

公民回報災害訊息綜整研判技術

Citizen Reporting: Disaster Integrating and Analysis Technology

主管單位：國家災害防救科技中心

劉致灝 蔣佳峰 張智昌

摘要

在科技的發展下帶動民眾不同的生活使用習慣，傳統的通訊方式藉由電話、電視與廣播進行資訊的散佈，而資訊科技發展下形成目前最大的資訊傳遞管道與環境，即是社群網路傳播管道。現今民眾利用社群網路服務，於災害發生當下傳遞各類型的災害訊息，有效的將現場第一時間的災害訊息傳遞出來。另外，在社群環境上有更積極參與的社群網路公民，於 2016 尼伯特颱風期間，自行建立災害訊息回報地圖，讓參與社群網路的公民能快速的蒐整相關災害訊息，這樣的公民參與防災的力量，應要有效地納入中央災害應變中心，作為政府應變期間，蒐整災害訊息資訊的來源之一。

為有效蒐集社群網路資料來源，蒐集熱門社群網站來源為優先處理的任務，從過去社群網路公民參與的軌跡，找出最多使用者的社群網路服務，利用網路攀爬技術來蒐整相關的社群網路，目前共建構 19 各主要來源，213 各來源頻道。之後透過文本分析技術擷取公民提供的災害訊息，導入中心應變作業流程，支援中央災害應變中心之災害訊息蒐整作業。

關鍵詞：社群網路、災情綜整、應變作業

Abstract

Driven by the development of technology, people change their lifestyles. Traditional methods of communication are usually telephone, television and radio. The development of information technology has built the largest channel to deliver messages in the social network. Now, people deliver all kinds of disaster-related messages during disaster by social network. This work can be done at the first time when any disaster event occurred. In addition, there were some active users in the social network creating disaster message reporting map when typhoon Nepartak hit Taiwan in 2016. This map can make users collect messages quickly. We consider the power of netizens participating information delivery, should be one of the source of gathering information by Central Emergency Operation Center (CEOC) during response phase.

In order to effectively collect social network information, the first task is to collect

popular social channels. According to the record of netizens participating message delivery, we can find popular social websites or forums. Through the crawling technic, we have built 19 major websites, including 213 channels. After that, any useful messages should be converted to disaster information by text mining. This approach is an important task supporting CEOC to integrate situation.

Keywords : Social network 、 situation integration 、 response phase

一、前言

災防科技中心透過社群網路作為資訊蒐集的管道，共攀爬19個主要社群網路平台(共213社群頻道來源)，將資料快速持續性的匯入社群資料庫，提供後續數據分析使用。因社群數據大多是純文字資料，對於大量的社群數據處理上，基於關鍵字詞作為資料過濾的準則，並且透過關鍵字詞的分析，能觀測社群網路當下最熱門的議題，同時藉由關鍵詞的設定與追蹤，也能呈現出網路議題的發展趨勢。最後利用關鍵詞過濾後的資料，再運用文本分析的技術，找出最有可能的災情描述文章，達到蒐整災情資訊的作業。

二、研究分析及成果

• 1.4-1 研發主動式資料擷取技術

依據過去災害事件累積之社群網站資料來源的經驗，災情資料來源大多被分享在臉書的社團網站內，並且分為私人與公開，而個人化的資料分享，則需透過臉書 Hashtag 的方式進行查詢與擷取。今年度在攀爬研發上，以模擬使用者行為的方式，解析網頁內容資料，來取得動態的臉書社團、活動及 Hashtag 之相關資料來源，並匯入本中心社群資料倉儲。同時增加 Hashtag 的動態設定方式，結合臉書 Hashtag 搜尋。蒐集內容包含作者、發文時間、內容及地點，並開發後台資料會送管理功能，提供應變時彈性使用。同時也強化整體攀爬系統的效能，社群資料庫查詢量可同時支援 800 萬筆資料查詢，攀爬資料即時性達到 10 分鐘，並且資料索引可同步 300 萬筆資料，而資料儲存以每月 500~700 萬筆計算，可支援一整年的資料攀爬。(圖 1)



圖 1. 社群攀爬技術與資料服務水準

透過社群資料庫的檢索，前台提供社群資料的分析與使用，在社群異質資料整合與資料洗滌，針對資料攀爬技術所擷取並解析轉化後的國內各大論壇、新聞、社群媒體等結構化資料，提供關鍵字檢索、欄位過濾等資料操作功能，並透過設定介面選擇有效篩選已蒐集的所有來源頻道(全來源)，或直接針對特定之來源頻道進行社群資料觀測(特定來源)。(圖 2)



圖 2. 查詢條件設定

查詢結果，將蒐集之來源資料以主/回文最新時間做為排序依據。本年度建立檢索引引擎能依據所指定之欄位進行排序，並研發相關度排序之方式，針對關鍵字搜尋結果其關鍵字命中相關度作為加權分數，並以該加權分數進行文章的排序。文章呈現資料的欄位包含標題、內文、時間、來源頻道等，如該

文章有圖片則列表額外呈現。同時依據網路提供的圖片內容，也利用自動化擷取的方式，擷取圖片網址來呈現相關資訊。(圖 3)



圖 3. 蒐集結果列表

前述針對特定的攀爬來源，如：臉書社團、活動等，今年度建構社群主動觀測來源索引架構，依據應變作業所需之優先觀測社群來源，建構快速資料索引機制，納入熱門社群觀測來源，以臉書活動/社團、Hashtag 以及 PTT 為主，並強化這些來源的攀爬能力，將資料即時性達高至 5 分鐘，增加 4 倍的資料索引效能(由 20 分降低至 5 分鐘)。整題索引架構如下圖 4 所示。

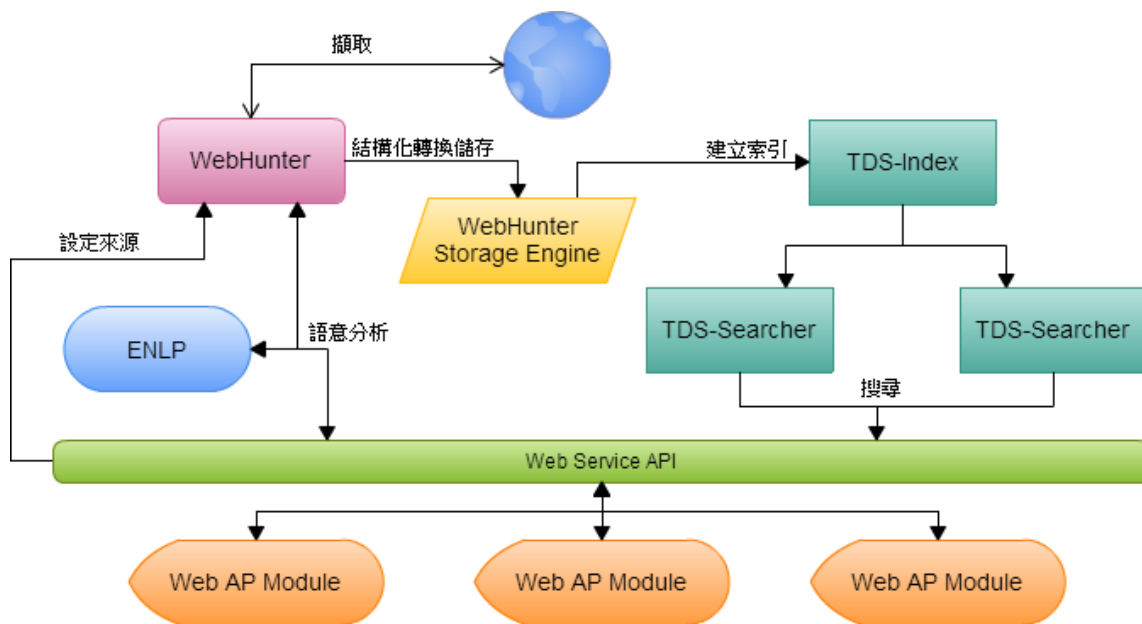


圖 4. 熱門社群來源資料索引架構

• 1.4-2 開發時序性災情儀表板

開發具時序性的社群熱門主動觀測儀表板：依據選定之特定來源頻道(三個來源頻道 Facebook 社團/活動、Facebook Hashtag、PTT)進行主動式的觀測，系統每 5 分鐘自動更新最新社群來源文章，並透過災害關鍵字詞，有效呈現當前社群網路上最新的災害訊息，建構本中心之社群網路輿情觀測站。於主動觀測方面，供查詢條件有統計主題(可複選)、統計方式(單選)、查詢主題(單選)等。資料呈現方式，以使用者設定的查詢條件(查詢主題)及既有所蒐集之來源資料，資料欄位包含標題、內文、時間、來源頻道等，時間範圍為 6 小時內之資料。若單筆資料有一張以上的圖片，圖片將會自動輪播。如下圖 5 所示。



圖 5. 社群網路輿情觀測站

其中上圖 5 中的輿情熱度觀測卡，藉由預先設定追蹤之關鍵字詞，系統依據不同時間區間計算當前網路於追蹤之關鍵字詞的討論熱度升降的百分比，其統計時間區間為 1 小時，並以顏色凸顯上升或下降之熱門聲量趨勢，藍色表示相較前一小時數值為上升狀態，紅色表示相較前一小時的數值為下降的狀態，綠色則與之前的數值相同沒有變動。並能夠一次設定多組關鍵字詞進行同步的觀測。

在本年度完成底層資料建置以及前台網站的功能開發，同時也考量外部系統的介接與整合，因此，開發標準化社群資料 API，以共通資料結構定義匯入

之社群資料，並透過資料篩選與過濾技術，建構標準化社群資料匯出 API，能有效的依據不同社群觀測主題擷取資料，以作為其他系統資料整合與呈現之應用。下圖 6 為利用 API 擷取不同觀測主題後之加值應用範例。

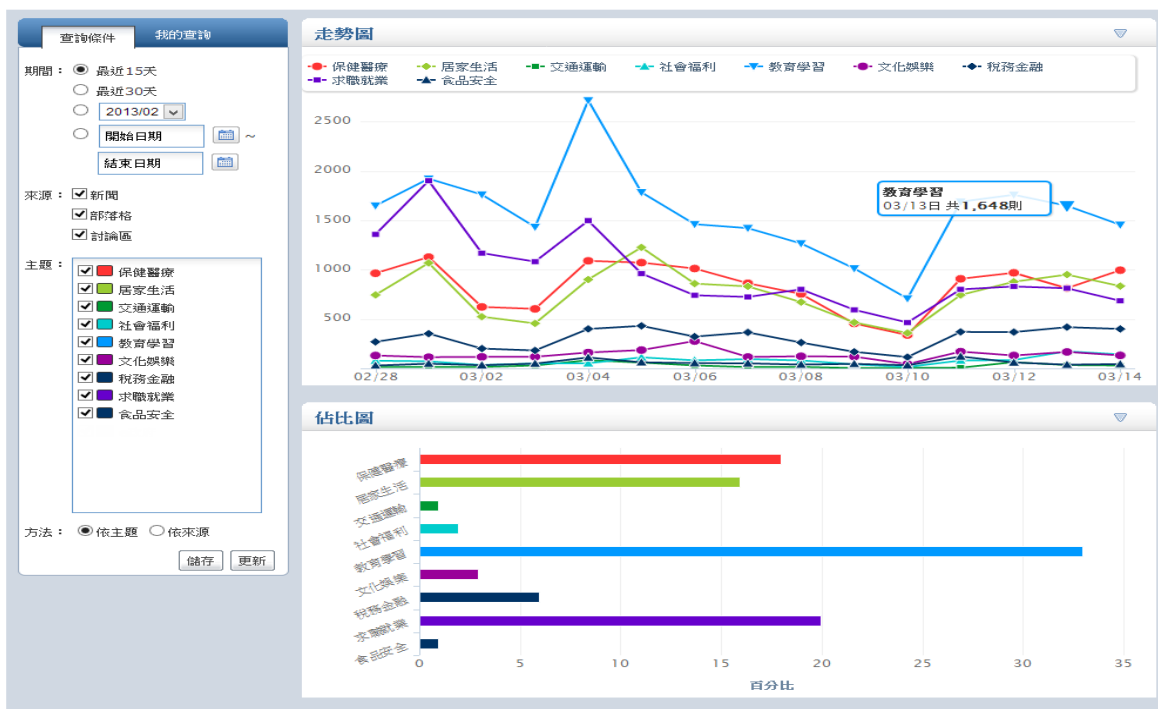


圖 6. 多主題面向社群觀測趨勢圖

1.4-3 研發社群網路關鍵字詞分析模式

在社群數據分析上，研發社群網路關鍵字詞追蹤分析機制，透過設定重要的災害關鍵字詞，以時序方式追蹤關鍵字詞的聲量變化，了解不同議題在社群網路發酵之情況，以呈現網路輿情的熱門程度。(圖 7)

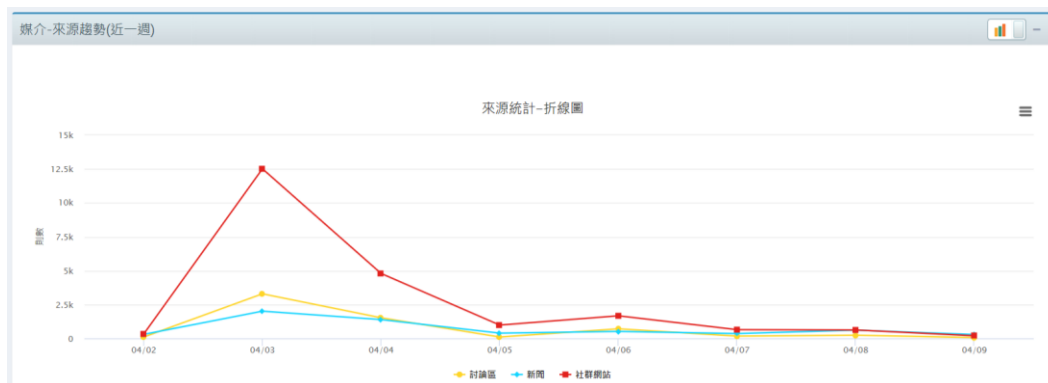


圖 7. 社群網路關鍵字詞聲量分析

另外在研發語意分析技術上，運用整理及分析非結構資料，運用無詞庫、支援非結構長短文處理技術，以更貼近人的思考方式解析文件，萃取其中關鍵

資訊、彙整重要主題、情緒分析等應用，呈現社群網路環境的多面向，並提供後續的加值應用。而權威詞的計算對於文章的使用有較大的影響，其權威詞哪種程度可代表文章的主旨或主要意涵，對於後續的模組分析有較大的關係，圖 8 為個種關鍵詞的分析說明。

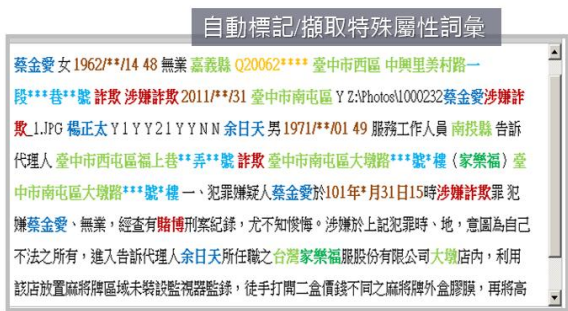


採用無詞庫斷詞法，應用字詞的位置、次數、上下文之關係，自動找出文件內重要的關鍵詞

圖 8. 權威詞分析技術

在文本分析的技術上，標記是必要的步驟，今年度針對文章屬性詞自動化標記技術，並利用機器學習技術，針對非結構化文件內既有特殊屬性的詞，依據人事時地物詞彙之分類，進行關鍵字詞的擷取與標記，並依此建立特殊屬性詞模型。(圖 9)

透過機器學習技術，針對非結構化文件內既有**特殊屬性**的人事時地物詞彙進行擷取，建立特殊屬性模型。



亦可透過標記訓練建立特殊屬性模型如水災辨識模型

姓名	王大明
組織名	家樂福
地址(點)	台北市忠孝東路四段
銀行帳戶	088888888888888
身分證號	E123456789
車牌	1234-AB
EMAIL	123@gmail.com
日期	2015年5月12日
金額	新台幣300元
物品(刑)	金飾、壓制器、電源線、狗鍊、空彈殼、海洛英
事件(刑)	行車不慎、毆打、踹開、詐騙、遭竊、攻擊、催討

圖 9. 屬性詞分析技術

三、結論與建議

- 研發主動式資料擷取技術：
透過研發特殊熱門來源(臉書社團、活動及hashtag)，提供更彈性且即時的資料攀爬功能，也強化整體攀爬系統在資料處理上的效能。
- 開發時序性災情儀表板：
底層資料的建置完成後，依據通用的資料分析，建置前台網站功能，並提供即

時的字詞分析，藉由動態的設定，能建置網路輿情觀測儀表板，呈現網路議題的發展趨勢與資料，運用在災害事件分析上，能快速觀測網路上對於災害事件的討論熱度等。

- 研發社群網路關鍵字詞分析模式：

運用文本分析技術以及機器學習技術，完成關鍵詞分析與自動化標記功能，快速產置文章內各類型詞性。該項成果也已運用到社群網路數據的地理資訊分析上。