社會輿情收集機制與警學知識庫建置規劃初探

古永昌[1]*、廖學進*、古美惠**、黃澡備***
*元智大學資訊管理學系
**文化大學企業管理學系
***中央警察大學學生總隊
unicon@mail.cpu.edu.tw

摘要：警察爲維護社會治安的基石，輿論則是人們對社會現象的表述。雖然悠悠之口難杜，但諸事的發生皆可見其先兆。人們對社會現況的意見已不僅表現在其人際交往中，更進入網路無垠的世界，且更無顧忌。倘若能事先掌握社會輿論的動向，便能在其擴大而引發實際事件前先制訂因應對策，消弭事件發生的可能性。本研究從資訊管理的觀點，發展一社會輿情收集機制與探討警學知識庫建置架構，提供警察機關了解、掌握社會現況脈動，並建議透過警學知識庫的建置將過去歷史文獻或案例彙整、分析處理及應用，供作警察人員習得執法經驗、知識的參考。

關鍵詞：輿情收集、警學知識庫、開源資訊

綱目：

一、前言

警察爲維持社會治安的中堅力量，透過警察勤務部署，及社區聯防機制互助，使社會治安達於安和樂利之境。然而當勤務佈署或警民聯防機制因社會環境變遷或犯罪型態轉變而未及查覺並隨之因應時，社會治安將每況愈下，除影響人們對警察執行公權力信賴感與滿意度外，並影響警察人員士氣與警政工作推動，尤其在傳播媒體對治安事件刻意渲染下，對拾回人們信任，將是雪上加霜。因此，如何對社會治安現況進行情報偵搜，有效地回饋至治安策略上，將是警政當局應積極正視的課題。

在資訊科技及相關基礎建設漸趨完備下，各式各樣的網路平台相繼建立，並形成功能、內涵各有特色的線上社群，如電子佈告欄(bulletin board system, BBS)、網路論壇(Internet forum)、部落格(bolg)、虛擬城市(virtual world/city)、即時通訊(Internet message, IM)、線上遊戲(massively multiplayer online role-playing game, MMORPG)及拍賣或購物網站等，使人們生活不再侷限於現實社會中，更有許多事務是透過網路來完成，如購物、學習、交友、資訊分享、申辦文件或掛號等，亦有一種現實與虛擬相互交織的現象在默默地進行著，潛移默化人們的意識與行爲。Wang 等人(2007)認爲，線上社群(online communities)正逐漸地從根本上改變人們分享資訊和溝通的方式，而且正深刻地衝擊舉凡經濟、文化、人際

[1] 古員中央警察大學教務處技士，目前就讀元智大學博士班。

互動與人們生活中的每一層面(Wang et al, 2007)。個體不管在真實世界或虛擬世界中皆能與其三五好友或所屬團體(社群)進行資訊交流，然而所不同的是，在真實世界中多屬語言上的交談，並侷限在交談個體的人際網絡中；而在虛擬世界中則會留下大量的文字訊息，透過網路快速傳遞或無遠弗屆的特性，在網路中流傳，其流傳層面或所造成的影響將難以估計。由於網路爲開放環境，個體將其生活經驗、對某事物的看法意見，描述紀錄於網路社群或個人網頁中供他人分享，並亦引發他人的共鳴與討論，甚至成爲網路中重要或流傳久遠的焦點話題或事件，形成網路中一股「輿論」熱潮。

警察工作係立基於社會現況或因應未來社會發展而推動，而「輿論」則爲人們對社會現狀的表述、討論及回響。凡事件發生之前必有先兆，故在警察工作推動的同時，亦必須觀察社會輿論變化情形，順時應勢，方可竟其功。在現行的相關機制中，收集社會輿論的機制，除用於犯罪偵查之情報偵查外，一般則是透過對電視媒體的錄製觀看或由專人針對平面媒體(如報紙、雜誌或文宣等)作過濾，耗時費工。警察學爲研究社會的學問，資訊管理則是透過資訊技術適當地將實際的工作與人連結在一起，且加上資訊社會本質之故，現實中多數的訊息將大量地轉成文字化或視覺化進入資訊網路中，以及現實與虛擬世界的資訊已有揉合及同步現象，故若能從資訊管理角度發展網路訊息偵搜機制，或可對收集社會輿情提供一個快速、有效的管道，迅速掌握社會現況的脈動及有效制訂後續因應之策。

二、社會輿情與網路開源資訊

「社會輿情」是一定時期、一定範圍的民眾對社會現實的主觀反映，屬於群體性的思想、心理、情緒、意見和要求的綜合表現，亦是社會發展現況的溫度計或晴雨表。它基於社會現實，具有相對獨立性、發展、傳播或變化等特性[2]。隨著網際網路的飛速發展，網絡媒體已被公認爲是繼報紙、廣播、電視之後的「第四媒體」，網絡成爲反映社會輿情的主要載體之一。網絡環境下的輿情訊息的主要來源有：新聞評論、BBS、聊天室、部落格、聚合新聞(RSS)等線上社群，社會輿情在網路上表達快捷、訊息多元，方式互動，流傳久遠與廣泛，具備傳統媒體無法比擬的優勢。由於網路開放性與虛擬等特質，決定了網路中社會輿情的特點[3]，包括：

1. 直接性：透過線上社群，使用者可以立即發表意見，下情直接上達，民意表達更加順暢；
2. 突發性：網絡輿論的形成往往非常迅速，若存在一個熱點事件(hotspot event)，及加上情緒化意見，很容易就造成輿論嘩然；
3. 偏差性：由於發言者身分隱蔽，並且缺少規則限制和有效監督，網絡自然成爲一些網民發洩情緒的空間，並可能受到偏激用語的刺激，在網路上或現實社會中發生偏差行爲，此特點亦是在網絡上更容易出現庸俗、灰色的言論原因之一。

至於在網路中「社會輿情」的散佈或被相信程度，Charron 等人(2006)則指出，個體寧願從網路上其他個體(如網友)或線上社群中取得資訊而非透過正式機構發佈的資訊

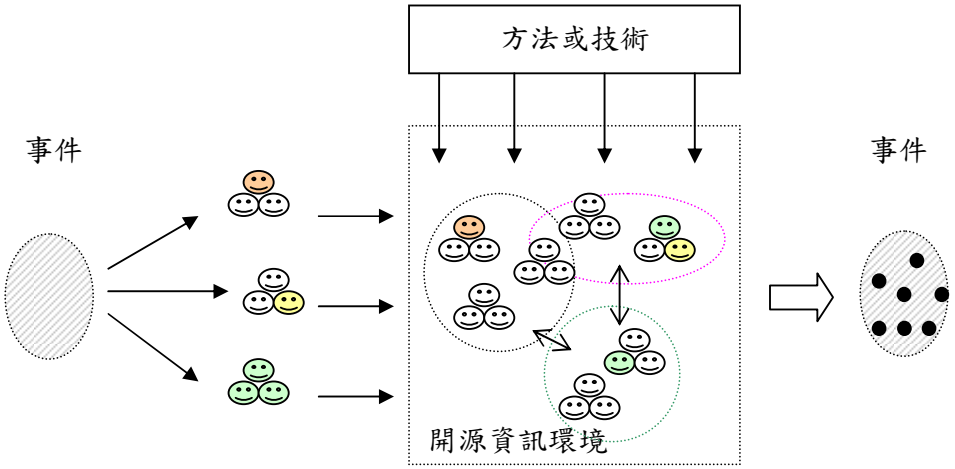
[2] http://www.arq.gov.cn/Article/Class24/Class94/200604/5016.html
[3] http://www.gmw.cn/content/2007-01/30/content_539690.htm

(Charron et al, 2006)，因此 Wang (2007)認爲應著重「開源資訊(open source information)」的收集。近年來，開源資訊變得愈來愈有價值，包括網路新聞(news sites)、討論群組(discussion boards)或聊天室(chat rooms)、部落格等，通常都提供某些事件或活動的報導與指引。如企業在網路中行銷自己時通常會介紹自己的技術能力、合作廠商等訊息，這些資訊便可以用來分析與了解企業目前的競爭實力如何，以及提供其未來願景的預測。因此「開源資訊」可以視爲「某種資訊的內涵，並用來尋找及規劃其他更有價值意涵」(Noble, 2004)。

一份美國官方的報告(Department of the Army, 2006)則對「開源資訊」作了定義並據而發展「開源情報(open source intelligence, OSINT)」：

1. 開放資源：任何人或群體提供資訊(不管是否與隱私有關)，這些資訊或關係，或兩者未受到保護而公開被揭露
2. 公開且可獲得資訊：資料、事實、指示、說明或其他公開可供一般大眾查看、或因要求而提供、可被看見或聽見，以及公佈的資訊等

因此，對於「開源資訊」，不管是基於何種動機去看待，任何人都可以透過某些方法或技術去取得及過濾分析，而「社會輿情」即爲開源資訊中的一種形式，在「開源資訊」的基礎下，透過某些方法或技術收集、分析「社會輿情」內涵應屬可行。圖一爲事件原本隱晦不明，經過人們相互傳遞、交流，形成輿論並在網路開源資訊環境中擴散，再透過某些方法或技術分析若干屬性或特徵，並經過重組後可以在程度上勾勒出及清楚化該事件的內涵的示意圖。



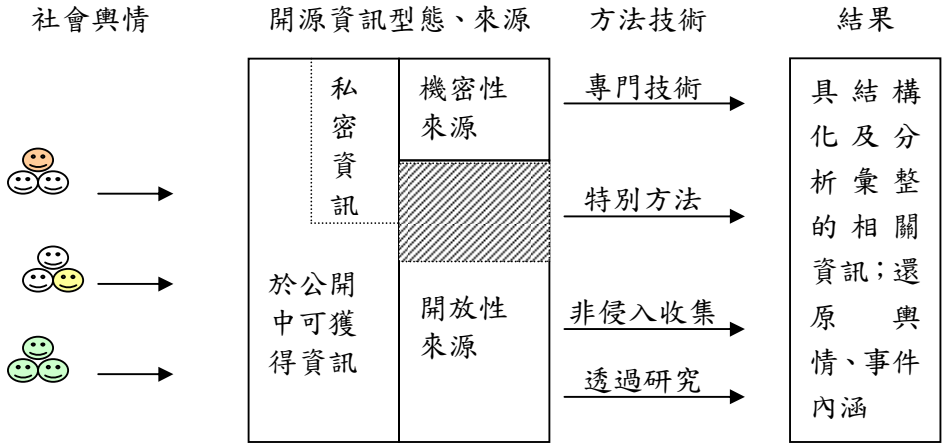
圖一：社會輿情事件揭露過程示意圖

圖二則是在開源資訊中對社會輿情分析處理架構圖，並說明如下：

1. 在開源資訊中可獲得的資料包括可公開的或私密的資訊。可公開與私密資訊中存在的「虛線」代表私密資料可能會因爲疏忽而變成可在公開環境中取得的私密資料。
2. 機密性資訊與公開性資訊中常存有模糊地帶，即圖三中灰色地帶，代表者機密性或開放性資訊的來源可能是混雜在一起的。
3. 在方法技術上，對機密性來源需要專門的技術(如駭客侵入或情報偵蒐技術等)去取得；對開放性來源一般則採取非侵入性的一般收集或透過研究分析資料而得；灰色

地帶部分則需要採取特別的方法或管道去取得(如利用ＯＯ銀行網頁設計疏失時，在該疏漏未補正前即下載銀行內存戶個人資料或 96 年 4 月 13 日發生之 FOXY 事件)。

4. 結果部分則是經過某段時間的收集與分析彙整後即可得到某事件相關資訊，並得還原該事件輪廓、內涵。



圖二：開源資訊中社會輿情處理架構圖
（修改自：Department of the Army, 2006)

三、自動化訊息收集與處理

總括來說，網路中「社會輿情」應是大量且零散分佈在不同的線上社群上。就理論上而言，不管是文字、聲音、影像或照片等多媒體形式之資訊內容均可被收集，但就技術層面而言，除了文字在收集後較易自動化分析外，其他形式即使被收集了，其自動化分析部分尚須投入大量的研究心力方有所成。故在本研究中將著重以「文字」層面的自動化收集與處理機制作說明。

(一)網頁抓取與中文斷詞

網頁抓取(Web Crawler)是一套軟體或程式，透過自動化的方法或程序，在網際網路中透過標準的 http 協定 (Http Protocol)搜尋超文字連結與接收網站文件(Cheong, 1996)，常被用來從網站上抓取網頁的資訊，如文章的標題、內容和作者等。一般來說，也常被稱爲 Web Spider 或 Web Robot，功能包括了個人化搜尋、收集網頁、備份檔案及網站統計(Chau&Chen, 2003)。因此，使用 Crawler 可以幫助研究者自動收集網路中事先設定好的標的資訊(target information)，將其傳送回來後再分門別類地儲存在資料庫中，供後續的研究與分析。標準化功能的 Crawler 客製化產品雖已問市，但由於研究標的多有不同，研究者多採自行撰寫或修改自網路下載附有原始碼的 crawler 程式，以符合研究需求。在本研究中，則是自行開發 Crawler 的程式，可以讓使用者輸入關鍵字及選擇標的資料來源，再透過入口網站搜尋並抓取相關的網頁資料後自動存入預設資料庫中。

在中文的文章裡，詞與詞之間是沒有明顯的區隔，詞可由一個以上的相鄰中文字所組成。與英文文件的相異點在於：中文文件如果沒經過中文斷詞技術前置處理，將無法拿來作後續文件分析處理，因此發展適當的中文斷詞技術是重要的。過去有許多的斷詞技術不斷的被發表出來，而最常見的中文斷詞技術主要可以分爲三大類：詞庫式斷詞(word identification)(Fan&Tsai, 1988; Sproat&Shih, 1990)、統計式斷詞法(statistical word identification)(Chen&Liu, 1992)與綜合前面兩種斷詞方法的混合式斷詞法(hybrid word identification)(Nie et al, 1996)。詞庫式斷詞法通常配合詞庫或辭典一起運作，根據一些規則逐步排除不可能的詞語組合，達到較好的斷詞結果，但由於受到詞庫品質的影響，當句子中出現新生的詞彙，將使斷詞正確性降低；若要提高斷詞正確性，則應不斷新增詞庫詞彙，如此則會大幅降低實際斷詞時的效率(Chen et al, 2000)。統計式斷詞法是基於對語料庫(corpus)語詞統計訓練，以鄰近字元同時出現的頻率(bi-gram)高低作爲斷詞依據，優點在於執行效率高，但多只能處理二字詞和單字詞。混合式斷詞法則是綜合以上兩種方法，利用詞庫找出不同組合的詞彙，再利用詞彙的統計資訊找出最佳的斷詞組合，如 Yeh 與 Lee (1991)提出以聯併(Unification)爲基礎的斷詞法，先利用詞典搜尋可能的斷詞組合，接著利用構詞規則簡化斷詞組合，再以一階馬可夫機率模式排列出所有可能的結果，然後依照機率值排列所有可能的斷詞組合，最後使用 HPSG 剖析器(Head-driven Phrase Structure Grammar Parser)逐一過濾這些斷詞組合，確認該斷詞組合是否合於文法(Yeh&Lee, 1991)。

目前國內最常使用於學術研究的中文斷詞系統則爲中央研究院開發的中文 CKIP 斷詞系統(Ma&Chen, 1992)，該系統提供中文斷詞與詞性標註的服務，使用者可以免費試用其簡易版線上分詞系統或申請帳號自行透過程式連線到該伺服器處理中文文件。該系統包含一個約 10 萬詞的詞彙庫，每週固定更新與維護資料庫，因此本研究選定該斷詞系統作爲中文文件前置處理的斷詞工具。

(二)文本特徵計算與相似度比對

1.特徵詞萃取

一般而言，透過web crawler自網路上抓取下來的網頁，可用「文本(texture)」稱之。文本須經前述步驟，從所有文章中探勘出具有代表性之輿情特徵詞彙，接著再判斷其特徵詞彙是否具有代表性或是鑑別力。因此，透過CKIP中文斷詞系統自動標記斷詞後的詞彙特性，進行詞性合併的動作，以擷取出具有意義之特徵詞彙。如Liao et al則利用CKIP進行線上社群情緒意見的探勘，並以下列規則作分析，如

(1) VH(狀態不及物動詞)

例如：安全(VH)

(2)D(副詞) + VH(狀態不及物動詞)

例如：顯然(D)　失職(VH)

(3) VH(狀態不及物動詞) + VA(動作不及物動詞)

例如：嚴厲(VH)　違約(VA)

2.字詞權重計算

在擷取文本特徵詞並存入資料庫後，接著則必須計算該特徵詞權重，一般最常使用的爲Salton與Gill(1983)所提出的TF*IDF(Term Frequency* Inverse Document Frequency)，其計算方式如下：

$$W_{ij} = t_j f * \log(id_i f) \tag{1}$$

說明：

d_i ：第 i 篇文件

$d_i = (W_{i0}, W_{i1}, W_{i2}, \cdots, W_{ij})$

W_{ij} ：代表特徵詞 t_i 出現在文件 d_i 之權重

$t_j f$ ：詞彙 $t_j f$ 在文件 d_i 中出現的次數

$id_i f$ ：N/df_i (所有文件／詞彙 t_j 在所有文件中出現過的次數)

TF (Term Frequency)：表示詞彙出現的頻率。代表某一詞彙出現在文件或資訊內容的相對頻率，用以測量該詞彙在文件中的相對重要性，出現頻率愈高愈能代表該詞彙為關鍵字；

IDF (Inverse Document Frequency)：以詞彙出現在其他文件的多寡來衡量詞彙的重要性，當詞彙出現在許多文件中，則表示其代表性低。反之，則詞彙具有代表性。

3.文本相似度比對與分群

向量空間模型(Vector Space Model, VSM)最早是由Salton與Gill(1983)所提出的，向量空間模型是以文本向量爲基礎，而建立詞彙－文件矩陣 (Term-Document Matrix)爲其核心，可以利用來計算文章之間向量的相似度來進行群聚 (Cluster) 或分類 (Classification)的處理，它亦是以文件索引向量(index vector)與關鍵詞的重要性(term significances)的計算當作評估索引正確與否的權重參考依據。故在本研究以計算過的特徵詞權重當作向量空間模型的基底，然後計算文本的相似度，找出有高度關聯之文本，並據以進行分群，以後到相類似文件彙聚後的內涵，此種文本彙聚所得出的內涵亦是了解該類文本所描述輿論或事件方向。詞彙－文件矩陣如下所示(Salton&Gill, 1983)。

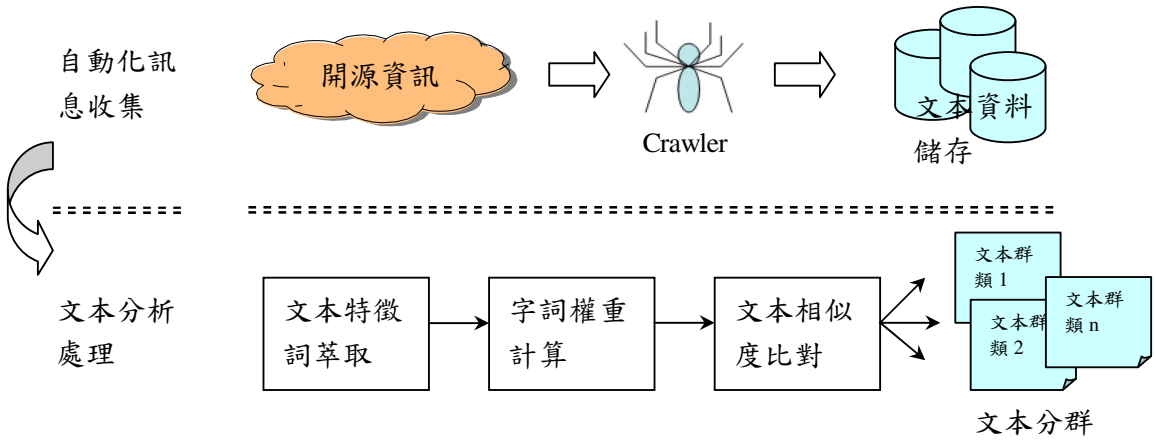
$$\begin{array}{c} \\ Doc_1 \\ Doc_2 \\ \vdots \\ \vdots \\ Doc_m \end{array} \begin{array}{c} \begin{array}{cccccc} Term_1 & Term_2 & Term_3 & \cdots & \cdots & Term_n \end{array} \\ \begin{bmatrix} W_{11} & W_{12} & W_{13} & \cdots & \cdots & W_{1n} \\ W_{21} & W_{22} & W_{23} & \cdots & \cdots & W_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_{m1} & W_{m2} & W_{m3} & \cdots & \cdots & W_{mn} \end{bmatrix} \end{array}$$

文本相似度計算在資訊處理的相關研究是最常被使用的技術，舉凡文件檢索、分群和分類等。表一列出常用的相似度的計算公式(Salton, 1988)。

表一：常見相似度測量方去的計算公式

相似度測量方法 $Sim(X,Y)$	計算字詞向量	計算字詞權重
Inner product	$\lvert X \cap Y \rvert$	$\sum_{i=1}^{t} X_i \times Y_i$
Dice coefficient	$2\frac{\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert}$	$\frac{2\sum_{i=1}^{t} X_i \times Y_i}{\sum_{i=1}^{t} X_i^{\,2} + \sum_{i=1}^{t} Y_i^{\,2}}$
Cosine coefficient	$\frac{\lvert X \cap Y \rvert}{\lvert X \rvert^{\frac{1}{2}} \bullet \lvert Y \rvert^{\frac{1}{2}}}$	$\frac{\sum_{i=1}^{t} X_i \times Y_i}{\sqrt{\sum_{i=1}^{t} X_i^{\,2} \bullet \sum_{i=1}^{t} Y_i^{\,2}}}$
Jaccard coefficient	$\frac{\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert - \lvert X \cap Y \rvert}$	$\frac{\sum_{i=1}^{t} X_i \times Y_i}{\sum_{i=1}^{t} X_i^{\,2} + \sum_{i=1}^{t} Y_i^{\,2} - \sum_{i=1}^{t} X_i Y_i}$

下圖三為自動化訊息收集與分析處理示意圖，其各階段處理方式則如上所述。



圖三：自動化訊息收集與分析處理示意圖

四、警學知識庫建置規劃

警學爲研究社會現象，了解時代脈動的學問，社會在流行什麼？民眾在議論什麼？或治安在民眾心中感受如何？均是警學應觸及的範疇。由於社會現象層面廣泛，無法面面俱到；警察工作亦有其既訂目標，力有未逮，故發展自動化社會輿論收集機制，與建立彈性化警學知識庫(曾榮汾，2006)以爲運作基礎，在快速掌握社會輿情，添補警力不足的需求下實有其必要性。

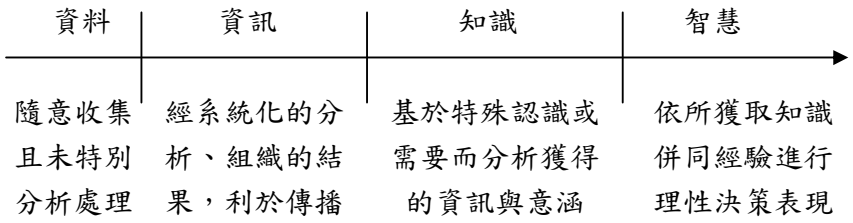
從資訊管理的角度來看，知識庫(knowledge base 或 knowledgebase, 簡稱 KB)是一基於知識管理與特別需求(special need)目的的資料庫，它能提供了電腦化的收集、組織與對知識的萃取過程4。在說明警學知識庫建置規劃之前，首先必須先對資料(data)、資訊(information)、知識(knowledge)與智慧(wisdom)作說明，圖四爲人們對一事實分析處理及理解過程示意圖：

1.資料：其形式往往爲對某一事實的描述、調查數據，如號碼、文字、圖像、聲音等，是一種未經詳細分析、處理且適合於儲存於或透過電腦處理的資料

2.資訊：爲一確切的知識或對某人或某事的分析處理結果，並可用於對事實或認知事物的傳遞，係在取得資料後作進一步有系統、有組織且具有意義的分析結果，並有利於事實或知識的傳播，如電話簿

3.知識：一般爲具有特殊認識或因需要而擁有的資訊、事實、思想真理或原則，能夠清楚明確地表達某情況或事實的資訊，並能透過熟悉、認識或經驗上的理解而了解所有資料、事實真相在時間或空間上的意涵

4.智慧：對所獲取的知識或經驗，作出明智的決定和判斷與理性行爲的決策表現；它必須積累知識或生活在某一活動領域中取得的經驗，綜合集成後所爲的決策結果



圖四：人們對一事實分析處理及理解過程示意圖

警學知識庫爲一基於社會治安需要，了解社會現狀及分析社會輿情等目的而建立的資料庫管理系統，並可分作二個層面作設計，其一爲此知識庫的核心功能；另一則爲涵括的層面。首先先說明此知識庫的核心功能至少應包括：

1.**社會焦點話題或敏感話題識別**：根據訊息來源、發言評論數量、時間及密集程度等參數，識別一定時間範圍內的熱門話題，並利字詞權重及語義分析技術，辨別出敏感或焦點話題，了解輿論現況；

4 http://en.wikipedia.org/wiki/Knowledge_base

2.**輿情主題追蹤**：對已分析的特定主題在文本自動收集(crawler)時作自動追縱，收集相同輿情主題的後續文本，並對正在分析的文本，經分群技術後依權重設定主題；

3.**輿論傾向性及趨勢分析**：對每個話題、發表文本的觀點，透過文本分群技術，進行類別或傾向性分析、統計，勾勒出話題範疇與輪廓，同時亦須分析某個主題在不同的時間段內，人們所關注的程度及趨勢分析；

4.**自動摘要**：對各類主題、輿情傾向及發展現況作定時、定期自動摘，提供有關單位分析報表；

5.**突發事件分析**：對突發事件或瞬時間有大量言論彙集的事件進行跨時間、跨空間綜合分析，獲知事件發生的全貌並事先因應事件可能發生的變化；

6.**警示系統**：對突發事件、涉及內容安全的敏感話題及時發現並作警示；

7.**統計報表與綜合瀏覽**：根據輿情分析的結果產生例行報表，供相關需求單位作綜合瀏覽與關聯檢索。

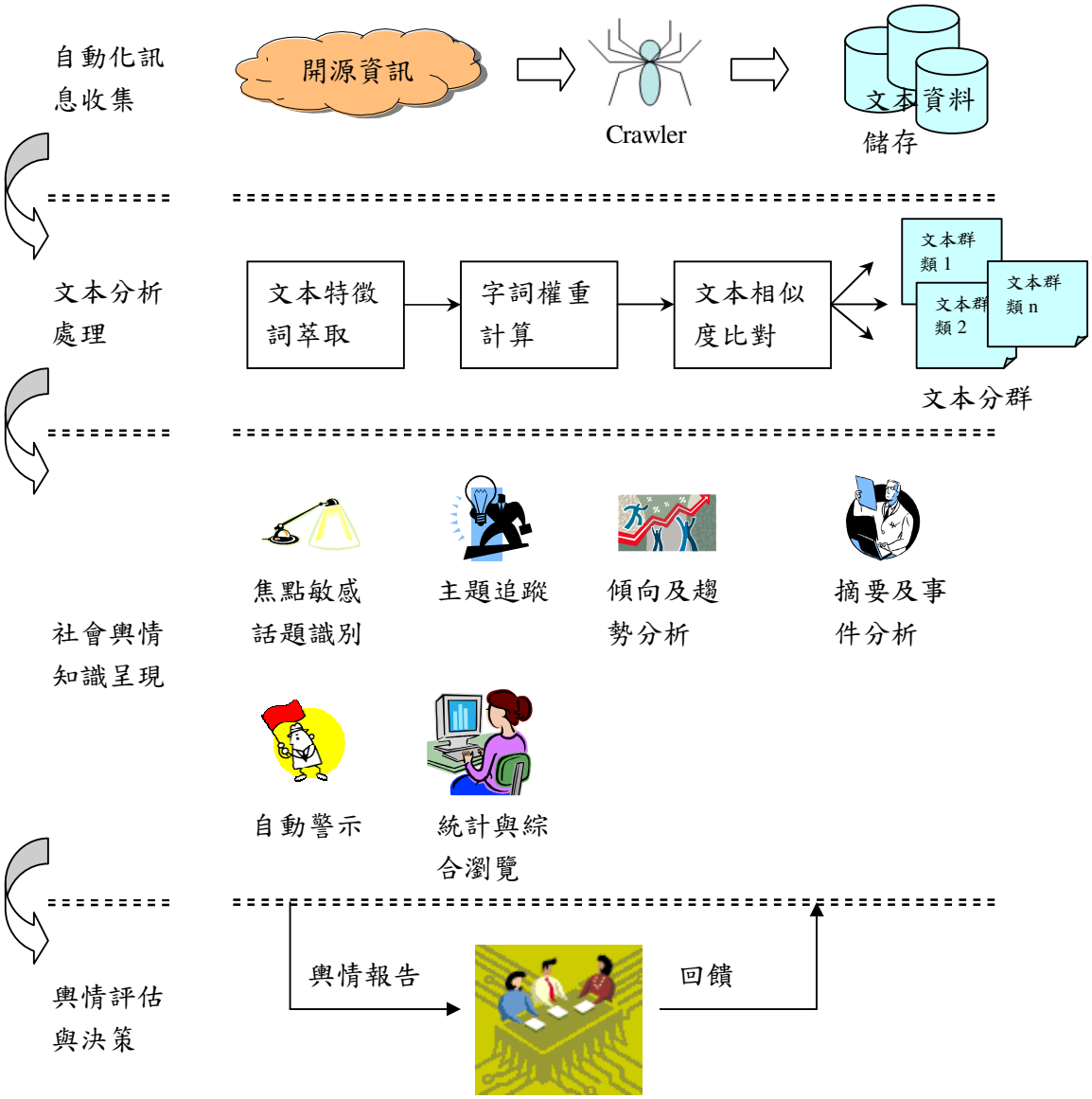
在前述就「開源資訊」的說明中，即闡明「開源資訊」的基本特質爲「開放的環境中」及「可公開取得」，且當開源的資訊經過沉澱、累積與修正則成爲書本及歷史材類；經過學習後則成爲人們的知識與經驗，並表現在人們的行爲上，故在警學知識庫的第二個層面上，應該過去的歷史文獻或案例資料、警察人員的執法經驗、判斷，以及社會輿情即時現象的判斷、評估作思考。圖五爲輿情資訊處理流程架構雛型，圖六爲警學知識庫主要功能模組，並在設計上應包括下列子系統：

1.**歷史文獻彙整**：基於鑑古知今的概念與學習參考，不同時代有不同時代的警學重要文獻，這些文獻均是當代在社會治安及偵查案件的重要知識，雖時代不同或有作法不同，但其中的精神當有參考、學習之處。且在前述自動化收集技術的支援下，只要能夠將過去的歷史文獻作數位化，即可透過本文所提的機制架構進入警學知識庫，供作綜合應用。

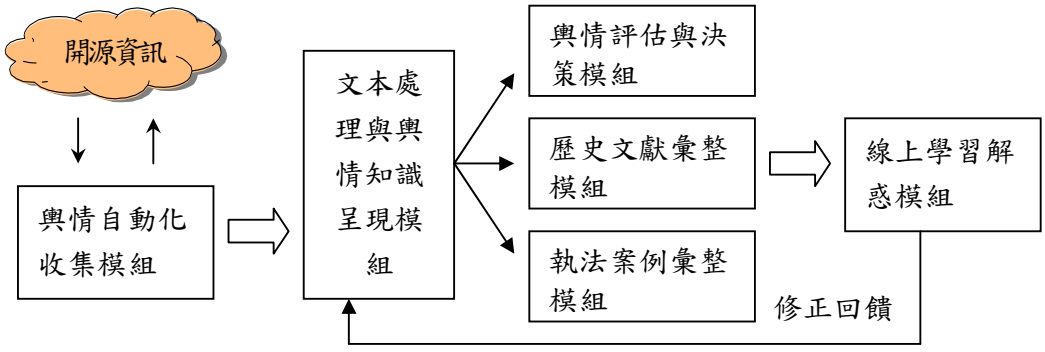
2.**執法案例彙集**：經驗是需要被傳承的，尤其是在隨著社會治安變遷即有不同的偵查執行方法。如過去詐騙案件是以「人」的媒介，必須清查人際關係，而現在則變成以「電話」作媒介，必須清查通聯紀錄，在可預見的未來則會改變成以「網路」作媒介。隨著時間的推演，執法經驗在學理指引及實務驗證操作下，從不熟悉變成嫻熟，其過程需要數年，但社會變遷快速，新的警察人員在學習上無法參酌過去案例，容易形成缺漏或在執法上稍顯不嚴謹與生疏，資深的警察人員又疲於因應接踵而至的案件，無法有效地傳承其經驗，故透過執法案例，將偵辦經驗紀錄至警學知識庫，形成執法案例知識是必要的，將可透過警學知識庫發展學習環境，供新進人員教育訓練時的參考。

3.**輿情即時判斷評估**：收集分析社會輿情的目的在於透過資訊技術與資訊管理了解社會現況及輿論焦點，供治安決策參考，但事實上，所收集分析得來的輿論並非全與治安有關，故應存在一判斷評估機制作過濾，且此判斷評估結果亦可回饋至系統中供作修改參考。

4.**線上疑難解惑**：知識的目的在於能夠解決疑難，故透過前三項的功能，提供線上即時的疑難解惑，提供警察人員在執行時的參考方案。



圖五：輿情資訊處理流程架構雛型



圖六：警學知識庫主要功能模組與分析

五、結論

警察爲維護社會治安的基石，而輿論則是人們對社會現象的表述，警察工作的思維已漸由過去著重在偵查面，而轉爲預防面。須知悠悠之口難杜，諸事的發生皆可見其先兆，人們對社會現況的聲音已不僅表現在其人際交往中，更進入網路無垠的世界，且更無顧忌，倘若能事先掌握社會輿論的動向，便能在其擴大而引發實際事件前先制訂因應之對策，消彌事件發生的可能性。

在開源資訊的概念上，任何好的、壞的訊息都能在網路中尋得，但其關鍵僅在於技術能力能否滿足，在警察工作繁重之下，再要求警察人員顧及社會輿論，無疑是萬鈞添石，更加重其工作負荷，甚而影響工作士氣。故若能在資訊管理觀念與技術的協助下，發展一可行且自動化收集處理至輿論知識呈現，並結合歷史文獻與執法案例作爲知識、經驗傳承之警學知識庫實屬必要，本文所勾勒之功能、內涵或可作爲系統發展參考。

在未來研究方面，本文所提之功能、內涵，僅爲初步概念，其細節尚有許多知識與技術須再探究，如發展快速有效的訊息收集方法或工具、文本斷詞、斷句及權重計算的精確性、輿論知識呈現的方法與理論建構、案例的實際操作與數位化技術，以及系統發展完成後之教育訓練等問題均須依賴大量的研究心力投入，故期望未來能有更多的研究人力加入警學知識庫的研究範疇，以完善功能並促進警察工作效率，共同維持良善、和諧的社會。

參考文獻

1. 曾榮汾，「整理《折獄龜鑑》的觀念析述」，中央警察大學第四屆通識教育學術研討會論文集，2006.
2. Department of the Army, *Open Source Intelligence*, Official Report, 2006.
3. Charron, C., Favier, J., and Li, C., "Social Computing: How Networks Erode Institutional Power, and What to Do about It," *Forrester Customer Report*, Feb., 2006.
4. Chau, M. and Chen, H. C., "Personalized and Focused Web Spiders," *Web Intelligence*, pp. 197-217, 2003.
5. Chen, J. S., Hsieh, C. L., and Hsu, F. C., "A Study on Chinese Word Segmentation Genetic Algorithms Approach," *Information Management Research*, Vol. 2, No. 2, pp. 27-44, 2000.
6. Chen, K. J. and Liu, S. H., "Word Identification for Mandarin Chinese Sentences," *Proceeding of COLING-92, 14th Int. Conf. On Computational Linguistics*, pp. 101-107, 1992.
7. Cheong, F. C., "Internet Agents: Spiders, Wanderers, Brokers, and Bots," *New Riders Publishing*, Indianapolis, Indiana, USA, 1996.

8. Fan, C. K., and Tsai, W. H., "Automatic Word Identification in Chinese Sentence by the Relaxation Technique," *Computer Processing of Chinese and Oriental Languages*, Vol. 2, No. 4, pp. 33-56, 1988
9. Liao, X. J., Ku, Y., Ku, A., and Chiu, C., "Sentiment Mining of Virtual Community," *The 18th International Conference on Information Management (ICIM2007)*, 2007. (accepted)
10. Ma, W. Y. and Chen, K. J., "Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff," *URL: http://dats.ndap.org.tw/docu/achievements/paper-92/92-10.pdf*, 1992.
11. Nie, J. Y., Brisebois, M. and Ren, X., "On Chinese Text Retrieval," *Proceeding of SIGIR*, pp. 225-233, 1996.
12. Noble, D. F., "Assessing the Reliability of Open Source Information," *Proceedings of the 7th International Conference on Information Fusion*, 2004.
13. Salton, G. and Gill, M., "Introduction to Modern Information Retrieval," McGraw-Hill, New York, USA, 1983.
14. Salton, G., "Automatic Text Processing," Addison-Wesley Publishing Company, Boston, MA, USA, 1988.
15. Sproat, R. and Shih, C. W., "Statistical Method for Finding Word Boundaries in Chinese Text," *Computer Processing of Chinese and Oriental Languages*, Vol. 4, No. 4, pp. 336-351, 1990.
16. Wang, F., "Open Source Intelligence and National Security in Internet Age," *299th Xiang-Shan Science Conferences*, pp. 33-35, 2007.
17. Wang, F., Carley, K. M., Zeng, D., and Mao, W., "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligence Systems*, Vol. 22, No. 2, pp. 79-83, 2007.
18. Yeh C. L. and Lee, H. J., "Rule-Based Word Identification for Mandarin Chinese Sentences-A Unification Approach," *Computer Processing of Chinese and Oriental Languages*, Vol. 5, No. 2, pp. 97-118, 1991.