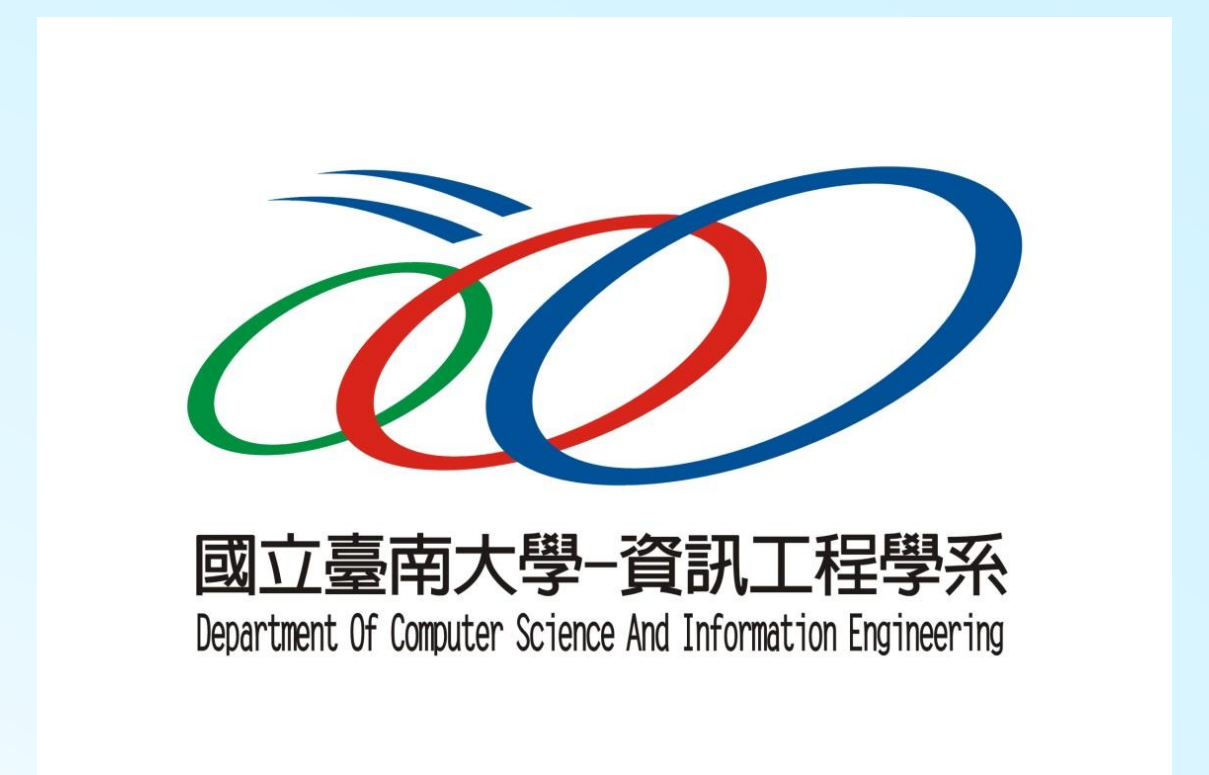


運用語意分群技術於專利文件之探勘分析



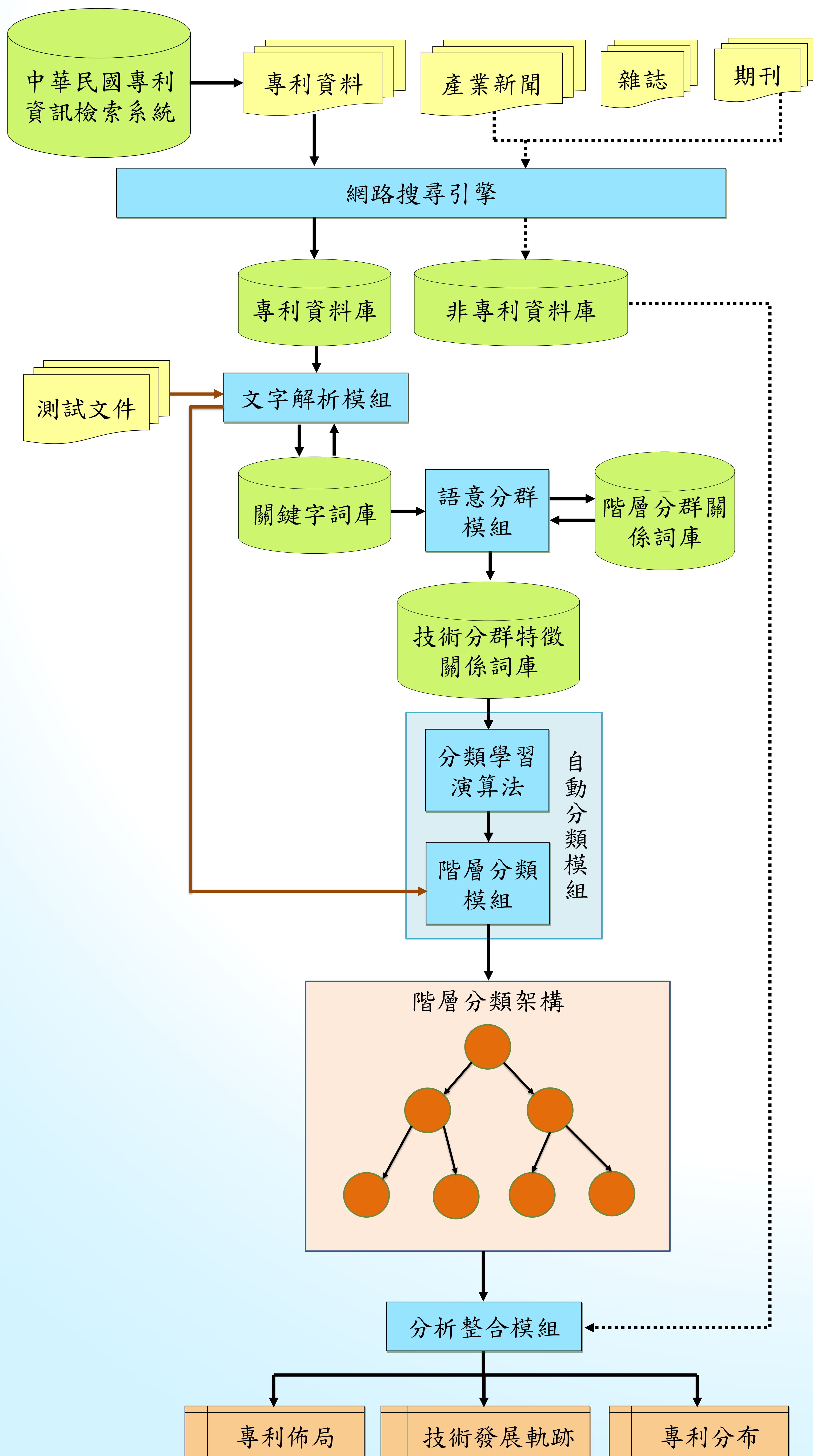
國立臺南大學資訊工程學系 106級畢業專題展
第十組：黎桂如



摘要

隨著知識經濟的蓬勃發展，專利技術的取得已經成為企業維護自身利益之最佳工具，而這些專利技術目前已經形成龐大的專利資料庫。專利資料庫除了可以讓我們了解他人的技術範圍，以避免侵權外，藉由分析專利資料庫更可以幫助我們預測競爭對手的技術發展動向與研發趨勢。而為了有效率的進行專利技術分析，專利地圖是常用的輔助工具，意即利用圖表幫助技術研發者快速了解目前技術發展之方向。但在建構專利地圖的過程當中，若使用人工方式進行分析與分類，會消耗大量人力與時間。因此，本研究希望能結合文字探勘與語意分群技術，實作一檢索分析工具，運用網路爬蟲系統(Web Crawler)自動擷取所需資料，運用文件分析技術進行分析與分群，最後以相關專利地圖表格等方式呈現，幫助使用者能更快取得所需的資訊，減少專利分析過程中投入的人力與時間。

系統架構



網路搜尋引擎：利用自身撰寫的網路爬蟲程式至中華民國專利資訊檢索系統抓取網頁原始程式。

文字解析模組：對網頁原始程式進行冗字去除、字詞斷句、關鍵字截取，建立關鍵字詞庫。

語意分群模組：透過關鍵字詞庫分析專利文件，建立各元件之間的階層關係，並將關鍵字詞庫與階層分群關係詞庫作為分群特徵，建立技術分群特徵關係詞庫。

分類學習演算法：利用分類演算法，如SVM、貝氏分類法等進行文件分類，並建立階層分類模組。

階層分類模組：將經過解析後的測試文件利用階層分類模組進行階層分類。

分析整合模組：整合階層架構圖表，並參考非專利文件，將分類結果與專利相關資訊填入表格中，完成專利分析表。

實驗成果

利用專利文件之固定欄位之交叉展開，可得到一階與二階之專利分析表。一階專利分析表為固定欄位之分析結果，二階專利分析表為不同欄位之綜合分析結果。

專利檢閱分類表						
橫軸欄位	1.年度	2.區域	3.專利權人	4.發明人	5.功效	6.產業結構
A.年度	A-1	A-2	A-3	A-4	A-5	A-6
B.區域		B-2	B-3	B-4	B-5	B-6
C.專利權人			C-3	C-4	C-5	C-6
D.發明人				D-4	D-5	D-6
E.功效					E-5	E-6
F.產業結構						F-6

透過專利檢閱分類表，利用交叉分析展開新式專利分析表。專利分析表之橫軸為檢閱分類表的橫縱軸展開，縱軸為該項產品之元件組成架構。利用各元件與元件之間的階層作為分類的類別對專利文件進行分類，最後將每個類別的專利文件依照橫軸對應之資訊欄位填入表格中，最後進行數量統計，即可得到該產品之專利分析表。

A-1 技術發展結構對應年度分析表					
核准年度			1985-1990	1991-1995	1996-2000
第一階	第二階	第三階			
影像感測模組	第一光學單元	凸面物鏡	12	37	23
	第二光學單元	光學目鏡	7	9	6
	發光單元		26	16	11
	影像感測單元		26	36	39
	發光單元		13	22	21
處理單元			32	28	35