

社群媒體分析

### 啦啦隊女神「林襄」網 路情緒分析

指導教授:黃三益博士

第十組期中報告 N104220007 黃足維 N104220013 倪家瑋 N104220016 何軒育 N104220023 許耀文

#### Agenda



04 視覺化分析結果

資料分析過程

03



### 研究動機與目的

自然語言爬蟲是一種利用自然語言處理技術和爬蟲技術來自動從網絡上抓取和分析文本的方法。對於演藝人員的分析,自然語言爬蟲可以從網絡上收集與演藝人員有關的文章、新聞、評論等文本資料,並進行分析和提取有用的信息。大致可分為下列四種:

#### 情感分析:

使用自然語言處理技術來分析文章中的情感色彩,例如對演藝人員的評價、反應等,並將其分為正面、負面或中性。這可以幫助了解演藝人員在大眾中的聲譽和形象,以及公眾對其演出和言論的反應。

#### 實體識別:

使用自然語言處理技術識別文本中提到的 實體,例如演藝人員的名字、演出作品等 等。這能幫助建立一個包含演藝人員相關 信息的知識庫,方便進一步的分析和挖掘。

#### 主題分析:

使用自然語言處理技術和機器學習方法,從文本中識別和提取出與演藝人員相關的主題,例如演出作品、活動、社交媒體上的言論等等。這可以幫助了解演藝人員的職業生涯、生活和與其他明星之間的關係。

#### 詞頻分析:

使用自然語言處理技術和統計方法,計算文本中詞語的頻率,例如演藝人員的名字、演出作品、活動等等,並從中識別出關鍵詞和熱門話題。這可以幫助了解演藝人員在不同時期的受關注程度和熱度,以及與其他演藝人員之間的競爭關係。

## 資料集描述



資料來源:蘋果日報爬蟲, 娛樂時尚版面



搜尋關鍵字:林襄



搜尋日期: 2022/08/01~2023/04/13



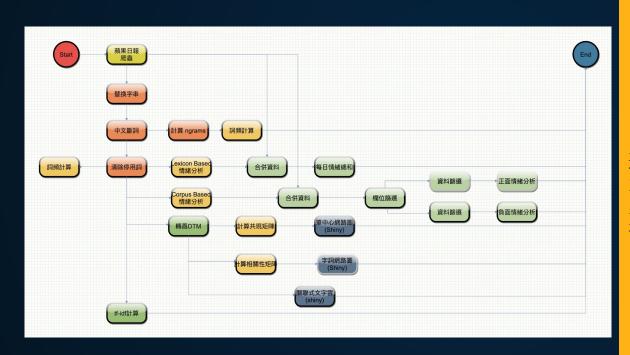
資料筆數: 132





## 資料分析過程

### 工作流程圖



透過工作流程平台爬蟲文章, 並進行報導內容的替換字串去除不 必要的文字和字詞,以利後續資料分 析與判讀能提高精準度。

## 資料前處理

蘋果日報 爬蟲



替換字串

#### 替換字串設定 **①** \n\n>>。 \n>>,

中文斷詞

#### 定義詞彙❸

林襄 100 啦啦隊女神 100 樂天啦啦隊 100 樂天女孩 100 全明星 100 樂天桃猿 100 全明星運動會 100 中信兄弟 100

#### 清除停用詞

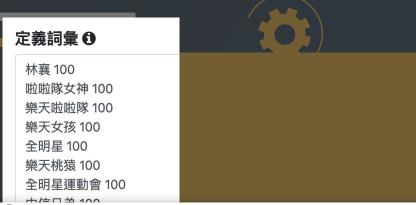


## 資料前處理-替換字詞

·替換字串 <sup>☞</sup> (10)		•
參數設定	Input - 7	任務結果
<b>異擇處理欄位 *</b>	替换字串記	
artContent	\/\n/u>> °	
	\n>>,	
選擇替换規則檔案 ➊		
請選擇	~	
	儲存更改	
替換字串 <sup>⑦</sup> (10)		×
參數設定	Input - 7	任務結果
充計資訊		
116 即代數量		
王務結果		
now 10 v entries		Search:
system_id :	result	÷
1	議事的下部居民並大學。「這是以前的快養了」並附上她打選學生時期所拍攝。向順友再年的影片,其相談與現下明顯不同,以及贴出她當時無形效顯而 嚴致「原来二次質勞是真的」然後現在變速樣。」與文書描述的、結果也公司被反於此样實證於不變的近。 希望原则一上她的影片,「低于女與反決 思,我都不解得。分假問題一人完全表做的原為其傳写了。而那戶的法之則皮影論。有人成「真何微假師,不入代她」、「讓因父雖然力法就是女生自討 期的無条行為」、「雖形沒薄,但如自己必是是天然的遊說就是無」、「人场的这與國」、「其實整形經濟光也沒付指不好」她比較顯於講話及覆風人看 不過程」、「上來都信在封據風外內的事情才對文」。他也有人認為「是 4 名判理的心理 一生平事,希明史國書一個、「歷刊的本學本學本學 便事,現代女態人整多整少而已」、「大板是忌妒她的整宜直接翻觸人生包」、「郑富德國师、現在不都這種模式」、「雖到口袋的才是裘真實的,其他 隨今門實」。	
		上節目時,穿上蕾絲蓋胸的上衣,若隱若現的巨乳讓人驚呼身材

- 文章内 \n\n>>換成。; \n>>換成,
- 任務結果:取代116筆

#### 資料前處理-中文斷詞





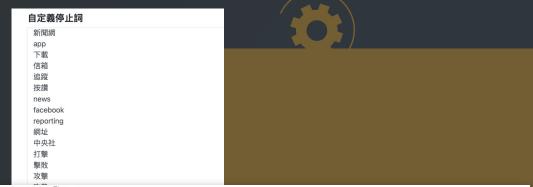
#### • 參數設定:

"林襄"會斷詞為"林"及"襄"

"啦啦隊女神"會斷詞為"啦啦隊"及"女神" 所以在定義詞彙增加權重:

林襄 100、啦啦隊女神 100、樂天啦啦隊 100、樂天女孩 100,等字詞。

### 資料前處理-清除停用詞





- 參數設定:自定義停用字42個
- 任務結果:清除30305個停用字

#### 情緒分析-Lexicon Based

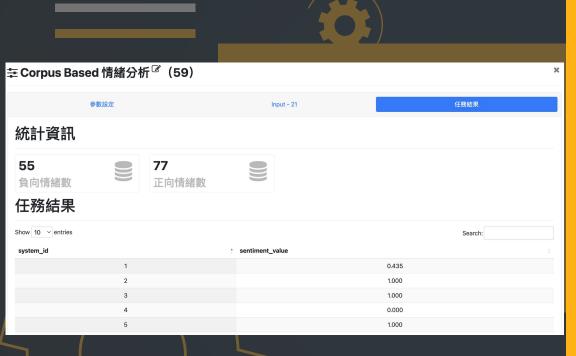






- 將"老婆"定義為正面詞彙。代表許多粉絲 將林襄視為最親密的愛人。
- 前三名正面字詞:希望、喜歡、開心
- 將"抽菸"定義為負面詞彙,主要是林襄曾因抽菸行為引起粉絲負面觀感
- 前三名負面字詞:問題、受傷、壓力

## 情緒分析-Corpus Based



- 最大正向情緒77
- 最大負向情緒55

### 關鍵字詞網路圖

選定詞類主題為中心,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 選擇次層詞類 設定網路節點數量 依關聯強度篩選鏈結 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

從字詞網路圖選擇"林襄"這個詞彙,然後 設定網路點數量為20,依關聯強度篩選連 結設為0.5,

找出跟"林襄"有關係的前10個詞彙分別是 "粉絲"、"節目"、"啦啦隊"、"女神"、"台 北"、"喜歡"、"希望"、"日本"、"人氣"、"身 材"。

#### 關聯式文字雲





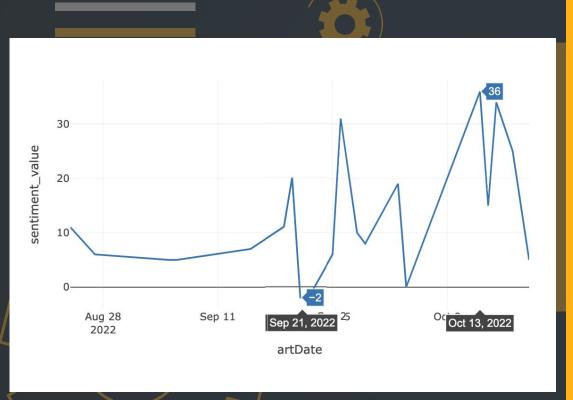
# 視覺化分析結果

## 詞頻計算-文字雲

樂天桃猿

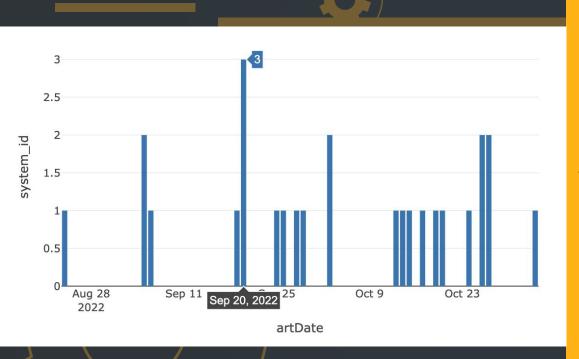
文字雲可以看到詞頻計算的結果,林襄、啦啦隊、女神、啦啦隊女神、粉絲等都是詞頻較高的字詞。

### 每日情緒總和-折線圖



- 利用折線圖可以看到粉絲每天對林襄的正面/負面情緒並對照出他們心情的喜怒哀樂(每日情緒總和)
- 2022/9/21 情緒分數相對較低甚至出現負數;2022/10/13 情緒分數最高。

### 正面情緒分析-直條圖



利用直條圖來看正面情緒分析,在 2022年9 月20日的正面情緒最多,主因是因為樂天 女孩趴要舉辦的消息公布。

#### 分析結論

透過關鍵字篩選及分析後可以發現社群媒體能影響的層面非常廣且具有高度的可變性。

我們以林襄抽菸的負面新聞為例, 在新聞爆發的前一天, 其正向情緒分數頗高

,爆發後隔天則直線下降,但隨著網民的跟風或是時間發酵後,受其鐘擺效應

影響又會快速的回到正軌, 可謂成也蕭何敗也蕭何,

所以我們的首要目標要訂立在可以快速的分析數據找出關鍵指標後,應用在變 化快速的網路社群媒體世界。



**THANKS!** 



字詞網路圖

詞彙關聯度範圍

本次繪圖所呈現的節點數:78

0

字體大小

