



0206 花蓮地震事件之社群網路 災情蒐整與分析應用

劉致灝／國家災害防救科技中心 助理研究員

蔣佳峰／國家災害防救科技中心 佐理研究員

資訊科技演進帶動新興的資訊傳遞模式，智慧型手機的普及與社群網路的蓬勃發展，也提升了大數據資料分析的重要性。使用者於社群網站頻繁且大量的交換各類資訊，在重大事件發生之時，資訊將很快地被散佈在社群網路之中，若災害發生期間，能即時將這些情資加以蒐整，對於災情的掌握應有一定程度上的助益。本文將說明自社群網站蒐整災情資訊的方法，以關鍵字詞、文章聲量為基礎，使用 2018 年 2 月 6 日的花蓮地震作為範例，呈現情資蒐整的分析與結果。

簡介

資訊科技演進與手持式 3C 設備普及，帶動新興資訊傳遞模式，早期網路主要提供網頁給使用者瀏覽之用，此做法偏向「線上的圖書館」之形式，而現今能透過社群網路，提供使用者一個互相交換資訊的管道，大量的資料在網路環境中逐漸產生，間接帶動大數據 (Big Data) 資料研究的興起。大數據 (Big Data) 是近期熱門的研究議題，資料的基本定義包含 3V 的特性^[1]：(1) 大量 (Volume)、(2) 多樣性 (Variety)、(3) 即時性 (Velocity)。

社群網路 (Social Web) 已經成為網路使用者最常使用的服務^[2]，根據財團法人台灣網路資訊中心調查顯示^[3]，台灣網路使用者每天花 3.25 小時上網，其中使用社群服務的佔總時間的 64.32%，成為上網瀏覽項目的第一名，由此可知社群網路的熱門度。目前廣泛使用的社群網站如：臉書 (Facebook, FB)、推特 (twitter)、批踢踢 (PTT) 等。使用者發布資訊分享給好友或其他使用者瀏覽，一般而言，任何使用者都能對於資訊提供回饋與心得，達到資訊流通的互動性，使用者也能加入社群服務的社團，定期獲取特定主題的資訊。

另外，根據 EMC Digital Universe Study 預測結果顯示，每年社群網路資料量約兩年成長一倍，到 2020 年會成長至十倍^[4]，表示網路資料成長的幅度。這些資料具備不同的類型，如：檔案、影像、多媒體等，具備資料的多樣性，同時這些資料會隨著時間大量且快速的產生，要能夠即時性的蒐集與彙整這些資訊，才能產生出應用的價值，這也是資料的即時性，而我們正是希望能利用這樣龐大的資訊交流管道，達到災害訊息蒐集的目標。

一般而言，民眾在敘述災情時稍微會潤飾一番，例如：民眾在網路上談到「路樹、招牌或電線杆掉落」、「橋梁斷裂無法通行」、「路面淹水深度近腳踝」，甚至地震災害造成「大樓、房屋倒塌」或「有人遭到掩埋、失聯」等，一篇文章開始討論當下的災害事件時，除了原作者敘述的「正文」外，亦有其他讀者的「留言」，若能從這些訊息中擷取出來，並整理成災情資訊，也許能成為一個新興評估災害情勢的管道。

然而社群網路上的資訊變動速度較快，考驗資料蒐整的速度，可靠的資料來源獲取不易，需要投注人力資源找尋，要有效率地分析潛藏於資訊之中的災情內容，提供災害決策應變應用尚有許多待解之難處^[5-7]。

根據全台災害應變單位或災害應變中心，於災害應變期間蒐整設置於各地之感測器，將水文、坡地等觀測資訊作為研判資料，並提供給指揮官做為決策輔助之用^[8]，根據完整之災情通報流程，這些觀測資料需進一步人為確認、勘誤後，始能進行資訊發布，其過程需消耗大量時間與人力資源，相對於社群網路產生的災情訊息，由於屬民間流言或街坊八卦，內容的正確性有待證實，且敘述方式之品質參差不齊，但這類訊息之優勢為：訊息產生速度相當快速、傳遞範圍亦相當廣泛，每當災害事件發生之時，位於災區之熱心民眾，會主動將最新資訊散佈於社群網路中，也有賴智慧型手機的普及，這些消息能在事發後的短時間內，讓不在場的廣大使用者們也能知曉。

基於以上觀點，本研究針對社群網路作為主要資料來源^[9]，自使用者對於災情內容所發表的文章、留言進行初步的分析，並建構「社群網路災情蒐整流程」，並以關鍵字詞與文章聲量為基礎，針對災害發生時期的資訊進行分析，藉由探索與攀爬社群網路即時貼文資訊，呈現災情研判上的實質應用效益，也期望能做為一套標準化的災情資訊蒐整流程，能即時提供災害應變情資研判期間新的災情參考資訊。

社群關鍵字詞之趨勢分析

針對使用者於社群網路發布的文章的常見特性，本文利用社群網路來源及趨勢分析找出社群網路上使用者的習慣及災情資訊的變化，同時針對災情內的時空地理資訊進行擷取與綜整^[10]。本文會依次介紹社群

網路災情綜整流程及社群網路來源及趨勢分析、災情時空綜整機制及案例成果展示，以2018年2月6日深夜，於花蓮發生之強烈地震（以下簡稱為花蓮地震），並呈現出社群網路於災情研判上所帶來的助益。

社群網路災情綜整流程

由於社群網站屬於提供給使用者自由發表言論的管道，其抒發的內容並無多加限制，且沒有統一的格式，屬於自由文本（Free Text），即無格式、無限制書寫方式的文章內容存在於網路上，而新文章與新留言隨著時間與事件的進度（進展）而增加。基於以上特性，我們應對來自社群網路的資訊須有一些認知：社群網路上的災害資訊屬即時性，任何重大事件產生的短時間內即出現大量相關資訊，如同新聞快報般，越是具有較高的談論價值的內容，產生速度與傳播範圍隨之提升，其目的為激起民眾熱烈的討論，進而產生更進一步的詳細、最新資訊。

我們在事件發生階段的第一目標，乃是迅速獲取災害資訊，並將災情資訊中的「地理資訊」、「時間資訊」兩種重點項目進行展示，提供給情資研判做為參考資料。以下將說明一套「社群網路應用於災害情資蒐整」的流程，藉由常見的訊息處理的方法：先將資料蒐集匯入後，透過文本分析技術（關鍵字詞分析）取出重要災情資料，再將災情資訊配合時空資訊彙集成災情地圖。同時藉由關鍵詞的分析，能觀察災情趨勢分析來監測災害事件的發展軌跡，圖1為自社群網路蒐集資訊並加以綜整的流程圖。



圖 1 社群網路災情綜整流程

目前具有各式各樣的社群網站使用愛好者，不同年齡層與居住地區的使用者，其選擇也不盡相同，歐美國家盛行使用推特 (Twitter)、Reddit，以台灣為例，臉書、IG (Instagram) 與 PTT 較為被廣泛使用，屬使用者族群較多的社群服務。災害期間，在災害發生地點附近的民眾，透過大多透過智慧型手機，拍攝或錄製最新的情形，並上傳至社群網站上，發布文字、圖片與影片訊息，如：「某某大樓倒了」、「某某大橋斷了」，並附上大樓倒塌的照片。這樣的訊息一旦出現在社群網站上，其他災區使用者便開始跟進，回饋更多的災情資訊。

本研究藉由網路爬蟲，以臉書的開放式文章與 PTT 作為主要資訊來源，並事先建立既定的資料格式，將社群網站的資訊，爬取至資料庫中，以提供後續資料分析使用。這些資訊在不同的社群網站上的呈現格式不盡相同，我們則是將所有能蒐集的資料一併進行統整，使用統一的格式儲存這些第一手資訊，為有效過濾文章，我們以常見災害相關字詞作為搜尋基礎，從資料庫中挑選可能與災害相關的文章，藉由文章數量變化作為使用者對於該事件的關注度，同時代表災害事件的進展，關鍵字詞的過濾蒐尋亦有助於災情的地理資訊與災害描述資訊的過濾。

以文章數量作為社群聲量高低的標準，除了能凸顯該社群網站自身的熱門程度，且能代表災情規模或影響程度，「來源頻道分析」可透過「使用者數量」與討論各式事件「文章數量」的活躍程度，評價社群網站的影響力；「災情趨勢分析」可以從討論的相關文章內容，看出災害造成的嚴重度及時間分布。當相關文章數量達到高峰，又這些文章提及的地理詞資訊，顯示充分的時空間資訊，針對資訊內容擷取地理、時間及災害情資的資訊相對較為順利，整理後以地理分布、時間排序提供災情綜整資訊。

社群網路來源分布與災害趨勢分析

從社群網路爬取資料並儲存到單一資料庫的工作上是一項較困難的工作，攀爬技術需要針對不同社群網站提供的 API 或是網頁結構來進行分析，透過不同的攀爬策略去擷取相關資訊。資料庫收集各來源網站，包含：社群網站、各大討論區及個人部落格 (如撲浪、Xuite) 等類型網站，並透過索引 (Indexing) 將

儲存的資料有效管理。關鍵字是常見的過濾機制，以取得災害相關的社群文章^[7]。配合關鍵字詞的時間分布的統計，可分析：(1) 社群網路頻道來源分布、(2) 災害趨勢分析。

社群網路頻道來源分布，以收集的社群網站來源為基礎，基於災害相關的資訊量，作為該社群網路來源的指標性評量，計算方法可概略的統計某一種災害類型，其常見之關鍵字詞 (如地震相關事件則常使用「倒塌」、「震度」、「芮氏規模」、「掩埋」等相關字眼)，出現於該社群網站上的討論文章數量，也可視為此網站在該災害事件的討論熱度，通常稱為社群網路聲量。

依照來源社群網站於某一組災害關鍵字詞，所獲得的社群聲量，能凸顯該社群網站可靠度。一般而言，社群網路來源聲量越大，代表該社群網站的使用者較多，也表示對於災害議題的討論度可較為熱烈，重要資訊潛藏在其中的機會也相對較高。我們以花蓮地震發生後一星期內，根據地震相關字詞過濾後，社群網路來源分布結果，詳如圖 2 所示。

圖 2 為花蓮地震後一星期，熱門社群網站之聲量累積分析，時間區間為 2 月 6 日當天 00:00 至 2 月 13 日 23:59，共收集 367,311 筆資料。以 FB 及 PTT 為最本次事件主要社群類災情資訊來源。從數據上也顯示大部分的網民經常性使用的社群網站，表示 FB 的粉絲團是使用者主要的資訊交換媒介。

社群網路是一個龐大的資料傳遞管道，使用者主要發表各類型文章，以求瀏覽者透過留言的方式進行互動，一個新文章的出現其新鮮度較高時，在短時間內能匯集較多使用者觀看，討論量也相對較高。

當災害發生之時，首位目擊民眾且將此資訊上傳至社群網站中時，這訊息如同連鎖反應一般，激起現

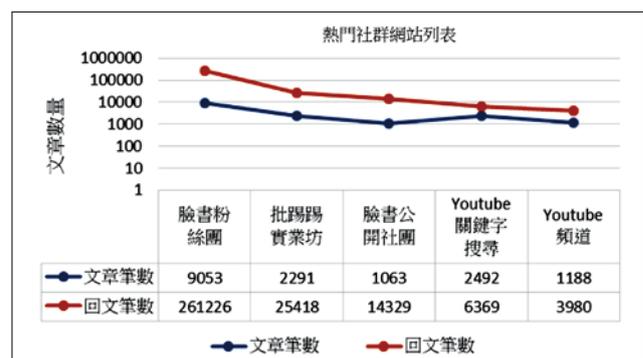


圖 2 熱門社群網站 (社群頻道來源聲量分析)

場民眾與社群網路上的觀看者熱烈的討論，災區內的民眾開始爭相發布自己獲得的資訊，活絡此事件的討論熱度，

圖 3 顯示，經地震相關的關鍵字詞過濾後，花蓮地震發生期間後六小時內，於 FB、PTT 等社群網站的聲量統計。

從圖中可以看到，相關文章數量於 2 月 7 日的午夜零時出現最高峰，而花蓮地震實際於 2 月 6 日晚間 11：50 分左右發生，在下一個時間區間，即 2/7 的 0：00 至 1：00 之間，與相關文章數目急遽升高，顯示社群網路在此事件發生後的反應結果，雖然社群網路沒辦法在事發的當下立刻反映討論熱度，原因為此類文章是依賴使用者主動發布資料迅速與否而定，但我們估略發生時間並在實際時間後一小時內，獲取足夠數量的文章數量，作為資訊倉儲或災情分析之用。

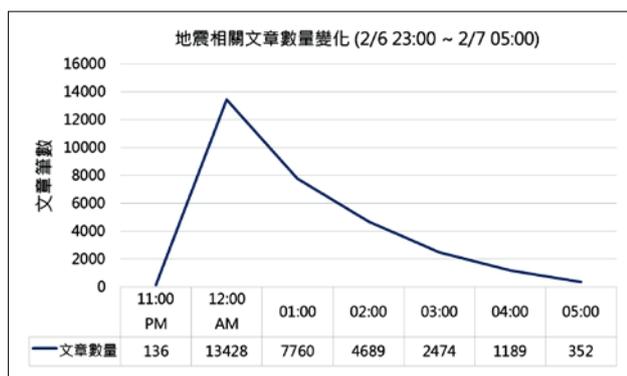


圖 3 2/6 23：00 至 2/7 5：00 花蓮地震相關文章統計

關鍵字詞整理流程

由前面所述，使用者於社群網路發布的資訊多以自由文本，陳述品質不一、用字遣詞較通俗的文章內容存在於網路上，而新文章與新留言隨著時間與事件的進度（進展）而增加。較為棘手的是，中文句子需經過許多事前處理，重要的關鍵字詞才能一一浮現，此階段的第一目標，為有效篩選句子，將不必要的常見字過濾，並正確的將句子切割成詞類的集合。

本文將說明自社群網路蒐集資訊，並將關鍵字詞應用於災害情勢追蹤的流程，藉由常見的訊息處理步驟：「資料蒐集」、「文本分析」、「詞類比對」等 3 個步驟，並結合災情趨勢分析來追蹤災害事件的發展軌跡，圖 1 為自社群網路蒐集關鍵字詞並加以綜整的流程圖。

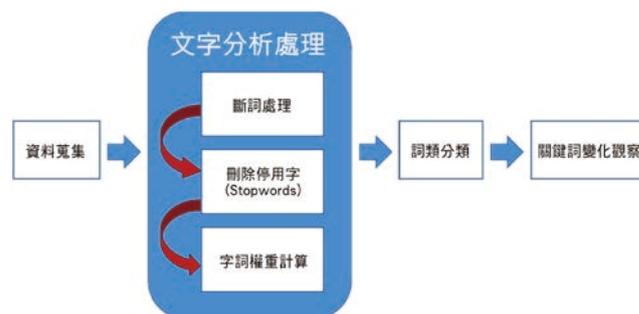


圖 4 關鍵字詞整理流程圖

首先藉由網路爬蟲，以臉書的開放式文章與 PTT 作為主要資訊來源，並事先建立既定的資料格式，將社群網站的資訊，爬取至資料庫中，以提供後續資料分析使用。

來自不同的社群網站的資料經過正規化、統一資料格式並儲存後，我們以常見災害相關字詞作為搜尋基礎，從資料庫中抓取災害相關資訊，藉由文章數量變化作為使用者對於該事件的關注度，同時代表災害事件的進展，關鍵字詞的過濾蒐尋亦有助於災情的地理資訊與災害描述資訊的過濾。

現今網路上常用的斷字斷詞工具種類相當之多，但是支援中文語法斷詞的目前以中央研究院所提供的斷字斷詞工具，以及結巴斷詞較為人所知，我們使用後者做為本研究分析文章的應用工具。

過濾文章的部分主要分為兩個階段，(1) 去除特殊、標點符號與 (2) 去除常見字詞，即停用詞 (Stopwords)。標點符號與特殊符號可藉由事先建立列表的方式進行篩選，停用詞如：「你」、「我」、「嗎」、「為什麼」、「應該」等過於口語的字詞，這些不足以做為關鍵字，但在出現頻率卻是相當之高，但在過濾上較為困難，因這些詞類數量龐大且較無跡可循，較為常見的作法為：使用網路上他人建立的停用字列表，或經統計計算蒐集的文章內容，將在單一文章內的頻率較高且廣泛出現於各文章的常見字進行過濾，以上兩種做法皆由其問題存在，前者受制於提供者對於選取字詞的嚴謹程度，無法保證過濾效果，後者則因為特定類型出現之常見詞類，如「地震」、「警報」、「震度」等內容，因為特殊時期造成使用者集中討論於此類事件，導致該時間區間內此類字詞廣泛出現於文章中，造成這類極具代表性的詞類卻被捨棄之情形發生。

根據以上所述，並結合下一階段的權重計算，可使用以下三種做法，計算詞類的重要程度：(1) TF-IDF、(2) 卡方檢定 (Chi-Square)、(3) 資訊獲利 (Information Gain)。

做法 (1) 主要計算該字詞於單一文章的詞類 (term frequency, tf)，以及在各文章出現的文章頻率 (Document Frequency, df) 倒數的乘積，若 tf 越高表示此詞在單一文章內越重要，df 則表示是否過於常見，可能為停用字之傾向。

而做法 (2) 與 (3) 較偏向計算詞類特徵重要度，以分布在目標類別的文章與否決定字詞鑑別度，以本研究為例，我們需要找出常見於地震相關卻不常出現在其他類文章的字詞，根據詞類出現是否計算分布程度，藉此取出地震相關的重要字詞。

由圖 5 與圖 6 的關鍵字變化，可以觀察事件發生的初期，民眾較為關切的主要為地震本身的相關資訊，在前幾名的詞類內容可以解讀出：(1) 花蓮受影響較大、(2) 是否還有餘震的可能、以及 (3) 統帥飯店可能有人受困等。

而兩天後的關鍵字排序則顯示出，民眾在情勢趨於穩定後，開始檢討此災難為何造成如此損害，以及

後續的救援，老屋、強震、家園等字眼顯示房屋結構與抗震之間的關連度，以及各地方人士是否出面，對於花蓮地區進行捐款或對此事件進行評論等。

花蓮地震案例災情綜整與成果展示

「花蓮地區 6 日深夜發生規模 6.0 強震，花蓮市震度達到 7 級，各地傳出不少災情。據了解，花蓮市內有兩棟大樓倒塌，包括花蓮市公園路的統帥飯店一樓整個坍塌，飯店「二樓變一樓」，另外在商孝街「阿官火鍋」所在的商旅大樓也整棟傾倒。目前花蓮已經成立一級應變中心，國軍花東防衛指揮部也動員救災。」以上節錄自 ETtoday 對於花蓮地震的相關報導，短短的文字蘊含了大量的地理與時間資訊，以下將針對擷取地理與時間資訊進行詳細說明。

在擷取地理與時間資訊部分，藉由關鍵字詞過濾出地震相關文章，並針對發文者對於災情的描述內容中，擷取出時間、地點及災情描述，在擷取社群網站上文章的時間資訊上，通常分為兩類，(1) 文章中明確說明事件發生之時間，此類能提供較為精確的事件發生時間、(2) 此文章被發布在社群網站上面的「發布時間」，此類時間與實際災害發生的時間會有一定的落差，但通常與實際發生時間相距不遠。

地點資訊在文章中較為直觀，通常是以文字描述區域、大眾感興趣的地點 (POI, Point of interest) 等，然而要能夠輸入到地理資訊系統中，必須先經正規化即地點資訊的數值化，將文字轉換為經緯度。因此，擷取出可能為地理相關的描述資訊後，須透過定位服務轉換為經緯度再定位至電子地圖上。而災情資訊往往伴隨描述災害類型 (例：倒塌、淹水、土石流等) 及發生的損害等，將這些資訊擷取後呈現在地理資訊系統上，如圖 4 所示。

圖 7 中顯示 2 月 6 日大樓倒塌的畫面，在 2 月 7 日晚間 00:06 發布到 Facebook 上，其地點描述資訊為花蓮商校街 2 號之 1，而災害描述是房屋倒塌。右側的地圖上顯示透過定位轉換的經緯度資訊，再繪製到地圖上。

將收集到的災情皆定位至在地圖上，就可以大略介定地震造成的影響範圍，也成為災情的地理空間分布，

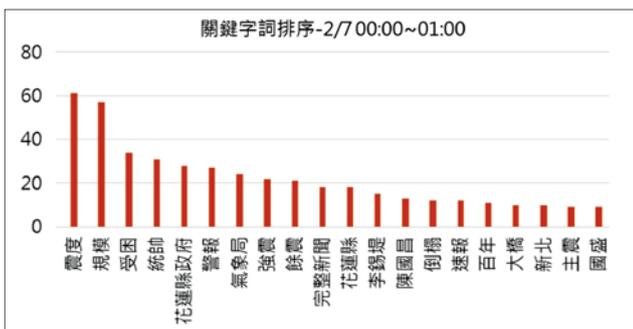


圖 5 2/7 00:00 至 01:00 花蓮地震的關鍵字詞排名

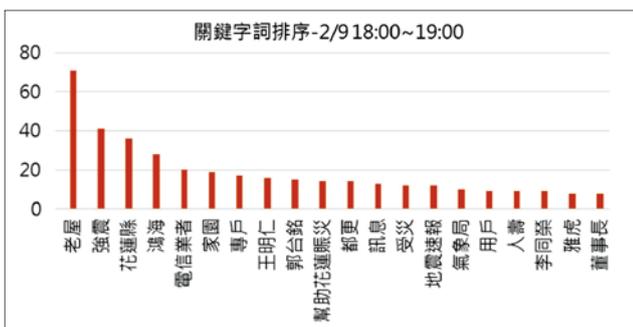


圖 6 2/9 18:00 至 19:00 花蓮地震的關鍵字詞排名

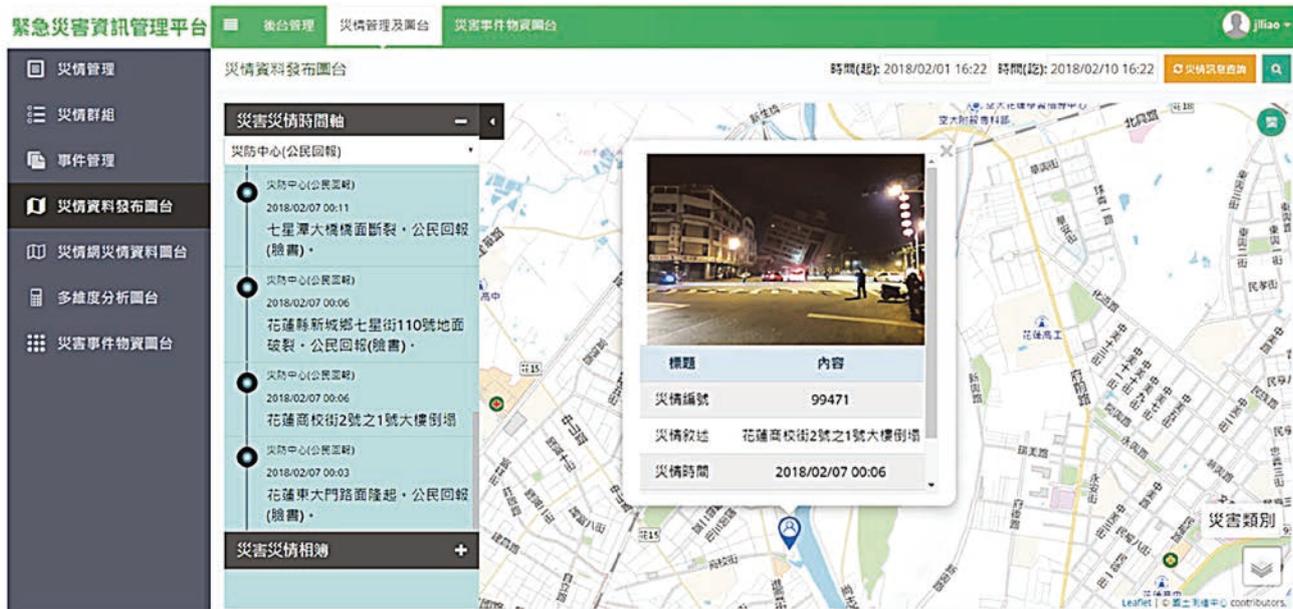


圖 7 花蓮地震災害情資



圖 8 花蓮地震影響主要區域分布

圖 8 顯示花蓮地震造成的災情範圍，可看出集中於花蓮地區，此類地理空間資訊有助於災害的影響範圍，對於救災人力的派遣、附近醫院的支援等，具有一定的參考效益。

自文章取得的災害地點資訊進行轉換、定位之後，依照時間資訊進行排序，呈現出花蓮地震的災情綜整，同時將完整的災情展示在社群網路災害綜整平台，並以瀏覽的方式提供災害時序情資。提供應變期間災害情資研判之參考依據，能夠讓情資研判時，對於災害情報的掌握更加齊全，並也能夠過現場的災情照片更快速的進行災情的判釋，如圖 9 所示。



圖 9 0206 花蓮地震災情時序

結論

本文提出自社群網路應用於災害情資蒐整，並以社群網路聲量為基礎、關鍵字詞作為過濾條件，進行社群網路來源及災情情勢分析，也呈現出台灣地區使用者普遍使用的社群網站偏好，以及災情發展趨勢，後續佐以簡單計算公式，快速計算詞類權重以取得代表性關鍵字，並觀察隨事件進展之熱門關鍵字對應變化，這些文字留言與圖片將整併為災情資訊，萃取出地理與時間的資訊後，以時序圖的方式描繪整起災害事件輪廓，呈現於社群網路災害綜整平台中。

針對災情時間、地理資訊蒐整，以 2018 年 2 月 6 日花蓮地震為案例，提供空間及時間的災情資訊，顯示出社群網路於災情資訊綜整能呈現出該災害事件的時序。從地理空間的分布上，亦可以看出災害影響的有效範圍。從此案例可觀察出，社群網路可作為一個有效的資訊蒐集來源，分析文章並加以蒐整的成果對於災害防救應用上有一定的價值，對於救災單位在掌握災情上能做為一項新興的參考內容。

參考文獻

1. 大數據, available at: <https://zh.wikipedia.org/wiki/%E5%A4%A7%E6%95%B8%E6%93%9A>
2. 社群網站的使用行為：創市際調查報告, available at: <https://rocket.cafe/talks/78006>
3. 臺灣寬頻網路使用調查, available at: <http://technews.tw/2014/08/20/twnic-online-behavior-survay-more-mobile-sulfing/>
4. EMC. The Digital Universe of Opportunities. Available at: <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf>
5. Haewoon Kwak, Changhyun Lee, Hosung Park, Sue Moon. (2010), What is Twitter, a Social Network or a News Media?. Proceeding of the 19th international conference on World Wide web, New York, USA, pp.591-600.
6. Andreas M. Kaplan, Michael Haenlein. (2010), Users of the World, Unite! The Challenges and Opportunities of Social Media. Volume 53, Issue 1, January-February 2010, pp. 59-68.
7. Pantti, M and Wahl-Jorgensen, K and Cottle, S. (2012), Disasters and the Media. Peter Lang, New York, pp. 248.
8. 林祺岳、蘇文瑞、楊鈞宏 (2015), WebGIS 緊急上圖框架應用於災害應變決策輔助系統, 台灣地理資訊學會年會暨學術研討會。
9. Opview 社群口碑資料庫, available at: <http://www.opview.com.tw/socialDB.html>
10. Zheng Xiang, Ulrike Gretzel. (2010), Role of Social Media in Online Travel Information Search. Volume 31, Issue 2, April, pp.179-188. 



土木水利 雙月刊

向您約稿

本刊出版有關土木水利工程之報導及論文，以知識性、報導性、及聯誼性為主要取向，為一綜合性刊物，內容分工程論著、技術報導、工程講座、特介、工程新知報導及其他各類報導性文章及專欄，歡迎賜稿，來稿請 email: service@ciche.org.tw 或寄 10055 台北市中正區仁愛路二段 1 號 4 樓，中國土木水利工程學會編輯出版委員會會刊編輯小組收，刊登後將贈送每位作者一本雜誌，不再另致稿酬；歡迎以英文撰寫之國內外工程報導之文章，相關注意事項如後：

- 工程新知及技術報導，行文宜簡潔。
- 技術研究為工程實務之研究心得，工程講座為對某一問題廣泛而深入之論述與探討。工程報導為新知介紹及國內外工程之報導。
- 本刊並歡迎對已刊登文章之討論及來函。
- 工程論著及技術研究類文章，由本刊委請專家 1~2 人審查，來文請寄電子檔案，照片解析度需 300dpi 以上。
- 文章應力求精簡，並附圖表照片，所有圖表及照片務求清晰，且應附簡短說明，並均請註明製圖者及攝影者，請勿任意由網站下載圖片，以釐清版權問題。