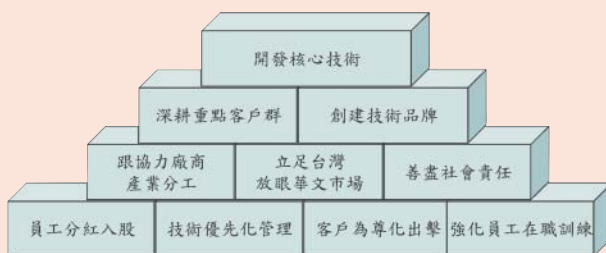


國立台灣師範大學／ 威知資訊股份有限公司

關聯詞網及其視覺化技術研發計畫

公司小檔案

- ★成立日期：民國 91 年 12 月 30 日
- ★負責人：鄭浩
- ★資本額：新台幣 27,500,000 元
- ★員工人數：50 人
- ★經營理念：



計畫緣起

在知識探勘的過程中，「詞彙關聯性」代表著資料庫或文件庫中，每一句子、段落，所記載的詞彙之間的關係，如因果關係 (causality)、共伴關係 (coupling) 以及主題事件上的關聯 (event relatedness) 等知識或情報。而一套可自動解析詞彙間相關性且可將其回饋予使用者的機制，除了可幫助使用者迅速過濾不相關的資訊，亦可有效協助使用者在龐大的資料中，針對有相關意義的訊息或規律，發掘有用的資訊或知識。因此，為了能針對大量的電子資料進行高效率的知識探勘，詞彙關聯性的相關應用研發，一直都是知識探勘領域中相當熱門的議題。

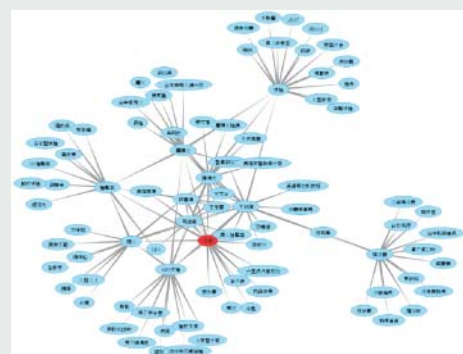
本研發團隊於本計畫中除了透過關鍵詞擷取、共現性分析 (co-occurrence analysis) 等文字探勘技術，研發一套效率高、應用範圍廣泛的詞彙關聯性計算方法之外，更將研發一套網絡圖形視覺化的方式來展現詞彙的關聯性，亦即研發一套互動高的關聯詞網，來簡單呈現複雜的網絡關係，以方便使用者進行瀏覽、探索其中隱而未顯、潛在有用的情報與知識，供使用者進行後續應用、決策的重要參考依據。

新產品簡介

本計畫技術將成為第 2 代 KSP 之子系統，名為「關聯脈絡圖子系統」，其系統功能規格如下所示：

1. 可針對各式資料來源進行分析、整合，並提供檢索機制供使用者進行查詢。其資料來源可包括：
(1) 一般電子文件 (如：doc、.ppt、.xls、.pdf、.xml、.txt...等)。

- (2) 壓縮檔案 (如：rar、.zip...等)。
 - (3) 網頁文件 (如：html、.jsp、.php...等)。
 - (4) 資料庫 (如：MS SQL Server、My SQL、Oracle...等)。
 - (5) 群組軟體 (如：Lotus Notes、QuickPlace)。
 - (6) 電子郵件 (如：eml、Notes Mail...等)
2. 針對特定文件集合 (檢索結果的文件集，或專案專卷中的類別文件) 進行關鍵詞擷取、關聯詞分析，以及關聯關係的建構，並提供使用者一個門檻值的調整空間。
 3. 可由使用者動態挑選欲展開之圖形的中心詞彙。
 4. 以一個網絡圖形來呈現詞彙之間的關聯性，供使用者對宏觀的資料概念進行了解。如下示意圖所示，藉由使用者輸入之查詢詞為中心點 (紅色)，將與其相關之關聯詞彙以放射狀方式呈現，各詞彙間以細線為關聯鏈結，透過多層次關聯呈現方式，以瞭解各關聯詞彙之上下層關係。



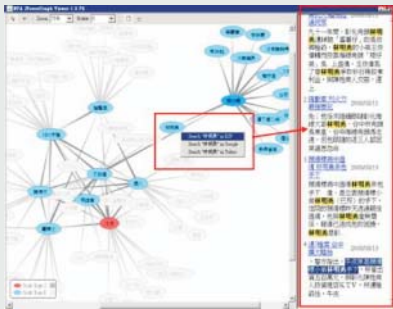
5. 使用者可點選有興趣的詞彙節點，系統可暫時將不相關的節點變淡，方便使用者集中注意力於重要節點的關聯探索。
6. 使用者可針對網絡圖形中的每個關聯性進行內部資料或是外部資訊的搜尋，以了解其中關聯性的詳細意義。

計畫創新重點

本計畫開發技術為一嶄新的視覺化詞彙關聯分析與探索之技術，供專業領域的應用，其為威知產品“第2代KSP之子系統－關聯脈絡圖子系統”。可應用的範圍，包括：刑事案件中各種關係人、犯案手法的關聯分析；教育領域中的文獻分析、引導式概念圖的教學應用等。

本計畫在技術面上，將強調技術上的效果與效率，使其可應用於各種領域；在產品面上，將強調使用上的便捷性，透過最新的AJAX技術，讓使用者有華麗介面、快捷的功能，以增加產品的價值。

下圖為計畫中的一種使用範例。使用者鍵入「牛皮」後，系統從刑案文件資料庫中檢索相關的案例，並分析其重要詞彙之間的關聯性，以圖形展示，方便使用者瀏覽。若發現某些關聯值得進一步探索，則可直接點選圖中的連結，搜尋相關的文件資料庫，以獲取詳細的資訊或情報。如此，員警透過新聞查找結果，利用系統進行詞彙關聯分析，而產生偵蒐地圖，可協助員警進行刑案線索的偵察與蒐集。



產學研各界之技術移轉及合作效益說明

本計畫各工作項目皆以台灣師範大學之研發人員作為核心技術來源，其餘工程師再以核心研發人員之研發概念、成果與透過技術指導與每週定期會議，協助各工作項目之實驗與測試，以找出最適方法。

目前台灣師大之研發人員主要負責整體系統所需使用之演算方法、搜尋時最適資料量的圍範調整與取用詞彙之門檻值，而威知資訊則將主要開發視覺呈現與人機介面設定之部分。整體技術傳承之流程為台灣師大先行著手進行演算方法與最適資料範圍等部分進行開發，待開發出雛型後，將雛型成果與技術移轉威知資訊，威知資訊開始開發介面呈現部分與應用程式並將其包裝成產品。

透過產學合作方式，不單單是增進產品開發速度時程，也可將學界研發之技術提供作為業界產品技術演進之一，提供一個技術交流的管道，進而擴大技術之效益。

新聘人力與效益

本研究團隊曾於計畫書中說明將依新聘人員之表現作為續聘標準，並預計將續聘本計畫中三分之二之大專畢業生。由於本計畫進行得非常順利，同時，本計畫所研發之成果將來亦將陸續應用於威知資訊之主力產品線中。為使相關之產品研發能夠持續精進，威知資訊規劃於本計畫結束後，將原服務於台灣師範大學之新聘人員吳正哲，改由威知資訊續聘，另威知資訊亦將續聘本計畫中原服務於威知資訊之新聘人員五人，總計本計畫結案後將由威知資訊續聘6位大專畢業生。

研發成果及衍生效益

本計畫所產生的關聯詞網及其視覺化呈現方法，未來主要為針對專門領域之應用，為使其達到更深層的服務效用，將針對各個不同領域之產業進行訪談，以調整服務內容，目前構想之主要應用領域包含警務單位、軍方單位、檢調單位、法務單位、學術研究單位、企業研發單位、傳統製造業、高科技產業以及各領域的服務業，其技術的應用範圍包含警政相關單位之案件偵辦、各大研究機構之文獻分析。未來可加入地理事件與時間事件之概念，將整體分析過程導向更廣泛之應用範圍。

產學合作計畫的目的，在於解決產界與學界技術的技術落差、矯正資源分配、研究能力等不足的問題，則其技術成果需能有效擴散至產業並被吸納。此外，技術或產品於進行推廣與應用後，技術提供者除可藉此控制及擴大既有市場外，亦可視技術或產品重要性與擴散程度獲取技術所衍生之報酬金以增加自身獲利。

本計畫預估於技術完成開發並將其成功產品化後，以每套第2代KSP之售價約3,000千元計算，估計98年可創造約3,000千元產值，估計99年可創造約27,000千元產值，估計至計畫結案次年底前將可創造約30,000千元之產值，同時本著此計畫需聘用多位大專畢業生，不僅可穩定就業率，更可提供首次就業的畢業生一個職場訓練的舞台。此外，藉由此計畫，可加速開發第2代KSP之子系統－關聯脈絡圖子系統之時程與其他相關應用產品之開發速度，估計在本計畫中減少合作業者約30%之人力於開發技術上。

專案執行重要心得

在開發視覺化產品時，除技術面外，最大的困難就在於使用者的呈現面，為了可以讓使用者有便捷的UI介面，研發團隊一直在探討什麼樣的圖形化介面可以同時在檢索效能與視覺效果皆達到最好的成效，透過各成員的努力與腦力激盪，才產出大家都滿意的結果。

在期中查訪時，很感謝審查委員給予研發團隊支持與肯定，希望未來還能夠有機會可以透過產學合作計畫，讓學界與業界在技術上有更多機會交流、切磋。