

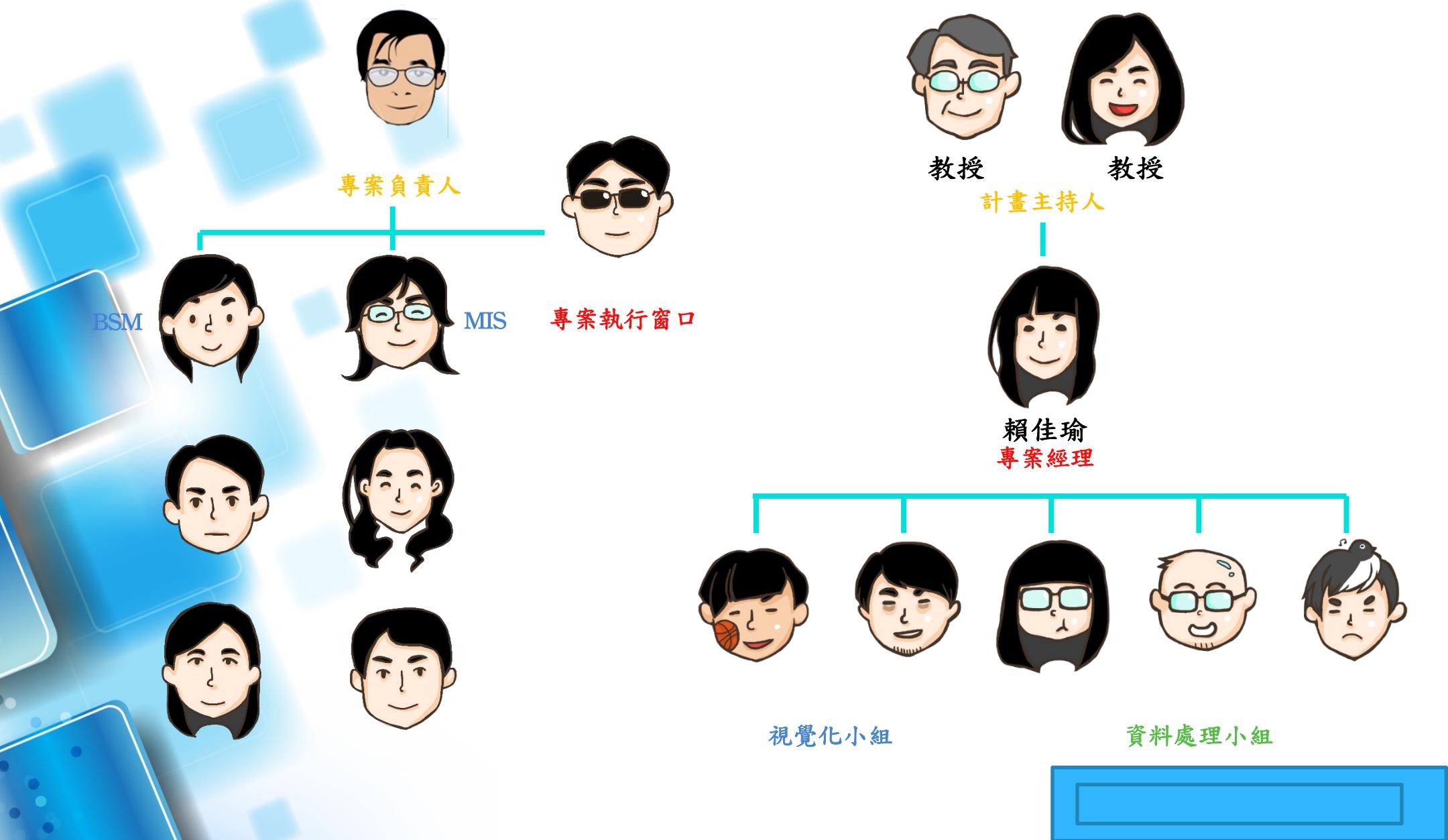
建構 Social Listening System

期末報告

報告大綱

- 專案說明
- 組織架構
- 計畫目標及介紹
- 期中進度簡介
- 計畫結果報告
- Demo展示
- 計畫效益評估
- 計畫未來展望

組織架構



計畫目標及介紹

計畫目標

本計畫旨在統整新聞及社群資料，分析後提供資訊供決策者參考。主任務如下：

- 新聞媒體與社群媒體之網路爬蟲資料
- 新聞媒體之文字處理分析
- 新聞媒體之文字議題與輿情分析
- 新聞媒體之視覺化設計
- 即時監控與警示系統
- 社群媒體之輿情分析
- 社群媒體之視覺化設計

計畫目標及介紹

系統建構流程

1. 網路資料擷取

-新聞媒體, 社交網站, PTT

4. 視覺化設計

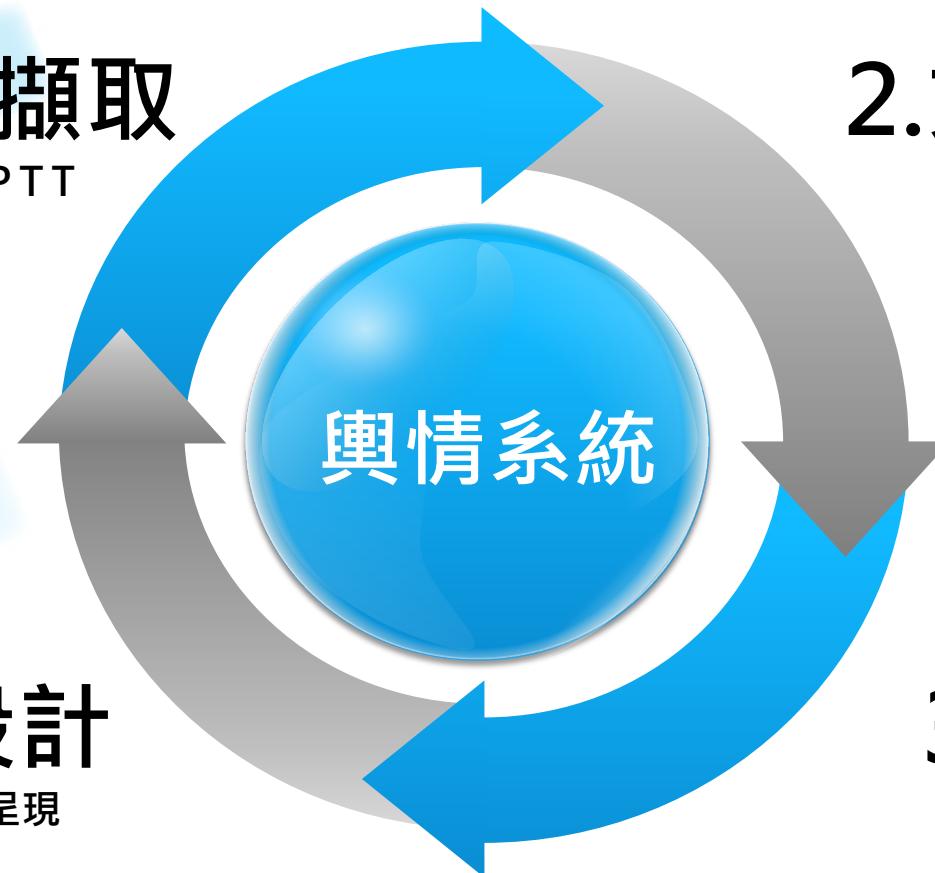
-看板與多維度資料表呈現

2. 文字資料處理

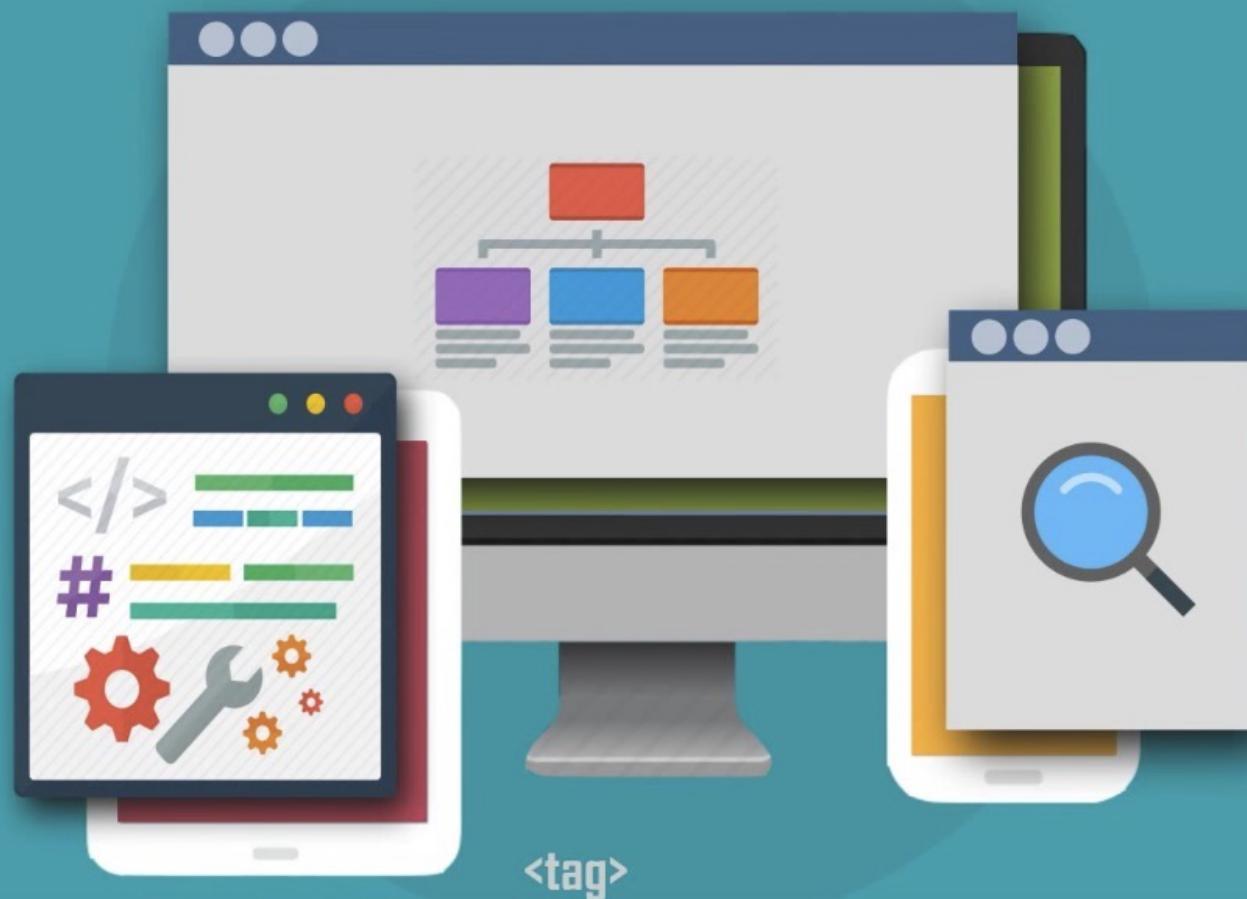
-資料清理/斷詞/標詞性

3. 文字資料分析

-重大議題分析與輿情分析



期中進度簡介



網路爬蟲

定期將新聞、社群平臺上的資料抓下來

期中進度簡介 (網路資料擷取)

媒體	起迄時間	內容	篇數
YAHOO! 奇摩新聞	2015/4/9~20 15/8/10	以日月光為關鍵字搜尋結果一百頁的資料	共 1159 筆，包含留言、斷詞斷句、詞性標記等資料
蘋果日報	2013/12/1~2 015/11/03	以日月光為關鍵字搜尋結果一百頁的資料	共 930 筆，包含留言、斷詞斷句、詞性標記等資料

期中進度簡介

我喜歡跑步 :)



我喜歡跑步



我 喜歡 跑步

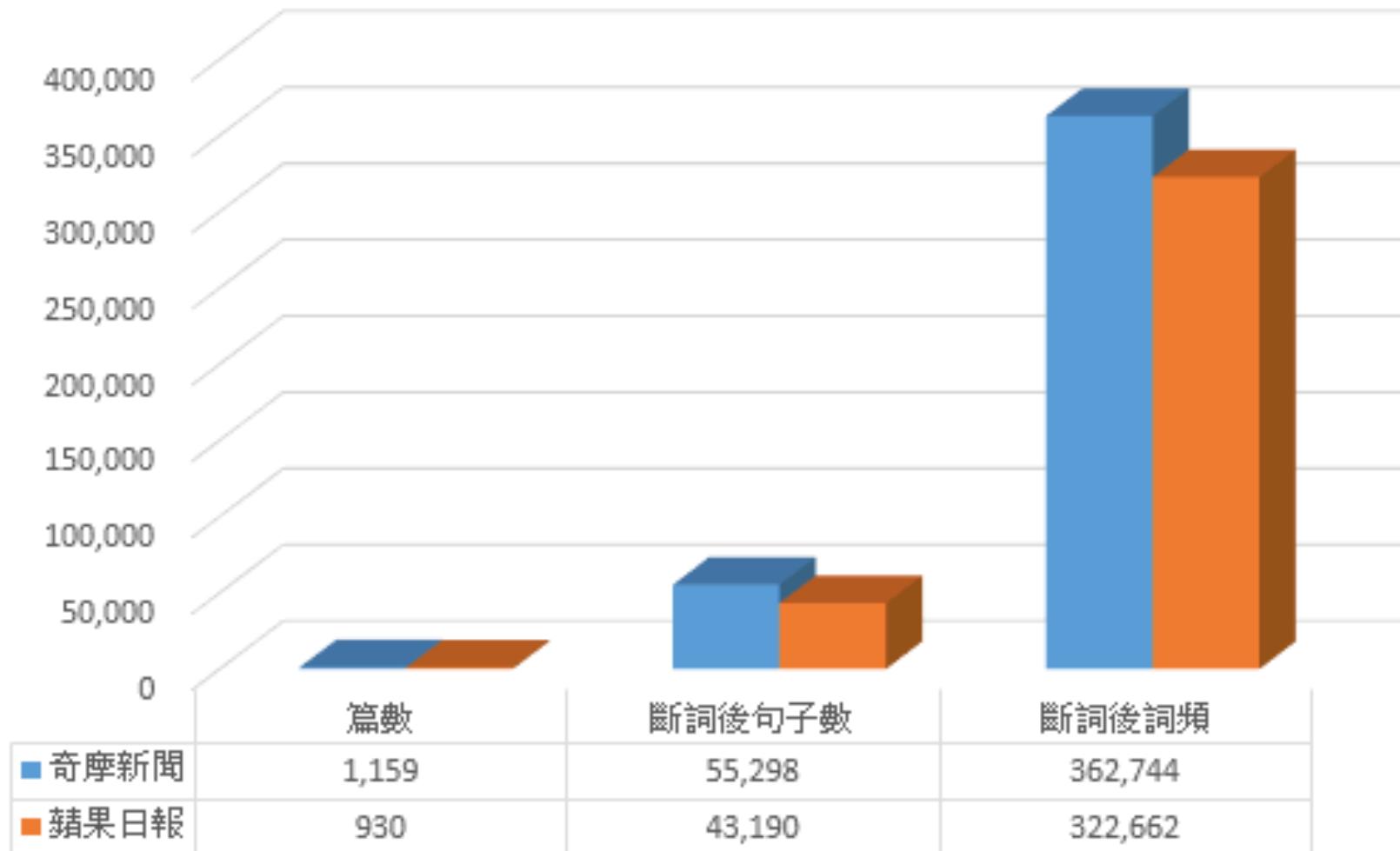


我(名詞) 喜歡(動詞) 跑步(名詞)

文字資料處理NLP

資料清理、斷句、
斷詞、標詞性

期中進度簡介 (文字資料處理)



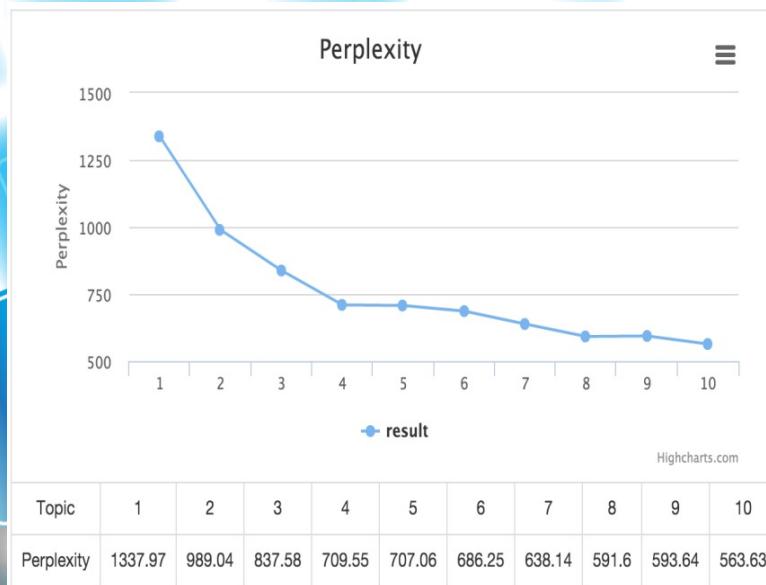
期中進度簡介



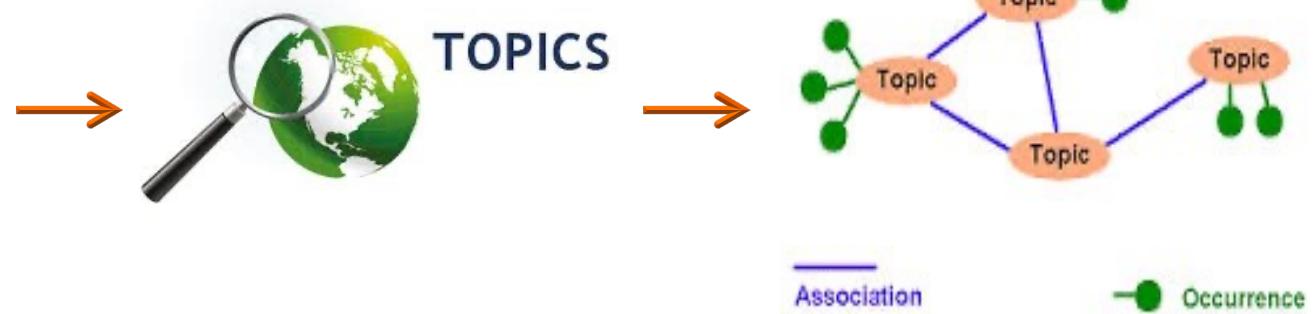
股市...正面

文字資料分析
LDA議題分析、
情緒字典建立、情緒分析

期中進度簡介 (文字資料分析)



使用混亂度分析主題數



設定主題數

每一主題關鍵字分布
主題分類器

期中進度簡介 (文字資料分析)

新聞稿、快報

國泰金
中信金
元大金
鴻海融資
發行
群創
交易
聯電
金額
資金
緯創
融券
友達

股市

股市
觀察
大盤
漲幅
震蕩
法人
反彈
短線
鴻海個股
大立光
下跌
上漲
跌幅

廢水

污水
環境
影音
鄉民
停工
社會
環保
廿
排放
H
數
張
環保局
後勁溪
政府
高雄市

全球供應鏈

訂單
矽品
產品
製程
客戶
需求
全球
營運
晶片
封裝
圓
供應鏈
全球
營運
集團
預估

大環境

員工
經濟部
科技
國際
發展
回
投資
認購
廠商
生產
管理
就業
再生
用水

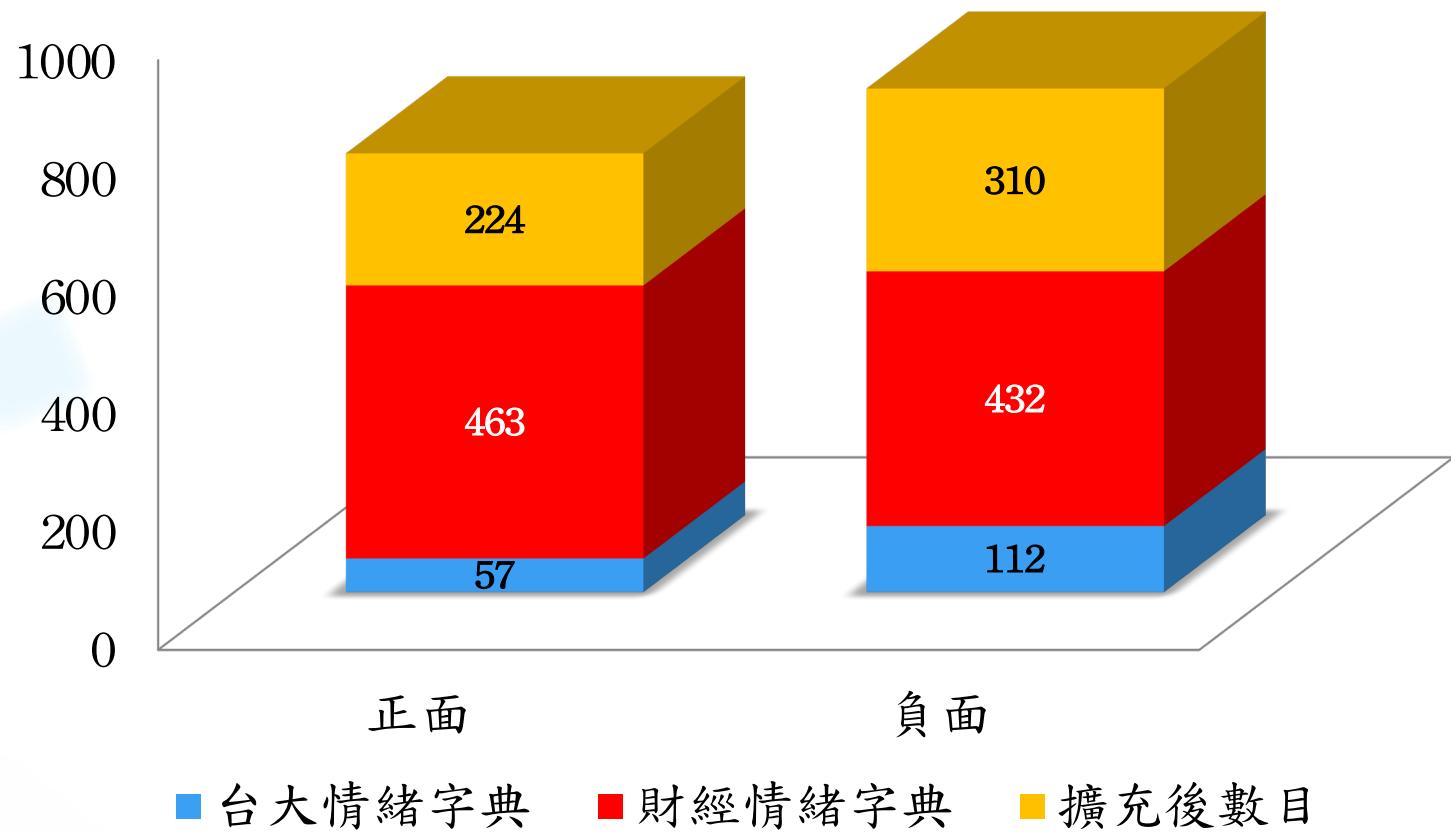
期中進度簡介 (文字資料分析)

定義一般
情緒字典

針對不同
主題擴充
情緒字典

彙整情緒
值計算

期中進度簡介 (文字資料分析)

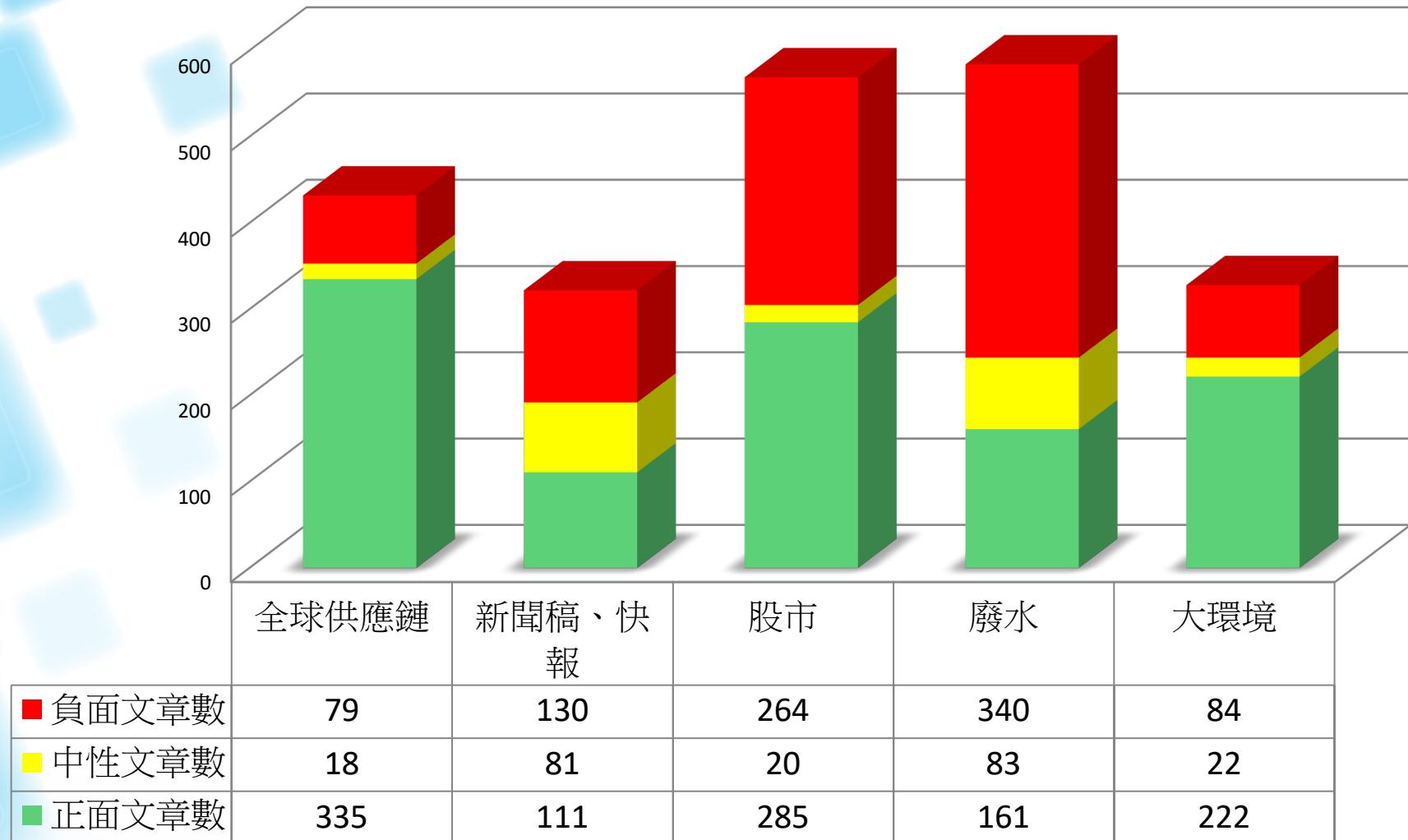


期中進度簡介 (文字資料分析)

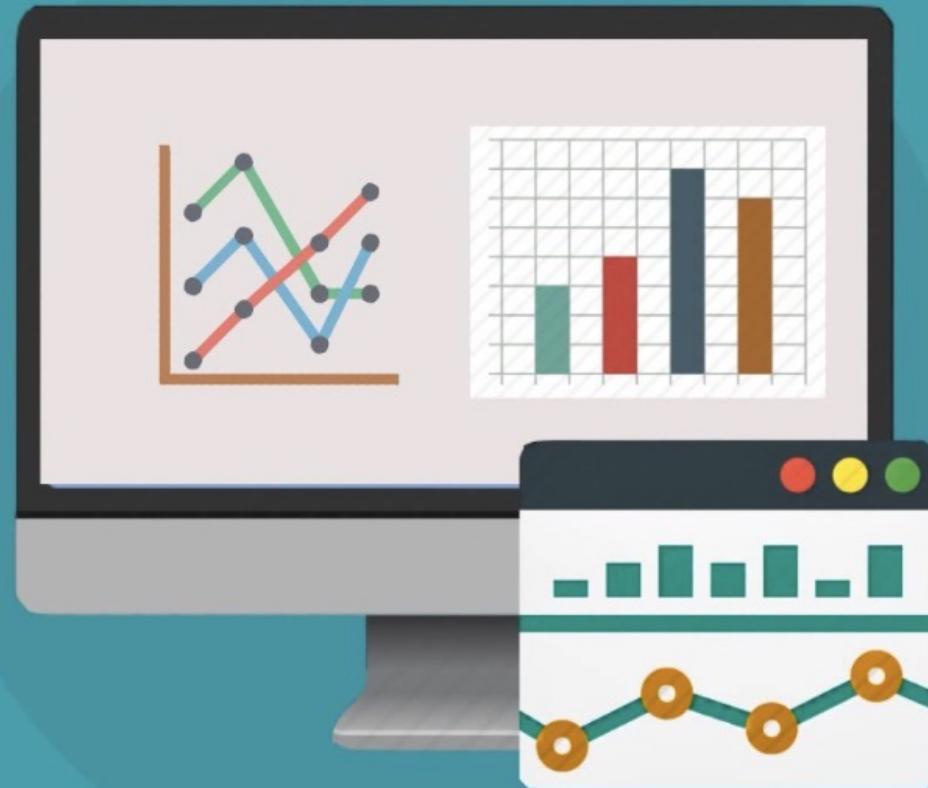
Total +11

矽品(2325)董事長林文伯今在線上法說會上表示，依照目前景氣以及台幣30元兌1美元的匯率來看，今年第1季的財測為，營收預估落在173。36~180。9億元，約季減4。2~9。6%。林文伯表示，今年第1季**高階封測產能利用率**比去年第1季要好很多，目前覆晶封裝以及晶圓凸塊的產線**利用率**維持**高檔**，第2、第3季也可以**維持**近滿載局面。展望金年，林文伯表示，許多研究機構對於今年經濟情勢轉趨**樂觀**，加上個人電腦產業將**不再衰退**，中低階智慧型手機的**熱銷**，將可**抵銷****高階**智慧型手機銷售趨緩的情況，整體的智慧型手機與平板電腦銷售量可以維持雙位數的**成長**，而半導體產業今年將有達高個位數的**成長**。林文伯說，矽品今年的**成長**動能主要來自於覆晶封裝FC CSP以及Bumping晶圓凸塊，應用產品以智慧型手機、平板電腦為主，另外，還有雲端運算、基地台、資料中心等應用也**增加**覆晶BGA的需求。至於市場關心日月光高雄K7廠部分製程**停工**所帶來的轉單效應，林文伯並未多提，僅表示，並沒有很明顯的資訊告知是來自於客戶的轉單，關於轉單無法評論。

期中進度簡介 (文字資料分析)

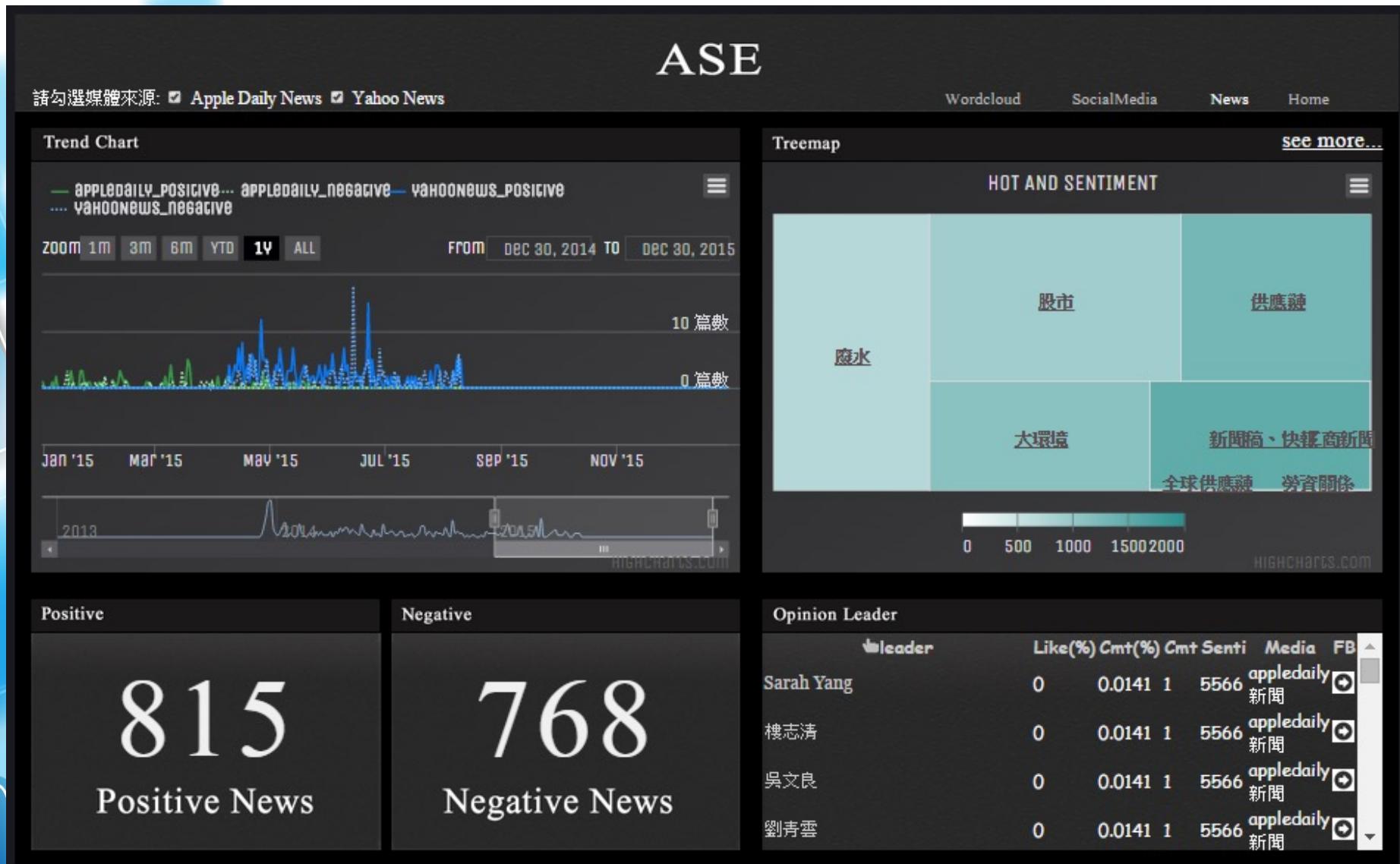


期中進度簡介



視覺化設計
看板與多維度資料表呈現

期中進度簡介 (視覺化設計)

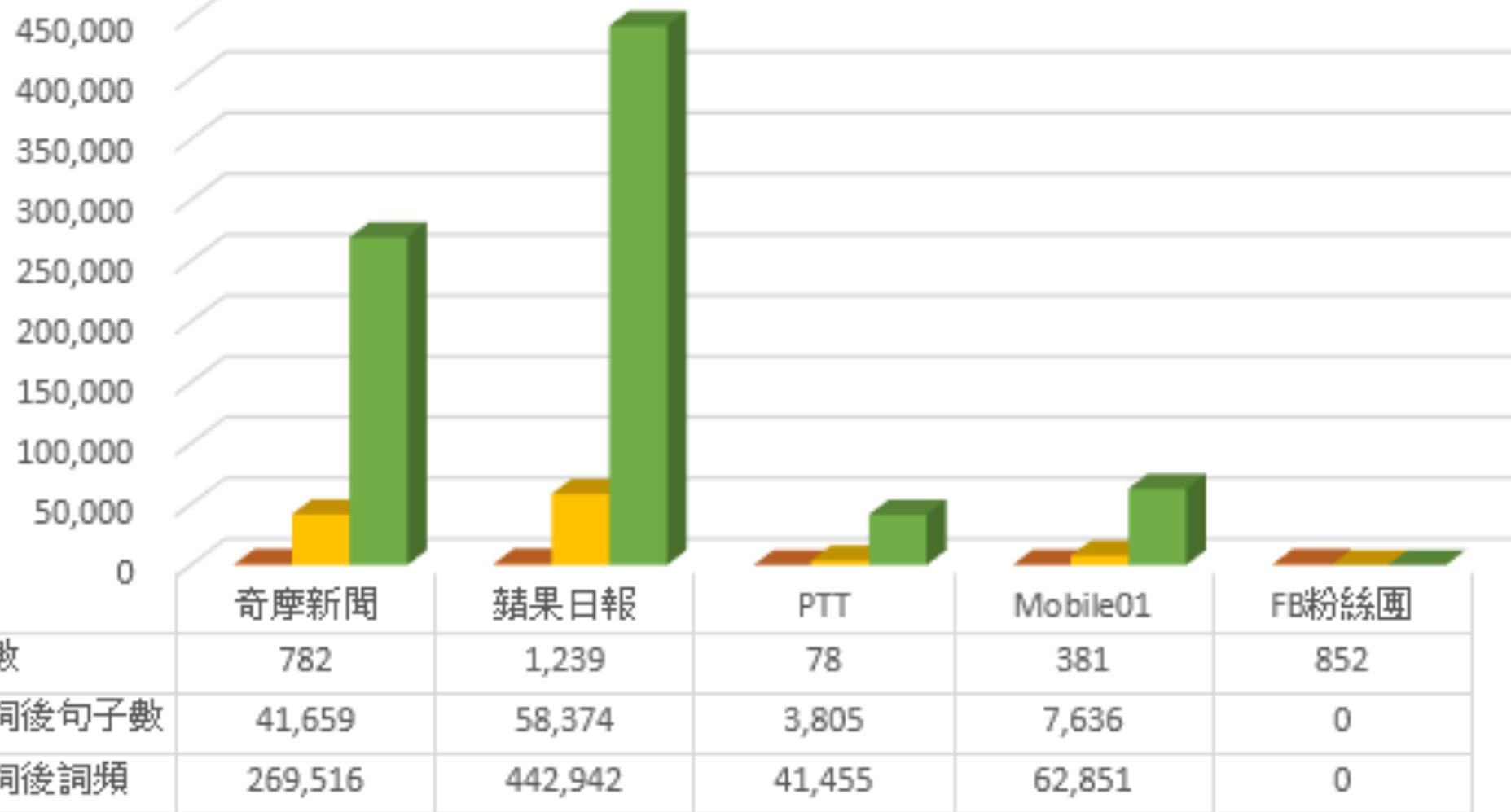


計畫期末結果報告

計畫結果報告 (網路資料擷取)

媒體	起迄時間	內容	篇數
YAHOO! 新聞 奇摩	2015/4/9~20 16/03/15	以日月光為關鍵字搜尋結果一百頁的資料	經過關鍵字排除以及去除純圖片共 782 筆，包含留言、斷詞斷句、詞性標記等資料
蘋果日報	2013/12/1~2 016/03/15	以日月光為關鍵字搜尋結果一百頁的資料	共 1239 筆，包含留言、斷詞斷句、詞性標記等資料
批踢踢實業坊	2015/11/1~2 016/03/15	版面包含八卦版及高雄版	共 143141 筆 (78 筆有關日月光)，包含留言
MOBLE	2004/10/19 ~2016/03/15	版面包含職場甘苦談、閒聊與趣味、投資與理財、新聞與時事	共 11000 筆 (381 篇有關日月光)，包含所有回覆
facebook	歷年所有文章	日月光粉絲專頁文章	852 篇文章，包含所有回覆

計畫結果報告 (文字資料處理)



計畫期末結果報告 (社群評論分析)

NLP :

在結巴斷詞系統中加入PTT常用詞庫加強斷詞效果



幫高調
懶人包
426
卡卡獸
台GG

新警察
打臉
法院認證
種花
魯蛇

22K
血汗
十萬青年
廢文
莫忘督割

計畫期末結果報告 (社群評論分析)

情緒分析：

加入表情符號以及髒話列表以擴充情緒詞庫，並同時以人工過濾常出現在社群媒體中的詞彙，找出更多情緒詞



表情符號



髒話

正面

高興 61
愉悅 80
愛 18
期待 22

負面

憤怒 37
哀愁 22
悲劇 20
沮喪 6

負面 111個

計畫期末結果報告 (社群評論分析)

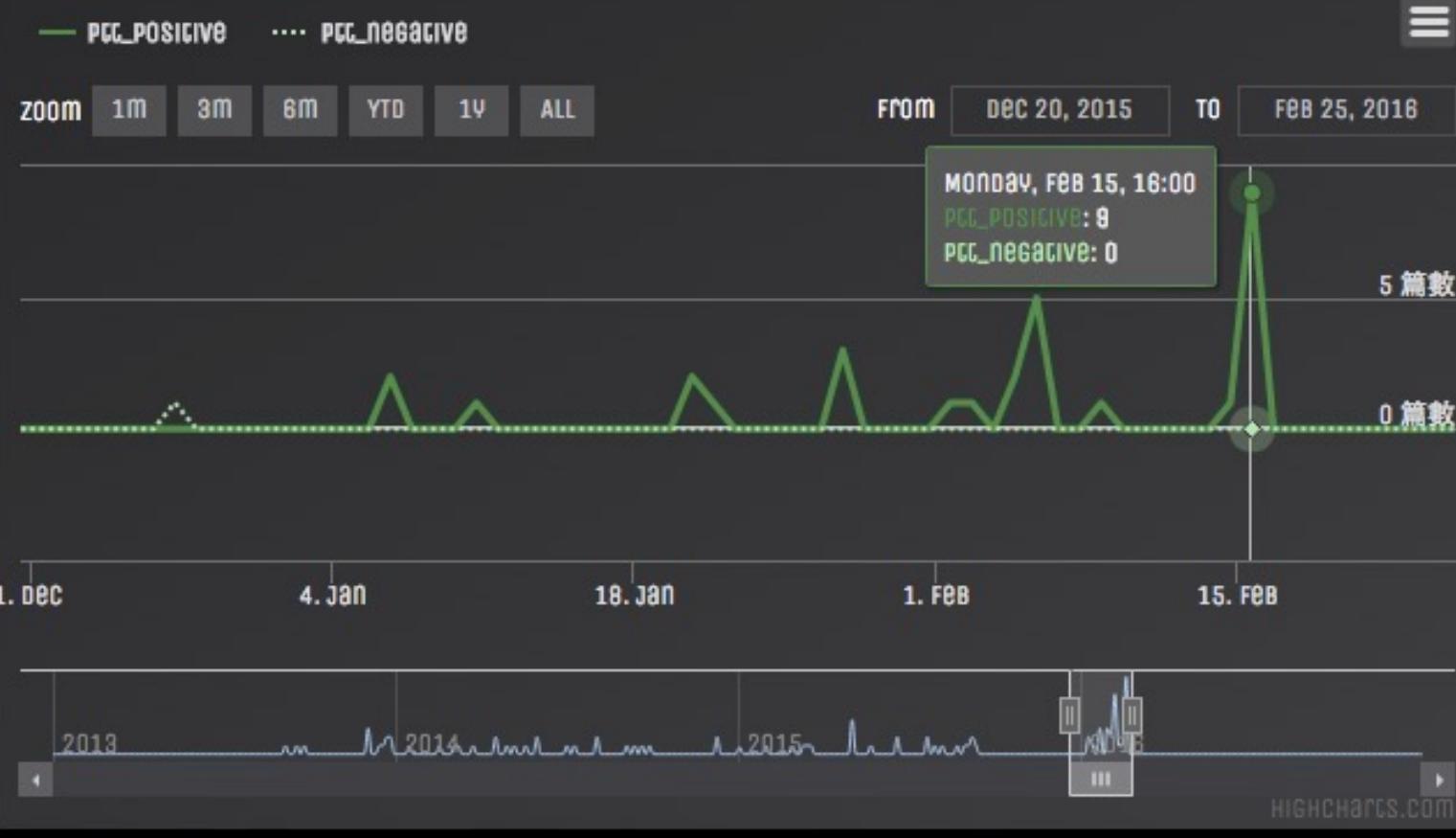
主題分類：

以新聞媒體建構的LDA模型分類器分類PTT和mobile01的相關討論串，找出輿情對於新聞主題之間的關係



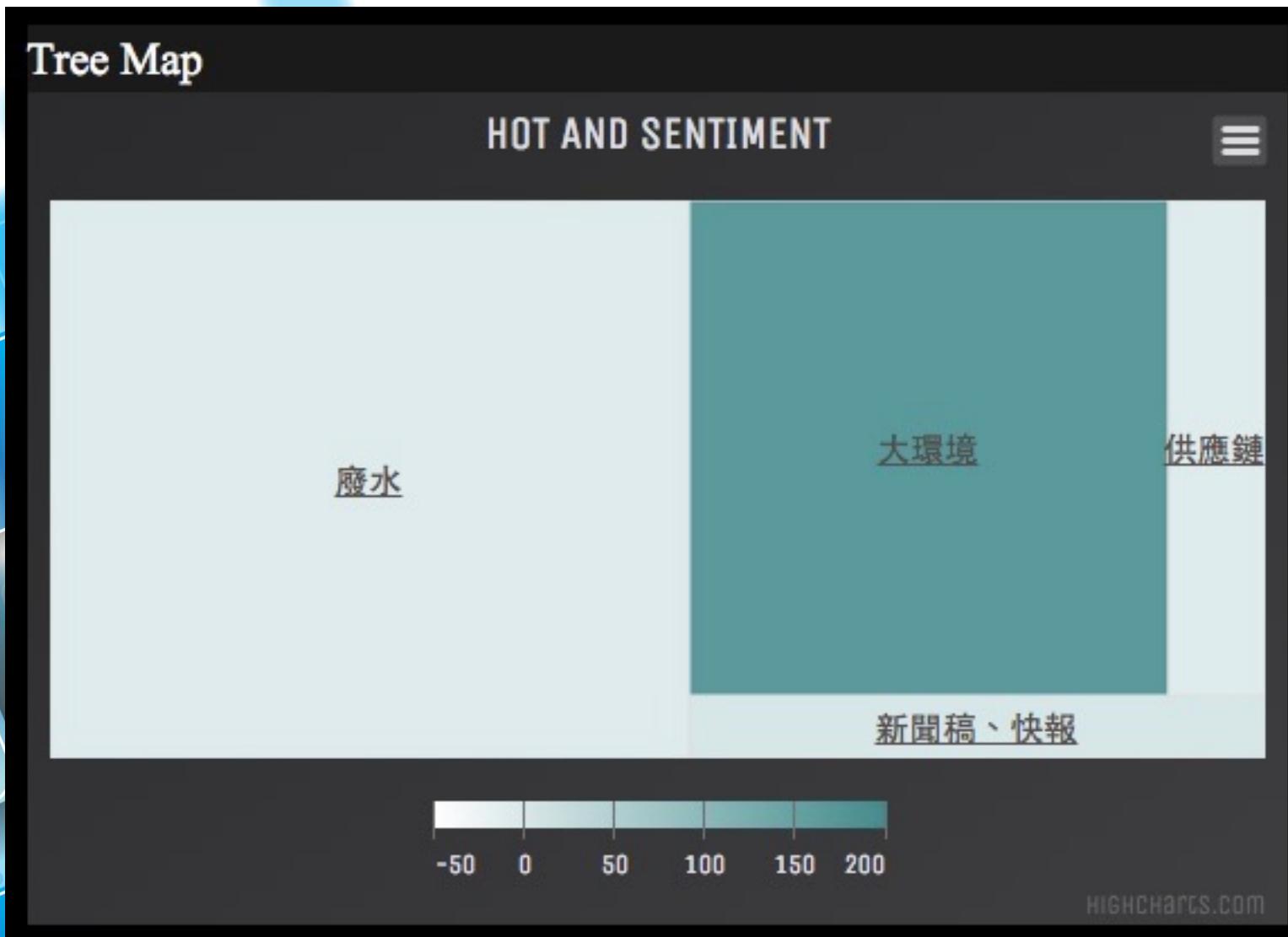
計畫期末結果報告(Visualization)

Trend Chart



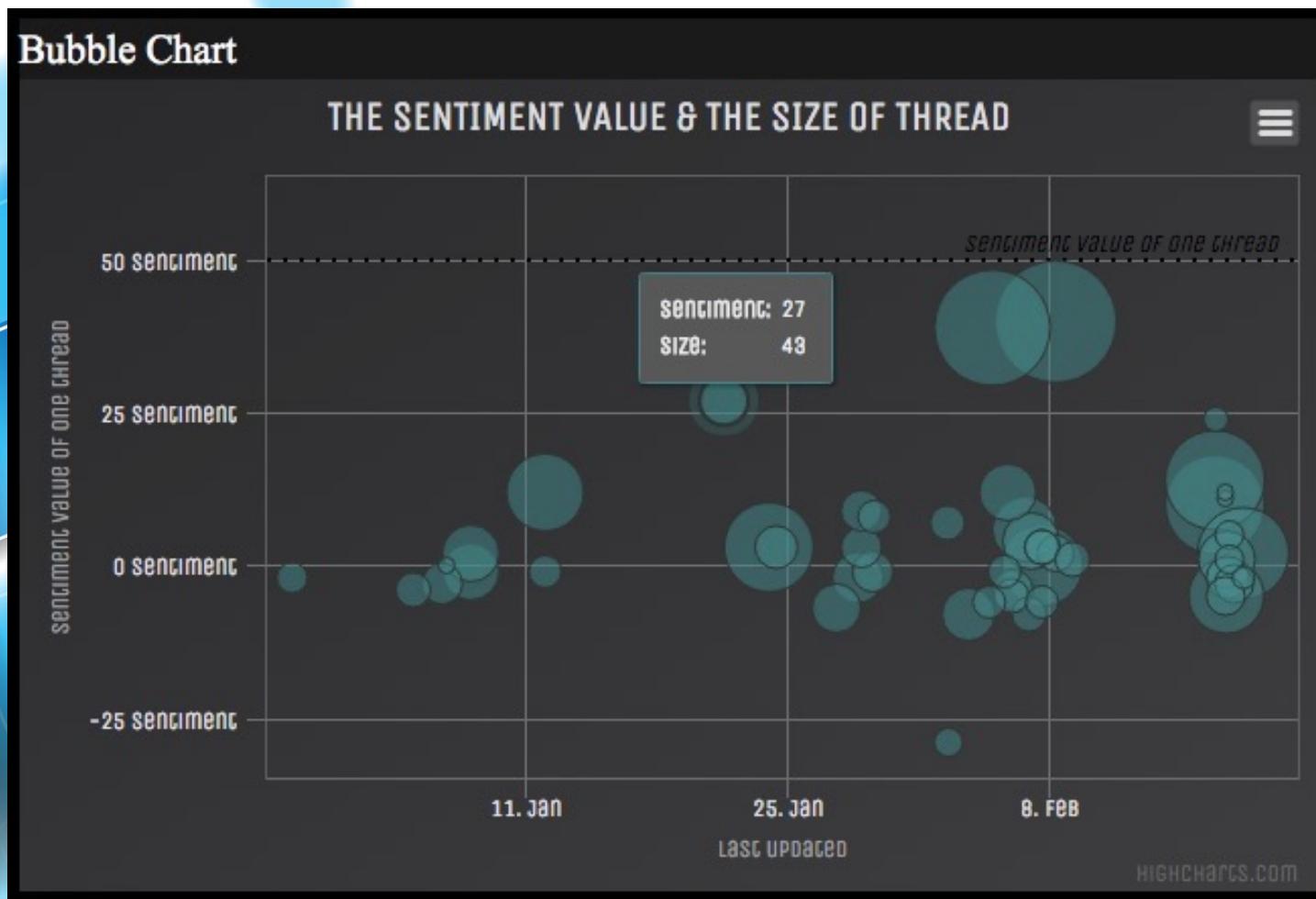
社群評論分析
整體趨勢

計畫期末結果報告(Visualization)



社群評論分析
整體主題

計畫期末結果報告(Visualization)



社群評論分析
文章重要性呈現

計畫期末結果報告(Visualization)

rusty44106

2007-01-24 11:01:00

樓主講的沒有錯 今天我們看到企業不斷的出走 資金不斷的外移 台灣的競爭力的確在耗損 企業有不得不走的壓力 我們不能一味的拿民族情感去要求企業留下來 上游的供應商和下游的客戶都在那裡 怎麼不走 他們有他們的員工和股東 這些都是他們要負責的..... 只是我擔心的不是現在 未來才是更艱鉅的挑戰 因為..... 畢竟現在還處鎖國階段 都已經出走成這樣了..... 那以後真的開放三通呢? 大大們 你要是台灣領導人 你會怎麼作呢? 這無關政治立場 但要是我 與其讓我的戰士餓死在牢裡 我還是會讓他們上戰場 至少悲壯得多.....

社群評論分析
個別評論

計畫期末結果報告(Visualization)



計畫期末結果報告(Visualization)

Timeline

利用時間軸的方式呈現新聞，並用顏色來表示其重要性



計畫期末結果報告(設定)

自動執行(Crontab)

在系統內設定好參數後便可自動執行程式

參數設定

The screenshot shows a web-based configuration interface for a news crawling system. At the top, there is a navigation bar with tabs: '文章爬蟲' (selected), 'NLP模組', '使用者字典', '情緒標記模組', 'LDA模組', '警示系統', and 'log'. Below the navigation bar, there is a section titled '資料來源' with radio buttons for selecting news sources: '蘋果日報' (selected), 'Yahoo奇摩新聞', 'Mobile01', 'Facebook FanPage', and 'PTT'. The main area contains several parameter settings:

- 查詢關鍵字:** 日月光
每輸入一個關鍵字換一行
ASE
- 關鍵字黑名單:** 家飾館
每輸入一個關鍵字換一行
股市
- 資料撈取時間間隔:** 15
以分鐘為單位
- 回應更新上限天數:** 21
更新設定值天數以內的新聞回應
建議設定14~28天

At the bottom right of the main area is a green button labeled '修改'.

計畫期末結果報告(設定)

Alert System

當新聞或社群評論數超過設定的門檻值，寄信到管理員

參數設定

文章爬蟲

NLP模組

使用者字典

情緒標記模組

LDA模組

警報系統

log

設定警報門檻
設定發送警報的條件

15 分鐘內 0 篇新聞
或
15 分鐘內 10 篇社群評論

修改

設定郵件寄送對象
警報郵件寄送對象

shake
user@example.com

新增

警報郵件寄送對象

peichi - peichi40233@gmail.com

系統效益

檢討過去

新聞及社群討論主題變化

新聞及社群數量變化

新聞及社群情緒變化

警示現況

即時顯示最新PO文

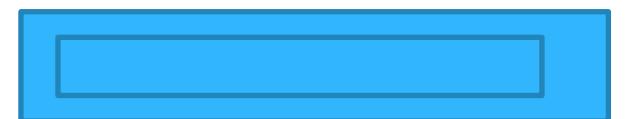
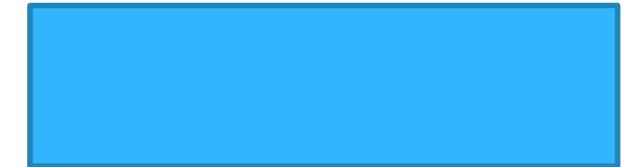
警示事件變化

展望未來

根據輿情訂定因應策略

計劃未來展望

- 將系統應用至英文新聞
- 動態主題塑模 Topic Evolution
- Entity(Aspect) Extraction
- 更詳細的情緒分析方法



Thank You

www.aseglobal.com