

影片網址

[https://youtu.be/EbVRE\\_jxb0s](https://youtu.be/EbVRE_jxb0s)



# 社群媒體分析 第十組 期末專題 性騷擾議題分析

N104020001 李采容

N104020007 郭育雯

N104020021 馮慧嬌

N114320009 林育伈

# Agenda



Step 01

動機和分析目的



Step 02

資料集  
描述與分析



Step 03

視覺化的分析  
結果與解釋



Step 04

結論

# 動機與分析目的

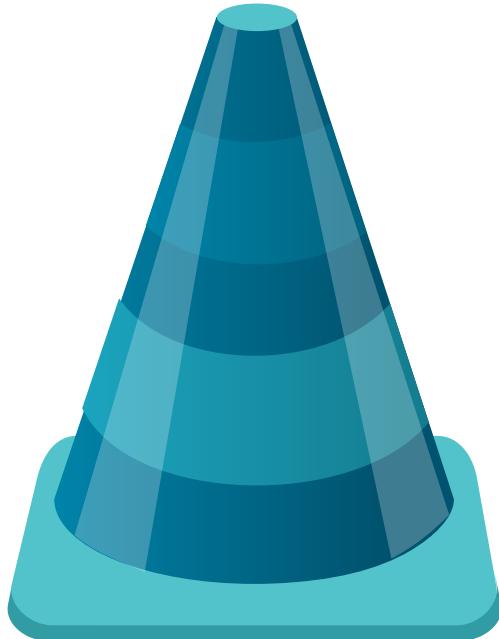


01

## 動機

性騷擾是一種不當的行為，包括言語、身體接觸或行為上的不當舉動，對於受害人造成困擾、恐懼和不適感。依據衛生福利部 2022 年統計資料顯示，近 5 年來性騷擾申訴件數逐年上升，性騷擾被害人 96% 為女性，且年齡為 30 歲以下者超過 60%。其實性騷擾可能會發生在工作場所、學校、社交場合或其他公共場所，且無論是男性或女性皆有可能成為受害者或加害者。2024 年台灣總統大選開跑之前，性騷擾議題意外成為了政治風暴。

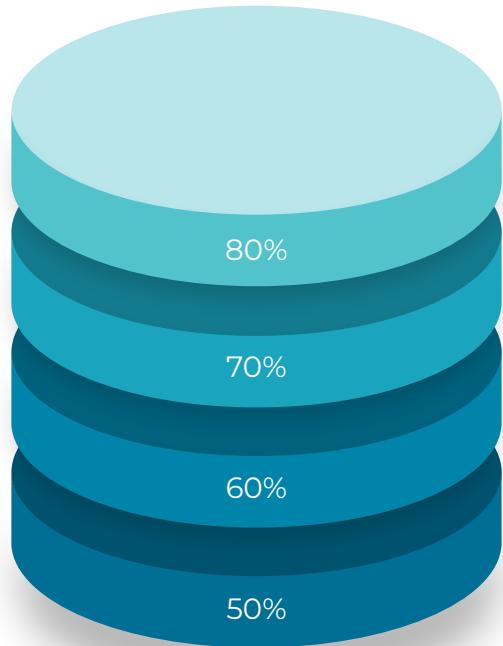
# 動機與分析目的



02

## 目的

性騷擾是一種嚴重的問題，應受到重視和譴責，現在有多個國家和地區制定了相關法律來禁止這種行為。本組希望能透過文字分析流程系統了解議題的參與討論狀況，探討民眾討論的情緒值，以及熱度分布狀況等情形。



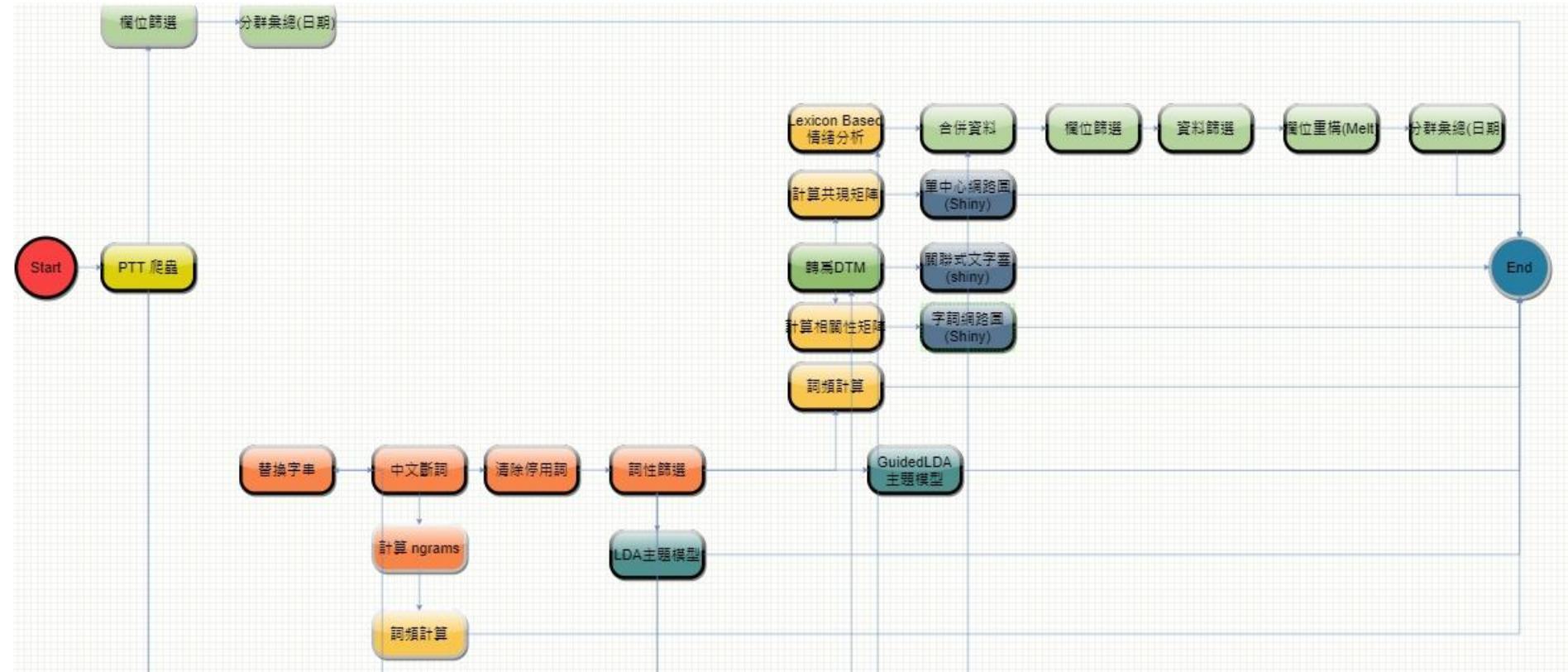
## 資料集描述

01 PTT八卦版、政黑版

02 2023/05/05 ~ 2023/06/08

03 關鍵字:性騷、性平

# 工作流程、視覺化儀表板名稱:Project2



參數設定

任務結果

## 選擇看板 \*

- GetMarry(結婚)
- Gossiping(八卦)
- graduate(研究所)
- GuineaPig(天竺鼠)
- Hamster(倉鼠)
- HardwareSale(硬體二手)
- HatePolitics(政黑)

## 搜尋起始日期

2023/05/05



儲存更改

## 搜尋關鍵字 ⓘ

性騷

性平

## 排除關鍵字 ⓘ

以換行區隔，e.g.

壽山動物園

猴子

...

## 搜尋結束日期

2023/06/08



**資料來源:** 選擇工作平台資料集「PTT爬蟲」，期間設定為 2023/05/05～2023/06/08，資料以八卦與政黑討論板為設定參數，主題為近期討論度較高的性騷與性平議題。

參數設定

任務結果

## 統計資訊

**10**

欄位數

**4104**

資料筆數



## 任務結果

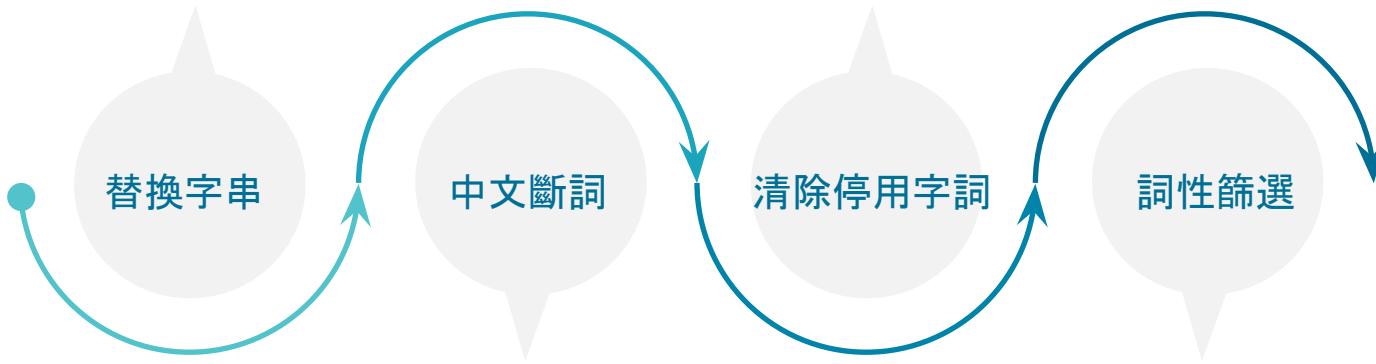
Show 10 entries

Search:

system_id	artUrl	artTitle	artDate	artPoster	artCategory	artContent	artComment	e_ip	insertedDate
1	<a href="https://www.ptt.cc/bbs/Gossiping/M.1683252674.A.708.html">https://www.ptt.cc/bbs/Gossiping/M.1683252674.A.708.html</a>	[新聞]加油站男員工建議紗裙女騎士「穿小短褲	2023-05-05 10:11:12	kent	Gossiping	1.媒體來源:\n聯合報\n\n2.記者署名:\n林孟潔 / 台北即時報導\n\n3.完整新聞標題:\n加油站男員工建議紗裙女騎士「穿小短褲更誘惑」性騷下場慘了\n\n4.完整新聞內容:\n一名男加...	[{"cmtStatus": "推", "cmtPoster": "kaitokid1214", "cmtContent": ":八卦-1", "cmtDate": "2023-05-05 10:11..."}]	111.249.157.234	2023-05-05 01:31:00

2023/05/05 ~ 2023/06/08 期間，統計資料筆數共計 4,104 筆。

# 資料清理



## 替换字串 (48)

參數設定

Input - 4

任務結果

選擇處理欄位 \*

artContent

選擇替換規則檔案 ⓘ

class.csv

替換字串設定 ⓘ

```
\n\n>> 。\n>> ,\n((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\&%\_\_/-~-]*?)>>\nSent from JPTT on my |w+>>
```

儲存更改

替換多餘字串(網址、PTT發文格式等)

## 替換字串 (48)

參數設定

Input - 4

任務結果

### 統計資訊

137624



取代數量

### 任務結果

Show 10 entries

Search:

system\_id

result

1

1.媒體來源:,聯合報。2.記者署名:,林孟潔 / 台北即時報導。3.完整新聞標題:,加油站男員工建議紗裙女騎士「穿小短褲更誘惑」性騷擾下場慘了。4.完整新聞內文:。一名男加油站員工前年向女機車騎士說「穿小短褲才對，很誘人」等語，被女方申訴性騷擾，擾，台北市政府認定成立性騷擾，裁罰1萬元，他不服提行政訴訟，台北地方法院判他敗，訴，可上訴。。女騎士2021年7月16日晚間10時許，騎車到加油站加油時，男員工見她穿粉紅紗裙內搭長，裙，稱「這件裙子裡面應該穿小短褲才對，很誘惑人，應該穿小短褲」、「很漂亮應該更，誘惑一點」等。。女騎士同年月18日向警局申訴性騷擾，經移送公司調查確認成立性騷擾，台北市政府依性，性騷擾防治法裁罰員工1萬元罰鍰；員工不服提訴願被駁回，再提行政訴訟救濟。。員工主張，女騎士沒有制止、反抗或表達不悅、驚慌，且加油站環境24小時人潮蜂擁，影，音監控，但女騎士毫無表達抗議；他認為稱讚客戶服飾是行銷常用套路，短褲是正常服飾，誘惑常用於形容美麗吸引力，必然與性或性別有關。。員工認為，他為了銷售而與客戶建立互動找話題尋找商機，而女騎士頭戴安全帽、口罩，，難判斷年齡相貌，稱讚服飾是唯一話題，過程10秒內要找到商機，實在來不及起邪念，請，求撤銷處分。。雖然函詢加油站後，員工確實是加油站個人銷售排名名列前茅，但法院根據採證光碟，確，認員工有講相關言論，內容也讓女騎士内心感受驚嚇、不舒服，構成性騷擾，市府裁罰沒，有違誤，判員工敗訴。。5.完整新聞連結 (或短網址)不可用YAHOO、LINE、MSN等轉載媒體：

2

板上有人po新聞時，若是性騷擾的社會案件時，下面推文都是八卦-1，但其實這類行為 西治板才一狗票吧？。每次看西治板的大多數文章，給我的感受就是，po文的人只要不尷尬 尷尬的就是看文的人。有沒有八卦？，

# 中文斷詞 (50)

×

參數設定

Input - 48

任務結果

選擇處理欄位 \*

result

定義詞彙 ⓘ

性騷擾 10000  
受害者 1000  
加害者 1000  
被害者 1000  
造浪者 1000

選取字典 ⓘ

-----請選擇-----

儲存更改

將不應被斷詞的字詞設定權重 (性騷擾、受害者、加害者等 )



## 清除停用詞 (52)

參數設定

Input - 50

任務結果

語言 \*

Chinese

是否清除單字元 ⓘ

是

清除英文字母 \*

否

清除換行符號 \*

是

清除html tag \*

是

使用預設停止詞

是

是否轉為小寫英文

是

清除數字 \*

是

清除特殊標點符號 \*

是

自定義停止詞

- ETtoday新聞
- 中時新聞網
- 聯合報
- 即時報導
- 轉錄網址

儲存更改

設定自定義停止詞(以新聞網名稱為主, 以及與主題無關之字詞 )



## 詞性篩選 (103)

x

參數設定

Input - 52

任務結果

語言 \*

Chinese

選擇保留詞性 \*

Noun  
Proper Noun  
Verb  
Adjective  
Adverb

儲存更改

為利於後續分析，只篩選名詞與專有名詞的部分作為分析來源。



# 資料分析

01

發文數量分析

02

詞頻計算、關聯式文字雲

03

情緒分析與視覺化圖表

04

單中心網路圖

05

字詞網路圖

06

主題模型

# 發文數量分析

## 欄位篩選 (27)

參數設定

Input - 4

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  
artUrl  
artTitle  
**artDate**  
artPoster  
artCategory  
artContent  
artComment

儲存更改

欄位篩選: 設定為system\_id & artDate

Input - 4

任務結果

參數設定

## 任務結果

Show 10 entries

Search:

system\_id

artDate

1		2023-05-05 10:11:12
2		2023-05-05 11:50:25
3		2023-05-05 11:55:47
4		2023-05-05 14:19:05
5		2023-05-05 19:42:39
6		2023-05-06 00:08:59
7		2023-05-06 00:48:42
8		2023-05-06 11:01:08
9		2023-05-06 16:25:43
10		2023-05-06 20:46:34

Showing 1 to 10 of 100 entries

Previous 1 2 3 4 5 ... 10 Next

欄位篩選結果依每日日期 &amp; 時間排序。

參數設定

Input - 27

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

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) \*

system\_id  
artDate

儲存更改

分群彙總: 以加總計算每日數量發文。

參數設定

Input - 27

任務結果

## 統計資訊

**34**

群組數量



## 任務結果

Show 10 entries

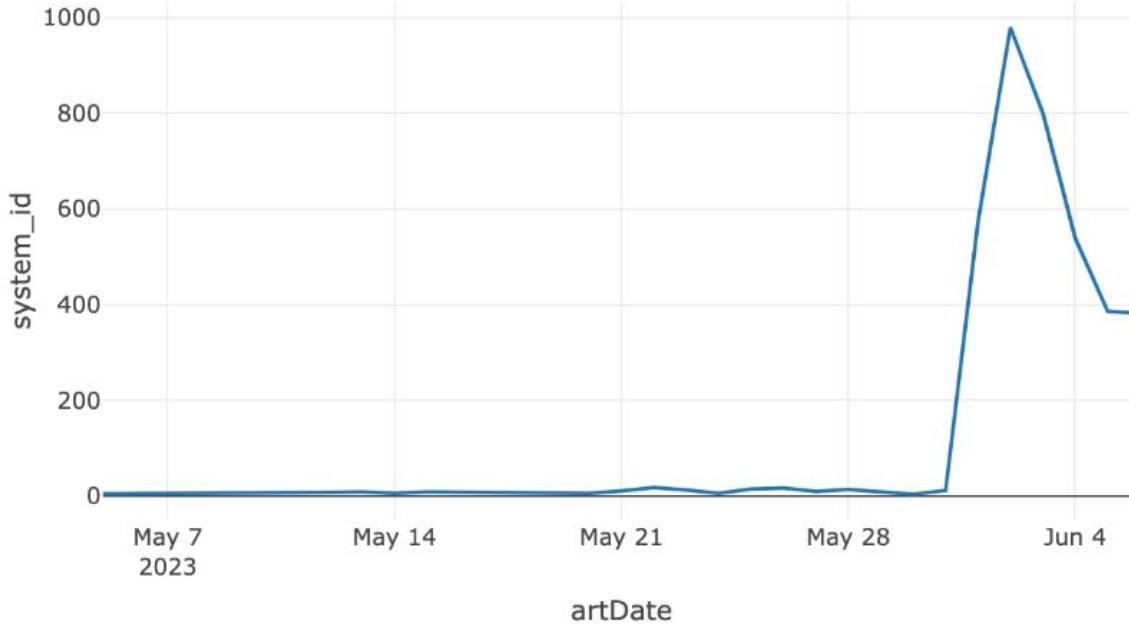
Search:

artDate

system\_id

2023-05-05	5
2023-05-06	5
2023-05-07	6
2023-05-08	5
2023-05-09	5
2023-05-10	5
2023-05-11	6

分群彙總後共計分為 34 個群組。



由圖中可發現，自五月底開始，台灣民進黨的基層黨工陳汎孫在「臉書」實名指控性騷擾事件爆發後，相關議題討論度明顯提升。

# 詞頻計算

## 詞頻計算 (60)

參數設定

Input - 103

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 ⓘ

1000

儲存更改

選取出現頻率最高的前1000個詞組來產生文字雲。

# 任務結果

Show  entries

Search:

n	Term
8122	性騷擾
5004	民進黨
1461	蔡英文
1420	台灣
1260	黨工
1203	國民黨
1167	受害者
1141	女性
1037	事件
912	賴清德

Showing 1 to 10 of 100 entries

Previous      ...  Next

## 文字雲



由圖中可發現，出現最多的名詞為民進黨、性騷擾、蔡英文等。由於相關實名指控先在民進黨 內部開始出現，之後有資深媒體人控訴國民黨民代性騷擾等，性騷擾議題與政治話題一同頻繁出現。

# 關聯式文字雲

轉為DTM (33)

參數設定

Input - 103

任務結果

## 保留詞彙 ?

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 ?

200

儲存更改



## 三 關聯式文字雲 (shiny) (43)

參數設定

Input - 33

任務結果

分群數 \*

20

迭代次數(最少250次)

1000

聚合演算法 \*

single

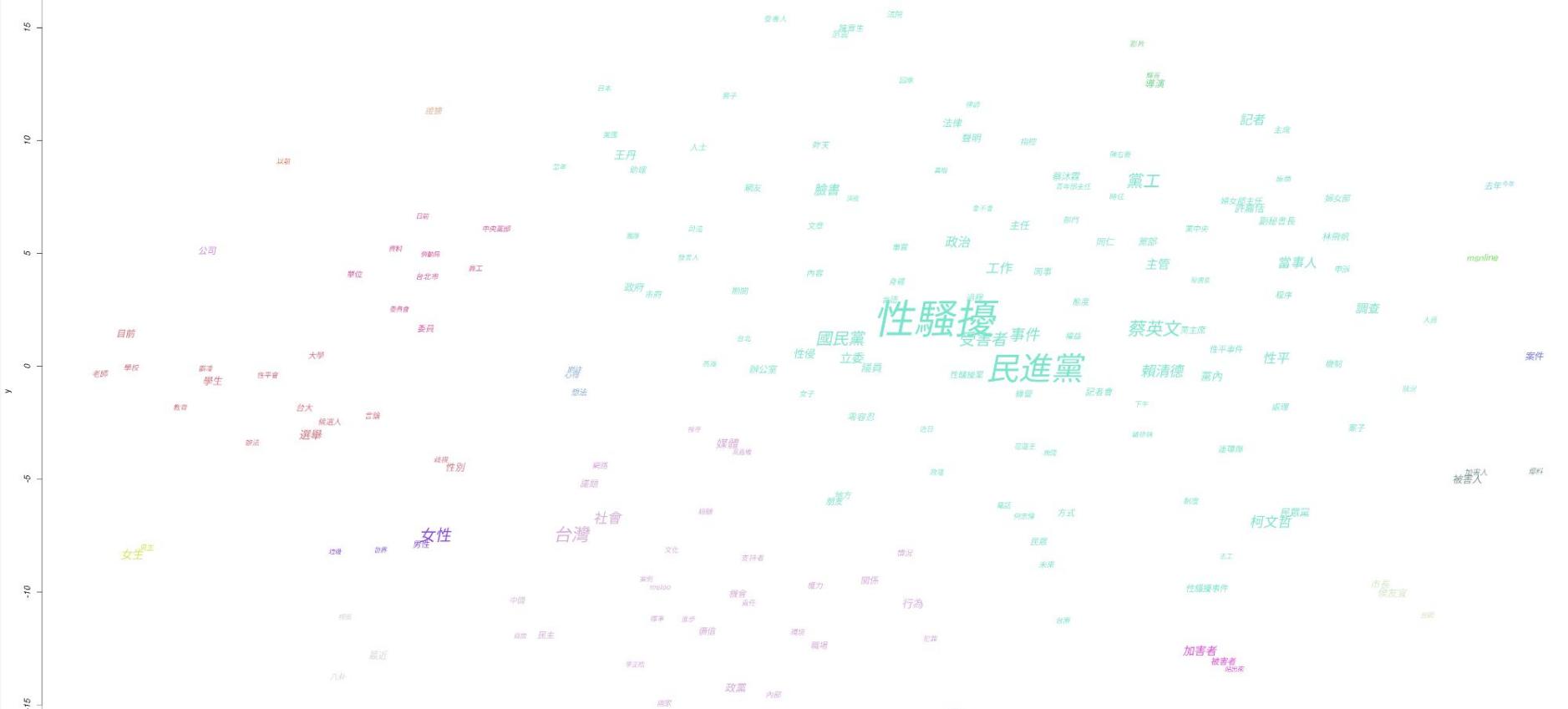
文字雲顯示字數 \*

1000

距離計算公式 \*

euclidean

儲存更改



由圖中可發現，綠色關聯式文字雲即為剛開始的民進黨與國民黨性騷擾事件，而左邊紅色則是陸續延燒的學術界性騷擾議題相關字詞。

# 情緒分析

## Lexicon Based 情緒分析 (21)

x

參數設定

Input - 103

任務結果

語言 \*

Chinese

載入預設情緒字典 \*

是

定義正面詞彙 ⓘ

以換行符號區隔，e.g.

詞彙

開心

高興

...

定義負面詞彙 ⓘ

笑死

儲存更改

使用Lexicon Based情緒分析字典，瞭解情緒詞彙、正面與負面詞彙的發生情形。

# 統計資訊

42  
最大正向情緒



0  
最小正向情緒



81  
最大負向情緒



0  
最小負向情緒



## 任務結果

Show 10 entries

Search:

system_id	positive_count	positive_words	negative_count	negative_words	sentiment_value
1	1	[誘人]	8	[性騷擾, 性騷擾, 性騷擾, 性騷擾, 性騷擾, 抗議, 騭嚇, 性騷擾]	-7
2	0	[]	3	[性騷擾, 八卦, 八卦]	-3
3	2	[操守, 幫忙]	4	[性騷擾, 詐騙, 犯罪, 犯罪]	-2
4	6	[獎金, 冠軍, 獎金, 作品, 朋友, 作品]	0	[]	6
5	2	[娛樂, 幸福]	1	[性騷擾]	1
6	3	[冠軍, 作品, 荣幸]	1	[性騷擾]	2
7	0	[]	1	[他媽的]	-1
8	0	[]	8	[性騷擾, 性騷擾, 性騷擾, 性騷擾, 壓力, 壓力, 性騷擾, 隔離]	-8
9	1	[勇氣]	10	[悲痛, 悲劇, 受害者, 暴力, 性騷擾, 辱罵, 性騷擾, 受害者, 威脅, 罪惡感]	-9
10	4	[體貼, 貢獻, 治療, 治療]	3	[侵犯, 性騷擾, 隔離]	1

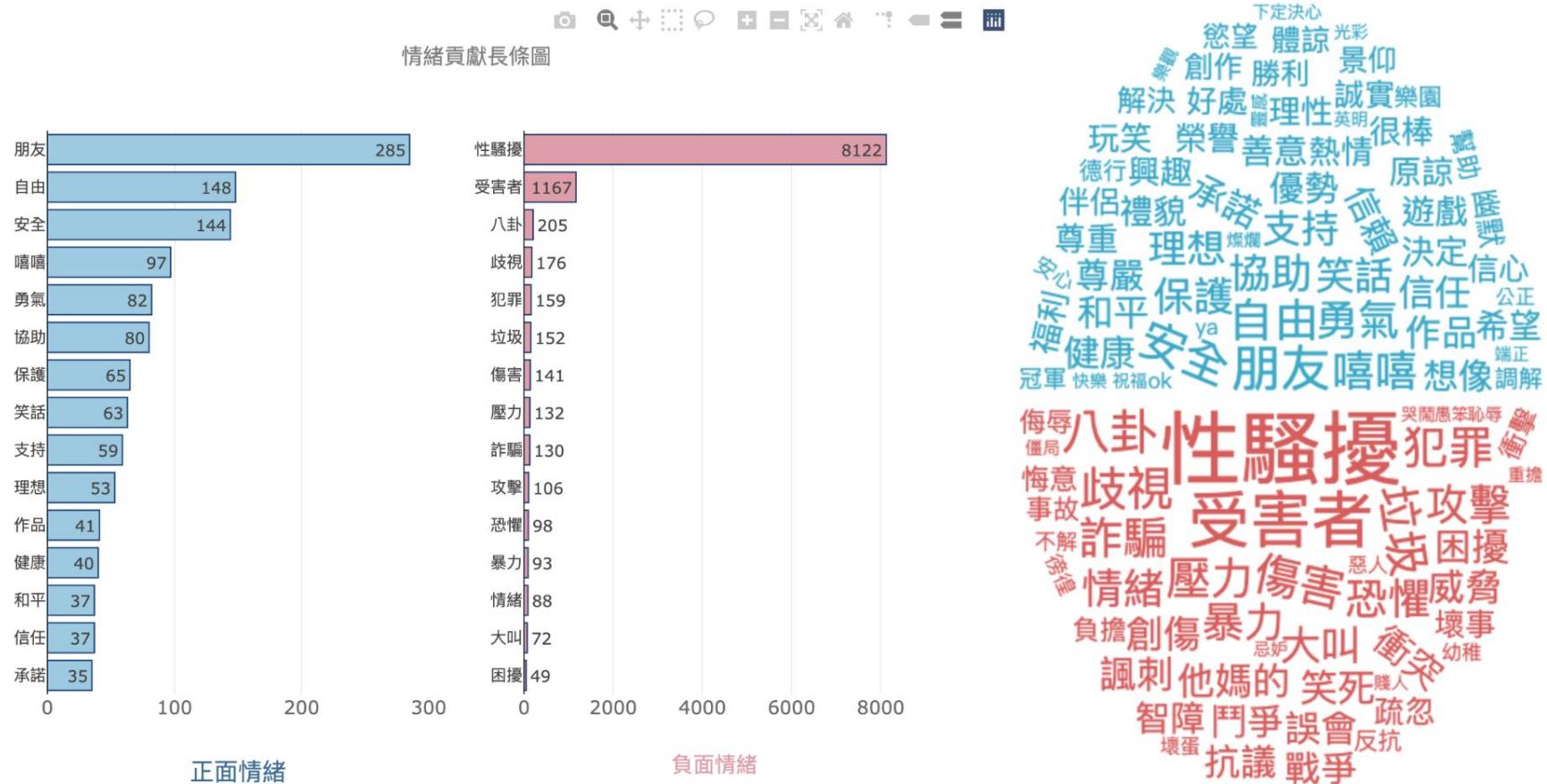
Showing 1 to 10 of 100 entries

Previous

1 2 3 4 5 ... 10 Next

因主題性質關係，負向情緒遠高於正向情緒。

# 視覺化



# 情緒分析的視覺化圖表

合併資料 (85)

參數設定      Input - 4      Input - 21      任務結果

**JOIN規則**

**新增規則**      **刪除規則**

left_key	right_key
system_id	system_id
-----請選擇-----	-----請選擇-----

儲存更改

This screenshot shows a user interface for joining datasets. At the top, there are tabs for '參數設定' (Parameter Settings), 'Input - 4', 'Input - 21', and '任務結果' (Task Results). Below this, under 'JOIN規則' (Join Rules), there are two buttons: '新增規則' (Add Rule) and '刪除規則' (Delete Rule). The main area displays a table with two columns: 'left\_key' and 'right\_key'. Both columns have dropdown menus. The first row shows 'system\_id' selected in both columns. The second row has dropdown menus labeled '-----請選擇-----' (Please Select). A large green button at the bottom right is labeled '儲存更改' (Save Changes).

將情緒分析結果與爬蟲資料合併

## 合併資料 (85)

參數設定

Input - 4

Input - 21

任務結果

### 任務結果

Show 10 entries

Search:

system_id	artUrl	artTitle	artDate	artPoster	artCategory	artContent	artComment	e_ip	insertedDate
1	<a href="https://www.ptt.cc/bbs/Gossiping/M.1683252674.A.708.html">https://www.ptt.cc/bbs/Gossiping/M.1683252674.A.708.html</a>	[新聞]加油站男員工建議紗裙女騎士「穿小短褲	2023-05-05 10:11:12	kent	Gossiping	1.媒體來源:\n聯合報\n\n2.記者署名:\n林孟潔 / 台北即時報導\n\n3.完整新聞標題:\n加油站男員工建議紗裙女騎士「穿小短褲更誘惑」性騷下場慘了\n\n4.完整新聞內文:\n\n一名男加...	[{"cmtStatus": "推", "cmtPoster": "kaitokid1214", "cmtContent": "111.249.157.234 :八卦-1", "cmtDate": "2023-05-05 10:11..."}]		2023-05-01 01:31:00
2	<a href="https://www.ptt.cc/bbs/Gossiping/M.1683258627.A.F54.html">https://www.ptt.cc/bbs/Gossiping/M.1683258627.A.F54.html</a>	[問卦]為何性騷新闻都是八卦-1不是八卦-2	2023-05-05 11:50:25	YU0121	Gossiping	板上有人po新聞時\n若是性騷擾的社會案件時\n下面推文都是八卦-1\n但其實這類行為 西洽板才一狗票吧?\n\n每次看西洽板的大多數文章\n給我的感受就是\npo文的人只 要不尷尬 尷尬的就是看文的人...	[{"cmtStatus": "→", "cmtPoster": "bill403777", "cmtContent": "118.168.145.229 :八卦噁心", "cmtDate": "2023-05-01 01:32:00"}]		2023-05-01 01:32:00

## 欄位篩選 (89)

參數設定

Input - 85

任務結果

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

- e\_ip
- insertedDate
- dataSource
- positive\_count
- positive\_words
- negative\_count
- negative\_words
- sentiment\_value

儲存更改

篩選正面字詞數量、負面字詞數量與情緒分數欄位。

## 三欄位篩選 (89)

x

參數設定

Input - 85

任務結果

### 任務結果

Show 10 entries

Search:

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value
1	2023-05-05 10:11:12	kent	1	8	-7
2	2023-05-05 11:50:25	YU0121	0	3	-3
3	2023-05-05 11:55:47	lwswjs	2	4	-2
4	2023-05-05 14:19:05	toebb	6	0	6
5	2023-05-05 19:42:39	bUnUd	2	1	1
6	2023-05-06 00:08:59	pttmovieLove	3	1	2
7	2023-05-06 00:48:42	TWipc	0	1	-1
8	2023-05-06 11:01:08	lolahjy	0	8	-8
9	2023-05-06 16:25:43	paericsson	1	10	-9
10	2023-05-06 20:46:34	Kitakami	4	3	1

## 資料篩選 (94)

x

參數設定

Input - 89

任務結果

條件式 \* ⓘ

```
{"OR": ["$sentiment_value>0", "$sentiment_value<0"]}
```

儲存更改

篩選出情緒小於0與大於0的資料(正面與負面情緒)。

## 資料篩選 (94)

×

參數設定

Input - 89

任務結果

### 統計資訊

665



移除數量

3433



保留數量

### 任務結果

Show  entries

Search:

system_id	artDate	artPoster	positive_count	negative_count	sentiment_value	result
1	2023-05-05 10:11:12	kent	1	8	-7	_selected
2	2023-05-05 11:50:25	YU0121	0	3	-3	_selected
3	2023-05-05 11:55:47	lwswjs	2	4	-2	_selected
4	2023-05-05 14:19:05	toebb	6	0	6	_selected
5	2023-05-05 19:42:39	bUnUd	2	1	1	_selected

## 欄位重構(Melt) (100)

x

參數設定

Input - 94

任務結果

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

artDate  
artPoster  
positive\_count  
negative\_count  
sentiment\_value

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

artDate  
artPoster  
positive\_count  
negative\_count  
sentiment\_value

轉換的欄位新的欄位名稱 \*

melt\_senti\_count

值的欄位名稱 \*

melt\_senti\_value

儲存更改

將正面、負面字詞數量與日期欄位進行重構。

## 三欄位重構(Melt) (100)

參數設定

Input - 94

任務結果

### 任務結果

Show 10 entries

Search:

system_id	artDate	melt_senti_count	melt_senti_value
1	2023-05-05 10:11:12	positive_count	1
2	2023-05-05 11:50:25	positive_count	0
3	2023-05-05 11:55:47	positive_count	2
4	2023-05-05 14:19:05	positive_count	6
5	2023-05-05 19:42:39	positive_count	2
6	2023-05-06 00:08:59	positive_count	3
7	2023-05-06 00:48:42	positive_count	0
8	2023-05-06 11:01:08	positive_count	0
9	2023-05-06 16:25:43	positive_count	1
10	2023-05-06 20:46:34	positive_count	4

得出每日發文時間的正面、負面字詞數量。

## 分群彙總(日期) (97)

×

### 參數設定

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\_senti\_count  
melt\_senti\_value

### Input - 100

### 任務結果

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\_senti\_count**  
melt\_senti\_value

儲存更改

將以上資料以單日進行加總。

## 分群彙總(日期) (97)

x

參數設定

Input - 100

任務結果

Show 10 entries

Search:

melt\_senti\_count

artDate

melt\_senti\_value

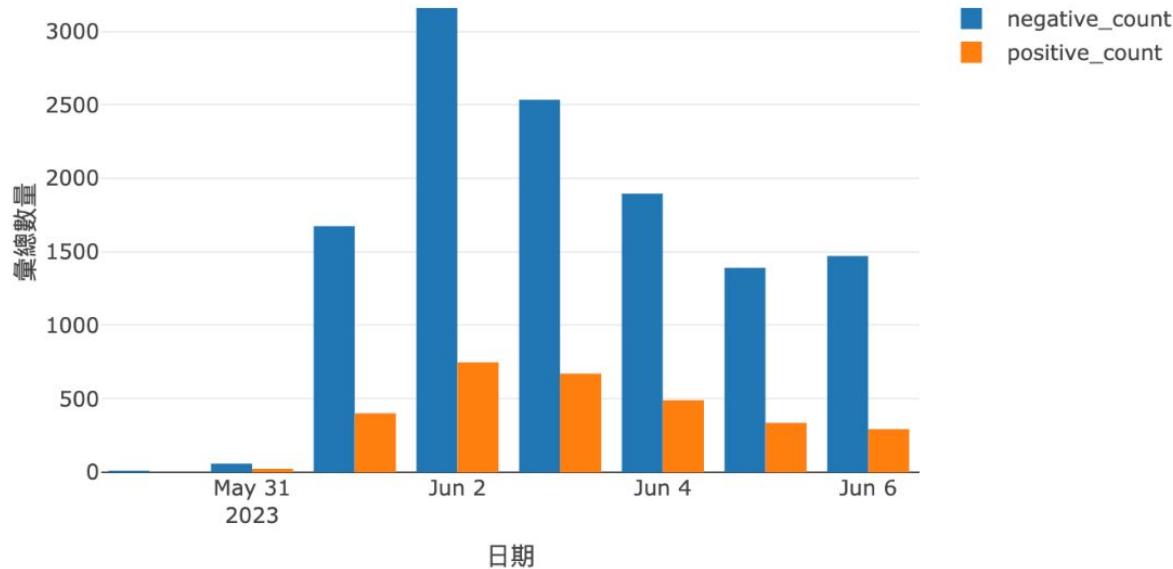
negative_count	2023-05-05	16
negative_count	2023-05-06	23
negative_count	2023-05-07	13
negative_count	2023-05-08	8
negative_count	2023-05-09	12
negative_count	2023-05-10	5
negative_count	2023-05-11	9
negative_count	2023-05-12	50
negative_count	2023-05-13	30
negative_count	2023-05-14	11

Showing 1 to 10 of 68 entries

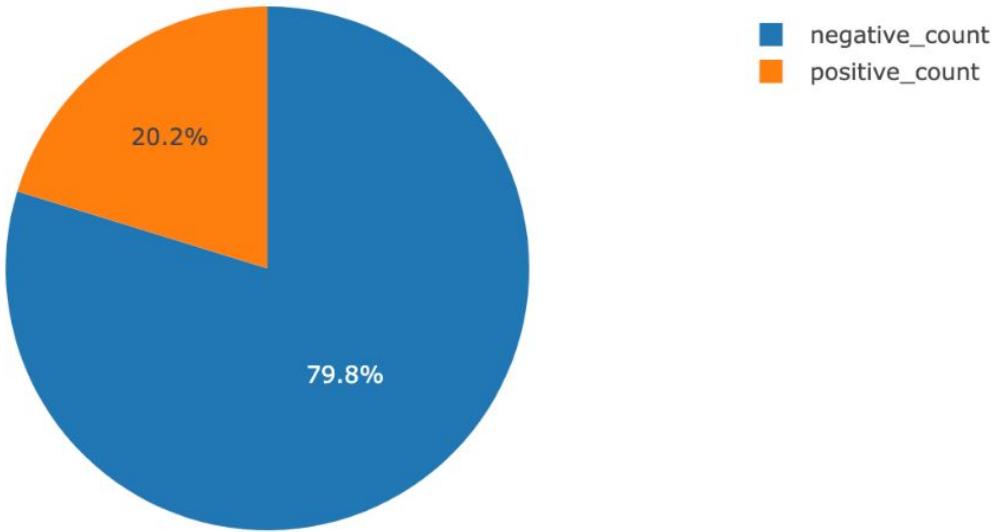
Previous

1 2 3 4 5 6 7 Next

## 每日正負向情緒彙總



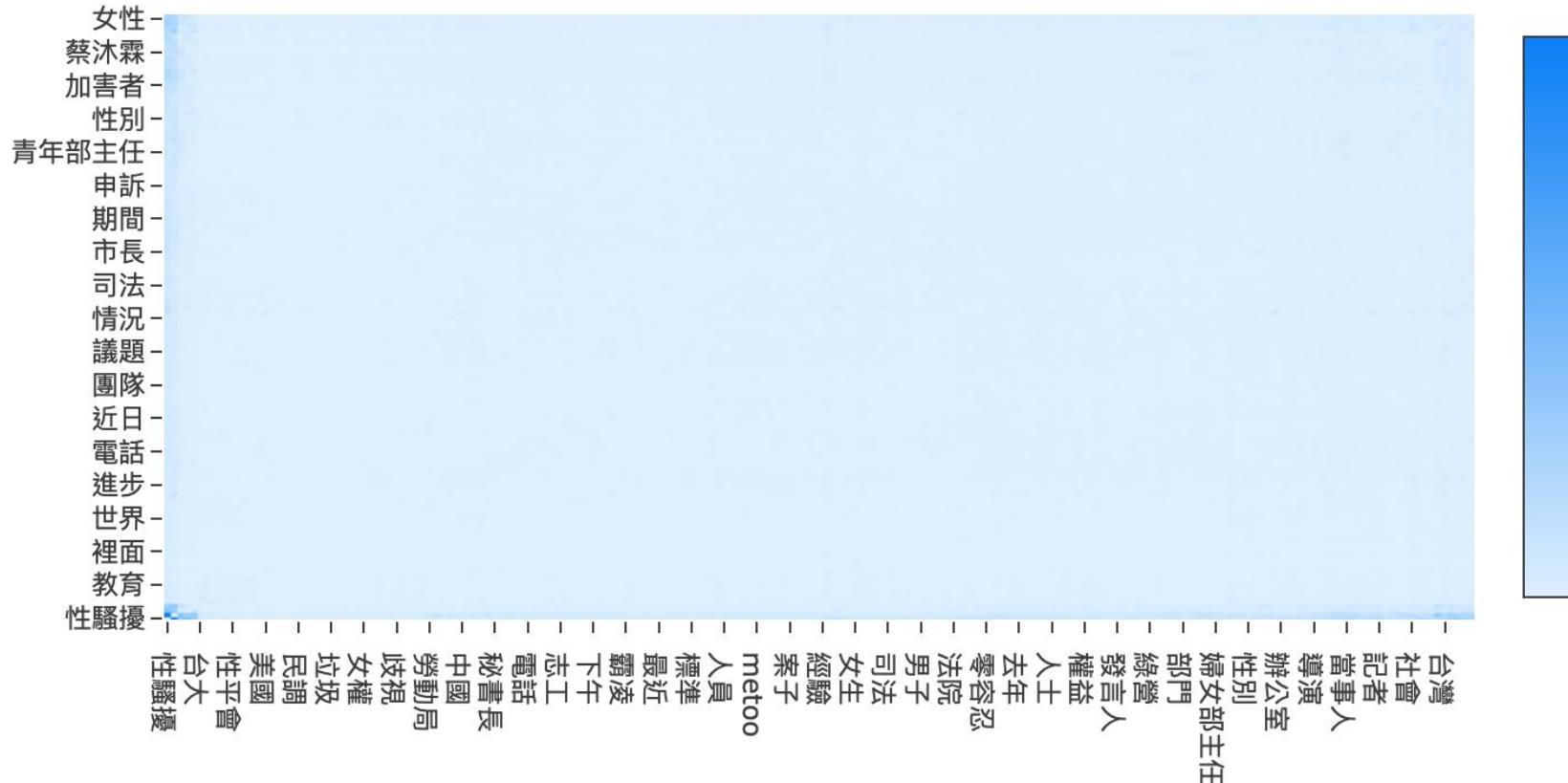
以長條圖呈現每日正負向情緒的匯總，可以發現六月初為負面情緒的最高峰。



以圓餅圖呈現正負向情緒分數的比重，可以發現負面情緒高達近80%。

# 計算共現矩陣

## 圖 森





# 單中心網路圖

## 單中心網路圖 (Shiny) (39)

參數設定

Input - 35

任務結果

選擇關鍵字典檔案 ?

class.csv

原始檔案元件 \*

48

DTM \*

33

共線/相關性矩陣 \*

35

儲存更改

選定請柬主題為中心，從文集中選出相關性最高的關鍵詞，並對它們進行社會網絡分析。

## 選擇詞類

民進黨

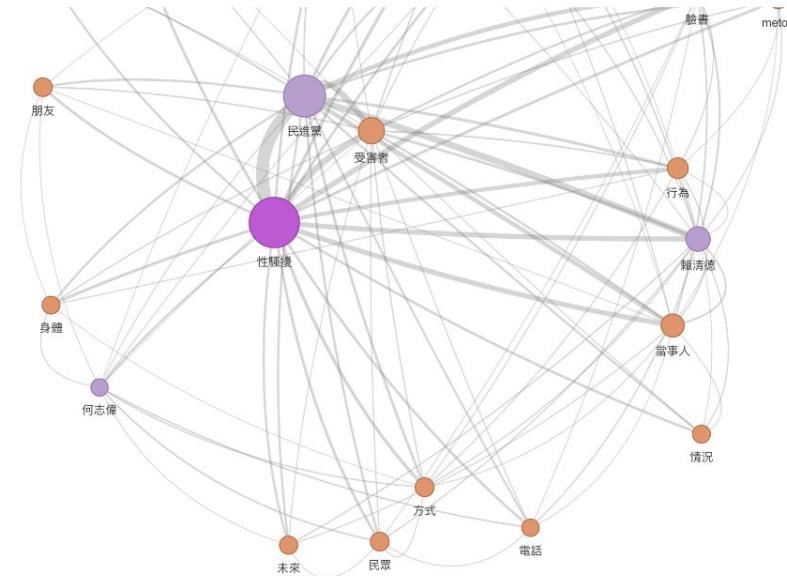
## 選擇次層詞類

何志偉

## 設定網路節點數量



## 依關聯強度篩選鏈結



以單中心網路圖可以看到性騷擾事件中心與其他字詞的關聯，多數皆為政黨政治相關字詞。

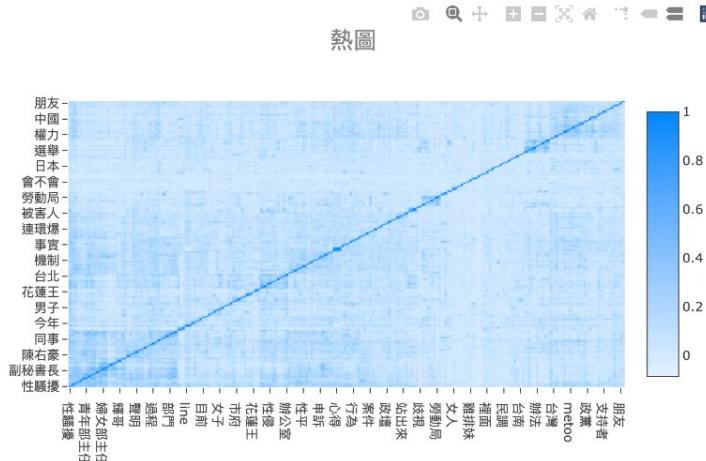
## 計算相關性矩陣 (37)

參數設定

Input - 33

任務結果

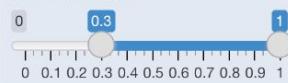
### 熱圖





# 字詞網路圖

詞彙關聯度範圍



字體大小



本次繪圖所呈現的邊 (edges) 數量:232

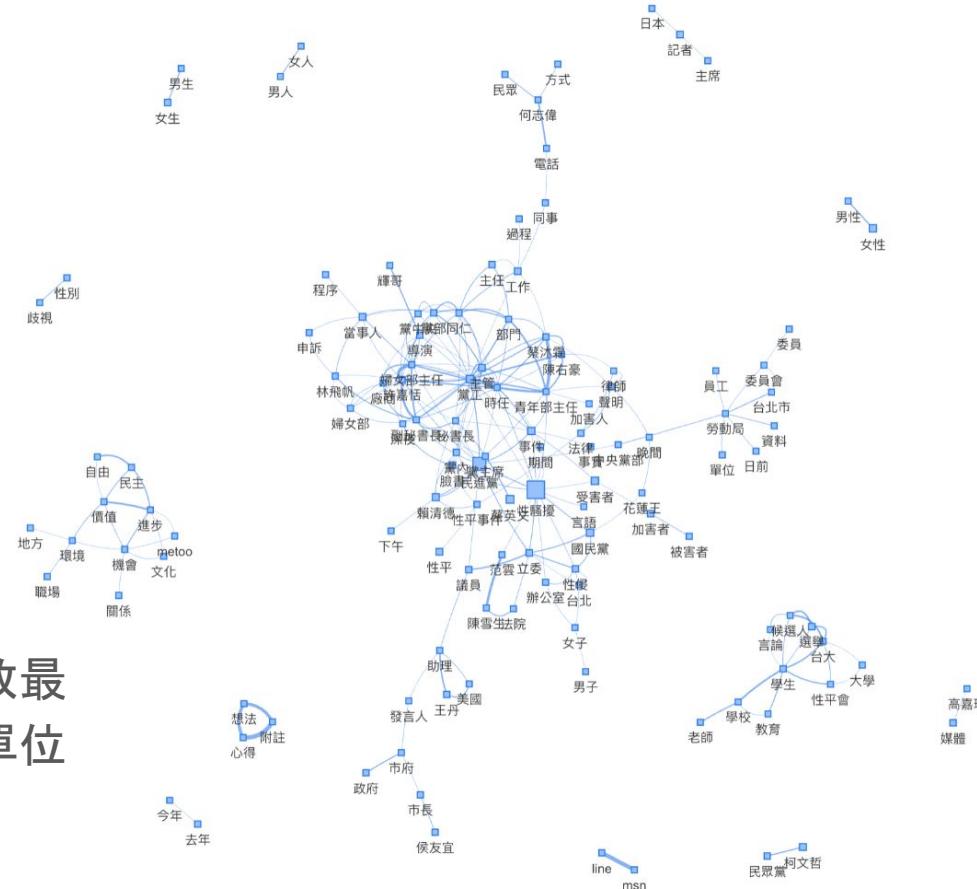
本次繪圖所呈現的節點 (nodes) 數量:125

請注意，每次更新參數後請再手動點選開始繪圖！

開始繪圖

因事件最初爆發為民進黨黨工爆料，故最密集的位置，字詞皆為民進黨與辦公單位等相關字詞。

network



# 主題模型

## GuidedLDA 主題模型 (45)

參數設定

Input - 103

任務結果

目標欄位 \*

result

主題數 \*

3

詞彙頻率下限 ⓘ

100

alpha

預設為主題數/50

主題種子字 ⓘ

民進黨,民進黨發言人,民進黨秘書長,民進黨青年部,民進黨主席,民進黨婦女部,民進黨立委,陳時中,何志偉,林男固,陳汶軒,何孟樺,陳右豪,賴清德,許嘉恬,蔡沐霖,蔡英文,林飛帆,范雲  
國民黨,國民黨立委,曾銘宗,曾柏文,花蓮王,侯友宜

迭代次數

100

主題保留關鍵字數量

20

詞彙頻率上限 ⓘ

0.6

Beta

預設為0.1

是否輸出字典

是

儲存更改

## 統計資訊

**60**  
字數 

**3**  
主題數 

**-1.779**  
主題連貫性(UMass) 

**0.007**  
主題連貫性(PMI) 

**0.489**  
主題連貫性(Cv) 

**156.78**  
混淆度 

## 沒有使用到的主題種子字

民進黨發言人, 民進黨秘書長, 民進黨青年部, 民進黨主席, 民進  
黨婦女部, 民進黨立委, 陳時中, 何志偉, 林男固, 陳汝軒, 何孟樟,  
陳右豪, 國民黨立委, 曾銘宗, 曾柏文, 花蓮王,

# GuidedLDA Vis

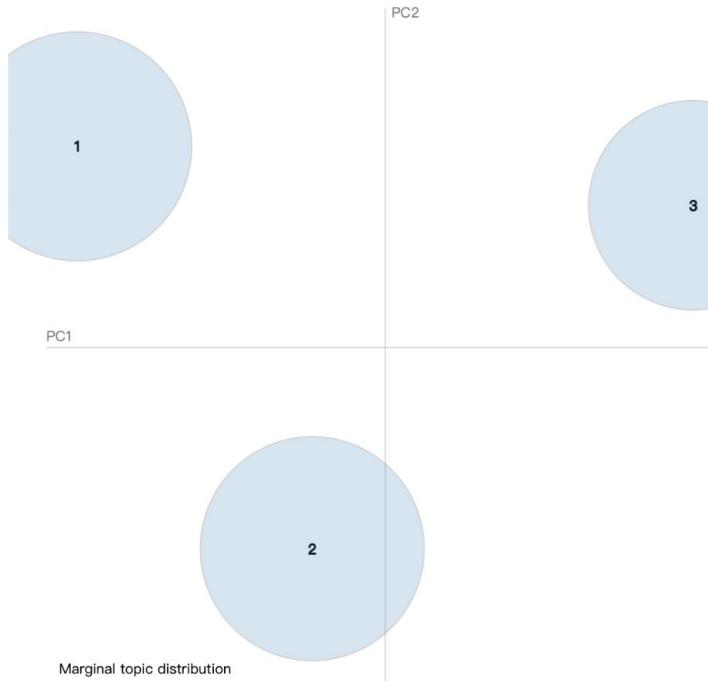
Selected Topic: 0 [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric:(2)

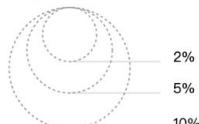
$\lambda = 0.6$



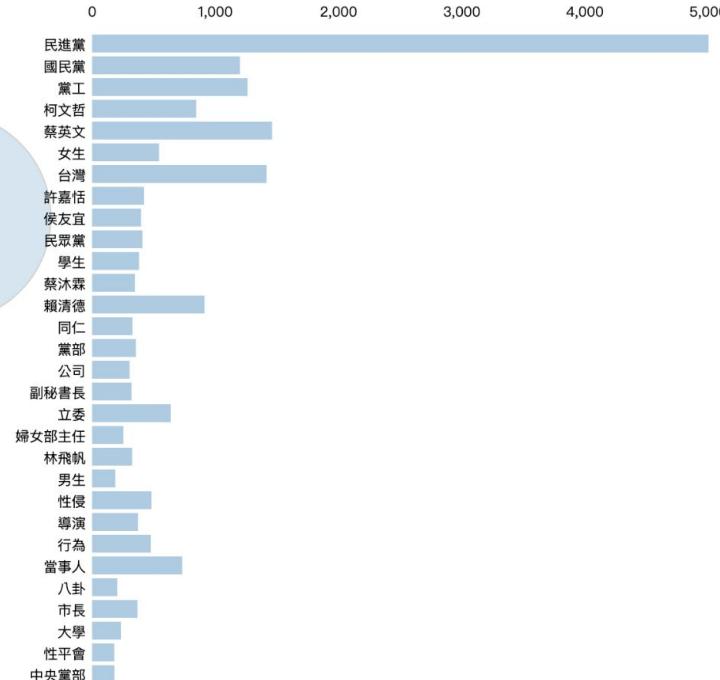
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Salient Terms<sup>(1)</sup>



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) \* [sum\_t p(t | w) \* log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)  
2. relevance(term w | topic t) =  $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Sievert & Shirley (2014)

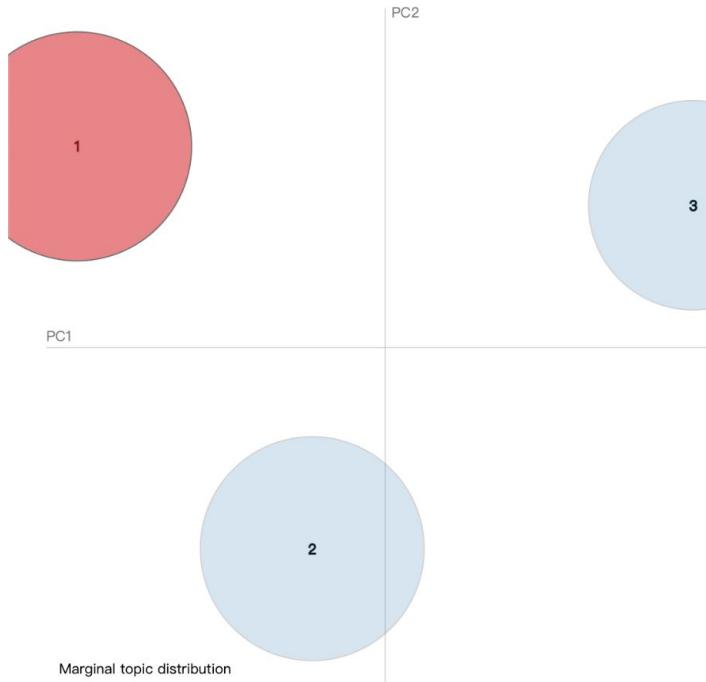
# GuidedLDA Vis

Selected Topic: 1 [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric:(2)

$\lambda = 0.6$

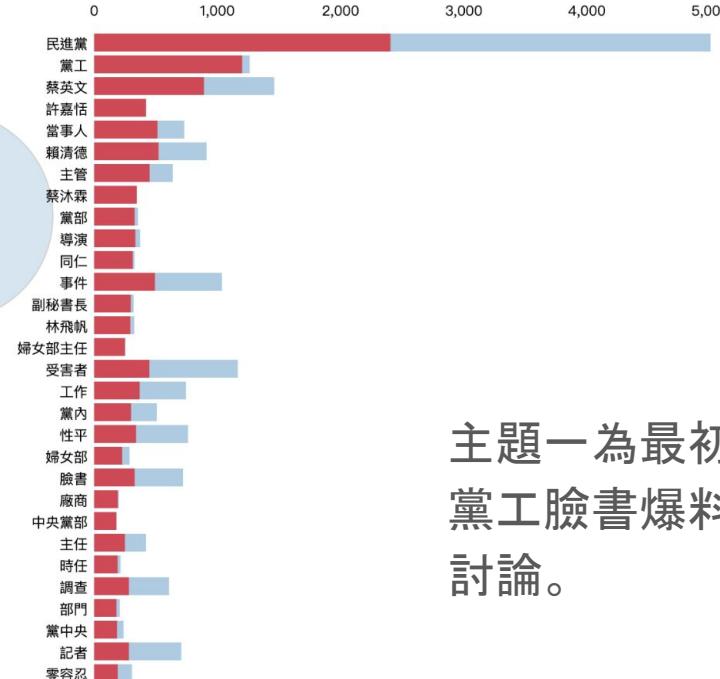
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (35.8% of tokens)



1. saliency(term w) =  $\text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$  for topics t; see Chuang et. al (2012)  
2. relevance(term w | topic t) =  $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Sievert & Shirley (2014)

主題一為最初的民進黨  
黨工臉書爆料相關事情  
討論。

# GuidedLDA Vis

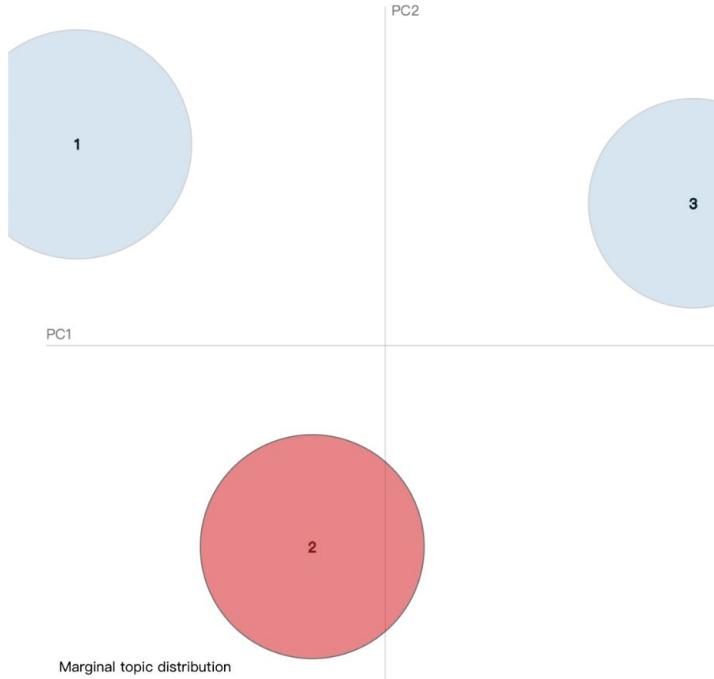
Selected Topic: 2 [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric:(2)

$\lambda = 0.6$

0.0 0.2 0.4 0.6 0.8 1.0

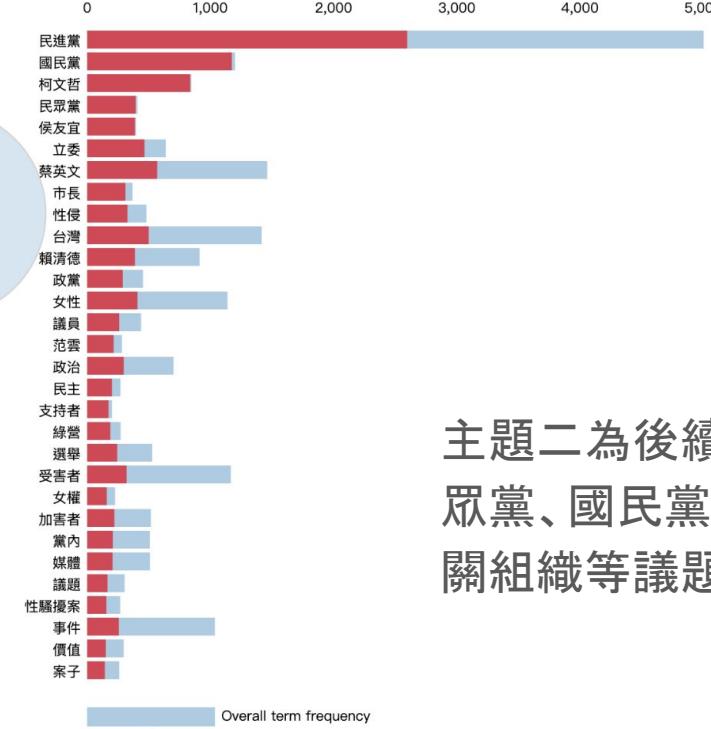
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 2 (34.2% of tokens)



1. saliency(term w) = frequency(w) \* [ $\sum_t p(t | w) * \log(p(t | w)/p(t))$ ] for topics t; see Chuang et. al (2012)  
2. relevance(term w | topic t) =  $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Sievert & Shirley (2014)

主題二為後續延燒的民  
眾黨、國民黨與性平相  
關組織等議題



# 結論



明年總統大選正式展開時，性騷擾議題突然成為一場意外的政治風暴，民進黨將如何全面回應這些指控，特別是年輕一代對他們的質疑，是大家關注的焦點。

透過此次專案可以發現網路上對於此次事件的情緒趨勢非常負面，而透過文字分析，我們也發現即使剛開始不清楚事件過程，也能透過主題模型與關聯式文字雲快速掌握事件的全貌，並針對討論相關議題進行分析或應對，在現今社群媒體普及發展的時代，讓我們能快速地針對多個主題與議題進行大方向的掌握，希望未來能將此技術應用於更多不同的主題與領域。