

文件摘要的應用與實作

系所／資訊工程學系

指導老師／顏秀珍

組員／張智歲、高偉峰、張莉敏、陳昱臻

在這資訊爆炸的時代，現代人每天透過各種不同的方式接收資訊，以新聞為例，網路上的新聞量龐大，閱讀起來費時，因此一套新聞文章摘要的系統便有存在的必要。本研究為探討文件摘要之方法並呈現摘要。透過我們的新聞檢索系統，使用者輸入關鍵字搜尋相關的新聞列表，並於標題旁點選摘要按鈕，即可輸出符合使用者需求的新聞文章摘要。如圖 1 所示，搜尋關鍵字颱風即出現相關新聞，以最下面紅色框起來新聞為例，點選摘要按鈕即可顯示摘要。

此摘要方法分為兩個階段，第一階段為資料前處理，第二階段為摘要提取。摘要提取部分分為：語句分群的研究方法與主題地圖相關的研究方法。其差異在於計算語句權重方式不同，我們將其互相比較，並評估何種摘要方法能獲得最大效益。

語句分群研究方法是將文章以關鍵詞的位置分為多個群 (Cluster)，計算群的重要性以作為句子權重的依據，將權重排序後即可提取出該文件摘要。主題地圖相關的研究方法則是以各關鍵

詞出現在句中的次數作為向量代表該句，經過處理後計算各語句間的相似度，排序出與較多語句有高相似度的句子做為摘要並結合標題分析處理。

圖 2 為利用類神經的文件分類去計算摘要後分類的準確度。將摘要後的文章進行分類，比較摘要後分類的準確度是否高於原文分類的準確度，也將各個摘要方法相互比對分類的準確度，挑選出較佳的結果呈現給使用者，讓使用者能夠方便閱覽各新聞摘要。

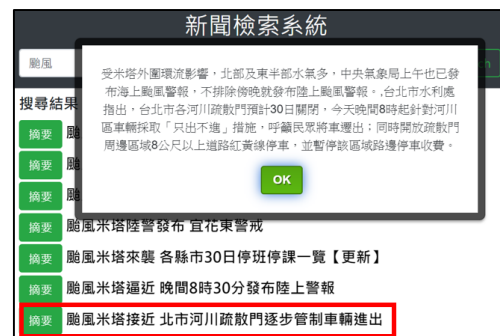


圖 1：新聞檢索系統介面

摘要方法	準確度
原始文章	0.7708
原文相似度	0.5250
LSA 與原文相似度	0.4750
語句無分群加入標題	0.6250
語句分群	0.7250
語句分群加入標題	0.6000
原語句陣列與主題地圖	0.5750
維度約化與主題地圖	0.8000
LSA 與主題地圖	0.6500
LSA 標題關鍵詞提高與主題地圖	0.5750
原語句陣列與標題關鍵詞提高及主題地圖	0.6500
原語句陣列與標題相似度	0.5000
LSA 標題相似度	0.5000

圖 2：摘要方法評估準確度