

第五章 結論

本文所提出的方法，搭配 Google news 中文新聞服務有效的擷取出新詞。根據觀察，新聞事件與新聞標題的出現，不一定每日皆不相同，往往一個新聞事件持續發酵好幾天，而 Google 提供的中文新聞服務不只是把當日的新聞標題歸類，而是把同一新聞事件出現在不同時間點的新聞標題皆歸類集中起來。如果不使用 Google 新聞服務而只是把每一天當日各家網路新聞媒體的所有新聞標題拿來套入本文提出的方法，其效率與正確率將是令人質疑的。本文提出的方法使用 Google 提供的新聞服務，無論是新聞標題的數量還是同質性皆非常充足，因此可以成功的擷取新詞並建立一新聞類專業詞庫。

包含本文提出的方法，所有新詞擷取技術雖然各有其優缺點，但正確率百分之百是一個共同追求的目標也是一個幾乎不可能的目標，本文提出考慮時間因素來淘汰詞庫裡的字詞的方法將能更有效提升新聞類專業詞庫的正確率，也大大降低人力維護的成本。

經由將詞庫裡的字詞更進一步的分類，可以提供未來的研究者在自然語言處理的便利性，例如新聞的追蹤與分類皆能透過使用本文所建立的新聞類專業詞庫，達到更加的正確率。現今科技發達，資訊的電子化有助於語言處理的相關研究。就本文提出的方法而言，新聞資料永遠不虞匱乏，使用新聞資料來建立詞庫並利用新聞資料的時間性來做進一步處理，完成一個幾乎不需要任何人力就可以自動更新自動維護的新聞類專業詞庫。