料可以分析。
- (2)資料前置處理:資料範圍再以時間今年113年上半年的資料進行分析。
並且因為資料龐大並且需要特殊的篩選條件，我們還是透過PYTHON的程式進行處理
- (3). 資料清理的部分有兩點
 1. 利用關鍵字的方法進行資料清洗:有”中央政府”、”民進黨”、”執政”、”福衛部”達到我們想進行分析的高相關性資料
 2. 常規清理: 資料格式對齊、替換字串、中文段詞、清除停用詞
- (4). 資料分析結果
 1. 計算詞頻，過程中我們可以看到透過設定已經可以篩選出需要的字詞。
 2. 單中心網路圖，我們先將資料轉為DTM後進行分析，可以使用圖形的方式了解每個字詞與字詞之間的關聯性。
 3. 字詞網路圖，再來使用字詞間關聯的分析可以得出四類的關聯群體。
 4. 而關聯式文字雲的部分成效較不彰，沒有辦法明顯看出其明顯的意義。
 5. 再來我們使用主題模型，使用GuideedLDA的方式進行分析，與網路圖的結果相仿，區分出四大類型主題分別是食安報告議題、蘇丹紅報導、寶林茶室報導、與瘦肉精議題。
 6. 最後是重點，我們用進階字典法情緒分析，得出的結果是正面與負面評論沒有明顯的差異，不論是以單純數量比較、日期發文時間區隔再做數量比較，正負頻論都是持平。

我們以這些分析方法方式，解讀政府與食安危機的負面情緒是否有強烈正相關以及關聯性是否足夠緊密，來說明我們對於食安危機與政府舉措，造成不信任的假設是否趨於為真，以分析結果來看並無顯著結果。

最後，在文字分析中我們可以藉由簡單的分析視圖來達成視覺化的效

果，以利於評估與假設的成立與研究方向的確立，並且可以此結果進行更多的假設做為參考，如反向假設如是否食安危機是兩岸危機、是否跟政黨交替有關係、是否因為疫情關係造成跟食安相關專業技能的職缺日漸增多，而引起社會現象化的食安危機等，TARFLOW的文字分析雖然沒有辦法使用到LLM的處理，但對於有明確假設目標以及有更精確的定義的研究者來說，一樣是套有助於完成分析的工具。