

計畫編號：MOHW108-CDC-C-114-134117

衛生福利部疾病管制署 108 年委託科技研究計畫

計畫名稱：自動化輿情監測系統與非結構性資料分析模式建置

108 年度/全程研究報告

執行機構：中華 R 軟體學會

計畫主持人：陳嬾如

研究人員：丘祐瑋

執行期間：108 年 1 月 1 日至 108 年 12 月 31 日

本研究報告僅供參考，不代表本署意見，如對媒體發布研究成果應事先徵求本署同意

目 錄

壹、研究報告中文摘要	4
貳、研究報告英文摘要	6
參、研究報告本文	8
(一) 前言	8
(二) 材料與方法	14
1. 輿情系統架構	14
2. 資料蒐集	20
3. 資料處理	43
4. 資料分析	49
5. 系統環境	51
(三) 結果	59
1. 監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之疫情輿情原計畫目標	59
2. 使用 Wikipedia 的資訊建立一多語系同義字典	73
3. 將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文	100
4. 使用名詞提取功能技術抽取內文中提及的輿情地點	101
5. 將各國疫情輿情發生地點，繪製在世界地圖上	102
6. 評估三個疫病字詞之 Precision 與 Recall 以評量輿情系統之確度原計畫目標	103
7. Python 與深度學習教育訓練	104
8. 評估 CDC 的通報資料與輿情所蒐集到的資訊與 Google Trend 做關聯分析	109
(四) 討論	111
1. 國外疫病輿情監測與分析	111
2. 比較 CDC, 輿情與 Google Trends 數據	114
3. 鼠疫疫情分析	117
4. 與 HealthMap 之比較	122
(五) 結論與建議	126
(六) 未來里程碑與展望	127
(六) 重要研究成果及具體建議	129
(七) 參考文獻	131

壹、研究報告中文摘要

大數據分析的魅力在於，可以從量大(Volume)、累積速度快(Velocity) 以及各式各樣的資料(Variety)之中挖掘數據的價值。而更可貴的是，在大數據分析中，如果能除了利用自身原有的數據，蒐集並分析其他外在的數據，我們便可以強化既有的分析結果，而得出許多意想不到但有用的資訊。以往在疫病資訊的分析中，多半都是利用醫院通報或是既有醫療體系、防疫體系提供的資料做數據分析。讓我們可以從這些數據可以勾勒出當前的疫情狀態。但除了專注在既有的數據外，如果能夠蒐集到外在潛藏的疫情資訊，例如網路上的輿論，或是外國的疫病資訊，便可以推論疫情是否有爆發或散佈的可能性，如果能快速掌握、處理國內外的疫病輿情訊息，並能迅速過濾、通報具有威脅的訊息，便能達到防患於未然的角色，因此，研究外來的輿情資訊，是否有當前的疫情有相關性，並探索其預測價值，便是本研究需要研究的主要課題。

本計畫為四年期的研究計畫，在 108 年度的計畫中，為了能夠擴展系統的監測範圍，我們將整合語言翻譯功能，能蒐集到外國語以後，將該語文翻譯成繁體中文資料，並將增加五個國家的語系資料(例如:西班牙文、葡萄牙文、日文、韓文、越南語等外國語系的資料來源，並增加上述語系監測能力到輿情系統之中。為此，我們將增加上述語系來源的網站,但因為網站眾多，因此如果要蒐集所有網站資料，近乎是不可能，但防疫卻又不可能遺漏任何一個資料源，因此，在這年度計畫中，我們將抓取 Google News 各個國家的新聞源，並透過 Google Translate 做翻譯，如此一來，即能跨越個國家的藩籬，有效抓取所有跟疫病相關的輿情。而當各國的輿情都被蒐集以後，我們將使用 NER 技術，抓取文章中出現的地理區域名稱，並根據抓取到輿情資訊的發生地點，繪製成一世界地圖，以便權責人員快速發現有哪₄

些地的疫病輿情正在蔓延。另外，隨著因應業務上的輿情分類與標記需求，而為了強化疾管署內部人員的資料分析與文字探勘能力，該年會以文字分析為主軸，並加開 Python 深度學習相關課程，藉而提升同仁的數據分析技巧。

中文關鍵詞(至少三個)：

輿情分析、輿情監測、非結構性資料、文字探勘

貳、研究報告英文摘要

The charm of the big data analysis lies in its ability to mine the value of data from large Volume, Velocity, and Variety. What is more valuable is that in big data analysis, if we can collect and analyze other external data besides the original data, we can strengthen the current analysis results and get a lot of unexpected but useful information. In the analysis of epidemic information in the past, most of them are using hospital notification or an existing medical system, an epidemic prevention system to provide data analysis. It allows us to draw a picture of the current situation from these data. In addition to the existing data, if one can draw hidden epidemic information from Internet public opinion or overseas diseases related news, one can predict the possibility of epidemic outbreaks. If one can utilize a real-time opinion monitoring system to monitor epidemic news from domestic and abroad and can further filter out possible threatening messages, we can identify and prevent possible disease outbreaks. Therefore, in this study, we will research how public opinion information correlates with the epidemic information, and to explore its predictive value, is the main topic to research on this study.

In this year's plan, we need to expand the monitoring scope of our system; we will add both the data sources and the ability to process Spanish, Japanese, Portuguese, Korea, and Vietnamese on to the opinion mining system. However, it is nearly impossible to collect all news sources from different countries of

different languages. Therefore, we will leverage google news to scrap disease-related news of the different countries. However, to overcome the language differences problem, we will use google translate API to translate the news articles. Moreover, since we have collected data from different countries, we can then use the name entity extraction technique to extract the location name from the news, and then visualize the data into a world map. Therefore, the user can quickly find out where has most disease-related news.

Keywords: Text Mining, Unstructured data, Opinion Mining

參、研究報告本文

(一) 前言

疾病防治對全世界而言，都是不可忽視的重要問題。身為官方單位的疾病管制署(下稱疾管署)，更是應為防疫把關的第一線，因此如何快速掌握疫情的相關情資，並能根據相關訊息進行預判，避免疫情可能的擴散與爆發，便是疾管署扮演的重要角色。然而隨著近年來全球化的趨勢，國與國之間密切交流，互通有無，而疫病也會隨著人口的遷移，輕易的將原屬地區性的疾病，升級成國際傳染病。因此國界與地域已逐漸不再是防止疫病傳染的界線，因此對於疫病情資的掌握，也不能只侷限於國內的資訊。如何能快速蒐集、分析全世界的疫情訊息，並建立模式以判斷疫情潛藏的威脅，便是當代防疫的重要課題之一。

所幸近年大數據與雲端技術的發展，讓大量數據的蒐集與分析已不再成為障礙。而大數據分析技術的應用，也將公衛資訊分析推向了新的境界與領域，近年來，無論是臨床或公衛的學門上，都紛紛有先進提出利用數據探勘等技術，將分析成果用在醫療、防疫、衛生等研究議題上。

近年來，有許多利用社交媒體資料嘗試回答臨床問題和流行病學問題。衛生資訊學中的資料“傳統上”是從醫生、診所、醫院等機構收集的，但最近世界各地的人們開始在互聯網上記錄衛生資訊。這些資料可能來自 Twitter、internet 查詢資料(例如 Google 搜索資料)、留言板或人們在網路論壇上或任何地方發佈資訊。這種形式的大資料帶來了額外的挑戰，如文本挖掘和處理潛在的假新聞或訊息，這些研究可能可以讓我們在公衛領域上有許多新發現與突破。以下是幾個過往知名的研究結果：

使用搜索查詢資料跟蹤流行情況

該研究的重點使用 Google (google.com)和百度(baidu.com)的搜尋指數是否可以用來預測流行病的發生和傳播。以往，美國疾病控制與預防中心(CDC)通常需要一到兩周的時間才能確認流感疾病大流行，但學者試圖利用檢索查詢資料，發現他們可以比傳統的疾病預防控制中心更快地向公眾發佈疫情流行的訊息。這項研究可以幫助醫生和醫院瞭解疫情發生的時間和地點，並即時更新最新動態，使他們能夠更快地採取行動阻止疾病的傳播，同時幫助已經感染的患者。

Ginsburg 等學者開發了一種自動化方法以分析來自 Google 的大量搜索指數，以跟蹤特定人口間是否有流感的大流行。他們對 Google 的搜索指數進行研究，使用來自於 Google 搜尋 5 年的歷史記錄(2003 年到 2008 年)，以及 CDC 的歷史資料。他們根據 2003-2007 年的資料建立模型，並對 2008 年的資料進行了驗證。將該研究成果與疾病控制和預防中心流感的報告相比，發現平均相關性為 0.97。通過該作者的方法，可以在 CDC 報告前 1 至 2 周即可預先發表報告。

Yuan 等人則使用百度搜索查詢資料開發了一個類似的系統，目的是跟蹤中國各地的流感疫情。作者從百度的資料庫(<http://index.baidu.com/>)中收集了自 2009 年 3 月至 2012 年 8 月 H1N1 流行期間的資料，並將其結果與中國衛生部(MOH)的結果進行了比較。與美國疾病控制與預防中心相似，中國衛生部的發布結果通常都會在疫情爆發後的一到兩周才公布爆發的訊息。如透過百度每天搜索相關疫情資料，便可以提供近乎即時的警示

從這些研究中發現，搜索查詢資料能快速、準確的檢測疫情是否發生的有效工具，甚至可以擴展到對疫情的跟蹤。

使用 twitter 發布資料來跟蹤疫情

與前面研究目標類似，學著們試圖檢測和跟蹤流感疫情，但是研究人員使用的是 Twitter 上的資料，而不是搜索指數。Twitter 資料相對於搜索指數的一個優勢是，Twitter 帖子帶有上下文(而不是不知道為什麼一個人要搜索一個主題)。Twitter 是一個社交網站，它允許用戶發布任何他們喜歡的資訊，字數不超過 140 個字(即 tweets)，目前在全球擁有大約 5.54 億用戶，每天大約有 5800 萬條 tweets。有這麼大量的資料，稱有可能會從中找到可能的流感疫情資訊，但是資料中可能會有雜訊，因此只有通過資料採擷技術和分析才能從中找到有用的資訊。另一個問題是可以我們無法從 Twitter 的貼文確定用戶的年齡,因此可能會阻礙我們準確預測哪個年齡層有流感疫情爆發的情形產生。

透過分析推文的內容，作者表示可以準確估計疾病控制中心的流感爆發流行的指數，確認我們的確可以使用 Twitter 資料確實能夠檢測到流感爆發流行。

Achrekar 等人設計了一個被認為是社交網路支援的流感趨勢(SNEFT)系統，該系統持續監控推特，目的是檢測和追蹤伊犁疫情的傳播。這項研究使用了一群 Twitter 用戶的貼文資料，這些用戶從 2009 年 10 月 18 日開始對流感關鍵字發表評論。SNEFT 系統使用網路爬蟲爬取包含 H1N1,和豬流感等相關 tweet 資訊，取得位置和相對關鍵字頻率等相關資料。還收集了美國疾病控制與預防中心的流感疫情報告以及其他與流感有關的資料。

作者發現，他們之前使用 2009- 2010 年的 tweet 資料做關聯分析，相關係數為 0.98，但自 2010-2011 年以來，該係數大幅下降，下降為 0.47。由於 H1N1 的資料中有許多雜音，因此作者則使用了文本分類技術，確認過訊息是否真為疫情相關訊息，進而減少相關雜音，以確定一條 tweet 是否真的與流感事件相關，進而提高預測的準確率。

這研究成果令人印象深刻，表明 Twitter 資料可以用於即時檢測和跟蹤類似流感的流行。這兩項研究都表明，我們可以使用網路上的貼文預測正在發生的流感疫情，這樣的預測方法準確率高，且誤差非常小，另外能比疾控中心公布的結果早上 1-2 周。這一系列的研究顯示出如果可以創建一個系統，持續跟蹤 twitter 上的帖子，進而即時創建一個全球的流感疫情地圖，將可以達到地區性與全球性疫情預警之效。

根據以上研究，如果我們能建立輿情分析平台，即時蒐集網路輿情我們便能夠以言論數據的角度了解疾病的傳染模式外，疾管署亦可以利用輿情數據做為感測器，檢測流感等季節性流行病在人群中的傳播行為。若能透過輿情早期發現可能的流行病，便可以利用輿情作為早期預警，提醒人民應該為即將流行的流感做好準備，以避免疾病爆發，造成區域間的大流行。

因應分析需求，我們在這四年期的計畫中，搭建一即時輿情分析平台，希望能夠透過即時搜尋國內外各種對疫病的討論與相關新聞，方能讓疫病於爆發早期時，疾管署便能第一手掌握消息。如能利用該輿情資訊建立監測與預測模式，便能利用該輿情預測疫情爆發的可能性，以達早期預防之效。另外，若能掌握各個消息來源，疾管署則可以在謠言廣泛散佈之前，先行闢謠消毒，對民眾宣導正確的防疫概念，避免民眾受到謠言的影響，導致不必

要的損失。藉由這兩方面雙管齊下，疾管署便可以利用該輿情系統，達公共治理之效。

近年來深度學習與資料科學的興起，引領大家對人工智慧擁有多所期待，而若疾管署的同仁皆有資料分析的基本底子，則在面對內部大量的數據與外在大量的輿情數據時，方有能力處理與關聯，從中找出對既有業務有益處的重要資訊。為了能讓內部同仁具有資料處理的知識，我方即提供疾管署一系列深度學習課程，希望內部同仁能學習 Python 語言的基礎與了解機器學習與深度學習等相關知識後，便有能力利用 Python 語言建立深度學習模型，期盼之後能從數據中挖掘出更多價值。

根據計畫的規劃下，我們這今年完成了以下目標：

- 為了能夠監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之輿情資料，我們將在抓取這些語系在 GoogleNews 中出現的新聞資料到監測系統中。
- 使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。為了確認翻譯的品質，我們亦會增加 precision 與 recall 之評量，確保字詞的翻譯品質。
- 為了避免權責人員沒辦法有效解讀日、韓、西班牙、葡萄牙、越南語系以及非洲等國之資料，我們將在會將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文，並將原文及翻譯文一同匯入系統中。
- 蒐集到各國輿情後，我們將使用名詞提取功能(Named Entity Extraction)技術抽取內文中提及的輿情地點。
- 由於建立分類引擎需要牽涉許多領域知識，因此我們將搭配 Python 程式語言的進階教學，讓負責人員可以動手建立適合的模型，以因應業務上

的輿情分類與標記需求。課程內容包含: Python 機器學習、Python 與深度學習、詞向量的建置與深度學習在文字上的應用。並於每次教學完後，提供作業與期末考予學員，確立學員能真正從教育訓練的過程中有所學習。

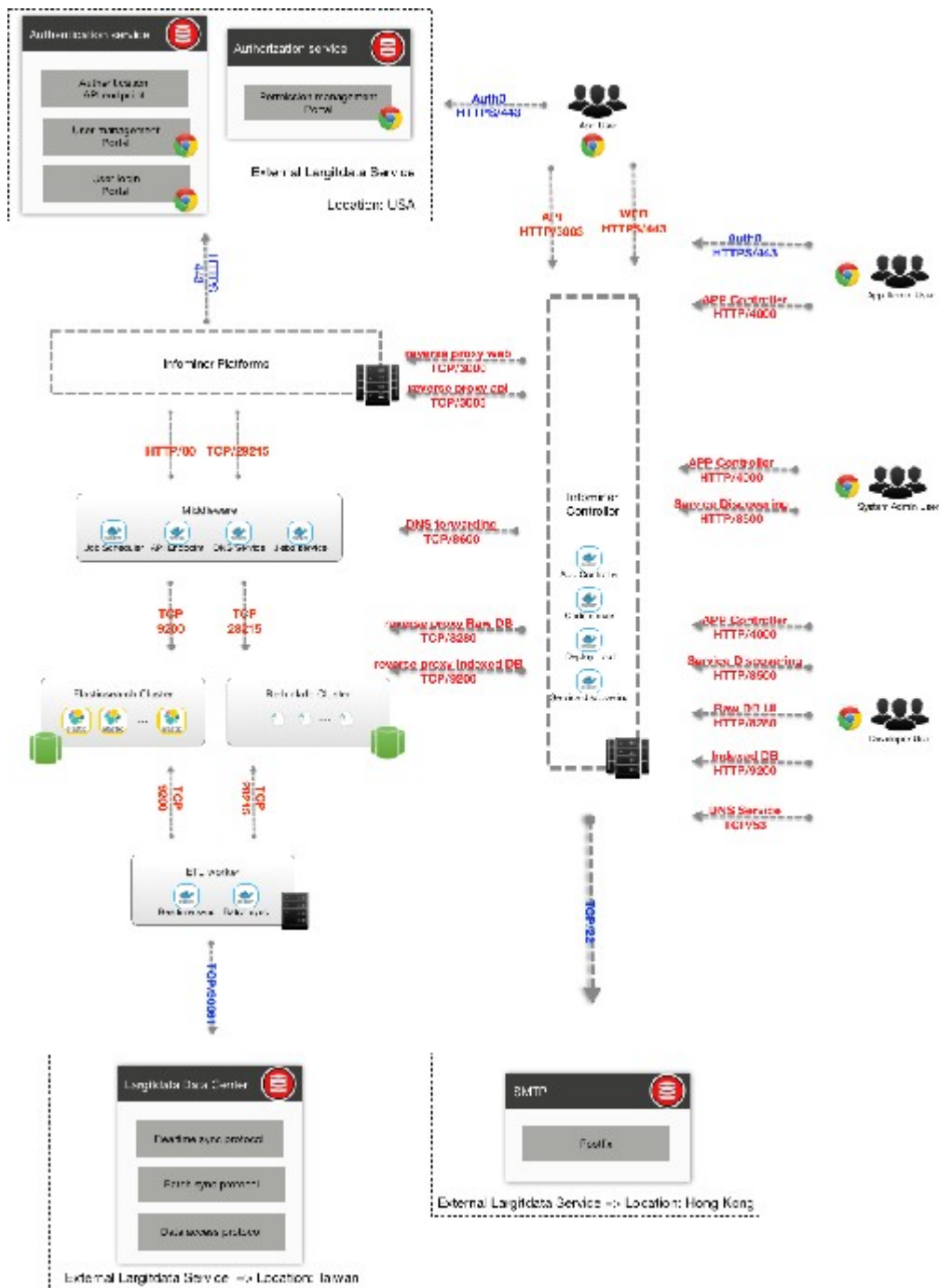
- 根據各國輿情發生的地點，將地點根據聲量繪製在世界地圖上，協助權責人員快速了解輿情的發生地。
- 評估三個疫病字詞在百度、Google Alert 與輿情系統搜尋之 Precision 與 Recall 以評量輿情系統之確度。
- 會將 CDC 的通報資料與輿情所蒐集到的資訊與 Google Trend 做關聯分析，評估系統是否能在重大疫情發生前，扮演預警的角色。

(二) 材料與方法

為了能夠快速蒐集國內外疫病各大頻道相關輿情訊息，並從非結構化的輿情資料之中分析出價值資訊，我們在第一年的計劃中建立了一個輿情觀測平台。並將該平台建構於雲端平台(Google Cloud Platform)上。透過該雲端平台，各部門的分析人員只須透過瀏覽器，便可以瀏覽並取用輿情分析數據，用做報表或決策依據。以下將詳列建立該輿情分析系統的架構與方法。

1. 輿情系統架構

由於量大、積累速度快、格式龐雜的輿情訊息皆符合一般對巨量資料的認定，為了能夠讓系統能即時蒐集、儲存、索引、分析並呈現輿情分析結果，我們架構了一分散式平台架構，以期能以該架構迅速處理並分析海量的輿情訊息。系統架構如下：



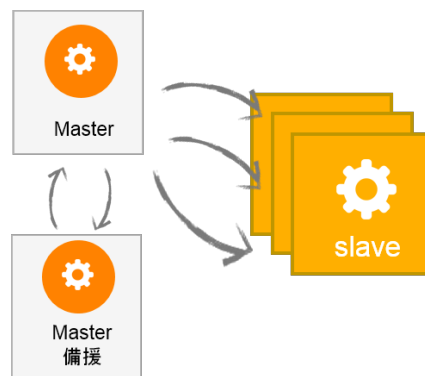
圖一、系統架構

以下將分為四個元件（分散式網路爬蟲、分散式儲存架構、搜尋引擎與應用程式）分別說明各元件的主要功能：

分散式網路爬蟲

本研究以 Python 建立分散式網路爬蟲，根據輿情資訊來源，系統將自動蒐集輿情資訊的主文、評論數、按讚數、回文數等不同指標資料。由於資料量龐大，為了能夠加速資料爬取的速度，將會架設分散式資料擷取模組爬取 PTT, Facebook、各大新聞媒體及其他疫病資訊源。為了能夠協調各分散式系統的運作，採用主從式架構，建立分散式爬蟲系統，概念如下圖。

圖二、主從式爬蟲架構



根據實機測試，每台機器可以每天抓取三十萬頁網頁，透過四台（一主三從）的部屬，將可每天抓取約 100 萬網頁的資料量。由於該架構屬於分散式架構，所有的工作統籌都由主節點分配，因此當如果有增加資料抓取的需求時，只需加機器便可擴增資料量的抓取。

除了抓取網頁內容外，必須使用網頁剖析程式方能抽取網頁中重要的資料，但往往目標網頁格式會有所變動，因此在抓取的過程之中，我們必須要保留網頁的原始檔，以免目標格式變更後，導致資料遺失。由於系統必須先針對網頁建立反向索引，而反向索引是根據詞庫內的字詞所建立，如果有新增字詞，則資料便必須重建索引，方能讓使用者可以根據新字詞檢索文字內容，也因如此，我們必須要保留原始網頁內容，以供後續檢索

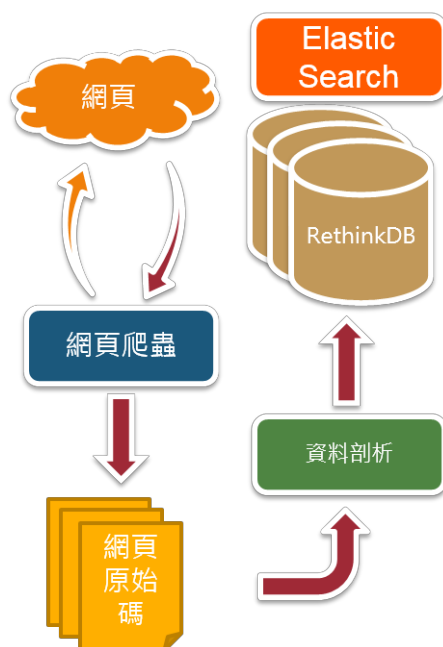
用。因此，系統會先在磁碟空間上預儲存一份資料源，而後再透過剖析程式，抽取重要資料。如果抓取資料頁面格式有變，該系統將會自動發出警示至權責人員信箱。

分散式儲存架構

為了因應抓取的資料屬於非結構化資料，不同來源資料要存儲的欄位可能不一致，因此我們採用了 RethinkDB 做為分散式資料儲存引擎，確保系統能以無綱要(Schema-Free)的 JSON 模型儲存大量的非結構化資料。由於該架構亦採用分散式架構，因此當使用者若需要加大儲存空間時，只需要增添機器便可以增加系統儲存容量。

搜尋引擎

而為了能夠加速資料搜索的速度，我們在寫進 RethinkDB 資料庫的同時，寫入一份資料至 ElasticSearch，使用者之後便可以透過其搜尋功能篩選輿情資訊。流程圖如下圖。



圖三、資料索引機制

由於搜尋引擎是以反向索引機制(Inverted Index)索引資料，當文章進入搜尋引擎後，系統將透過比對字典內的字詞，便可以將輸入文章分詞，並針對各詞建立索引，因此，使用者只須透過關鍵字，便可以快速查詢到跟關鍵字相符合的疫病相關文章。

應用程式

為了能夠在雲服務上即時分析非結構化數據，我們在分析程式端部署了多台虛擬機器，部分機器用做資料 ETL(Extract, Transform, Loading)，而應用及分析程式則部署在另外兩台機器上。該應用程式端負責分詞，並透過文字探勘機制（分群、詞性標註、分類、語意分析、情緒判斷）處理及分析爬蟲所搜集到的輿情資料，之後可透過網頁介面呈現文字探勘後的分析結果。

2. 資料蒐集

在 106 年度的計畫中，為了能夠廣泛搜索國內外的疫情資訊，因此已佈建部分網路爬蟲蒐集國內各來源的輿情資訊。於 107 年度為了能擴展輿情系統的既有能力，於是整合更多繁體、簡體中文與英文媒體的疫情訊息來源。在 108 年度的計畫中，為了能夠擴展系統的監測範圍，我們將整合語言翻譯功能，能蒐集到外國語以後，將該語文翻譯成繁體中文資料，並將增加五個國家的語系資料(例如:西班牙文、葡萄牙文、日文、韓文、越南語等外國語系的資料來源，並增加上述語系監測能力到輿情系統之中。為此，我們將增加上述語系來源的網站,但因為網站眾多，因此如果要蒐集所有網站資料，近乎是不可能，但防疫卻又不可能遺漏任何一個資料源，因此，在這年度計畫中，我們將抓取 Google News 各個國家的新聞源，並透過 Google Translate 做翻譯，如此一來，即能跨越個國家的藩籬，有效抓取所有跟疫病相關的輿情。108 年的資料來源可見下表：

表一：108 年資料來源

網域	名稱	來源	地區
udn.com	udn 部落格	blog	台灣
xuite.net	隨意窩	blog	台灣
roodo.com	樂多日誌	blog	台灣
pixnet.net	痞客邦	blog	台灣
medium.com	Medium	blog	國際
zjwjw.gov.cn	浙江省衛生和計劃生育委員會	cdc	大陸
yncdc.cn	雲南疾控資訊網	cdc	大陸
xizang.gov.cn	西藏自治區人民政府	cdc	大陸
tj.gov.cn	天津政務網	cdc	大陸
tibetcdc.cn	西藏自治區疾病預防控制中心	cdc	大陸
sxcdc.cn	山西省疾病預防控制中心	cdc	大陸

xzwsjsw.cn	西藏自治区卫生和计划生育委员会	cdc	大陸
wsjsw.gov.cn	上海市卫生和計劃生育委員會	cdc	大陸
sdcdc.cn	山東省疾病預防控制中心	cdc	大陸
sxcdc.com	陝西省疾病預防控制中心	cdc	大陸
qhwhjw.gov.cn	青海省卫生和計劃生育委員會	cdc	大陸
sdwsjs.gov.cn	山東省卫生和計劃生育委員會	cdc	大陸
nhfpc.gov.cn	中華人民共和國國家衛生健康委員會	cdc	大陸
sxwsjs.gov.cn	山西省卫生和計劃生育委員會	cdc	大陸
jshealth.com	江蘇省疾病預防控制中心	cdc	大陸
lncdc.com	遼寧省疾病預防控制中心	cdc	大陸
nmcdc.com.cn	內蒙古自治區綜合疾病預防控制中心	cdc	大陸
lndoh.gov.cn	遼寧省國土資源局	cdc	大陸
jxwst.gov.cn	江西省卫生和计划生育委员会	cdc	大陸
nxcdc.org	寧夏疾病預防控制中心	cdc	大陸
hubei.gov.cn	湖北省卫生和計劃生育委員會	cdc	大陸
gzhfpc.gov.cn	貴州省卫生和計劃生育委員會	cdc	大陸
gdwst.gov.cn	廣東省卫生和計劃生育委員會	cdc	大陸
hnwsjsw.gov.cn	河南省卫生和計劃生育委員會	cdc	大陸
gdemo.gov.cn	廣東省人民政府應急管理辦公室	cdc	大陸
hunan.gov.cn	湖南省人民政府	cdc	大陸
gsws.gov.cn	甘肅省卫生和計劃生育委員會	cdc	大陸
hncdc.cn	湖南省疾病預防控制中心	cdc	大陸
jxcdc.cn	江西省疾病預防控制中心	cdc	大陸
hljcdc.org	黑龍江省疾病預防控制中心	cdc	大陸
www.gov.cn	中華人民共和國中央政府網	cdc	大陸
jswst.gov.cn	江蘇省卫生和計劃生育委員會	cdc	大陸
hljhfp.gov.cn	黑龍江省卫生和計劃生育委員會	cdc	大陸
gzscdc.org	貴州省疾病防預控制中心	cdc	大陸

gxzf.gov.cn	廣西壯族自治區衛生和計劃生育委員會	cdc	大陸
gxcdc.com	廣西壯族自治區疾病預防控制中心	cdc	大陸
gscdc.net	甘肅省疾病預防控制中心	cdc	大陸
fujian.gov.cn	福建省疾病預防控制中心	cdc	大陸
cqcdc.org	重慶市疾病預防控制中心	cdc	大陸
cqwsjsw.gov.cn	重慶市衛生和計畫生育委員會	cdc	大陸
gd.gov.cn	廣東省疾病預防控制中心	cdc	大陸
bjchfp.gov.cn	北京市衛生和計劃生育委員會	cdc	大陸
cdctj.com.cn	天津市疾病預防控制中心	cdc	大陸
ahwjw.gov.cn	安徽省衛生和計劃生育委員會	cdc	大陸
bjcdc.org	北京市疾病預防控制中心	cdc	大陸
ahcdc.cn	安徽省疾病預防控制中心	cdc	大陸
scwst.gov.cn	四川省衛生和計劃生育委員會	cdc	大陸
nxws.gov.cn	寧夏回族自治區衛生和計劃生育委員會	cdc	大陸
nmg.gov.cn	內蒙古自治區政府門戶網站	cdc	大陸
moa.gov.cn	中華人民共和國農業部	cdc	大陸
hainan.gov.cn	海南省人民政府	cdc	大陸
cdc.zj.cn	浙江省疾病預防控制中心	cdc	大陸
fjhfpc.gov.cn	福建省衛生和計劃生育委員會	cdc	大陸
jl.gov.cn	吉林省人民政府	cdc	大陸
hebwst.gov.cn	河北省卫生和计划生育委员会	cdc	大陸
hbcdc.cn	湖北疾控	cdc	大陸
chinacdc.cn	中國疾病預防控制中心	cdc	大陸
xinhuanet.com	新華網	chinanews	大陸
sina.com.cn	新浪台灣	chinanews	大陸
wsj.com	華爾街日報(中國)	chinanews	香港
nikkei.com	日本經濟中文網	chinanews	大陸

jfdaily.com	解放網	chinanews	大陸
huanqiu.com	環球網	chinanews	大陸
ftchinese.com	F T 中文網	chinanews	大陸
guancha.cn	觀察者網	chinanews	大陸
chinanews.com	中國新聞網	chinanews	大陸
people.com.cn	人民網報	chinanews	大陸
facebook.com	臉書	fans	台灣
wsj.com	The Wall Street Journal	foreign	國際
washingtonpost.com	Washington Post	foreign	國際
nytimes.com	紐約時報	foreign	國際
latimes.com	Los Angeles Time	foreign	國際
pia.gov.ph	PIA	foreign	國際
economist.com	經濟學人	foreign	國際
bbc.com	BBC	foreign	國際
antaranews.com	安塔拉新聞	foreign	國際
nhk.or.jp	NHK	foreign	國際
bloomberg.com	彭博新聞社	foreign	國際
pcdvd.com.tw	P C D V D 數位科技討論區	forum	台灣
mobile01.com	Mobile01	forum	台灣
jkforum.net	捷克論壇	forum	台灣
gamer.com.tw	巴哈姆特電玩資訊站	forum	台灣
eyny.com	伊莉討論區	forum	台灣
dcard.tw	DCARD	forum	台灣
cool3c.com	癮科技	forum	台灣
ck101.com	卡提諾論壇	forum	台灣
sogi.com.tw	SOGI 手機王	forum	台灣
wealth.com.tw	財訊	news	台灣
twreporter.org	報導者	news	台灣
thenewslens.com	關鍵評論網	news	台灣
tvbs.com.tw	TVBS 官方網站	news	台灣
theinitium.com	端傳媒	news	台灣

setn.com	三立新聞網	news	台灣
upmedia.mg	上報	news	台灣
udn.com	聯合新聞網	news	台灣
pts.org.tw	公視	news	台灣
peopo.org	PeoPo 公民新聞	news	台灣
nexttv.com.tw	壹電視 NextTV	news	台灣
ttv.com.tw	台視	news	台灣
technews.tw	TechNews 科技新報	news	台灣
people.com.cn	人民網	news	大陸
newsmarket.com.tw	上下游 News&Market 新聞市集	news	台灣
newtalk.tw	新頭殼 newtalk	news	台灣
kairos.news	風向新聞	news	台灣
pansci.asia	PanSci 泛科學	news	台灣
moneydj.com	MoneyDJ 理財網	news	台灣
gvm.com.tw	遠見雜誌	news	台灣
ltn.com.tw	自由時報電子報	news	台灣
gamme.com.tw	卡卡洛普 Gamme	news	台灣
ftv.com.tw	民視 F T V	news	台灣
eracom.com.tw	年代電視台	news	台灣
ebc.net.tw	東森電視	news	台灣
chinatimes.com	中國時報	news	台灣
www.digitimes.com.tw	電子時報	news	台灣
turnnewsapp.com	時報週刊	news	台灣
squarespace.com	Duda	news	台灣
crntt.com	中國評論新聞	news	台灣
ctv.com.tw	中視	news	台灣
cdnews.com.tw	中央網路報	news	台灣
businesstoday.com.tw	今周刊	news	台灣
commonhealth.com.tw	康健雜誌	news	台灣
cnyes.com	cnYES 鉅亨網	news	台灣
epochtimes.com.tw	台灣大紀元	news	台灣

cctv.com	央視網	news	大陸
e-info.org.tw	台灣環境資訊協會	news	台灣
bcc.com.tw	中廣	news	台灣
businessweekly.com.tw	商業周刊	news	台灣
asahichinese-f.com	朝日新聞中文網	news	台灣
appledaily.com	蘋果日報	news	台灣
agriharvest.tw	農傳媒	news	台灣
anntw.com	台灣醒報 Awakening News Networks	news	台灣
yahoo.com	Yahoo	news	台灣
storm.mg	風傳媒	news	台灣
nownews.com	NOWnews 今日新聞	news	台灣
hinet.net	HiNet	news	台灣
my-formosa.com	美麗島電子報	news	台灣
yam.com	yam 蕃薯藤	news	台灣
udn.com	經濟日報	news	台灣
twse.com.tw	公開資訊觀測站	news	台灣
line.me	line	news	台灣
cw.com.tw	天下雜誌	news	台灣
cts.com.tw	華視全球資訊網	news	台灣
cna.com.tw	中央社	news	台灣
ptt.cc	批踢踢實業坊	ptt	台灣
weibo.com	微博	weibo	大陸
youtube.com	Youtube	youtube	台灣
wxhk.org	香港地下天文台		台灣
will-news.info	willnews 分享健康		台灣
wednesdayjournal.net	WednesdayJournal		香港
zytzb.gov.cn	中國統戰部		香港
vos.com.cn	海峽之聲網		大陸
yicai.com	第一財經		大陸
vietnamplus.vn	Vietnam+		大陸

sina.com.cn	新周刊的博客	大陸
verywed.com	veryWed 非常婚禮	台灣
ukrinform.org	烏克蘭國家通訊社	國際
taiwan368.com.tw	中國時報_台灣 368 品味生活	台灣
yna.co.kr	韩联社	大陸
ydn.com.tw	國防部青年日報社	台灣
udn.com	聯合影音	台灣
tsrus.cn	俄罗斯新闻	大陸
yahoo.com	Yahoo 股市	台灣
sina.com.cn	徐靜蕾的博客	大陸
yam.com	蕃新聞	台灣
travelrich.com.tw	旅遊經	台灣
yahoo.com	雅虎香港	香港
teepr.com	TEEP 亮新聞	台灣
xzwjw.gov.cn	西藏自治区卫生健康委员会	大陸
jl.gov.cn	吉林省卫生健康委员会	大陸
williamlong.info	月光博客	大陸
yahoo.com	Yahoo 奇摩理財	台灣
techbang.com	T 客邦	台灣
wenweipo.com	香港文匯網	香港
yahoo.com.tw	Yahoo 奇摩汽車機車	台灣
sohu.com	搜狐-谢国忠	大陸
unwire.pro	UNWIRE.PRO	香港
walkerland.com.tw	WalkerLand 窩客島	台灣
shzfzz.net	上海政法综治	大陸
twmotor.com.tw	超越車訊官方網站	台灣
stheadline.com	頭條日報	香港
tvb.com	TVB 討論區	台灣
gd.gov.cn	广东省卫生健康委员会	大陸
sh.gov.cn	上海市卫生健康委员会	大陸
worldjournal.com	世界新聞網	大陸

sputniknews.cn	俄羅斯衛星通訊社	大陸
tssdnews.com.tw	台灣新生報	台灣
weddingday.com.tw	WeddingDay 好婚市集	台灣
sheitc.gov.cn	上海市經濟和信息化委員會	大陸
toutiao.com	今日头条	大陸
sofree.cc	香腸炒魷魚	台灣
thestandnews.com	立場新聞	香港
voachinese.com	美國之音	大陸
singpao.com.hk	成報	香港
techapple.com	TechApple	香港
sohu.com	徐靜蕾的博客	大陸
seozac.com	SEO 每天一貼	大陸
sohu.com	搜狐-時尚先生	大陸
urcosme.com	UrCosme	台灣
synergy-edu.com	synergy	台灣
scmp.com	South China Morning Post	香港
twitter.com	Twitter	國際
81tech.com	国防科技網	大陸
tnews.cc	台灣新聞網	台灣
scwjxx.cn	四川省衛生健康信息中心	大陸
sohu.com	搜狐	大陸
wantgoo.com	玩股網	台灣
turnnewsapp.com	翻報	台灣
takungpao.com.hk	大公網	香港
thestar.com.my	The Star Online	國際
sciencenet.cn	科學網	大陸
vietnamnews.vn	Viet Nam News	國際
read01.com	壹讀	台灣
indiatimes.com	RISING STAR	國際
taihainet.com	台海網	大陸
thestandard.com.hk	The Standard 英文虎報	香港

rfa.org	自由亞洲電台	大陸
utimes.today	灣仔日報	香港
prius-c-club.com	U-CAR	台灣
thepaper.cn	澎湃	香港
theedgemarkets.com	The Edge Markets	國際
reuters.com	Reuters	國際
ustv.com.tw	USTV 非凡電視台	台灣
post852.com	852 郵報	香港
techweb.com.cn	TechWeb	大陸
stheadline.com	星島頭條網	香港
rav4-club.tw	RAV4 Club	台灣
ptt01.cc	ptt01	台灣
u-car.com.tw	U-CAR 討論區	台灣
pctech.com.hk	PC TECH 電腦技術	香港
sspai.com	少數派	大陸
on.cc	On.cc 東網	香港
spp.gov.cn	中共最高人民檢察院	大陸
qinghai.gov.cn	青海省卫生健康委员会	大陸
plus28.com	Plus28 論壇	台灣
twpowernews.com	勁報	台灣
on.cc	東方日報	香港
sina.com.cn	郎咸平的博客	大陸
nytimes.com	紐約時報中文網	大陸
southcn.com	南方网	大陸
nmg.gov.cn	内蒙古自治区卫生健康委员会	大陸
twgreatdaily.com	GreatDaily	台灣
okgo.tw	完全台灣	台灣
ntdtv.com	中國電視新聞網	大陸
sogou.com	搜狗新聞	香港
rti.org.tw	Rti 中央廣播電台	台灣
life.tw	LIFE 生活網	台灣

singtao.ca	加拿大星島日報	香港
pchome.com.tw	pchome 股市	台灣
qq.com	騰訊網	大陸
theguardian.com	TheGuardian	國際
nextmag.com.tw	台灣壹週刊	台灣
rfi.fr	世界之聲	大陸
881903.com	881903	香港
shinmin.sg	新明日報	國際
pchome.com.tw	PChome 新聞	台灣
natgeomedia.com	國家地理雜誌中文網	台灣
tvb.com	無綫新聞	香港
popdaily.com.tw	波波黛莉的異想世界	台灣
ndrc.gov.cn	中華人民共和國中央人民政府	大陸
rthk.hk	香港電台網站	香港
passiontimes.hk	熱血時報	大陸
haiwainet.cn	中国南海新闻网	大陸
myweb.wang	玩圈	香港
mfa.gov.cn	中國外交部	大陸
parstoday.com	Pars Today	國際
mod.gov.cn	中華人民共和國國防部	大陸
tvbs.com.tw	食尚玩家	台灣
mof.gov.cn	中華人民共和國財政部	大陸
nova.com.tw	NOVA 資訊廣場	台灣
mcu.edu.tw	銘報即時新聞	台灣
manilatimes.net	The Manila Times	國際
ulifestyle.com.hk	晴報	香港
wallstreett.cn	华尔街头条	香港
luckydog.tw	Luckydog	台灣
oursogo.com	SOGO 論壇	台灣
hellotw.com	你好台灣網	香港
mamibuy.com.tw	媽咪拜 MamiBuy	台灣

sinchew.com.my	印尼星洲日報	國際
mingpao.com	明報香港新聞網	香港
new7.com.tw	新新聞	台灣
now.com	Now 新聞	香港
sabinahuang.com	Sabina Huang	台灣
metrodaily.hk	都市日報	香港
ner.gov.tw	國立教育廣播電台	台灣
jiemian.com	界面新聞	大陸
menews.info	MeNews	香港
5550555.com	真晨網	台灣
lihkg.com	吹水台	大陸
miui.com	MIUI 米柚論壇	台灣
news.gov.hk	香港政府新聞網	香港
managertoday.com.tw	經理人	台灣
krt.com.hk	國度復興報(香港)	香港
kyodonews.net	共同網	大陸
marieclaire.com.tw	美麗佳人	台灣
ithome.com.tw	iThome	台灣
163.com	網易新聞	大陸
pcdiy.com.tw	PCDIY!	台灣
thejakartapost.com	The Jakarta Post	國際
ltn.com.tw	汽車頻道-自由電子報	台灣
iscarmg.com	isCar	台灣
nakuz.com	樂古	香港
csdn.net	銘毅天下	大陸
incar.tw	癮車報	台灣
lifeonea.co	壹 A 新聞	台灣
hkcnahk	香港新聞網	香港
msn.com	msn 台灣	台灣
36kr.com	36Kr	大陸
sina.com.cn	平民眼中的未來的博客	大陸

sina.com.cn	孔庆东的博客	大陸
hkcd.com	香港商報	香港
itpromag.com	IT Pro Magazine	香港
jrj.com.cn	金融界	大陸
nst.com.my	New Straits Times	國際
jiangxi.gov.cn	江西省卫生健康委员会	大陸
kbs.co.kr	KBS WORLD Radio	國際
hk01.com	香港 01	香港
moh.gov.sg	ministry of health	大陸
indianexpress.com	The Indian EXPRESS	國際
jiangsu.gov.cn	江苏省卫生健康委员会	大陸
mpfinance.com	明報財經網	大陸
investor.com.tw	財訊快報	台灣
haiwainet.cn	HWW 海外網	大陸
hybridclub.tw	HybridClub.TW	台灣
ioh.tw	IOH 開放個人經驗平台	
mmtimes.com	The Myanmar Times	國際
geekpark.net	極客公園	大陸
inmediahk.net	香港獨立媒體網	香港
metroradio.com.hk	新城電台	香港
hongkongcard.com	香港信用卡優惠網	大陸
intalking.com	美妝 IN TALKING	台灣
memehk.com	謎米香港	香港
ifanr.com	愛范兒	大陸
kochinews.co.jp	高知新聞	國際
idn.com.tw	自立晚報	台灣
match.net.tw	match 生活網	台灣
www.gov.cn	中國政府網	大陸
kkp.org.hk	公教報	香港
fujian.gov.cn	福建省卫生健康委员会	大陸
hqck.net	九尾网	大陸

feedly.com	南韓青瓦台政府	國際
kknews.cc	KKNews	台灣
ftv.com.tw	民視地方	台灣
hkgolden.com	香港高登	大陸
feedly.com	新加坡政府	國際
appledaily.com	香港蘋果新聞	香港
fingermedia.tw	指傳媒 Fingermedia	台灣
hket.com	香港經濟日報	香港
feedly.com	Vive Forum	國際
infzm.com	南方周末	大陸
18hall.com	拾捌堂	香港
icook.tw	愛料理	台灣
feedly.com	瑞典政府	國際
hexun.com	和訓網	大陸
feedly.com	Intro Version	國際
hongkongfp.com	HKFP	香港
feedly.com	印度政府	國際
feedly.com	Econsultancy	國際
hksmedia.com	港股策略王	香港
foxnews.com	Fox News	國際
freshweekly.tw	鮮週報	台灣
feedly.com	英國政府	國際
feedly.com	全國法規資料庫	台灣
hkej.com	信報	香港
fengmin.com.tw	鳳鳴廣播電台	台灣
feedly.com	洲政府	國際
feedly.com	加拿大政府	國際
google.com	墨西哥疾管	國際
feedly.com	彰化縣政府	台灣
google.com	韓國疾管	國際
feedly.com	白宮	國際

feedly.com	THE Verge	國際
feedly.com	Silver Screen and Roll	國際
gamasutra.com	Gamasutra	國際
feedly.com	中央銀行全球資訊網	台灣
google.com	法國疾管	國際
feedly.com	日本政府	國際
feedly.com	OECD	國際
saigon-gpdaily.com.vn	ai Gon GP Daily	國際
feedly.com	PoliticalForum.com	國際
feng.com	威鋒	大陸
2000fun.com	2000Fun 遊戲資訊網	台灣
feedly.com	BBC UK	國際
feedly.com	International Student	國際
2000fun.com	2000fun 遊戲資訊網	香港
feedly.com	TED-Ed	國際
feedly.com	TNTvillage	國際
feedly.com	No More Mister Nice Blog	國際
feedly.com	證券投資信託暨顧問商業同業公會	台灣
fashionguide.com.tw	FashionGuide	台灣
feedly.com	Hubspot	國際
feedly.com	OnePlus Community	國際
ettoday.net	ET Fashion	台灣
feedly.com	Get Rich Slowly	國際
feedly.com	環保署	台灣
feedly.com	美國政府	國際
feedly.com	臺南市政府	台灣
feedly.com	International Economic Law and Policy Blog	國際
feedly.com	XDA-Developers Android Forums	國際
feedly.com	臺灣期貨交易所	台灣
feedly.com	SpyParty	國際
feedly.com	交通部	台灣

feedly.com	Economist's View	國際
eonline.com	EOnline	國際
feedly.com	Tomorrow Corporation	國際
feedly.com	立法院	台灣
feedly.com	Introversion	國際
feedly.com	證券商業同業公會	台灣
feedly.com	桃園縣政府	台灣
feedly.com	Money Saving Expert	國際
feedly.com	行政院全球資訊網	台灣
eurogamer.net	Eurogamer	國際
feedly.com	Health Care Conversation	國際
feedly.com	內政部營建屬	台灣
ecoast.com.tw	花蓮電子報	台灣
feedly.com	International Women's Forum	國際
epochtimes.com	大紀元	大陸
feedly.com	distractionware	國際
feedly.com	The Forum on Education Abroad	國際
ettoday.net	ETtoday 寵物雲	台灣
indiatimes.com	The Economic Times	國際
feedly.com	Websblog	國際
dwnews.com	多维博客	台灣
coolaler.com	滄者極限	台灣
feedly.com	Weebly	國際
enews.tw	ENEWS	台灣
dw.com	DW	大陸
feedly.com	交通部觀光局	台灣
elle.com	ELLE Taiwan	台灣
depkes.go.id	Kementerian Kesehatan	大陸
contentparty.org	Content Party	台灣
feedly.com	Tim blog	國際
cri.cn	国际在线	大陸

feedly.com	法源法律網	台灣
e2.com.tw	壹凸新聞	台灣
dcview.com	DCView 數位視野	台灣
feedly.com	MENTALITYWOD	國際
dwnews.com	多維新聞網	香港
ciaa.org.tw	中華民國保險代理人商業同業公會	台灣
chinapost101.com	China Post e-News	台灣
ctwant.com	CTwant	台灣
feedly.com	Hootsuite	國際
chinagfw.org	功夫网与翻墙	大陸
eastmoney.com	东方财富网	大陸
ettoday.net	Ettoday 新聞雲	台灣
china.com	中華網	大陸
corolla-altis-club.tw	台灣 Altis 論壇	台灣
feedly.com	Born Fitness	國際
cssn.cn	中國社會科學網	大陸
chinadaily.com.cn	China Daily	大陸
daai.tv	大愛電視	台灣
dahe.cn	大河网	大陸
china5e.com	中國能源網	大陸
chosun.com	朝鮮日报网	大陸
feedly.com	EnglishUSA	國際
cntimes.info	大華網路報	台灣
cgs.gov.cn	中國地質調查局	大陸
coolshell.cn	酷殼	大陸
8world.com	8 頻道新聞	國際
cdns.com.tw	中華日報新聞網	台灣
feedly.com	保發中心	台灣
cheers.com.tw	Cheers 快樂工作人雜誌	台灣
ctitv.com.tw	中天快點 TV	台灣
ce.cn	中国经济网	大陸

channel-auto.com	ca 汽車頻道	台灣
carnews.com	CARNEWS 車訊網	台灣
ctitv.com.tw	中天 必 PO TV	台灣
carstuff.com.tw	CarStuff	台灣
businessinsider.com	Business Insider	國際
caixin.com	財新網	大陸
chinadaily.com.cn	中國日報網	國際
coco01.today	coco01	台灣
cardu.com.tw	卡優新聞網	台灣
brotherhood.space	香港玩樂網	台灣
ccps.gov.cn	中共中央黨校	大陸
buycartv.com	Go 車誌 BuyCarTv	台灣
china.com.cn	中國網	大陸
chinaeducenter.com	中國留學社	大陸
cnr.cn	央廣網	大陸
bldaily.com	美麗日報	台灣
beauty321.com	beauty 美人圈	台灣
eastday.com	東方網	大陸
capital-hk.com	Capital	香港
bomb01.com	bomb01	台灣
7car.tw	7Car 小七車觀點	台灣
www.gov.cn	中國政府網	大陸
bjnews.com.cn	新京報網	大陸
81.cn	中國軍網	大陸
cambodiadaily.com	The Cambodia Daily	國際
bbc.com	BBC News 中文	香港
bo6s.com.tw	波新聞	台灣
taiwan.cn	中國台灣網	大陸
chbcnet.com	華廣網	大陸
bannedbook.org	中國禁聞網	大陸
8891.com.tw	8891 汽車討論區	台灣

businessmirror.com.ph	BusinessMirror	國際
aqnews.com.cn	安慶新聞網	大陸
baby-kingdom.com	親子王國香港討論區	台灣
babyhome.com.tw	babyhome 親子討論區	台灣
cdcmoh.gov.kh	Ministry of health	大陸
bharian.com.my	BH online	國際
bastillepost.com	巴士的報	香港
bworldonline.com	BusinessWorld	國際
atanews.net	亞太新聞網 ATA News	台灣
asiatimes.com.my	亞洲時報	大陸
buzzhand.com	Buzzhand	台灣
bc3ts.com	爆料公社	台灣
autonet.com.tw	AutoNet 汽車日報	台灣
appledaily.com	蘋果新聞網	台灣
arats.com.cn	海峽兩岸關係協會	香港
bella.tw	Bella.tw 儂儂	台灣
android-hk.com	Android-HK	香港
am730.com.hk	am730	香港
businesstimes.com.sg	The Business Times	國際
auto007.com.tw	007 汽車網	台灣
aboluowang.com	阿波羅網	大陸
aimpact.com	AIMPACT 鋒投	香港
go.com	ABC News	國際
auto-online.com.tw	汽車線上	台灣
aotter.net	電獺少女	台灣
zaobao.com.sg	早報	大陸
yzzk.com	亞洲週刊	大陸
wed168.com.tw	婚禮情報	台灣
wearn.com	聚財網	台灣
vjmedia.com.hk	輔仁文誌	香港
uwants.com	U W A T S	台灣

udn.com	發燒車訊 聯合新聞網	台灣
taiwantimes.com.tw	台灣時報	台灣
taiwannews.com.tw	Taiwan News Online	台灣
taiwannet.com.tw	台灣產經新聞	台灣
catholic.org.hk	天主教香港教區	香港
stdaily.com	中国科技网	大陸
scmp.com	South China Morning Post	國際
qianlong.com	千龙网	大陸
pchome.com.tw	PChome 基金	台灣
nx.gov.cn	宁夏回族自治区卫生健康委员会	大陸
now.com	now 新聞	香港
money-link.com.tw	富聯網	台灣
mamilove.com.tw	媽咪愛	台灣
lifeonea.com	生活知識+	台灣
ksnews.com.tw	更生日報	台灣
From the Desk of the Director-General of Health		
kpkesehatan.com		
Malaysia		大陸
kocpc.com.tw	電腦王阿達	台灣
karayou.com	卡拉娛樂網	台灣
joylah.co	Joy 啦	台灣
inquirer.net	inquirer	國際
i-cable.com	H K Cable TV	香港
huaxia.com	華夏經緯網	大陸
hkpost.com.hk	香港ポスト	國際
gwytb.gov.cn	国台办	大陸
greatnews.com.tw	大成報	台灣
google.com	越南疾管	國際
google.com	葡萄牙疾管	國際
google.com	日本疾管	國際
ebc.net.tw	東森財經新聞	台灣

feedly.com	Forum Blue And Gold	國際
feedly.com	Openknowledge	國際
feedly.com	HuffPost	國際
feedly.com	SnapPages - Blog	國際
feedly.com	雲林縣政府	台灣
feedly.com	證券交易所	台灣
feedly.com	新北市政府	台灣
feedly.com	南投縣政府	台灣
yahoo.com.tw	Yahoo 奇摩旅遊	台灣
google.com	迦納疾管	國際
vogue.com.tw	VOGUE 時尚網	台灣
stnn.cc	星島環球網	大陸
getit01.com	GETIT01	台灣
udnbkk.com	泰國世界日報	國際
thesaigontimes.vn	Saigon Times	國際
sina.com.tw	新浪台灣	台灣
feedly.com	EnglishUSA	國際
taiwanhot.net	台灣好新聞	台灣
transcendadmissions.com	創陞教育顧問	台灣
reuters.com	路透中文網	大陸
sienta-club.tw	Sienta Club	台灣
tvbs.com.tw	女人我最大 TVBS 官網	台灣
sanook.com	sanook	國際
feedly.com	中華民國保險代理人商業同業公會	台灣
mirrormedia.mg	鏡週刊	台灣
tanews.org.tw	台灣動物新聞網	台灣
taiwan.cn	中國台灣網	香港
ifeng.com	鳳凰網	大陸
newstaiwandigi.com	台灣好報	台灣
stats.gov.cn	中共中央黨校	大陸
feedly.com	信託業商業同業公會	台灣

feedly.com	壽險公會	台灣
nakuz.com	樂古論壇	台灣
hifidiy.net	HIFI 音响网	香港
cnr.cn	央广军事	大陸
eprice.com.tw	ePrice 比價王	台灣
ctee.com.tw	工商時報	台灣
instagram.com	instagram	國際
turnnewsapp.com	Girl 愛女生	台灣
boxun.com	博訊新聞網	香港
civilmedia.tw	公民行動	台灣
feedly.com	MarketingProfs	國際
feedly.com	財政部	台灣
feedly.com	行政院金融監督管理委員會	台灣
feedly.com	Power Line blog	國際
bnext.com.tw	數位時代	台灣
beijing.gov.cn	北京市卫生健康委员会	大陸
discuss.com.hk	香港討論區	大陸

另外，為了因應能夠監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之輿情資料，我們將抓取這些語系在 Google News 中出現的新聞資料到監測系統中。

為了能夠有效監控外國輿情資料，因此在輿情系統之中，我們必須要加入外國輿情之資料。但由於外國輿情新聞來源過於眾多，難以一一寫爬蟲抓取各國來源資料，因此我們從 Google News 下手，希望利用 Google 的新聞平台，能讓我們廣泛的搜集全世界與疾病相關之輿情。

我們的作法如下：

1. 進到 Google News 平台



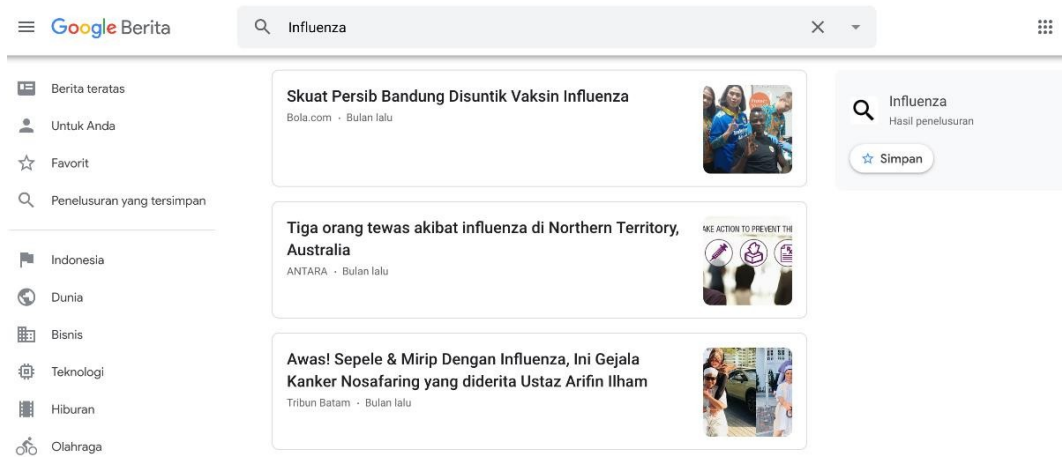
圖四、Google News 平台

2. 切換語系



圖五、語系切換

3. 下關鍵字搜尋該語系的新聞



圖六、特定語系新聞

4. 使用網路爬蟲抓取該頁面

3. 資料處理

多語系同義詞典

於 108 年度的計畫，因為多了外國疫情的資訊需要偵蒐，因此我們必須取得疾病相關的翻譯字詞，才有辦法從 Google News 檢索其他語系的疾管資料。因此我們使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。

由於此次要監控的輿情來自眾多國家，而輿情也分屬不同語系，因此要如何精準的搜尋出跟疾病相關的外文輿情變成是一大挑戰。所幸維基百科上面列有各專業名詞的不同語系翻譯，因此我們便抓取維基百科上與疾病相關之字詞與對應語系的翻譯詞，我們便可以建立出一個外文的同義辭典，以便我們準確於 Google News 中搜尋出外文疫情資訊。

我們的作法如下：

- a. 進到維基百科平台
- b. 下關鍵字搜尋與疾病相關之條目

- 分類索引
- 特色內容
- 新聞動態
- 近期變更
- 隨機條目

- 說明
- 說明
- 維基社群
- 方針與指引
- 互助客棧
- 知識問答
- 字詞轉換
- IRC即時聊天
- 聯絡我們
- 關於維基百科
- 資助維基百科

- 其他專案
- 維基共享資源

- 列印/匯出
- 下載為 PDF

流行性感冒 [編輯]

維基百科，自由的百科全書

流行性感冒（Influenza），通常簡稱為**流感**，為一種由**流感病毒**造成的**傳染性疾病**^[1]。流感的**症狀**可輕可重^[4]，最常見者為**高燒**、**流鼻涕**、**喉嚨痛**、**肌肉痠痛**、**頭痛**、**咳嗽**和**疲倦感**。患者通常在接觸病毒2天後發病，症狀大多在一週內會解除，但咳嗽可能持續超過兩週^[1]。孩童可能會**噁心**和**嘔吐**，但這在成人並不常見；噁心和嘔吐更常發生在與流感病毒無關的**感染性腸胃炎**，有時會不精確地稱此為**腸胃型感冒**（stomach flu）^[5]。流感可能的併發症包括**病毒性肺炎**、**次級細菌性肺炎**、**鼻窦感染**以及造成其他疾病惡化（如**氣喘**或**心臟衰竭**）^{[4][2]}。

可感染人類的流感病毒有A、B、C三型^[2]。病毒通常由咳嗽，打噴嚏和說話產生的**飛沫傳播**^[1]，近距離接觸時尤其容易發生^[6]。此外，病毒也可藉由接觸到受染污的物體表面、再碰觸口或眼睛後傳播^{[4][6]}。受感染的患者無論在發病前後均可能具有傳染性^[4]，喉嚨、痰液或鼻黏膜等檢體的病毒測試則可作為確診的依據。目前已有數種**快篩方法**，然而其靈敏度為診斷（即病毒感

注：然而其靈敏度為診斷（即病毒感

流行性感冒

同義詞 流感



圖七、疾病相關之條目

- c. 切換語系

其他語言 ⚙️

- العربية
- English
- ★ Español
- हिन्दी
- 🇮🇩 Bahasa Indonesia
- ★ Jawa Influenza – 印尼文
- ★ Português
- Русский
- اردو

圖八、語系切換

d. 抓取切換語系後的條目



圖九、其他語系條目

語言翻譯

為了避免權責人員沒辦法有效解讀日、韓、西班牙、葡萄牙、越南語系以及非洲等國之資料，我們將在會將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文，並將原文及翻譯文一同匯入系統中。

我們透過 Google Translate 偵測抓下來的文章語系後，並透過 Google Translate 將標題跟內文翻譯成繁中與英文，讓權責人員可以根據繁中跟英文檢視外國疾管文章。

使用者可以到 InfoMiner 即時輿情分析平台中「即時搜尋」下的「國外疾管」檢視翻譯過的疾病相關輿情。

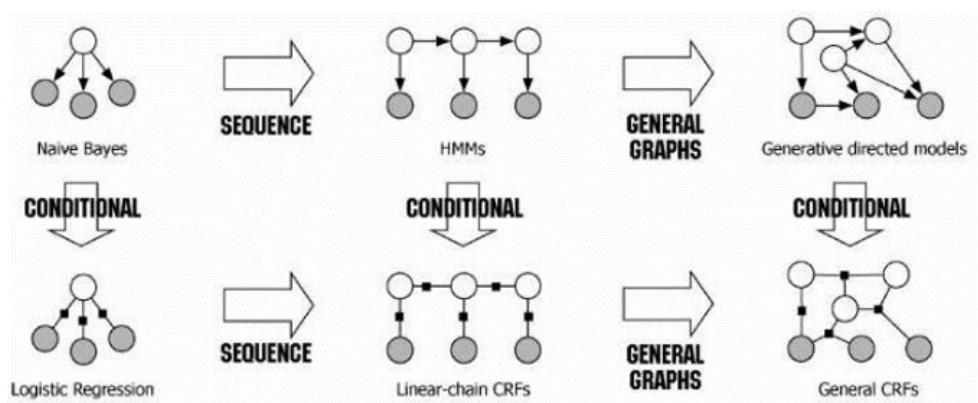


圖十、翻譯內容

命名實體識別

蒐集到各國輿情後，我們將使用名詞提取功能(Named Entity Extraction)技術抽取內文中提及的輿情地點

我們一開始使用條件隨機域技術辨識文本中的命名實體(e.g. 人名、位置名稱等)，條件隨機域是條件概率分布模型，表示的是給定一組輸入隨機變量 X 的條件下另一組輸出隨機變量 Y 的馬爾可夫隨機場，用於標注或分析序列資料，透過該演算法，我們便可利用從屬關係分析和詞性分析提取所有名詞。



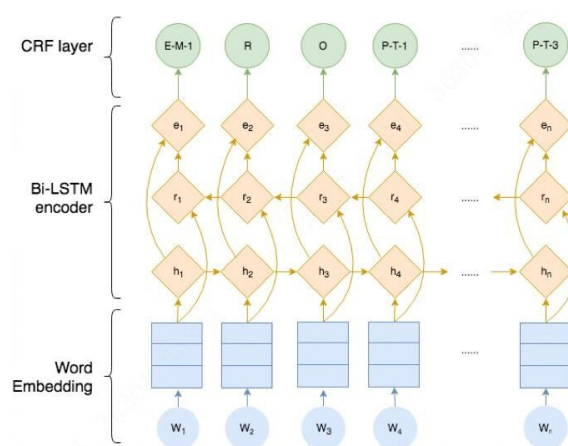
圖十一、條件隨機域

我們將辨認出來的地點以關鍵字詞與標籤方式放置於搜尋結果之後，以便之後做地理視覺化時，可以用地點資訊呈現輿情位置。

由於 2019 年 9 月，中研院馬教授有提供 ckiptagger，提供免費的 NER 模型，因此我們即有將既有的 NER 模型改成使用 ckiptagger 所使用的 bilstm 模型，以提高辨識的精準度。

以往的分析模型是將詞的表示組合成句子的表示，可以採用相加的方法，即將所有詞的表示進行加和，或者取平均等方法，但是這些方法沒有考慮到詞語在句子中前後順序。而如果改用 LSTM 模型可以更好的捕捉到較長距離的依賴關係。因為 LSTM 通過訓練過程可以學到記憶哪些資訊和遺忘哪些資訊。但是 LSTM 無法編碼從後到前的資訊。捕捉語義依賴關係，而 BiLSTM 則做了雙向編碼，因此可以更好的捕捉雙向的語義依賴。

圖十二、bilstm



4. 資料分析

準確度與召回率分析

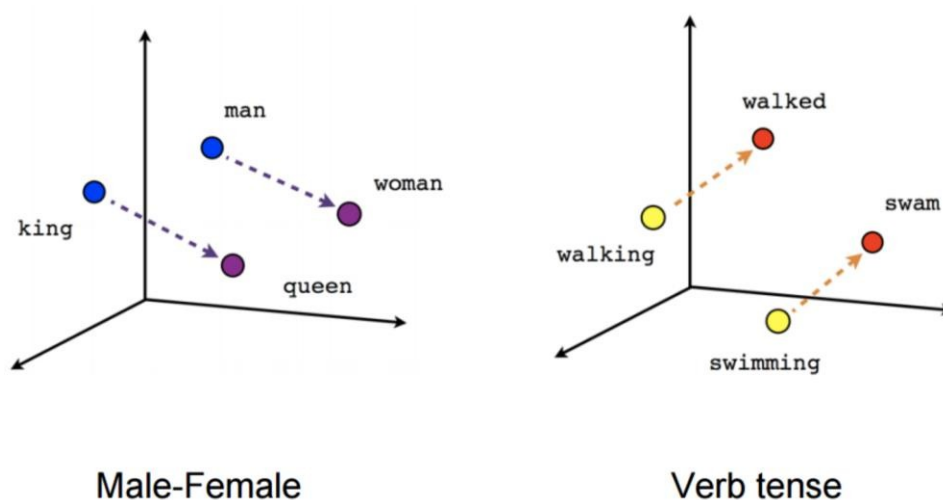
由於此次要監控的輿情來自眾多國家，而輿情也分屬不同語系，因此要如何精準的搜尋出跟疾病相關的外文輿情變成是一大挑戰。所幸維基百科上面列有各專業名詞的不同語系翻譯，因此我們便抓取維基百科上與疾病相關之字詞與對應語系的翻譯詞，我們便可以建立出一個外文的同義辭典，以便我們準確於 Google News 中搜尋出外文疫情資訊。

為了能夠有效評估該翻譯字詞是否能協助我們抓到正確的輿情新聞，因此我們用 recall 與 precision 兩個指標來評估，以下是 recall 與 precision 的完整定義精確率(precision)的公式是 $P = TP / (TP+FP)$ ，它計算的是所有"正確被檢索的結果(TP)"占有"實際被檢索到的(TP+FP)"的比例。召回率(recall)的公式是 $R = TP / (TP+FN)$ ，它計算的是所有"正確被檢索的結果(TP)"占有"應該檢索到的結果(TP+FN)"的比例。

我們抽樣一天所抓取的 6/11 流感新聞 315 篇，發現約有 58 篇跟流感毫無關係，多數是跟那斯達克的股票新聞有關，推測是可能該文章有提到製藥公司，而該公司有製作跟流感相關的疫苗，因此才納入搜集範圍之中。以此數據評估：準確率落在：81.6%而在評估召回率，Google News 的召回率約為 31% 我們發現如果有一則新聞被重複轉載^{0%}，Google 只會紀錄一篇，而且 Google 並非全面性的涵蓋所有新聞。

文章分類

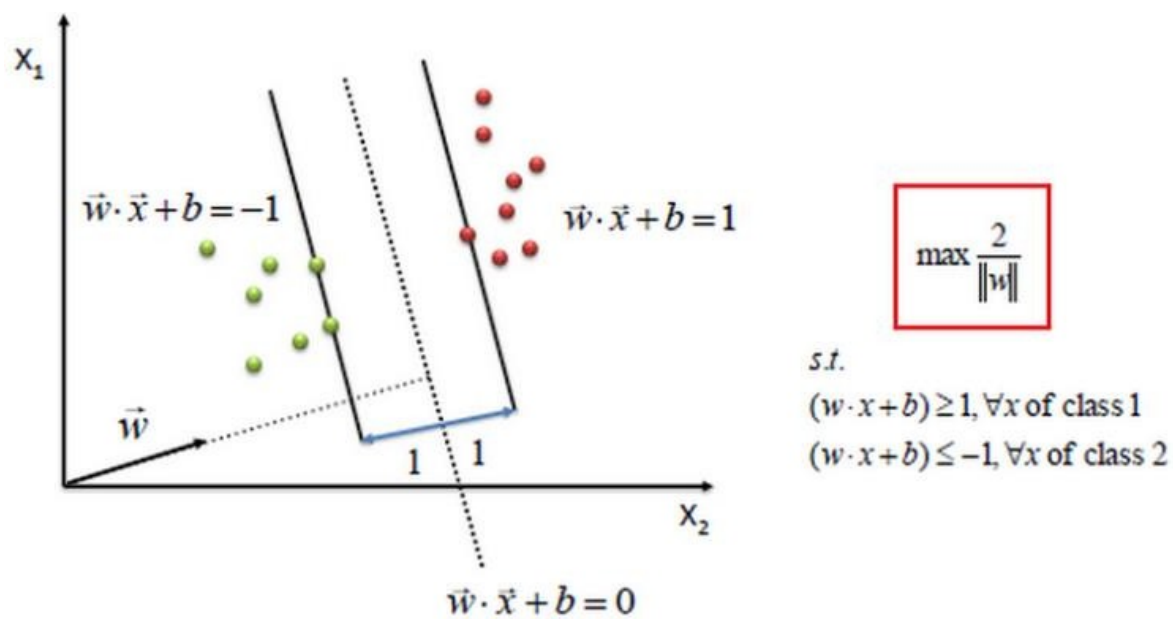
為了避免使用者收到太多不相關的文章，本系統將會使用機器學習的分類方法學習使用者的標準，讓系統自動分類文章。在本研究中，我們將採用 Word2Vec 方法先訓練字詞間的關係。



圖十三、word2vec

在建立 Word2Vec 的過程中，我們即會訓練出文字的詞向量，而詞向量除了可以減少文字的表示空間亦可以捕捉詞與詞之間的關係。

而為了能夠得到更好的分類效果，我們將句子中所有的詞向量取平均，進一步得到句向量，而後就可以透過句向量，套用 SVM 分類模型，即可準確的判斷該則資料是否屬於疫情相關的輿情，或只是單純的雜音。



圖十四、SVM 分類方法

5. 系統環境

在此節將說明，建立「即時輿情分析系統」所需之軟硬體技術。

軟體部分

1. NGINX 伺服器：

Apache 與 Nginx 伺服器是世界上最常見到的兩個網頁伺服器軟體。兩者綜合起來服務了超過網路上約%的流量。這兩者不但皆可以服務大量的網路請求，並可協同多種程式語言（Java, Python, PHP）建構完整的網路應用程式。

在國內，大多數應用程式還是使用 Apache 所建構的，但 Apache 伺服器的缺點在於無法同時乘載成千上萬個 HTTP 請求時，而讓網頁應用程式能同時承載成千上萬個請求的問題也被稱為 C10K 挑戰。為了解決這個問題，Igor Sysoev 於

2002 年發明了 Nginx 伺服器，並在 2004 年推出了第一個版本，希望透過非同步，事件驅動架構解決 C10K 問題。Nginx 後來逐漸成為網頁伺服器的主流之一，特點在於它使用較輕量的資源而且可以分散佈署在低端的硬體上，可以快速服務靜態的資源，並可以動態傳遞請求至其他應用程序。

在本規劃中，為了能讓「輿情分析系統」可以乘載大量的請求，故使用 Nginx 做為底層的網頁服務器。

2. RethinkDB :

網路輿情資料的特性是龐大且非結構化，為了能有效儲存符合這樣特性的資料，我們選用了非關聯式資料庫 RethinkDB 存放爬蟲蒐集到的輿情資料。RethinkDB 如同 MongoDB，可以使用 JSON 格式儲存資料，並可透過分散式的架構，使用多台機器乘載大量資料。但 RethinkDB 有一項其他 NoSQL 所沒有的特點，提供了主動推送更新(Push-Query)功能。

當以往前端頁面要更新呈現內容時，必須使用輪詢 (Polling) 的方式不停的對後端資料庫發出請求，以查詢最新的資料。如果輪詢 (Polling) 的頻率過於頻繁，將會對資料庫造成負擔，但若減少輪詢 (Polling) 的頻率，則實際資料跟資料庫中的內容會有落差。而 RethinkDB 的主動推送更新功能解決了這問題。

3. Ubuntu :

本系統將建構在 Ubuntu 16.04 上，Ubuntu 是以桌面應用為主的 GNU/Linux 作業系統，除了具有免費、開源的優點外，相較於其他 Linux 作業系統有下列好處：

1. 對使用者相對親善，Ubuntu 隱藏了許多 Linux 操作的細節，使用者可以透過圖形化介面操作作業系統，並透過其完備的套件管理工具安裝及管理各式應用程式
2. 具有完備的程式管理套件，使用者可透過套件管理工具安裝及管理各式應用程式
3. 擁有相對成熟的開源社群，使用者可以輕易透過社群尋求幫助
4. 具商業公司 Canonical 做技術支援，被回報的軟體缺陷，都會在下個新版本被修正。

4. PostgreSQL

PostgreSQL 是可以媲美 Oracle 的開源資料庫，由全世界超過 1000 名貢獻者所維護。PostgreSQL 只提供單一版本，不像 MySQL 有區分社區版、商業版與企業版。基於自由的 BSD/MIT 許可，任何組織可以使用、複製、修改和重新發佈程式碼。

PostgreSQL 相較於其他資料庫而言，有相當高的可靠度、資料一致性、安全性與完整性，企業可以使用 PostgreSQL 打造相對穩健、安全的企業資料倉儲。另外，PostgreSQL 的文件非常完整，使用者可以從線上找到大量免費的線上手冊，或可以從歸檔文件中找尋舊版本的文件。

而除了資料庫本身功能外，多數企業使用者通常會將該開源產品是否

容易維護納入考量之一，而由於 PostgreSQL 具有強大的社群及獲得些許商業公司的支持，使用者還是可以透過社群或商業公司的管道獲得使用上的協助。

5. Python 程式語言：

Python 是一通用的直譯式、交互式、物件導向高階程式語言，由於其非常簡單易用，具有廣大的社群支持，還有完整的套件管理工具，Python 已經被廣泛運用在自動化腳本、物聯網及資料分析上。而由於 Python 的幾個套件可以大幅簡化資料 ETL(Extract, Transformation, Loading)的工作，並且可以透過其高階分析套件 Pandas 及 scikit-learn 進行敘述性統計與機器學習，因此在本專案上將會使用 Python 做為系統開發的主力工具。以下將簡介我們會使用到的套件：

1. Requests:

Requests 套件讓我們可以使用 Python 對遠端伺服器下達 REST (包含 GET, POST, PUT, DELETE)的操作，因此我們可以使用該 Requests 對網頁伺服器遞送 GET 與 POST 請求，蒐集伺服器回應資訊，進而取得網路輿情頁面原始碼。

2. Scrapy:

Scrapy 是一 Python 爬蟲框架，內建許多爬蟲所需的支援與函式庫，讓使用者可以輕鬆開發、維護一多線程(Multithread)的網路爬蟲。

3. BeautifulSoup4:

取得頁面原始碼後，為了能萃取重要資訊如：文章主題、作者、主文、推文等資訊，我們可以使用 BeautifulSoup4 剖析頁面原始碼，並可透過 css selector 或 XPath 等操作抽取頁面中關鍵資訊。

4. Pandas:

為了能夠快速清理及分析資料，在本系統中使用了 Pandas(Python for Data Analysis)套件，Pandas 套件能將 csv, Json, html 等格式快速轉變成 DataFrame 格式，之後使用者可以透過類似 SQL 操作對資料作敘述性統計。

5. Jieba:

為了能夠對中文資料斷詞，以及可以從中挑出重要的人、事、時、地、物等名詞，在本系統中使用了 jieba 套件，該套件使用 Trie Tree 等資料結構去生成句子中文字所有可能組合，接者使用動態規劃找出組合中最大概率路徑，這個路徑就是基於詞頻的最可能分詞結果。而在新詞辨識部分則使用了隱馬可夫模型的 Viterbi 演算法。該套件是目前 Python 界最受歡迎的中文斷詞工具。

6. Scikit Learn:

在本研究中，會使用機器學習技術分群、分類輿情資訊。而 scikit learn 套件內建有回歸分析、資料分類、資料分群與降低維度等功能，使用者便可在整理好資訊後，使用機器學習套件快速分析研究資料。

7. django:

django 提供一 MTV 架構，讓使用者可以用模組化方式架構一完整的網頁應用程式。另外，django 有提供一個完整的後台操作介面，使用者便不需要另外編寫管理介面，即可透過後端的管理介面操做資料庫。由於本計劃將大多使用 javascript 做前端頁面的呈現，因此本計劃將使用 django 搭建一個 API 服務，做為前端畫面與後端資料庫的接口，供前端頁面存取資料庫內容，並呈現到地圖上。

8. Networkx:

Networkx 是可以用來建構複雜網路模型的 Python 套件，除了可以透過該套件建構一網路模型外，該套件內建許多圖型探勘的演算法，使用者可以使用該套件找出子網路模組(module)或探勘圖型的重要 Javascript 程式語言：

相較於使用 Python 開發伺服器端（後端）的 API，我們必須使用客戶端程式語言 Javascript 呈現頁面資訊，使用者才能與前端介面進行互動。而在本計劃中，為了讓使用者能享受互動體驗，並觀看相關動畫，我們將使用 Javascript 撰寫所有前端頁面的呈現內容(包含地圖座標的標記，前端互動的表單，圖表的呈現，前端動畫)，而我們除了使用原生的 Javascript 外，會搭配以下套件完成我們的頁面製作。

9. JQuery:

Jquery 是一跨平台的 Javascript 套件，主要的目的是為了簡化 Javascript 的撰寫，讓使用者可以更方便的透過 DOM 操作頁面元素、建立動畫與處理非同步事件。使用 JQuery，我們便可以 Write Less & Do More 完成所有頁面製作。

10. Leaflet:

為了製作互動式的輿情地圖，我們除了將使用 Openstreetmap 提供的圖資外，我們可以使用 Leaflet 建構互動地圖。Leaflet 是由 Vladimir Agafonkin 所開發的免費且開源的 Javascript 套件，套件本身可以讓使用者在地圖上任意添加點、線、面等資訊，也支援主要作業系統與移動裝置，非常適合套用在互動地圖的設計。

11. Chart.js:

Chart.js 是個以 HTML5 的 Canvas 為基礎的圖表插件(Plugin)，可以用圖表視覺化呈現數據，支援動畫效果，且可以運作在所有支持 HTML5 的瀏覽器上。

6. ElasticSearch：

ElasticSearch 是建構在 Apache Lucene 的開源搜尋引擎，但比較起其前輩 Lucene，Elasticsearch 安裝簡單、以 JSON 做為模型，使用者可以透過 JSON 建立模型、索引資料、查詢資料，以及可以佈署在分散式架構等優點，讓 ElasticSearch 成為當前最熱門的開源搜尋引擎。

7. Redis：

Redis 是一存放在記憶體中的非關聯式資料庫。不同於另外一個記憶體的儲存方案 Memcached，其可存取較大物件與支援 SQL 語法，讓他成為記憶體資料庫的最佳選擇。因為該資料庫存放在記憶體中，因此可快速吞吐輸入與輸出資料，通常被用作儲存使用者的會話與購物車資訊，另外可以用來當作分散式叢集的佇列架構。

(三) 結果

1. 監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之疫情輿情原計畫目標

為了能夠監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之輿情資料，我們將在抓取這些語系在 Google News 中出現的新聞資料到監測系統中。

本研究成果

關於疾病管制相關的資料源，目前已增設墨西哥疾管、韓國疾管、法國疾管、越南疾管、葡萄牙疾管、日本疾管與迦納疾管等資訊。

使用者可以先進到輿情系統中，點選右上角的個人設置（紅色框框標註的綠色圓圈圖樣）



圖十五、進入個人設置

點擊綠色圓圈以後，下方將出現使用者設置、頻道來源、使用文件與聯絡我們等選單畫面，在此點選頻道來源等畫面，右方將出現頻道來源。點選「國外疾管」後，下方將列出跟過外疾病相關資料來源共 7 個。

	名稱	子版塊數量
+	墨西哥疾管	22
+	法國疾管	6
+	迦納疾管	10
+	韓國疾管	27
+	越南疾管	17
+	葡萄牙疾管	27
+	日本疾管	22

圖十六、檢視「國外疾管」來源

表二、國外疾管網站名稱與連結

來源	關鍵字
墨西哥 疾管	流感
	百日咳
	破傷風
	腸病毒感染併發重症
	漢生病
	多重抗藥性結核病
	西尼羅熱
	梅毒
	德國麻疹
	鼠疫
	登革熱
	霍亂
	茲卡病毒感染症

	天花
	急性病毒性 A 型肝炎
	傷寒
	急性病毒性肝炎未定型
	漢他病毒症候群
	流行性斑疹傷寒
	狂犬病
	急性病毒性 B 型肝炎
	小兒麻痺症/急性無力肢體 麻痺
韓國疾管	流感
	百日咳
	破傷風
	人類免疫缺乏病毒感染
	嚴重急性呼吸道症候群
	急性病毒性 C 型肝炎
	漢生病
	多重抗藥性結核病
	梅毒
	德國麻疹
	退伍軍人病
	鼠疫
	腸道出血性大腸桿菌感染 症
	登革熱
	霍亂
	淋病
	日本腦炎
	天花
	瘧疾
	急性病毒性 A 型肝炎

傷寒

	流行性腮腺炎
	流行性斑疹傷寒
	狂犬病
	急性病毒性 B 型肝炎
	炭疽病
	小兒麻痺症/急性無力肢體 麻 痺
迦納疾管	流感
	急性病毒性 A 型肝炎
	急性病毒性 D 型肝炎
	人類免疫缺乏病毒感染
	德國麻疹
	侵襲性 b 型嗜血桿菌感染 症
	茲卡病毒感染症
	腸病毒感染併發重症
	漢生病
	天花
法國疾管	流感
	副傷寒
	瘧疾
	急性病毒性肝炎未定型
	梅毒
	流行性腮腺炎
	登革熱
越南疾管	流感
	副傷寒
	瘧疾
	急性病毒性肝炎未定型
	梅毒

流行性腮腺炎

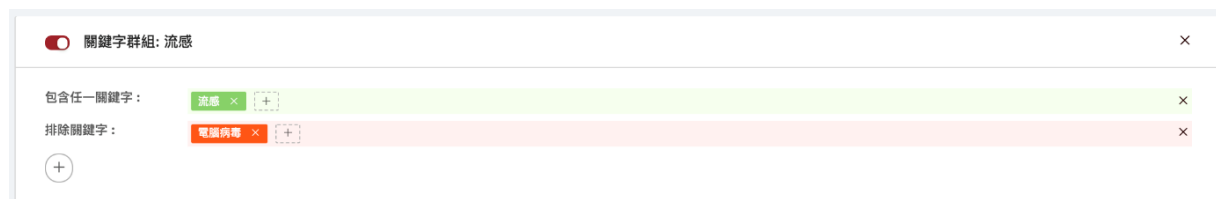
	登革熱
	嚴重急性呼吸道症候群
	流行性斑疹傷寒
	狂犬病
	霍亂
	急性病毒性 C 型肝炎
	急性病毒性 B 型肝炎
	茲卡病毒感染症
	漢生病
	日本腦炎
	小兒麻痺症/急性無力肢體 麻痺
葡萄牙 疾管	流感
	急性病毒性 E 型肝炎
	急性病毒性 D 型肝炎
	破傷風
	屈公病
	人類免疫缺乏病毒感染
	嚴重急性呼吸道症候群
	漢生病
	腸病毒感染併發重症
	多重抗藥性結核病
	梅毒
	德國麻疹
	白喉
	麻疹
	鼠疫
	登革熱
	霍亂
	茲卡病毒感染症

天花

	瘧疾
	急性病毒性 A 型肝炎
	漢他病毒症候群
	流行性斑疹傷寒
	狂犬病
	急性病毒性 B 型肝炎
	炭疽病
	小兒麻痺症/急性無力肢體 麻 痺
日本疾管	流感
	急性病毒性 D 型肝炎
	人類免疫缺乏病毒感染
	嚴重急性呼吸道症候群
	急性病毒性 C 型肝炎
	漢生病
	多重抗藥性結核病
	梅毒
	德國麻疹
	麻疹
	鼠疫
	登革熱
	霍亂
	日本腦炎
	瘧疾
	傷寒
	急性病毒性肝炎未定型
	流行性腮腺炎
	漢他病毒症候群
	流行性斑疹傷寒
	狂犬病

急性病毒性 B 型肝炎

由於已經將跟相關網站加入到監測範圍中，使用者便可以在系統上鍵入關鍵字，即可搜尋到跟該語系相關的疫病資訊。例如：使用者可以在關鍵字中建立「流感」關鍵字群組，接者在群組中設定流感（並排除掉電腦病毒等字樣），如下圖所示：



圖十七、設定關鍵字組

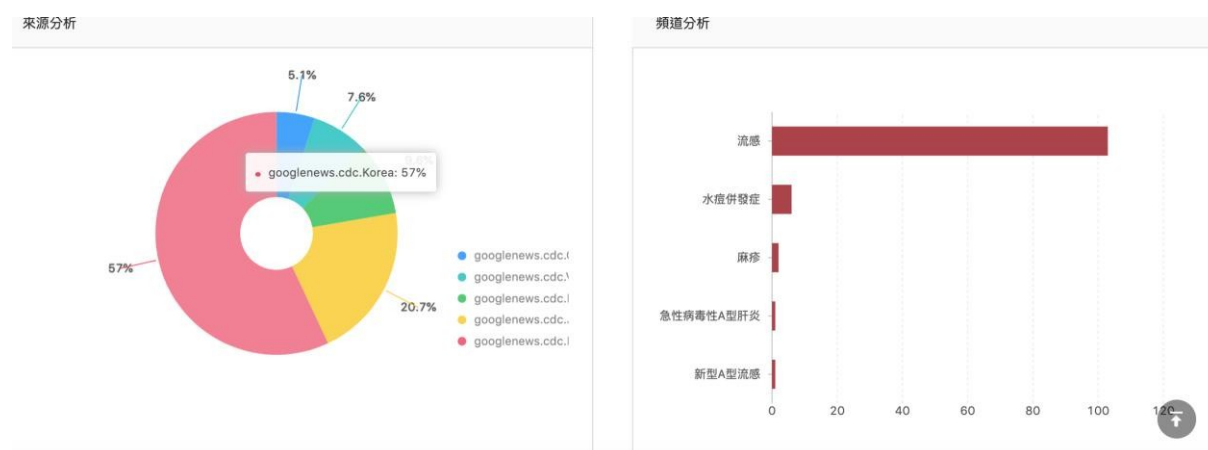
接者，將系統切換至關鍵字報表，使用者便可以在報表中，瀏覽跟流感相關的輿情聲量、頻道分析、來源分析、熱詞分析及文章表列。



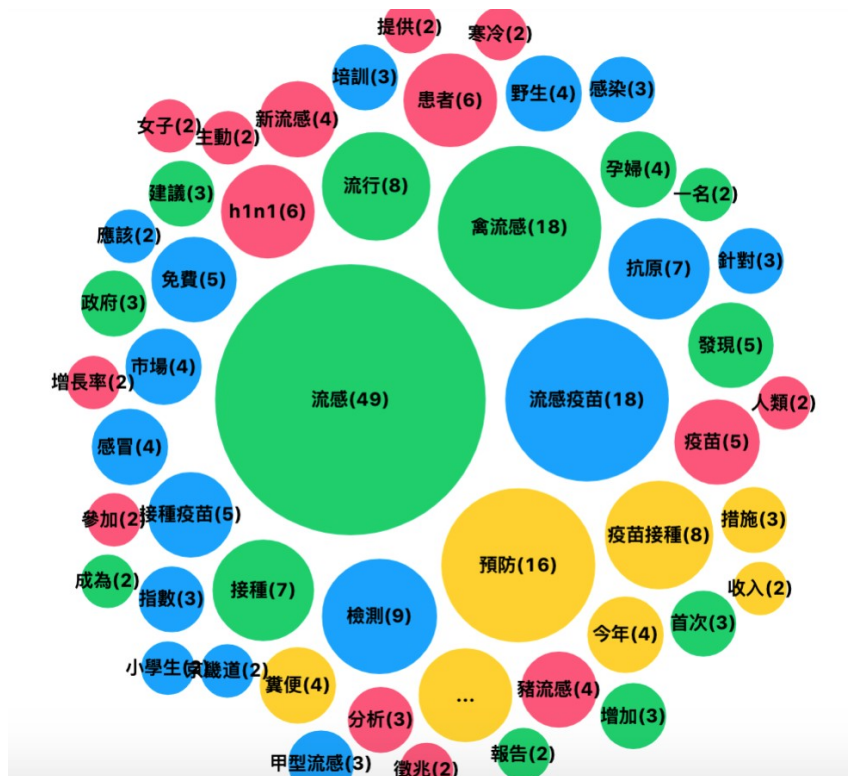
圖十八、設定關鍵字與指定頻道



圖十九、觀測輿情聲量



圖二十、分析輿情連結來源與輿情頻道來源



圖二十一、熱詞分析

文章列表 [下載文章](#) [議題關聯字分析](#)

There have been deaths from influenza A / H1N1: Experts warn of what to do immediately | 甲型/ H1N1 流感已導致死亡：專家警告立即採取行動 🔍

2019/11/14 07:05 國際 越南疾管 流感

Giới chuyên gia khuyến cáo, thời điểm hiện nay dịch cúm bắt đầu hoành hành nên mọi người cần hết sức cẩn trọng, tránh chủ quan dẫn đến hậu quả đáng tiếc

情緒: 中立 | 👍 0 | 👎 0 | 🗨️ 0 | 🔄 0 | 🌐 0 | 🌐 0 | 🌐 0 | 🔍 檢視文章 | 📄 分享連結 | ⌛ 過濾此類文章

Prime Minister Medvedev proposes vaccination to colleagues for Ebola vaccine instead of influenza | 總理梅德韋傑夫提議為同事接種埃博拉疫苗而不是 流感疫苗 🔍

2019/11/13 22:45 國際 日本疾管 流感

ロシアのドミトリー・メドベージェフ首相は、シベリアの学研都市ノボシビルスクの視察で、インフルエンザやエボラ出血熱の予防接種について冗談を言った。...

👍 0 | 👎 0 | 🗨️ 0 | 🔄 0 | 🌐 0 | 🌐 0 | 🌐 0 | 🔍 檢視文章 | 📄 分享連結 | ⌛ 過濾此類文章

Pneumonia, flu or cold: how to differentiate respiratory infections? | 肺炎、流感或感冒：如何區分呼吸道感染？ 🔍

圖二十二、文章列表

2. 使用 Wikipedia 的資訊建立一多語系同義字典 原計畫目標

使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。為了確認翻譯的品質，我們亦會增加 precision 與 recall 之評量，確保字詞的翻譯品質。

本研究成果

於 108 年度的計畫，因為多了外國疫情的資訊需要偵蒐，因此我們必須要取得疾病相關的翻譯字詞，才有辦法從 Google News 檢索其他語系的疾管資料。因此我們使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。

由於此次要監控的輿情來自眾多國家，而輿情也分屬不同語系，因此要如何精準的搜尋出跟疾病相關的外文輿情變成是一大挑戰。所幸維基百科上面列有各專業名詞的不同語系翻譯，因此我們便抓取維基百科上與疾病相關之字詞與對應語系的翻譯詞，我們便可以建立出一個外文的同義辭典，以便我們準確於 Google News 中搜尋出外文疫情資訊。

我們的作法如下：

- a. 進到維基百科平台
- b. 下關鍵字搜尋與疾病相關之條目

- 分類索引
- 特色內容
- 新聞動態
- 近期變更
- 隨機條目
- 說明
- 說明
- 維基社群
- 方針與指引
- 互助客棧
- 知識問答
- 字詞轉換
- IRC即時聊天
- 聯絡我們
- 關於維基百科
- 資助維基百科
- 其他專案
- 維基共享資源
- 列印/匯出
- 下載為 PDF

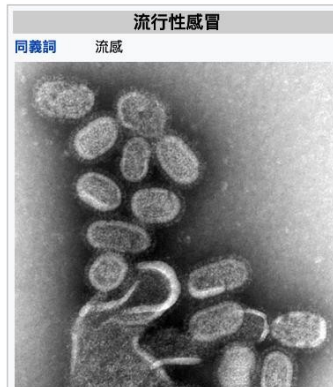
流行性感胃 [編輯]



維基百科，自由的百科全書

流行性感胃（Influenza），通常簡稱為**流感**，為一種由**流感病毒**造成的傳染性疾病^[1]。流感的症狀可輕可重^[4]，最常見者為高燒、流鼻水、喉嚨痛、肌肉痠痛、頭痛、咳嗽和疲倦感。患者通常在接觸病毒2天後發病，症狀大多在一週內會解除，但咳嗽可能持續超過兩週^[1]。孩童可能會噁心和嘔吐，但這在成人並不常見；噁心和嘔吐更常發生在與流感病毒無關的感染性腸胃炎，有時會不精確地稱此為腸胃型感冒（stomach flu）^[5]。流感可能的併發症包括病毒性肺炎、次級細菌性肺炎、鼻竇感染以及造成其他疾病惡化（如氣喘或心臟衰竭）^{[4][2]}。

可感染人類的流感病毒有A、B、C三型^[2]。病毒通常由咳嗽，打噴嚏和說話產生的飛沫傳播^[1]，近距離接觸時尤其容易發生^[6]。此外，病毒也可藉由接觸到受染污的物體表面、再碰觸口或眼睛後傳播^{[4][6]}。受感染的患者無論在發病前後均可能具有傳染性^[4]，喉嚨、痰液或鼻黏膜等檢體的病毒測試則可作為確診的依據。目前已有數種快篩方法：然而快篩仍有危險性（即便經感染，檢測結果仍顯示為未感染的



圖二十三、疾病相關之條目

a. 切換語系



圖二十四、語系切換

b. 抓取切换語系後的條目

The screenshot shows the Indonesian Wikipedia page for 'Influenza'. The title is 'Influenza' and it is noted as being from the Indonesian Wikipedia. A warning box states that the article is for general information and not medical advice. The main text describes influenza as a disease caused by RNA viruses, with symptoms like fever, cough, and fatigue. It also mentions that influenza can be spread through the air and contact with contaminated surfaces. A small image of influenza virus particles is shown on the right side of the page.

根據上述方法，我們目前已經根據常見疾管資料建立詞典，範例如下：

表三：字詞翻譯結果

語系	翻譯文	原文
韓文	원발성 수막뇌염 아메바성	福氏內格里阿米巴腦膜腦炎
韓文	광견병	狂犬病
韓文	큐열	Q熱
韓文	디프테리아	白喉
韓文	장출혈성 감염증 대장균	腸道出血性大腸桿菌感染症
韓文	매독	梅毒
韓文	인플루엔자	流感
韓文	C형 간염	急性病毒性B型肝炎
韓文	항생제 저항성	VISA/VRSA 抗藥性檢測
韓文	지카바이러스	茲卡病毒感染症
韓文	림프사상충증	淋巴絲蟲病

韓文	치쿤구니아열	屈公病
韓文	백일해	百日咳
韓文	파상풍	破傷風
韓文	야토병	免熱病
韓文	유비저	類鼻疽
韓文	결핵	多重抗藥性結核病
韓文	레지오넬라증	退伍軍人病
韓文	수두	水痘併發症
韓文	묘소병	貓抓病
韓文	인간면역결핍 바이러스	人類免疫缺乏病毒感 染
韓文	중동호흡기증후군	中東呼吸症候群冠狀 病 毒感染症
韓文	임질	淋病
韓文	탄저	炭疽病
韓文	메티실린 내성 황색포도상구균	社區型 MRSA
韓文	엔테로바이러스	腸病毒感感染併發重症
韓文	머릿니	頭蝨感染症
韓文	에볼라바이러스속	伊波拉病毒感感染
韓文	홍역	麻疹
韓文	소아마비	小兒麻痺症/急性無力 肢體麻痺
韓文	황열	黃熱病
韓文	장염	細菌性腸胃炎
韓文	발진티푸스	流行性斑疹傷寒
韓文	콜레라	霍亂
韓文	덴기열	登革熱
韓文	페스트	鼠疫
韓文	페렴 연쇄상구균	侵襲性肺炎鏈球菌感 染

		症
韓文	렙토스피라증	鉤端螺旋體病
韓文	클로스트리디움 보툴리눔	肉毒桿菌中毒
韓文	중증급성호흡기증후군	嚴重急性呼吸道症候群
韓文	리스테리아 모노키토게네스	李斯特菌症
韓文	일본뇌염	日本腦炎
韓文	파라티푸스	副傷寒
韓文	리프트 계곡 열	裂谷熱
韓文	풍진	德國麻疹
韓文	중증열성혈소판감소증후군	發熱伴血小板減少綜合症
韓文	쯔쯔가무시병	恙蟲病
韓文	브루셀라증	布氏桿菌病
韓文	천연두	天花
韓文	말라리아	瘧疾
韓文	B 형 간염	急性病毒性 C 型肝炎
韓文	라싸열	拉薩熱
韓文	크로이츠펠트-야코프병	庫賈氏病
韓文	장티푸스	傷寒
韓文	옴	疥瘡感染症
韓文	한타바이러스	漢他病毒症候群
韓文	라임병	萊姆病
韓文	간흡충	中華肝吸蟲感染症
韓文	H1N1	新型 A 型流感
韓文	나병	漢生病
韓文	볼거리	流行性腮腺炎
韓文	A 형 간염	急性病毒性 A 型肝炎
英文	Hepatitis A	急性病毒性 A 型肝炎
英文	Measles	麻疹

英文	Cholera	霍亂
英文	West Nile virus	西尼羅熱
英文	Trichinella spiralis	旋毛蟲感染症
英文	Intestinal parasite infection	常見腸道寄生蟲病
英文	Scabies	疥瘡感染症
英文	Whooping cough	百日咳
英文	Lymphatic filariasis	淋巴絲蟲病
英文	Amoebic dysentery	阿米巴性痢疾
英文	Marburg virus	馬堡病毒出血熱
英文	Anthrax	炭疽病
英文	Hepatitis C	急性病毒性 B 型肝炎
英文	Salmonellosis	沙門氏菌感染症
英文	Cryptococcosis	隱球菌症
英文	Hepatitis D	急性病毒性 D 型肝炎
英文	Melioidosis	類鼻疽
英文	Shigellosis	桿菌性痢疾
英文	GB virus C	急性病毒性肝炎未定 型
英文	Rubella	德國麻疹
英文	Henipavirus	亨德拉病毒感染症
英文	HIV	人類免疫缺乏病毒感 染
英文	Head louse	頭蝨感染症
英文	Malaria	瘧疾
英文	Chickenpox	水痘併發症
英文	Brucellosis	布氏桿菌病
英文	Plague	鼠疫
英文	Clonorchis sinensis	中華肝吸蟲感染症
英文	Hepatitis E	急性病毒性 E 型肝炎
英文	Lyme disease	萊姆病
英文	Streptococcus suis	第二型豬鏈球菌感 染症

英文	Streptococcus pneumoniae	侵襲性肺炎鏈球菌感 染 症
英文	Leptospirosis	鈎端螺旋體病
英文	Tetanus	破傷風
英文	Psittacosis	鸚鵡熱
英文	Paratyphoid fever	副傷寒
英文	Naegleriasis	福氏內格里阿米巴腦 膜 腦炎
英文	Yellow fever	黃熱病
英文	Scrub typhus	恙蟲病
英文	Middle East respiratory syndrome	中東呼吸症候群冠狀 病 毒感染症
英文	Angiostrongylus cantonensis	廣東住血線蟲感染症
英文	Chikungunya	屈公病
英文	Pneumocystis pneumonia	肺囊蟲肺炎
英文	Lassa fever	拉薩熱
英文	Rift Valley fever	裂谷熱
英文	Toxoplasmosis	弓形蟲感染症
英文	Ebolavirus	伊波拉病毒感染
英文	Gastroenteritis	細菌性腸胃炎
英文	Enterovirus	腸病毒感染併發重症
英文	Smallpox	天花
英文	Cat-scratch disease	貓抓病
英文	Acanthamoeba	棘狀阿米巴
英文	Typhoid fever	傷寒
英文	Syphilis	梅毒
英文	Antimicrobial resistance	VISA/VRSA 抗藥性檢 測
英文	Gonorrhoea	淋病

英文	Severe fever with thrombocytopenia syndrome	發熱伴血小板減少綜合症
英文	Polio	小兒麻痺症/急性無力肢體麻痺
英文	Influenza A virus subtype H1N1	新型 A 型流感
英文	Tuberculosis	多重抗藥性結核病
英文	Typhus	流行性斑疹傷寒
英文	Rabies	狂犬病
英文	Methicillin-resistant Staphylococcus aureus	社區型 MRSA
英文	Diphtheria	白喉
英文	Dengue fever	登革熱
英文	Leprosy	漢生病
英文	Zika virus	茲卡病毒感染症
英文	Influenza	流感
英文	Orthohantavirus	漢他病毒症候群
英文	Mumps	流行性腮腺炎
英文	Haemophilus influenzae	侵襲性 b 型嗜血桿菌感染症
英文	Hepatitis B	急性病毒性 C 型肝炎
英文	Enterohemorrhagic	腸道出血性大腸桿菌感染症
英文	Clostridium botulinum	肉毒桿菌中毒
英文	Legionnaires' disease	退伍軍人病
英文	Listeria monocytogenes	李斯特菌症
英文	Creutzfeldt - Jakob disease	庫賈氏病
英文	Tularemia	兔熱病
英文	Severe acute respiratory syndrome	嚴重急性呼吸道症候群

英文	Paragonimus westermani	肺吸蟲感染症
英文	Japanese encephalitis	日本腦炎
西班牙文	Hepatitis A	急性病毒性 A 型肝炎
西班牙文	Cryptococcosis	隱球菌症
西班牙文	Enfermedad por arañazo de gato	貓抓病
西班牙文	Staphylococcus aureus resistente a la meticilina	社區型 MRSA
西班牙文	Streptococcus pneumoniae	侵襲性肺炎鏈球菌感染症
西班牙文	Gonorrhea	淋病
西班牙文	Enfermedad de Creutzfeldt-Jakob	庫賈氏病
西班牙文	Varicela	水痘併發症
西班牙文	Psittacosis	鸚鵡熱
西班牙文	Viruela	天花
西班牙文	Tos ferina	百日咳
西班牙文	Hepatitis C	急性病毒性 B 型肝炎
西班牙文	Chikunguña	屈公病

西班牙 文	Encefalitis japonesa	日本腦炎
西班牙 文	Fiebre de Lassa	拉薩熱
西班牙 文	Hepatitis D	急性病毒性 D 型肝炎
西班牙 文	Sarampió n	麻疹
西班牙 文	Rubeola	德國麻疹
西班牙 文	Legionelosis	退伍軍人病
西班牙 文	Shigelosis	桿菌性痢疾
西班牙 文	Pediculus humanus capitis	頭蝨感染症
西班牙 文	Có lera	霍亂
西班牙 文	Hepatitis B	急性病毒性 C 型肝炎
西班牙 文	Melioidosis	類鼻疽
西班牙 文	Clonorchis sinensis	中華肝吸蟲感染症
西班牙 文	Hepatitis E	急性病毒性 E 型肝炎
西班牙 文	Difteria	白喉

西班牙 文	Leptospirosis	鉤端螺旋體病
西班牙 文	Influenzavirus A subtipo H1N1	新型 A 型流感
西班牙 文	Parotiditis	流行性腮腺炎
西班牙 文	Fiebre del valle del Rift	裂谷熱
西班牙 文	Fiebre tifoidea	傷寒
西班牙 文	Fiebre Tsutsugamushi	恙蟲病
西班牙 文	Fiebre paratifoidea	副傷寒
西班牙 文	Sí ndrome respiratorio agudo grave	嚴重急性呼吸道症候 群
西班牙 文	Sí filis	梅毒
西班牙 文	Streptococcus suis	第二型豬鏈球菌感染 症
西班牙 文	Virus del Zika	茲卡病毒感染症
西班牙 文	Virus de la inmunodeficiencia humana	人類免疫缺乏病毒感 染
西班牙 文	Toxoplasmosis	弓形蟲感染症

西班牙 文	Sarna	疥瘡感染症
西班牙 文	Enterovirus	腸病毒感 染併發重症
西班牙 文	Gastroenteritis	細菌性腸胃炎
西班牙 文	Té tanos	破傷風
西班牙 文	Meningoencefalitis amebiana primaria	福氏內格里阿米巴腦 膜 腦炎
西班牙 文	Acanthamoeba	棘狀阿米巴
西班牙 文	Virus del Nilo Occidental	西尼羅熱
西班牙 文	Brucelosis	布氏桿菌病
西班牙 文	Lepra	漢生病
西班牙 文	Resistencia a antibió ticos	VISA/VRSA 抗藥性檢 測
西班牙 文	Salmonelosis	沙門氏菌感染症
西班牙 文	Poliomielitis	小兒麻痺症/急性無 力 肢體麻 痺
西班牙 文	Carbunco	炭疽病

西班牙 文	Enfermedad de Lyme	萊姆病
西班牙 文	Tuberculosis	多重抗藥性結核病
西班牙 文	Virus de la hepatitis G	急性病毒性肝炎未定 型
西班牙 文	Dengue	登革熱
西班牙 文	Fiebre amarilla	黃熱病
西班牙 文	Tifus	流行性斑疹傷寒
西班牙 文	Henipavirus	亨德拉病毒感染症
西班牙 文	Clostridium botulinum	肉毒桿菌中毒
西班牙 文	Hantavirus	漢他病毒症候群
西班牙 文	Fiebre Q	Q 熱
西班牙 文	Rabia	狂犬病
西班牙 文	Malaria	瘧疾
西班牙 文	Haemophilus influenzae	侵襲性 b 型嗜血桿菌 感 染症
西班牙 文		

	Peste	鼠疫
西班牙 文	Gripe	流感
西班牙 文	Listeria monocytogenes	李斯特菌症
西班牙 文	Infecció n intestinal por parás itos	常見腸道寄生蟲病
西班牙 文	Elefantiasis	淋巴絲蟲病
西班牙 文	Tularemia	兔熱病
西班牙 文	Paragonimus westermani	肺吸蟲感染症
西班牙 文	Trichinella spiralis	旋毛蟲感染症
越南 文	Sốt vàng	黃熱病
越南 文	Quai bị	流行性腮腺炎
越南 文	Thủy đậu	水痘併發症
越南 文	Cúm	流感
越南 文	Viêm gan A	急性病毒性 A 型肝炎
越南 文	Giang mai	梅毒
越南 文	Sốt Lassa	拉薩熱

越南 文	Bệnh nhiễm khuẩn salmonella	沙門氏菌感染症
越南 文	Naegleriasis	福氏內格里阿米巴腦 膜 腦炎
越南 文	Toxoplasmosis	弓形蟲感染症
越南 文	Bệnh ghẻ	疥瘡感染症
越南 文	Bệnh giun chỉ bạch huyết	淋巴絲蟲病
越南 文	Chấy	頭蝨感染症
越南 文	Virus Tây sông Nin	西尼羅熱
越南 文	Phế cầu khuẩn	侵襲性肺炎鏈球菌感 染 症
越南 文	Bại liệt	小兒麻痺症/急性無 力 肢體麻痺
越南 文	Sốt rét	瘧疾
越南 文	Thương hàn	傷寒
越南 文	Chikungunya	屈公病
越南 文	Bệnh Creutzfeldt-Jakob	庫賈氏病
越南 文	Virus cúm A H1N1	新型 A 型流感

越南 文	Melioidosis	類鼻疽
越南 文	Clonorchis sinensis	中華肝吸蟲感染症
越南 文	Bạch hầu	白喉
越南 文	Viêm gan siêu vi B	急性病毒性 C 型肝炎
越南 文	Viêm gan siêu vi C	急性病毒性 B 型肝炎
越南 文	Đậu mùa	天花
越南 文	Bệnh Lê dương	退伍軍人病
越南 文	Angiostrongylus cantonensis	廣東住血線蟲感染症
越南 文	Viêm não Nhật Bản	日本腦炎
越南 文	Liên cầu khuẩn lợn	第二型豬鏈球菌感染症
越南 文	Bệnh dengue	登革熱
越南 文	Hội chứng hô hấp Trung Đông	中東呼吸症候群冠狀 病毒感染症
越南 文	Virus Zika	茲卡病毒感染症
越南 文	Bệnh Shigella	桿菌性痢疾
越南 文	Uốn ván	破傷風
越南 文	Sởi	麻疹

越南 文	Bệnh sốt thỏ	兔熱病
越南 文	Lậu mủ	淋病
越南 文	Virus Hanta	漢他病毒症候群
越南 文	Dịch hạch	鼠疫
越南 文	Phong củ i	漢生病
越南 文	Bệnh xoắn khuẩn và ng da	鈎端螺旋體病
越南 文	Bệnh sởi Đức	德國麻疹
越南 文	Ho gà	百日咳
越南 文	Bệnh tả	霍亂
越南 文	Sốt phát ban	流行性斑疹傷寒
越南 文	Bệnh Lyme	萊姆病
越南 文	Viêm dạ dày ruột	細菌性腸胃炎
越南 文	Bệnh than	炭疽病
越南 文	Sốt thung lũng Rift	裂谷熱
越南 文	Clostridium botulinum	肉毒桿菌中毒
越南 文	Lao	多重抗藥性結核病

越南 文	Bệnh mèo cào	貓抓病
越南 文	Staphylococcus aureus kháng methicillin	社區型 MRSA
越南 文	Haemophilus influenzae	侵襲性 b 型嗜血桿菌 感 染症
越南 文	Hội chứng hô hấp cấp tính nặng	嚴重急性呼吸道症候 群
越南 文	GB vi rút t C	急性病毒性肝炎未定 型
越南 文	HIV	人類免疫缺乏病毒感 染
越南 文	Kháng thuốc kháng sinh	VISA/VRSA 抗藥性檢 測
越南 文	Bệnh dại	狂犬病
越南 文	Sốt phó thương hàn	副傷寒
越南 文	Listeria monocytogenes	李斯特菌症
越南 文	Virus Marburg	馬堡病毒出血熱
越南 文	Trichinella spiralis	旋毛蟲感染症
葡萄牙 文	Malária	瘧疾
葡萄牙 文	Leptospirose	鉤端螺旋體病
葡萄牙 文	Gonorréia	淋病

葡萄牙 文	Salmonelose	沙門氏菌感染症
葡萄牙 文	Filariase linfática	淋巴絲蟲病
葡萄牙 文	Hepatite E	急性病毒性 E 型肝炎
葡萄牙 文	Varicela	水痘併發症
葡萄牙 文	Brucelose	布氏桿菌病
葡萄牙 文	Parotidite epidémica	流行性腮腺炎
葡萄牙 文	Varíola	天花
葡萄牙 文	Hepatite D	急性病毒性 D 型肝炎
葡萄牙 文	Hantavírus	漢他病毒症候群
葡萄牙 文	Vírus da zica	茲卡病毒感染症
葡萄牙 文	Hepatite A	急性病毒性 A 型肝炎
葡萄牙 文	Tifo	流行性斑疹傷寒
葡萄牙 文	Difteria	白喉
葡萄牙 文	Sarna	疥瘡感染症

葡萄牙 文	Rubé ola	德國麻疹
葡萄牙 文	Chicungunha	屈公病
葡萄牙 文	Ornitose	鸚鵡熱
葡萄牙 文	Criptococose	隱球菌症
葡萄牙 文	Febre amarela	黃熱病
葡萄牙 文	Có lera	霍亂
葡萄牙 文	Febre paratifoide	副傷寒
葡萄牙 文	Hepatite C	急性病毒性 B 型肝炎
葡萄牙 文	Streptococcus suis	第二型豬鏈球菌感染 症
葡萄牙 文	Doenç a por arranhadura do gato	貓抓病
葡萄牙 文	Febre Q	Q 熱
葡萄牙 文	Enterovirus	腸病毒感染併發重症
葡萄牙 文	Angiostrongylus cantonensis	廣東住血線蟲感染症
葡萄牙 文	Resistê ncia antibió tica	

		VISA/VRSA 抗藥性檢測
葡萄牙文	Vírus da imunodeficiência humana	人類免疫缺乏病毒感染
葡萄牙文	Pertussis	百日咳
葡萄牙文	Melioidose	類鼻疽
葡萄牙文	Sífilis	梅毒
葡萄牙文	Naegleríase	福氏內格里阿米巴腦膜腦炎
葡萄牙文	Vírus do Nilo Ocidental	西尼羅熱
葡萄牙文	Doença de Lyme	萊姆病
葡萄牙文	Febre tifoide	傷寒
葡萄牙文	Hepatite B	急性病毒性 C 型肝炎
葡萄牙文	Vírus de Marburg	馬堡病毒出血熱
葡萄牙文	Tétano	破傷風
葡萄牙文	Peste	鼠疫
葡萄牙文		麻疹

	Sarampo	
葡萄牙文	Amebí ase	阿米巴性痢疾
葡萄牙文	Doenç a dos legioná rios	退伍軍人病
葡萄牙文	Raiva	狂犬病
葡萄牙文	Carbú nculo	炭疽病
葡萄牙文	Staphylococcus aureus resistente à metilina	社區型 MRSA
葡萄牙文	Febre de Lassa	拉薩熱
葡萄牙文	Lepra	漢生病
葡萄牙文	Tuberculose	多重抗藥性結核病
葡萄牙文	Febre do vale Rift	裂谷熱
葡萄牙文	Streptococcus pneumoniae	侵襲性肺炎鏈球菌感染症
葡萄牙文	Pneumocistose	肺囊蟲肺炎
葡萄牙文	Dengue	登革熱
葡萄牙文	Shigelose	桿菌性痢疾

葡萄牙 文	Clostridium botulinum	肉毒桿菌中毒
葡萄牙 文	Gastroenterite	細菌性腸胃炎
葡萄牙 文	Ebolavirus	伊波拉病毒感染
葡萄牙 文	Doença de Creutzfeldt-Jakob	庫賈氏病
葡萄牙 文	Haemophilus influenzae	侵襲性 b 型嗜血桿菌 感 染症
葡萄牙 文	Poliomielite	小兒麻痺症/急性無力 肢體麻痺
葡萄牙 文	Acanthamoebí ase	棘狀阿米巴
葡萄牙 文	Influenzavirus A subtipo H1N1	新型 A 型流感
葡萄牙 文	Clonorchis sinensis	中華肝吸蟲感染症
葡萄牙 文	Gripe	流感
葡萄牙 文	Sí ndrome respirató ria aguda grave	嚴重急性呼吸道症候 群
葡萄牙 文	Listeria monocytogenes	李斯特菌症
葡萄牙 文	Toxoplasmose	弓形蟲感染症
葡萄牙 文		

	Tularemia	兔熱病
葡萄牙文	Paragonimus westermani	肺吸蟲感染症
葡萄牙文	Trichinella spiralis	旋毛蟲感染症
日文	麻疹	麻疹
日文	日本脳炎	日本脳炎
日文	チクングニア熱	屈公病
日文	H1N1 亜型	新型 A 型流感
日文	炭疽症	炭疽病
日文	オウム病	鸚鵡熱
日文	ヒト免疫不全ウイルス	人類免疫缺乏病毒感 染
日文	パラチフス熱	副傷寒
日文	ジカウイルス	茲卡病毒感染症
日文	チフス	流行性斑疹傷寒
日文	トキソプラズマ症	弓形蟲感染症
日文	破傷風	破傷風
日文	コレラ	霍亂
日文	ハンセン病	漢生病
日文	サルモネラ症	沙門氏菌感染症
日文	象皮病	淋巴絲蟲病
日文	クロイツフェルト・ヤコブ病	庫賈氏病
日文	A 型肝炎	急性病毒性 A 型肝炎
日文	ツツガムシ病	恙蟲病
日文	天然痘	天花
日文	肝吸虫	中華肝吸蟲感染症
日文	重症熱性血小板減少症候群	發熱伴血小板減少綜 合 症

日文	マラリア	瘧疾
日文	ライム病	萊姆病
日文	腸チフス	傷寒
日文	レジオネラ症	退伍軍人病
日文	C型肝炎	急性病毒性B型肝炎
日文	風疹	德國麻疹
日文	Q熱	Q熱
日文	リフトバレー熱	裂谷熱
日文	B型肝炎	急性病毒性C型肝炎
日文	ラッサ熱	拉薩熱
日文	梅毒	梅毒
日文	ペスト	鼠疫
日文	黄熱	黄熱病
日文	D型肝炎	急性病毒性D型肝炎
日文	ボツリヌス菌	肉毒桿菌中毒
日文	野兎病	兔熱病
日文	メチシリン耐性黄色ブドウ球菌	社區型MRSA
日文	デング熱	登革熱
日文	胃腸炎	細菌性腸胃炎
日文	中東呼吸器症候群	中東呼吸症候群冠狀 病 毒感染症
日文	インフルエンザ	流感
日文	ウェステルマン肺吸虫	肺吸蟲感染症
日文	クリプトコッカス症	隱球菌症
日文	ジフテリア	白喉
日文	西ナイルウイルス	西尼羅熱
日文	急性灰白髄炎	小兒麻痺症/急性無力 肢體麻痺
日文	リステリア・モノサイトゲネス	李斯特菌症
日文	百日咳	百日咳

	感染症の予防及び感染症の患者に対する 医 療に関する法律	疱疹 B 病毒感染症
日文	アメーバ赤痢	阿米巴性痢疾
日文	ハンタウイルス	漢他病毒症候群
日文	類鼻疽	類鼻疽
日文	流行性耳下腺炎	流行性腮腺炎
日文	疥癬	疥瘡感染症
日文	アカントアメーバ	棘状阿米巴
日文	水痘	水痘併發症
日文	結核	多重抗薬性結核病
日文	ブルセラ症	布氏桿菌病
日文	エンテロウイルス	腸病毒感染併發重症
日文	E 型肝炎	急性病毒性 E 型肝炎
日文	抗微生物薬耐性	VISA/VRSA 抗薬性検 測
日文	レプトスピラ症	鈎端螺旋體病
日文	インフルエンザ菌	侵襲性 b 型嗜血桿菌 感 染症
日文	猫ひっかき病	貓抓病
日文	肺炎レンサ球菌	侵襲性肺炎鏈球菌感 染 症
日文	ニューモシスチス肺炎	肺囊蟲肺炎
日文	淋病	淋病
日文	G 型肝炎	急性病毒性肝炎未定 型
日文	狂犬病	狂犬病
日文	重症急性呼吸器症候群	嚴重急性呼吸道症候 群
日文	マールブルグ熱	馬堡病毒出血熱

於建立多語系詞典後，系統便可以同時分析與檢索簡體與繁體中文的疫病輿情資料。由於本研究是利用維基百科建立同義詞典，可能會造成誤判的結果，因此在本研究中，有量測檢索的 Recall 與 Precision，作為檢視系統確度的評估標準。

為了能夠有效評估該翻譯字詞是否能協助我們抓到正確的輿情新聞，因此我們用 recall 與 precision 兩個指標來評估，以下是 recall 與 precision 的完整定義精確率(precision)的公式是 $P = TP / (TP+FP)$ ，它計算的是所有"正確被檢索的結果(TP)"占有"實際被檢索到的(TP+FP)"的比例。召回率(recall)的公式是 $R = TP / (TP+FN)$ ，它計算的是所有"正確被檢索的結果(TP)"占有"應該檢索到的結果(TP+FN)"的比例。

我們抽樣一天所抓取的 6/11 流感新聞 315 篇，發現約有 58 篇跟流感毫無關係，多數是跟那斯達克的股票新聞有關，推測是可能該文章有提到製藥公司，而該公司有製作跟流感相關的疫苗，因此才納入搜集範圍之中。以此數據評估：準確率落在：81.6%

而在評估召回率，Google News 的召回率約為 31%我們發現如果有一則新聞被重複轉載，Google 只會紀錄一篇，而且 Google 並非全面性的涵蓋所有新聞。

3. 將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文

原計畫目標

為了避免權責人員沒辦法有效解讀日、韓、西班牙、葡萄牙、越南語系以及非洲等國之資料，我們將在會將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文，並將原文及翻譯文一同匯入系統中。

本研究成果

我們透過 Google Translate 偵測抓下來的文章語系後，並透過 Google Translate 將標題跟內文翻譯成繁中與英文，讓權責人員可以根據繁中跟英文檢視外國疾管文章。

使用者可以到 InfoMiner 即時輿情分析平台中「即時搜尋」下的「國外疾管」檢視翻譯過的疾病相關輿情。



圖二十六、翻譯內容

4. 使用名詞提取功能技術抽取內文中提及的輿情地點 原計畫目標

蒐集到各國輿情後，我們將使用名詞提取功能(Named Entity Extraction)技術抽取內文中提及的輿情地點

本研究成果

蒐集到各國輿情後，我們將使用名詞提取功能(Named Entity Extraction)技術抽取內文中提及的輿情地點

我們一開始使用條件隨機域技術辨識文本中的命名實體(e.g. 人名、位置名稱等)，後來因為中研院馬教授有提供 ckiptagger，提供免費的 NER 模型，因此我們即有將既有的 NER 模型改成使用 ckiptagger 所使用的 bilstm 模型，以提高辨識的精準度。

我們將辨認出來的地點以關鍵字詞與標籤方式放置於搜尋結果之後，以便之後做地理視覺化時，可以用地點資訊呈現輿情位置。

5. 將各國疫情輿情發生地點，繪製在世界地圖上

原計畫目標

根據各國輿情發生的地點，將地點根據聲量繪製在世界地圖上，協助權責人員快速了解輿情的發生地。

本研究成果

將地理名詞從文章中提取出來後，我們即利用該地理資訊反查 Google，以取得地理的經度與緯度資訊，並將該資訊利用 leaflet 和 javascript 將資料標注在 OpenStreetMap 上，使用者可以透過該地圖，即可快速檢視國外疫情輿情的所在地。

使用者可以先進到輿情系統中，點選進階分析，找到聲量地理分析選項。



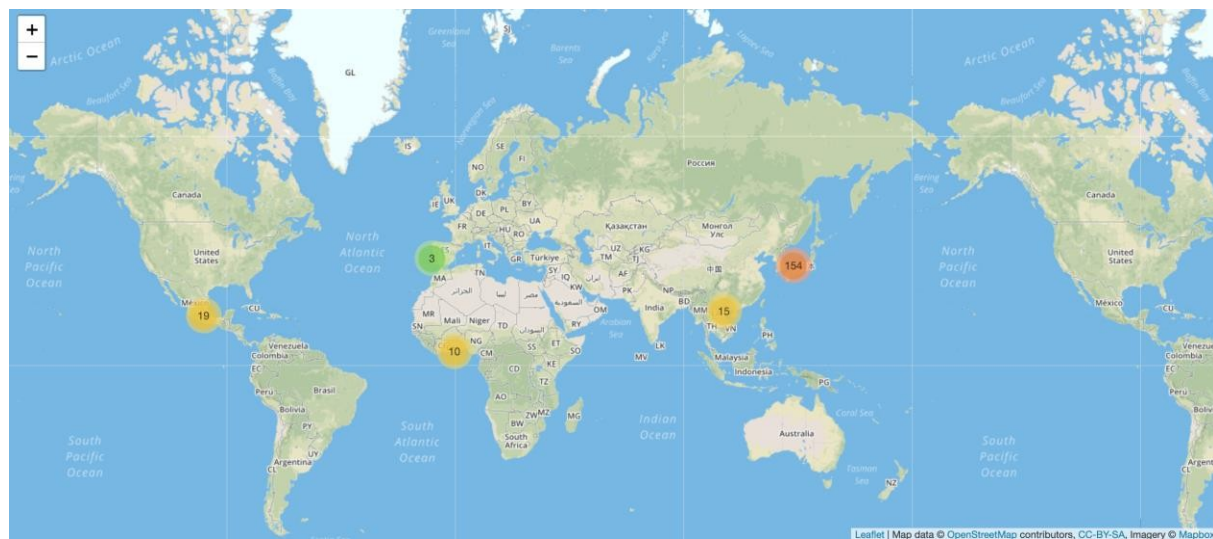
圖二十七、聲量地理分析

進到聲量地理分析後，將頻道切換至國外疾管，並將關鍵字設為流感相關字詞，即可以瀏覽到該地理區域的聲量分析。



圖二十八、檢視聲量地理分析結果

將畫面捲至下方，即可以看到相關疫情輿情的出沒地。



圖二十九、疫情輿情所在地

6. 評估三個疫病字詞之 Precision 與 Recall 以評量輿情系統之確度原計畫目標

評估三個疫病字詞在百度、Google Alert 與輿情系統搜尋之 Precision 與 Recall 以評量輿情系統之確度。

本研究成果

目前已評估流感、登革熱與屈公病等三字詞在百度、Google Alert 與輿情系統（中國疾管、中國新聞與外國疾管）搜尋之準確率與召回率以評量輿情系統之確度，精確度的評量表如下表所示：

	流感	登革熱	屈公病
中國疾管訊息 - 準確率	89.20%	97.80%	83.10%
外國疾管訊息 - 準確率	73.20%	83%	76.40%
中國疾管訊息 - 召回率	62%	76%	79.3
外國疾管訊息 - 召回率	31%	43%	35.50%

由於資料都有完整多數中國主流媒體訊息，因此在評量上，流感、登革熱與屈公病在中國疾病的召會與準確率較高。而外國疾管訊息多數用的是 Google News 的訊息，所以相關的召回率會比較低。而因為在搜索時，我們是針對翻譯成中文的資料進行查詢，但有時可能資料會翻譯錯誤，導致無法準確查找到正確的訊息，影響系統準確率。

7. Python 與深度學習教育訓練 原計畫目標

由於建立分類引擎需要牽涉許多領域知識，因此我們將搭配 Python 程式語言的進階教學，讓權責人員可以動手建立適合的模型，以因應業務上的輿情分類與標記需求。課程內容包含: Python 機器學習、Python 與深度學習、詞向量的建置與深度學習在文字上的應用。並於每次教學完後，提供作業與期末考予學員，確立學員能真正從教育訓練的過程中有所學習。

本研究成果

今年課程主軸為：Python 語言簡介、Python 語言與機器學習、Python 語言與深度學習等課程。

本課程目的在於降低初學者的學習門檻，也期望能減少專業使用者的程式撰寫時間。課程內容從 Python 軟體基本概念與資料整理/組織切入，著重資料的理解與統計/機器學習等基本觀念的建立，以整合式開發環境 Jupyter Notebook 進行實機操作。

為了能讓同仁充分吸收課程內容，上課時間規劃為每次三小時，授課日期自 108 年 4 月 19 日至 108 年 7 月 10 日。

表四、課程課綱

	課程時間	時數	課程內容大綱
Python 程式設計：機器學習			
1	108/04/19 (五) 09:00-12:00	3	Python 簡介與環境設定 Python 基本資料結構簡介
2	108/04/26 (五) 09:00-12:00	3	Python 迴圈與控制流程 Python 函數簡介
3	108/05/03 (五) 09:00-12:00	3	Python 物件導向程式設計

4	108/05/08 (三)	13:30-16:30	3	Python Numpy 簡介與操作實務
5	108/05/24 (五)	09:00-12:00	3	Python Pandas 簡介與操作實務
6	108/05/29 (三)	13:30-16:30	3	人工智慧概論 機器學習簡介
7	108/06/05 (三)	13:30-16:30	3	類神經網路簡介
8	108/06/12 (三)	13:30-16:30	3	使用 Numpy 建構類神經網路模型
9	108/06/19 (三)	13:30-16:30	3	TensorFlow 與 Keras 簡介
10	108/06/26 (三)	13:30-16:30	3	神經網路模型評估 神經網路模型調整參數 優化器簡介
11	108/06/28 (五)	09:00-12:00	3	卷積神經網路簡介與實務
12	108/07/03 (三)	13:30-16:30	3	遞歸神經網路簡介與實務
13	108/07/05 (五)	09:00-12:00	3	深度學習與文字探勘 - 訓練詞向量
14	108/07/10 (三)	13:30-16:30	3	深度學習與文字探勘 - 運用深度學習探勘文字

透過 14 天的教學內容，我們討論了如何使用 Python 處理資料、整理資料與分析資料的觀念，並懂得該如何建立深度學習模型。

於課程結束後，亦舉辦了期末考，考試題目涵蓋所有課堂上教學的內

容，並使用同仁一般在業務上會用到的資料做為測驗資料。測驗題目可連結到以下網頁：

(https://github.com/ywchiu/cdc_course/blob/master/Exam20190729.ipynb)

題目如下圖所示：

期末考題

乳癌預測資料

現有一資料集(如下表) `breast_cancer_data.csv`
(https://raw.githubusercontent.com/ywchiu/cdc_course/master/data/breast_cancer_data.csv)

欄位資料如下：

編號	變數名稱	敘述
1	id	ID
2	diagnosis	診斷
3	radius_mean	平均半徑
4	texture_mean	平均質地
5	perimeter_mean	平均周長
6	area_mean	平均面積
7	smoothness_mean	平均平滑度
8	compactness_mean	平均緊密度
9	concavity_mean	平均凹陷度
10	concave points_mean	平均凹陷值
11	symmetry_mean	平均對稱值

圖三十、期末考題

表五、期末考考題

編號	題	提	配
1	使用 Pandas 將資料讀取到一名為 df 的 DataFrame 中？	提示：使用 read_csv	(5 分)
2	請表列出該資料集的綱要(schema)？	提示：使用 info	(5 分)
3	請表列出該資料集的敘述性統計？	提示：使用 describe	(5 分)
4	請用列出資料型態為字串的欄位？	提示：dtypes: object	(5 分)
5	請移除 id 與 Unnamed: 32 兩個欄位？	提示：del or drop	(5 分)
6	請將 diagnosis 欄位的 M 轉成 1, B 轉成 0？	提示：map	(5 分)
7	請將特徵資料命名為 X, 目標(diagnosis) 資料命名為 Y?	提示：iloc	(5 分)
8	請將特徵資料標準化, 並命名為 scaled_X?	提示：StandardScalar	(5 分)
9	請將資料分為訓練資料集與測試資料集, 其中訓練資料集占 80%, 測試資料集占 20%?	提示：train_test_split	(5 分)
10	請使用 Keras 建立人工神經網路模型 (目標 y 為 diagnosis)？	提示：keras	(10 分)
11	繼第 10 題, 請計算出模型準確度(Accuracy)?	提示：accuracy	(5 分)
12	繼第 10 題, 請計算出混淆矩陣(Confusion Matrix)?	提示：confusion_matrix	(5 分)
13	請使用邏輯式迴歸(Logistic Regression) 針對特徵與目標建立一分類模型?	提示：linear_model	(10 分)
14	請使用隨機森林(Random Forest) 針對特徵與目標建立一分類模型?	提示：ensemble	(10 分)
15	請繪製類神經網路、隨機森林與邏輯式迴歸三模型 ROC Curve ?	提示：roc_curve	(10 分)
16	請比較類神經網路、隨機森林與邏輯式迴歸三模型的 AUC ?	提示：auc	(5 分)

經評量後統計考生資訊，共有 10 人參與考試，9 人通過考試。為了嘉獎通過的同仁，亦頒發了中華 R 軟體學會的訓練課程結業證書，以肯定同仁們的學習成果以及資料分析能力。



圖三十一、結業證書張樣

8. 評估 CDC 的通報資料與輿情所蒐集到的資訊與 Google Trend 做關聯分析原計畫目標

評估 CDC 的通報資料與輿情所蒐集到的資訊與 Google Trend 做關聯分析，評估系統是否能在重大疫情發生前，扮演預警的角色

本研究成果

目前已評估流感、登革熱與屈公病等三字詞在CDC 開放資料與Google Trends 及輿情新聞資料(台灣新聞)的相關係數，並將相關係數製表，放入報告中，以評估輿情系統是否能扮演預警的角色。

	輿情 v.s. CDC	輿情 v.s. Google Trends	CDC v.s. Google Trends
流感	-5.74%	71.11%	20.69%
登革熱	50.12%	81.07%	59.46%
屈公病	59.71%	73.82%	50.81%

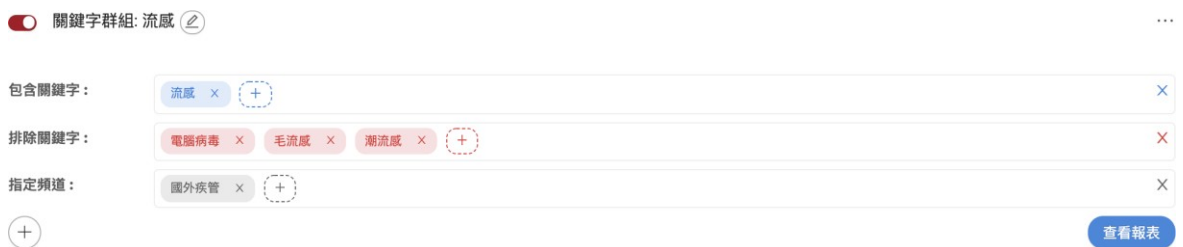
從輿情, Google Trends 與 CDC 的三數據比較，可以發現輿情跟Google Trends 的相關性相當高，代表可以使用輿情聲量揭露民眾對議題的關心程度。但在流感上，似乎相關性較低，目前尚不能根據輿情數量預測疫情是否爆發，但可以用聲量突波當作起漲指標

(四) 討論

1. 國外疫病輿情監測與分析

由於本計畫所建立的輿情分析平台收納了國外的輿情資訊，因而分析人員可以輕易使用本系統，觀測外國疫情資訊，亦可以使用該系統建立輿情分析報表，分析事件始末。以下以流感作為範例：

1. 首先先設定跟流感相關的關鍵字與頻道



關鍵字群組: 流感

包含關鍵字: 流感

排除關鍵字: 電腦病毒, 毛流感, 潮流感

指定頻道: 國外疾管

查看報表

圖三十二、設定關鍵字

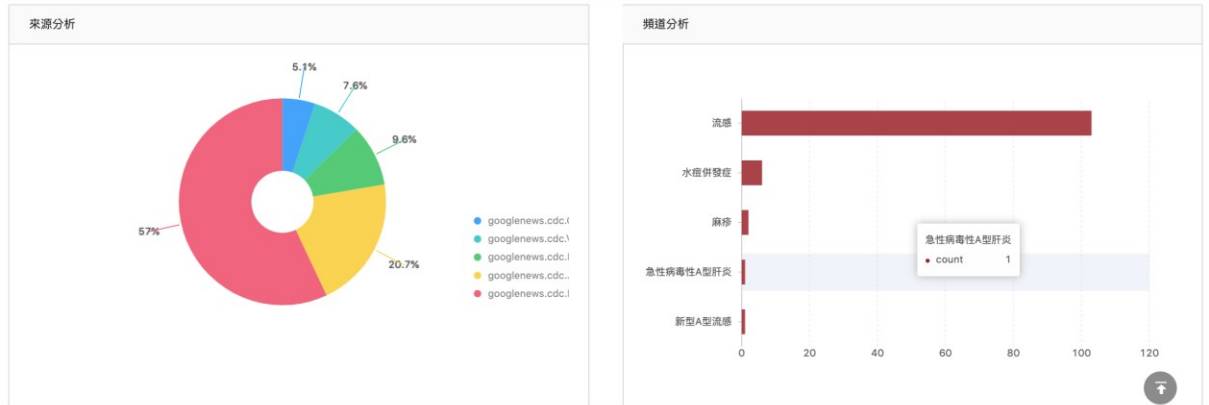
2. 點選查看報表，進到全頻道報表專區，觀看聲量



圖三十三、聲量報表

可以發現九月開始進入流感報導的高峰

3. 查看頻道分析



圖三十四、頻道報表

發現在該時段，韓國的流感疫情聲量遠大於其他國家

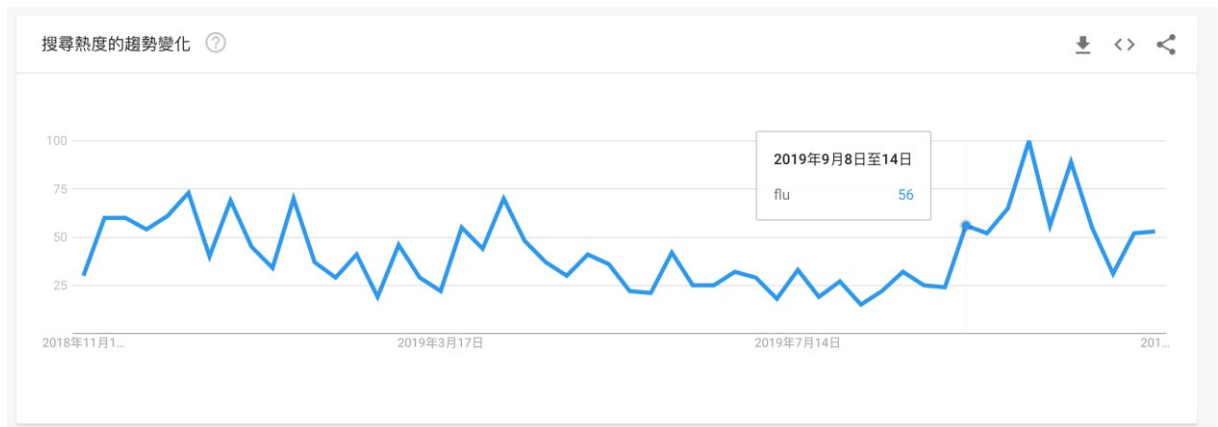
4. 比對聲量地理分析



圖三十五、聲量地理分析報表

發現韓國的疫情與情聲量明顯較為突出

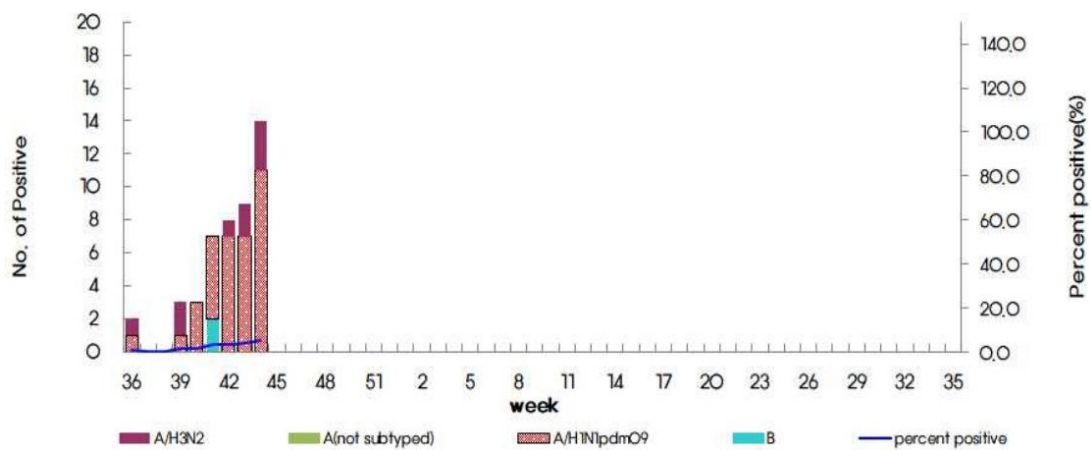
5. 抓取南韓近期的 Google Trend 上對流感的搜尋



圖三十六、Google Trend

發現相關關鍵字搜尋熱度也同時在 9 月初開始上漲

6. 對照南韓 CDC 的報告，可以發現從第 36 週，開始有一些確診病例傳出



圖三十七、韓國 CDC 報告

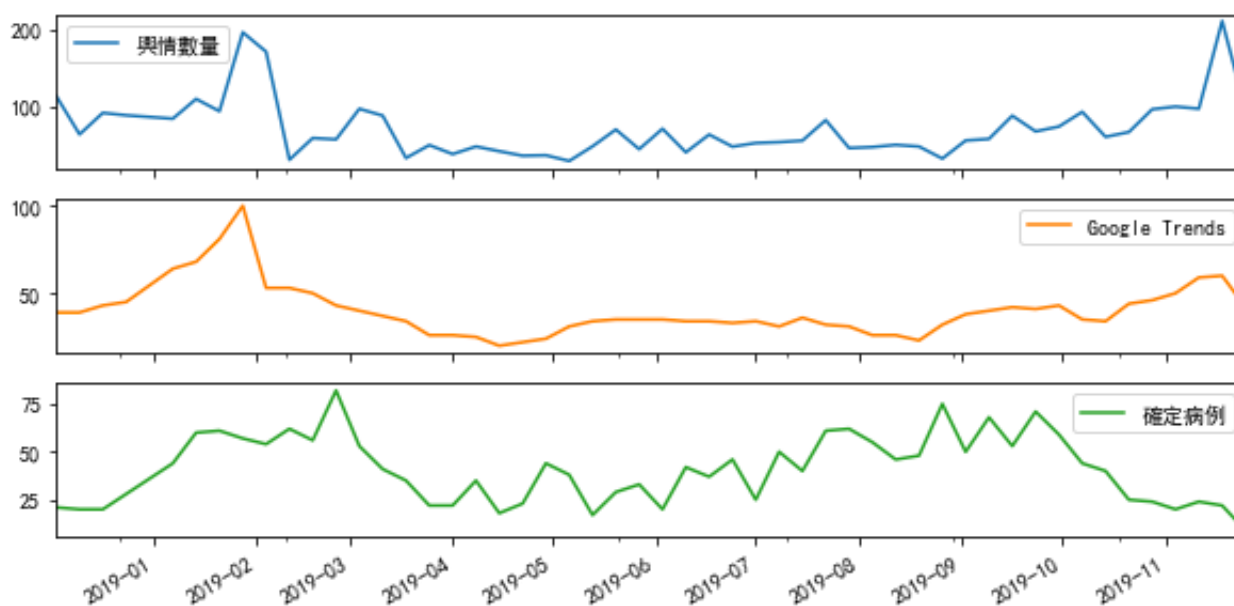
<http://www.cdc.go.kr/board.es?mid=a30504000000&bid=0033>

發現 9/13 ~ 9/20 開始也有疫情相關的統計數字。

從以上事件，可以觀測出透過輿情的蒐集，加上視覺化的呈現，我們可以掌握國外疫情輿情的脈動，進而知道該疫情是否有傳染或爆發的可能性，此時相關單位便可以作出預防行為，避免境外傳入相關傳染疾病。

2. 比較 CDC, 輿情與 Google Trends 數據 評估流感數據

為了評估 CDC 的通報資料與輿情所蒐集到的資訊與 Google Trend 做關聯分析，評估系統是否能在重大疫情發生前，扮演預警的角色。我們蒐集 2018/11/27 ~ 2019/11/27 的輿情、Google Trends 與 CDC 在流感上的數據。其中輿情數據的來源是來自 PTT, 臉書, 論壇, 部落格, 新聞與影音新聞(電視)，從數據線可以看出 Google Trends 與 輿情的走勢類似。

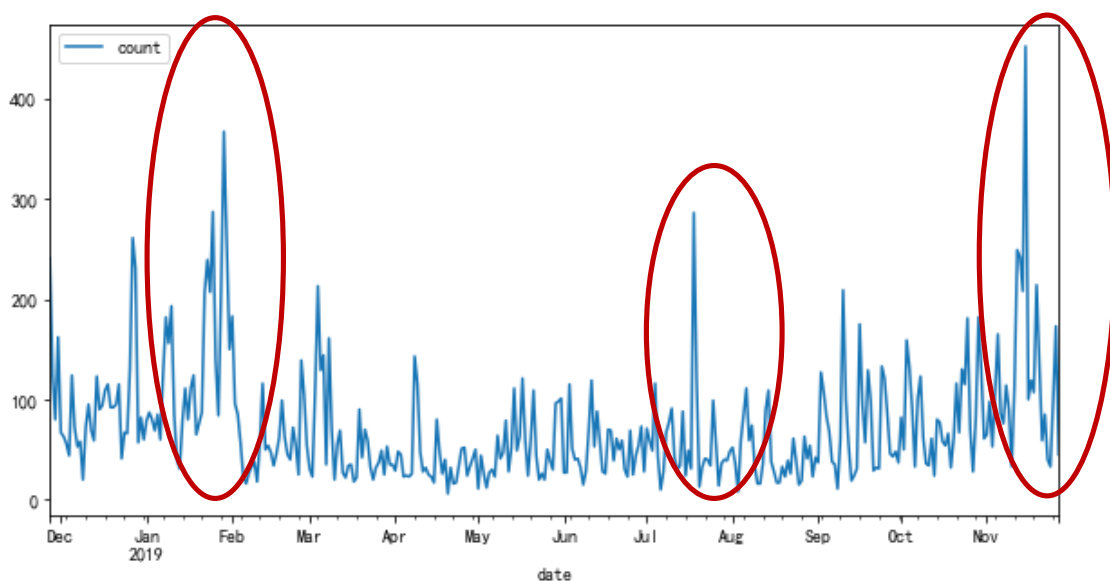


比較三者的相關係數 (Correlation Coefficient)，可以發現輿情跟 Google Trends 的相關係數相當高 (> 0.7)，但兩個數據與確定病例的相關性低，Google Trends 與輿情數量都相對確診人數為近乎無相關性

	輿情數量	Google Trends	確定病例
輿情數量	1.000000	0.711105	-0.057403
Google Trends	0.711105	1.000000	0.206947
確定病例	-0.057403	0.206947	1.000000

比較三者的相關係數 (Correlation Coefficient)，可以發現輿情跟 Google Trends 的相關係數相當高 (> 0.7)，但兩個數據與確定病例的相關性低，Google Trends 與輿情數量都相對確診人數為近乎無相關性。

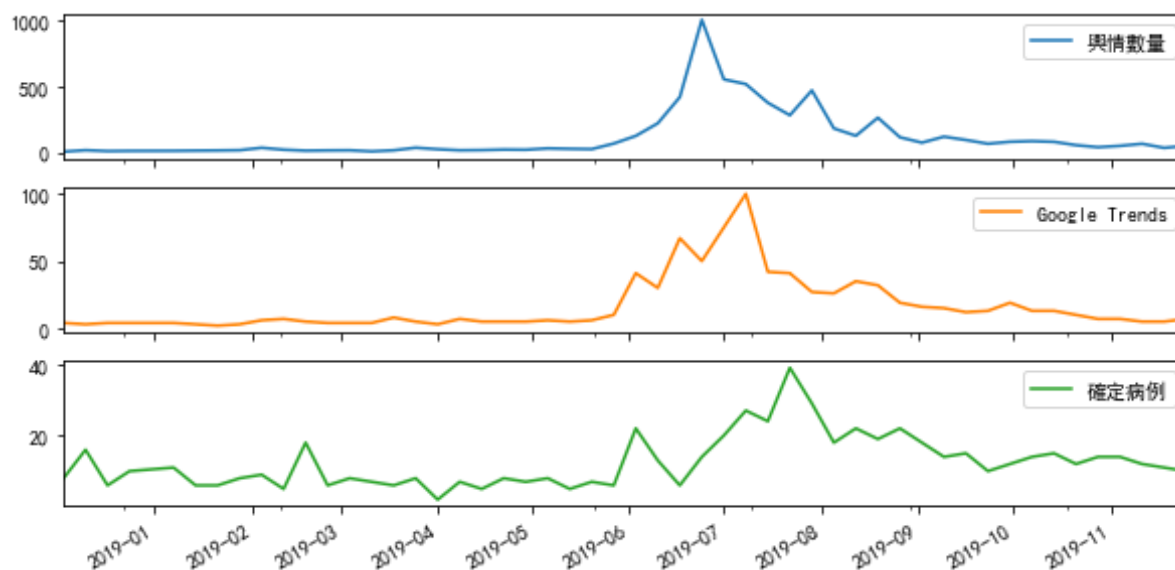
雖然輿情跟 CDC 的確診數量的相關性低，但可以發現，當輿情討論聲量達到高峰時，幾乎都為當年疫情開始爆發的起漲點，之後可以研究兩者之間是否相關，以判斷是否可以用輿情突波點來判斷流感是否有開始流行的現象。



此外比較三者的相關係數 (Correlation Coefficient)，可以發現輿情跟 Google Trends 的相關係數相當高 (> 0.7)，但兩個數據與確定病例的相關性低，Google Trends 與輿情數量都相對確診人數為近乎無相關性。

評估登革熱數據

於此列出 2018/11/27 ~ 2019/11/27 的輿情、Google Trends 與 CDC 的數據。其中輿情數據的來源是來自 PTT, 臉書, 論壇, 部落格, 新聞與影音新聞(電視), 從數據線可以看出 Google Trends、輿情與登革熱的走勢皆類似。

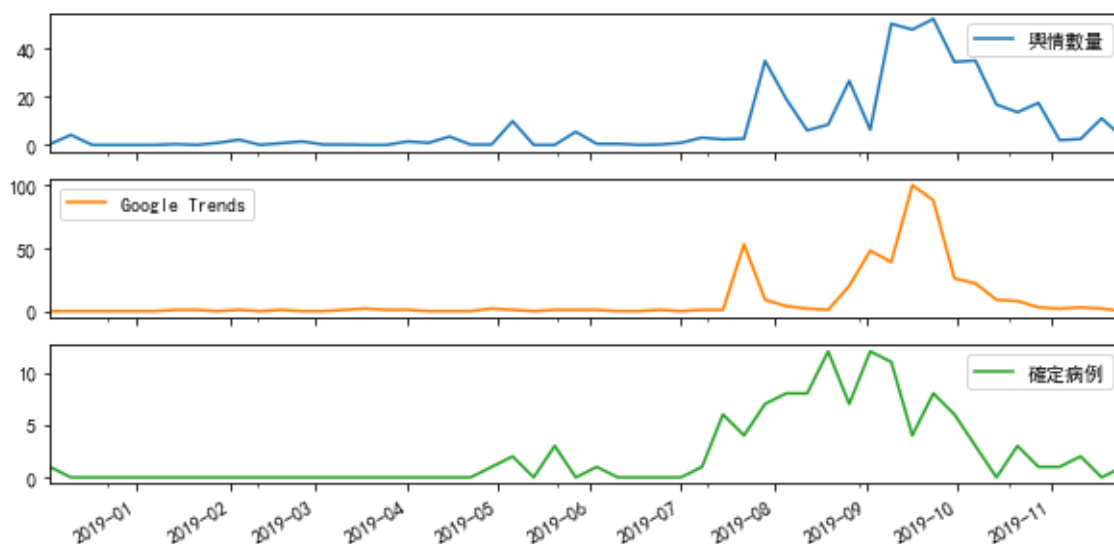


比較三者的相關係數 (Correlation Coefficient), 可以發現輿情跟 Google Trends 的相關係數相當高 (> 0.8), 但兩個數據與確定病例為中度相關。

	輿情數量	Google Trends	確定病例
輿情數量	1.000000	0.810722	0.501156
Google Trends	0.810722	1.000000	0.594553
確定病例	0.501156	0.594553	1.000000

評估屈公病數據

於此列出 2018/11/27 ~ 2019/11/27 的輿情、Google Trends 與 CDC 的數據。其中輿情數據的來源是來自 PTT, 臉書, 論壇, 部落格, 新聞與影音新聞(電視), 從數據線可以看出 Google Trends、輿情與屈公病的走勢皆類似。



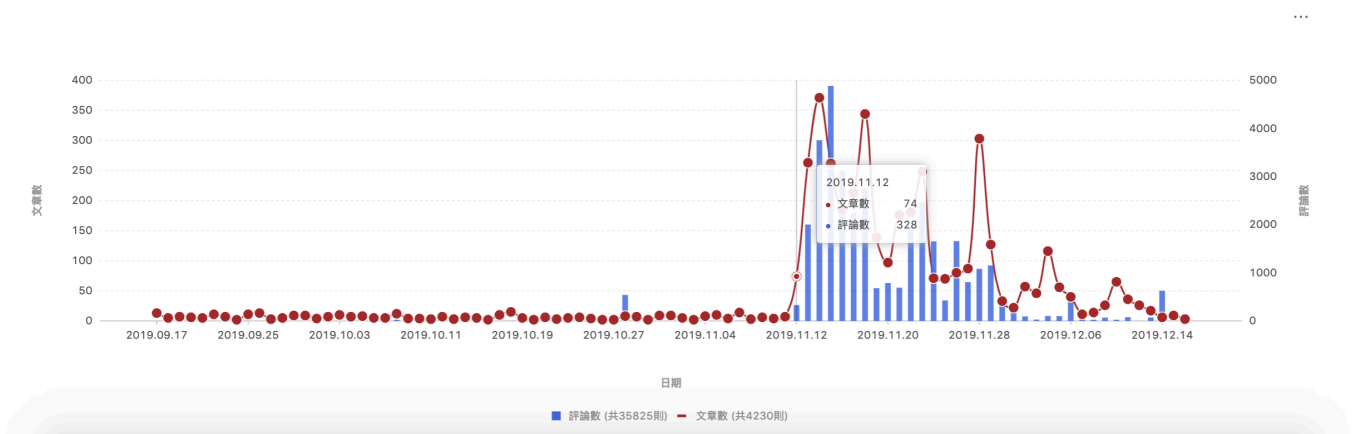
比較三者的相關係數 (Correlation Coefficient), 可以發現輿情跟 Google Trends 的相關係數相當高 (> 0.7), 兩個數據與確定病例為中度相關。

	輿情數量	Google Trends	確定病例
輿情數量	1.000000	0.738165	0.597149
Google Trends	0.738165	1.000000	0.508085
確定病例	0.597149	0.508085	1.000000

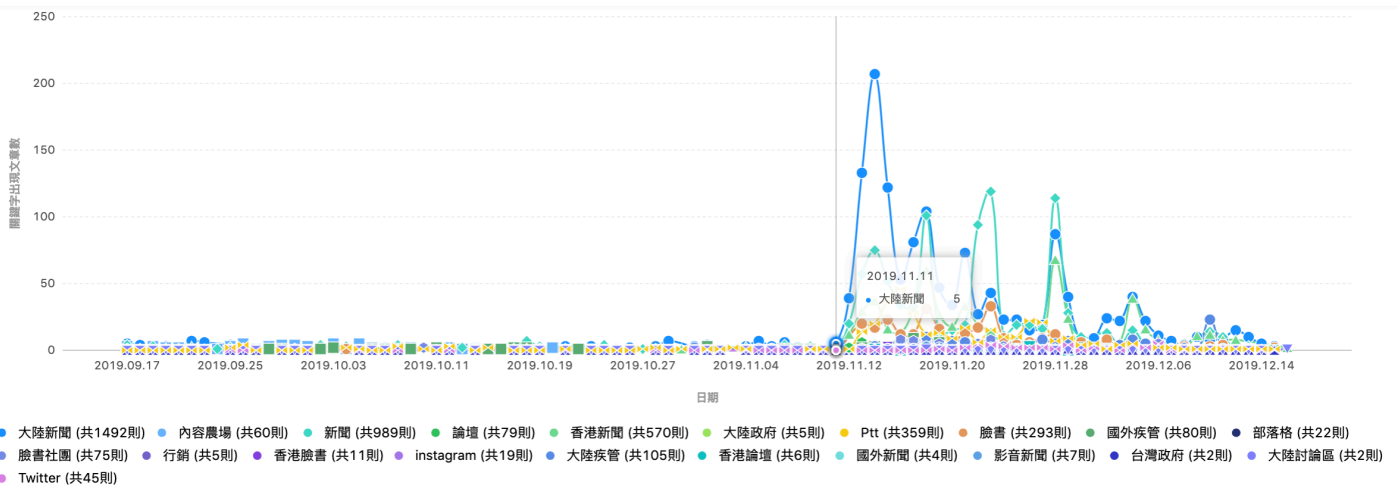
3. 鼠疫疫情分析

評估鼠疫數據

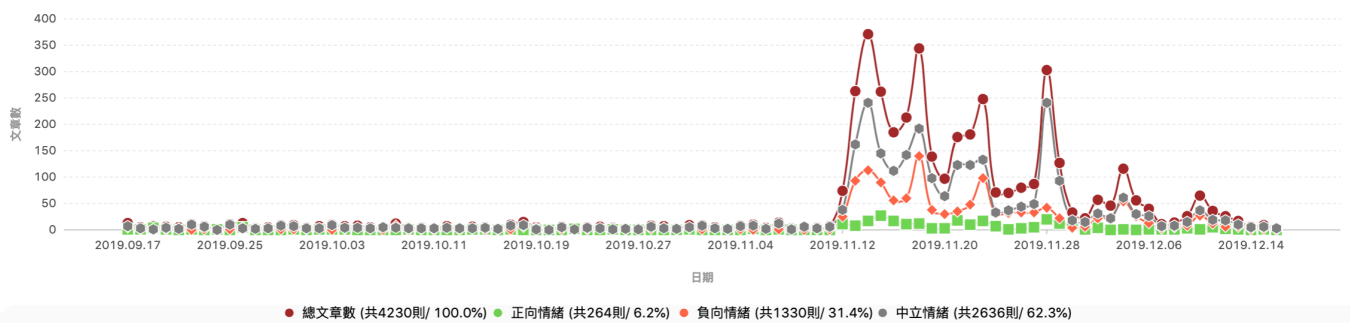
為了評估系統是否能夠確實蒐集到與鼠疫相關之訊息, 我們進入 InfoMiner 即時輿情分析系統, 鍵入鼠疫, 作為分析事件的標的。



從圖中可以發現從 11/12 號開始，鼠疫的相關討論數量開始竄升。
 但若以頻道分析來看，是以大陸新聞道討論量居多。



若以情緒分析來分析整起事件，發現多數是負面討論的事件。



若以熱詞分析來分析整起事件，發現相關討論的字詞有肺鼠疫、腺鼠疫、黑死病與內蒙古及北京還有跳蚤。

中國官方今晚證實，內蒙古有2人經診斷為「肺鼠疫」確診病例，目前患者已在北京得到妥善救治。中國媒體引述北京朝陽醫院相關負責人稱，「沒有必要恐慌，一切盡在掌控中」。北京市朝陽區政府今晚通報，11月12日，內蒙古自治區錫林郭勒盟蘇尼特左旗2人經專家會診，被診斷為肺鼠疫 確診病例。目前，患者已在北京市朝陽區相關醫療機構得到妥善救治，相關防控措施已落實。中國媒體經濟觀察網報導，有訊息指北京朝陽醫院急診科接診兩例患者，患者曾住進朝陽醫院負壓病房，其樣本經中國疾控中心檢測，12日中午確診為肺鼠疫 病例。根據有關部門的要求，所有開設發熱門診的醫療機構，重點篩查發熱門診就診患者。報導引述北京朝陽醫院相關負責人表示，「沒有必要恐慌，一切盡在掌控中」。當記者追問「一切盡在掌控中」是否意味「兩名患者已經脫離生命危險」時，上述負責人表示：「患者目前已經不在朝陽醫院，轉去了別的醫院。具體情況請等待官方正式通知。」鼠疫 是中國甲類傳染病，又稱黑死病。訊息傳出後，已在中國網路社群引發震驚與熱傳。據北京市朝陽區政府通報，鼠疫 發病急、病程短、死亡率高、傳染性強、傳播迅速。特別是敗血性鼠疫 和肺鼠疫 ，如果不加治療，病死率為30%至100%。鼠疫 潛伏期較短，一般為1至6天，但個別病例可達8至9天。通報稱，感染動物和肺鼠疫 患者，宿主動物常見有齧齒類動物和野生食肉動物，如灰旱獭、紅旱獭、喜馬拉雅旱獭和長尾黃鼠等。傳播途徑有二：一、蚤叮咬的傳播方式為鼠-蚤-人，即跳蚤叮咬病鼠後再叮咬人，或剝取染疫旱獭皮或剝食其他染疫動物，此類傳播方式常引起腺鼠疫 或敗血型鼠疫 。二、人-人傳播方式，即健康者接觸患有肺鼠疫 的病人後，經呼吸道吸入感染，此種方式感染的主要為肺鼠疫 。

[關閉](#)

若將該時間的文章做命名實體識別(NER)，並將辨認出來的位置做文字雲，可發現內蒙古被提到的次數相當的多，指出內蒙古就是鼠疫發源地。



若將地點資訊轉換成經緯度，並標注在地圖上，可以明顯發現，主要的鼠疫發源地來自於內蒙古，並且之後在北京都有陸續有相關的鼠疫輿情的發生。



4. 與 HealthMap 之比較

經檢視 HealthMap 的訊息，發現 HealthMap 整合了 GoogleNews、Baidu News, SohoInfo 等資訊，使用者可以於地圖上，設定區域與時間等條件後，檢索該地區近期的疫情輿情訊息。不過經比較，我們所蒐集的訊息。發現 HealthMap 明顯未揭露近期所有的疫情輿情訊息。

舉例來說，設定地區為台灣，時間為一星期(2019/12/09 ~ 2019/12/16)，HealthMap 只顯示在 2019/12/13 關於非洲禁運台後台灣禁止從印尼進口等新聞資訊。



而搜尋過 Google News 後，發現光是流感訊息，Google News 即有相當多相關訊息。

焦點新聞

為你推薦

我的收藏

已儲存的搜尋資訊

國際

地方新聞

商業

科學與科技

流感達高峰期 葛謨：速打疫苗

Yahoo奇摩新聞 · 3 天前



流感疫苗接種需求暴增4成 陳時中：成立調度中心加強配送

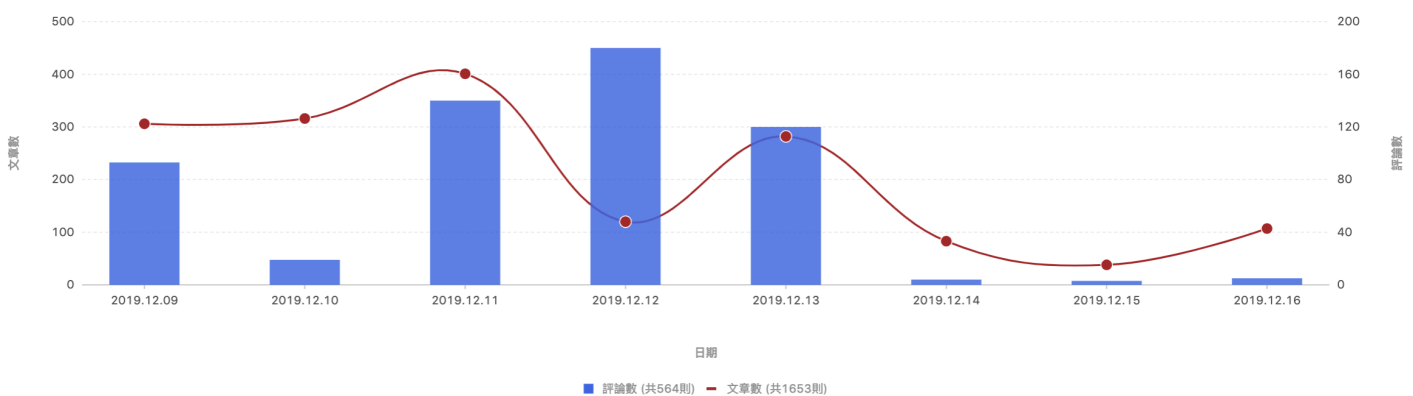
自由時報電子報 · 3 天前

流感高峰期 中醫師教你提升保護力

Yahoo奇摩新聞 · 3 天前



代表 HealthMap 可能只蒐集訊息內包含疫情等相關字眼關鍵字的新聞，而無法針對各式法傳疾病做訊息的蒐集與分析。因此對防疫需求而言，HealthMap 只能揭露全球有包含疫情相關字眼的區域資訊，而無法揭露何種傳染病在哪裡流行等資訊。而我們所建構的輿情平台，有包含各式法傳相關的搜尋關鍵字，因此單是搜尋台灣近期的流感新聞。即有 1,653 篇。若以分析而言，本計畫的輿情平台能完成較細緻的統計分析。



以下列出本輿情平台與 HealthMap 比較：

	本輿情平台	HealthMap
--	-------	-----------

資料來源(國內)	近三萬個頻道	Google News, Soho, Baidu News
資料來源(國外)	Google News, Soho, Baidu News	Google News, Soho, Baidu News
蒐集關鍵字	所有法傳相關字詞與翻譯字詞	只有疫情等關鍵字
分析功能	能產生聲量、來源分析、頻道分析、情緒、熱詞以及地理分析功能	只能利用地圖做大規模的觀測

(五) 結論與建議

以之前的研究分析的結論來看，有效利用輿情，的確可以利用其數據預測可能的疫情事件。因此如果能持續追蹤各個來源的疫情資訊，進而即時創建一個全球的流感疫情地圖，將可以達到地區性與全球性疫情預警之效。

但在蒐集輿情訊息時，往往輿情訊息帶有很多的雜訊，如何能夠過濾掉這些雜訊，只留下真正有用的數據，便會是在分析時的一大挑戰。往後如能更精進分類規則，利用建立分類模型行資料篩選，將可有望過濾掉大部分不重要的訊息，讓我們能夠建立精準的預測模式，以達警示或預測之效用。

從輿情, Google Trends 與 CDC 的三數據比較，可以發現輿情跟 Google Trends 的相關性相當高，代表可以使用輿情聲量揭露民眾對議題的關心程度。但在流感上，似乎相關性較低，目前尚不能根據輿情數量預測疫情是否爆發，但可以用聲量突波當作起漲指標。

(六) 未來里程碑與展望

在未來的計劃，將著墨於幾個重點：

1. 加廣國際疫情輿情的觀測範圍

我們將藉由增加更多的國家資料來源，以觀測更多國家的疫情輿情，並嘗試利用該輿情以推測外國的疫病對國內疫情是否在未來會有所影響。

2. 減少疫病輿情的訊息量

我們將藉由放入分類引擎，透過分類機制，來確定每個訊息與疫情的相關程度。方能讓我們可以透過系統化的方式減少疫情訊息的閱讀量，並加大疫情輿情的分析準確度。

3. 增加疫情輿情預警功能

我們藉由分析疫情輿情，觀察是否疫情輿情有上升的趨勢，藉由建立警示指標，當指標出現時，可以藉由發送警示，已達到警示之效，而為了確定指標的確信程度，我們亦將分析指標的準確度，以期之後達預警之效。

109 年計畫的主旨在於延伸前三年計畫所建立的輿情監測系統，希望能藉由增加 Google News 的來源，逐步蒐集全世界跟疫情相關之輿情，以期能透過該系統監測全世界跟疫情相關之輿情訊息。另外，為了能夠讓該系統也能夠成為疫情的預警平台，我們之後也會設定預警判斷模型，當疫情討論數超過警戒線時，系統即可自動發送相關警示。而為了驗證輿情是否能夠協助我們預判疫情的爆發，我們亦會比對疫情輿情與疾管署的數據，希冀能夠建立模型，及早發現疫情爆發的可能性。最後我們將另外開 R 語言等相關課程，藉而提升同仁的數據分析技巧。

鑒於以上考量，109 年的計畫執行項目如下所示：

1. 為了能夠監控更多國家的疫情輿情，我們將增加至少五個東協與南美國家於 Google News 出現的新聞資料，並加入至監測系統中。
2. 使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。
3. 將新增語系資料透過 Google Translate 翻譯，並將當前所檢索到的語句翻譯成中文與英文後，把原文及翻譯文一同匯入系統中。
4. 蒐集到各國輿情後，我們將使用名詞提取功能(Named Entity Extraction)技術抽取內文中提及的輿情地點
5. 根據各國輿情發生的地點，將地點根據聲量繪製在世界地圖上，協助權責人員快速了解輿情的發生地。
6. 加入 Twitter，增加社群媒體的資料來源，擴大資料蒐集的廣度
7. 建立疫情輿情的分類模型，能夠過濾掉輿情雜音，只留下跟疫情真正相關的輿情訊息。
8. 建立疫情模型預警系統，當疫情輿情超過警戒線時，我們會透過 Line 自動發送警示
9. 將疫情輿情的聲量與 CDC, Google Trends 的資料交叉比對，確立可以用過濾後的疫情輿情資訊預警疫情的爆發。
10. 由於建立分類引擎需要牽涉許多領域知識，因此我們將搭配 R 程式語言的進階教學，讓權責人員可以動手建立適合的模型，以因應業務上的輿情分類與標記需求。課程內容包含：R 語言基礎、R 機器學習。

為了能夠產出與貴署目標相符的輿情分析平台，在 109 年的計畫中，我們仍將持續另外安排時間訪談輿情相關使用者的使用經驗，以訪問內容做為該

系統改進之參考，以期署內同仁能使用系統更全面、更快速找到與工作相關之輿情熱點，並簡化輿情觀測所需之工。

(六) 重要研究成果及具體建議

根據計畫的規劃下，我們這今年完成了以下成果：

- 為了能夠監控日、韓、西班牙、葡萄牙、越南語系與非洲等國之輿情資料，我們將在抓取這些語系在 Google News 中出現的新聞資料到監測系統中。
- 使用 Wikipedia 的資訊建立一多語系同義字典，讓使用者得以檢索並歸納不同語系但同義的輿情資訊。為了確認翻譯的品質，我們亦會增加 precision 與 recall 之評量，確保字詞的翻譯品質。
- 為了避免權責人員沒辦法有效解讀日、韓、西班牙、葡萄牙、越南語系以及非洲等國之資料，我們將在會將抓取到關鍵字之資料以 Google Translate 或微軟翻譯將當前所檢索到的語句翻譯成中文與英文，並將原文及翻譯文一同匯入系統中。
- 蒐集到各國輿情後，我們將使用名詞提取功能 (Named Entity Extraction) 技術抽取內文中提及的輿情地點。
- 繳交期中報告。
- 由於建立分類引擎需要牽涉許多領域知識，因此我們將搭配 Python 程式語言的進階教學，讓權責人員可以動手建立適合的模型，以因應業務上的輿情分類與標記需求。課程內容包含: Python 機器學習、Python 與深度學習、詞向量

的建置與深度學習在文字上的應用。並於每次教學完後，提供作業與期末考予學員，確立學員能真正從教育訓練的過程中有所學習。

- 根據各國輿情發生的地點，將地點根據聲量繪製在世界地圖上，協助權責人員快速了解輿情的發生地。
- 評估三個疫病字詞在百度、Google Alert 與輿情系統搜尋 Precision 與 Recall 以評量輿情系統之確度。

本計畫在第三年的目標中，已增添許多外文輿情來源網站，並增加系統對外文的搜集、處理、分析的能力。而透案例分析，已可知道透過輿情系統與視覺化，我們便可以快速掌握全世界疫情正在爆發的地區，若能即時再過人出國前，提醒並告知防範的相關訊息，將有助於國人避免受到傳染。經這三年的開發下，雖然已經能透過系統，掌握國際性與區域性的輿情訊息，未來將增加探測的區域與分析語種，另外將會多與其他數據，例如 CDC 的通報數據或 Google Trends, Twitter 相較比較，便可以嘗試利用這些資料建立模型，以預警未來可能的疫情。

等到確認疾管署可以運用該系統，有效的探勘非文字資料，以探勘出具威脅的疫病訊息源，之後便可以結合疾管署系統與資料，發展出一全面性的疫情預警與防堵平台，讓權責人員能透過自動且系統化的方式，防堵疫情，保護國民健康。

(七) 參考文獻

- [1] Allcott H, Gentzkow M. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*. 2017.
- [2] FLETCHER, Richard, et al. Measuring the reach of “fake news” and online disinformation in Europe. Reuters Institute Factsheet, 2018.
- [3] Georgia State University. "Study uses social media, internet to forecast disease outbreaks." *ScienceDaily*. ScienceDaily, 19 January 2017.
- [4] Iafusco D, Ingenito N, Prisco F. The chatline as a communication and educational tool in adolescents with insulin-dependent diabetes: preliminary observations. *Diabetes Care*. 2000;23: 1853–1853.
- [5] McIver D, Brownstein J. Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time. *PLoS Comput Biol*. 2014;10: 1–8.
- [6] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457: 1012–1014. pmid:19020500
- [7] Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping Review on Search Queries and Social Media for Disease Surveillance: A Chronology of Innovation. *J Med Internet Res*. 2013;15: e147. pmid:23896182
- [8] PHILLIPS, Lawrence, et al. Using social media to predict the future: a systematic literature review. *arXiv preprint arXiv:1706.06134*, 2017.
- [9] ALESSA, Ali; FAEZIPOUR, Miad. A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling*, 2018, 15.1: 2.
- [10] 鄒函升, 新聞輿情與民意偵測追蹤之研究—大資料之研究取向, 資訊管理研究所, 國立政治大學, 2013, pp. 73.
- [11] 林婉茹, 利用全民健康保險資料庫建置門診類流感監測系統, 醫務管理學研究所, 長榮大學, 2009, pp. 94.
- [12] 柳姚仁, 運用 Facebook 公開資料監測類流感疫情, 資訊管理學系碩士在職專班, 淡江大學, 2015, pp. 43.
- [13] 林世然, 大數據分析應用於登革熱疫情趨勢之研究, 電機工程系博碩士班, 國立高雄應用科技大學, 2015, pp. 70.

- [14] 吳和生, 莊人祥, 張筱玲, 我國傳染病監測系統簡介, 學校衛生護理雜誌 (2010) 51-58.
- [15] K.J. Henning, What is syndromic surveillance?, *Morbidity and Mortality Weekly Report* (2004) 7-11.
- [16] R. Heffernan, F. Mostashari, D. Das, A. Karpati, M. Kulldorff, D. Weiss, Syndromic surveillance in public health practice, *New York City, Emerg Infect Dis* 10 (2004) 858-864.
- [17] L. Steiner-Sichel, J. Greenko, R. Heffernan, M. Layton, D. Weiss, Field investigations of emergency department syndromic surveillance signals—New York City, *Morbidity and Mortality Weekly Report* (2004) 184-189.
- [18] R.C. Jones, M. Liberatore, J.R. Fernandez, S.I. Gerber, Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction, *Public health reports* (2006) 133-139.
- [19] M. Kulldorff, R. Heffernan, J. Hartman, R. Assunção, F. Mostashari, A space-time permutation scan statistic for disease outbreak detection, *PLoS Med* 2 (2005) e59.
- [20] H. Chen, R.H. Chiang, V.C. Storey, *Business Intelligence and Analytics: From Big Data to Big Impact*, *MIS quarterly* 36 (2012) 1165-1188.
- [21] P. Zikopoulos, C. Eaton, *Understanding big data: Analytics for enterprise class hadoop and streaming data*, McGraw-Hill Osborne Media 2011.
- [22] 蕭元哲, 葉上葆, 電子化政府之使用行為分析: 檔案分析法, *Electronic Commerce Studies* 1 (2003) 207-224.
- [23] H. Chen, M.-T. Lin, 陳祥, 林明童, 我國 [電子化政府整合型入口網站] 使用者行為分析, *圖書館學與資訊科學* 28 (2002).
- [24] J. Caverlee, L. Liu, D. Buttler, Probe, cluster, and discover: Focused extraction of qa-pagelets from the deep web, *Data Engineering, 2004. Proceedings. 20th International Conference on, IEEE, 2004*, pp. 103-114.
- [25] Z. Nie, S. Kambhampati, A frequency-based approach for mining coverage statistics in data integration, *Data Engineering, 2004. Proceedings. 20th International Conference on, IEEE, 2004*, pp. 387-398.
- [26] O. Etzioni, The World-Wide Web: quagmire or gold mine?, *Communications of the ACM* 39 (1996) 65-68.

- [27] Y. Zhang, Y. Zhang, The Study on the Governmental Tactics of Persuasion of Network Public Sentiment, 2013 International Conference on Public Management (ICPM-2013), Atlantis Press, 2013.
- [28] R. Cooley, B. Mobasher, J. Srivastava, Data preparation for mining world wide web browsing patterns, Knowledge and information systems 1 (1999) 5-32.
- [29] 廖洲棚, 陳敦源, 蕭乃沂, 廖興中, 運用巨量資料實踐良善治理：網路民意導入政府決策分析之可行性研究, 國家發展委員會, 2014, pp. 95.
- [30] S. Shindelar, Big Data and the Government Agency, Public Manager 43 (2014) 52.
- [31] United Nations Global Pulse, Big Data for Development: Challenges & Opportunities, New York, 2012.
- [32] M.A. Pirog, Data will drive innovation in public policy and management research in the next decade, Journal of Policy Analysis and Management 33 (2014) 537-543.
- [33] W.-Y. Ma, K.-J. Chen, Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff, Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17, Association for Computational Linguistics, 2003, pp. 168-171.
- [34] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, Advances in knowledge discovery and data mining, (1996).
- [35] Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette; Smolinski, Mark S.; Brilliant, Larry (19 February 2009). "Detecting influenza epidemics using search engine query data". *Nature*. 457 (7232): 1012–1014. PMID 19020500. doi:10.1038/nature07634.