

協助解決 Google 難題的資訊萃取機制

An Information Extraction Scheme for Google-Hard Problems

¹ 謝育平

¹ 詹登淵

² 郭子文

¹Yuh-Pyng Shieh

¹Teng-Yuan Chan

²Tzu-Wen Kuo

¹ 銘傳大學資訊工程學系

¹Department of Computer Science and Information Engineering,
Ming Chuan University

² 銘傳大學財務金融學系

²Department of Finance,
Ming Chuan University

摘要

搜尋引擎在日常生活上已被廣泛的使用，可是搜尋引擎不能滿足使用者所有的資訊需求，依舊存在著許多使用 Google 無法快速解決的搜尋問題，本文稱之為 Google 難題(Google-Hard Problem)。為解決 Google 難題，本文建立萃取機制。該機制對第零階網頁和第一階網頁進行萃取，先將文本透過長詞優先法建立以詞為單位的句子，再運用 N-gram 演算法在詞序列中進行中文斷詞，提供一個加上關聯性比重的詞集列表來對使用者進行內容提示及精度關鍵詞推薦。另外利用詞夾子演算法(Term-Clip Algorithm)的特性與優點，以學習模式對文本探勘，然後找出與樣本詞相同屬性的同位詞來對使用者進行內容提示及廣度關鍵詞推薦。萃取機制幫助使用者加速搜尋的過程，並透過操作詞集來改變搜尋的方向，這是一種全新概念的搜尋操作模式。

關鍵字：詞夾子、萃取、中文斷詞

Abstract

Search engine has been an indispensable tool in our daily lives. But search engine cannot always satisfy users' information needs. There are still many problems that need much time to be solved by Google. Such problems are called "Google-hard problems" in this paper. Information extraction scheme was built to resolve the Google-hard problems. The scheme extracts information from depth-zero and depth-one web pages. First, text is split into sentences and sentences are split as a term sequence by longest term first method. Second, sentences perform Chinese word segmentation by N-gram method. The scheme provides an relationship term list. Another capability uses features and benefits of Term-Clip algorithm. Text mining is performed through learning mode to collect Appositional terms with the same properties as the query terms. The scheme helps users to speed up the search process. Users can handle the term list to change search direction. This is a new concept of search modes.

Keywords: Term-clip. Chinese term segmentation. Extraction

1. 簡介 (Introduction)

現今是個網路資訊發達的時代，資訊數位化伴隨著網路普及化逐漸提高，生活上的任何疑難雜症的文章都會發佈於網路平台，人們開始利用網路搜尋資訊尋求解答，搜尋引擎因此誕生於世界上。在1994年，開始有搜尋引擎的出現，Carnegie Mellon大學中Michael Mauldin所進行的研究項目Lycos，此研究項目正是第一個網路摘要搜尋引擎的誕生。在同一年代還有Colorado大學所建立的World-Wide Web Worm，當時搜尋的資料量約11萬頁，到2006年時，Google可以供索引的網頁已經成長到260億頁之多。人們對於求知慾望的需求造就龐大的搜尋市場，在各家競爭搜尋市場之下，又以Google為最成功。根據網路市場研究機構comScoreInc公布統計[14]，Google在2009年7月全球網路搜尋市場中佔有68%，單單在七月份全球網路使用人總共執行1136億次搜尋，搜尋引擎的使用已成為網路運用的一部分。在現今網路平台上搜尋各式的資訊已非難事；往地下可找到挖掘地底隧道的技術[22]，往天上可以找到人造衛星的系統[17]。在搜尋引擎上千億次數的搜尋過程之中傳輸了龐大資訊，使用者從成千上萬的字中找尋答案與相關訊息。

1.1 資訊需求(Information Needs)

使用者在操作搜尋引擎的過程是一種交錯式動作模式，在開始使用搜尋引擎前人們心中會有想要獲得的資訊需求，依照個人的知識背景嘗試著聯想出關鍵字，搜尋引擎就會根據關鍵字回應搜尋結果，因資訊需求與搜尋結果之間，產生符合情況有所不同，以下提出幾個在 Google 上的搜尋範例：

案例 A(菜市場名字問題)：當要查詢的人名字是一個菜市場名時，搜尋引擎出來的結果通常就是一堆不相關的人，若搜尋以聯考的菜市場名冠軍“雅婷”為例，找出來的有氣象主播、攝影師、民宿...等，就需要閱覽文章去除掉不相關要find人的關鍵字。

案例 B(筆電筆比較問題)：當想要購買筆記型電腦時，需要一些規格才知道哪一些符合自己的功能需求，若搜尋流行的筆電以“小筆電”為例，找出來的有貝殼機、口袋筆電、超薄筆電...等，這些雖然都是筆記型電腦，但是需要對個人需求進行規格資訊確認，所以要鍵入更細節或是型號進行搜尋。

案例 C(價錢與位元問題)：當想要尋找處理器的價格時，若以直接反應就在關鍵字加入“元”，搜尋結果卻是出現是32位“元”處理器的訊息，搜尋的方向偏向的是技術方面並非價格，所以重新想一個關鍵字，若以“售價”當作關鍵字就會出現是價格方面的搜尋結果。

案例 D(消毒劑藥名)：當H1N1流行起來想要買殺死H1N1病毒的消毒劑，額外特殊的需要是要對人無害，一開始不知道商品名稱時，直覺反應在搜尋引擎輸入“H1N1病毒”“殺死”，搜尋結果出現「台大病毒崩」與「矽世代奈米矽土」...等商品，接著要針對特殊的需求資訊找尋相關規格。

案例 E(跳電開機問題)：當電腦在跳電之後呈現關機狀態，希望找一個方法或是設定在跳電後是開機狀況時，若在搜尋引擎輸入“跳電”與“開機”，搜尋結果會出現會是關於跳電之後電腦無法開機，這些結果並非所需的資訊需求，重新思考關鍵字的方向。

案例 F(本體論報告問題)：當要了解一個題目作報告時，一開始對題目不知道太多資訊，若以“本體論”為例，搜尋結果中發現本體論有哲學的、人工智慧的...等，因此需要再從其中方向進一步搜尋報告的內容。

案例 G(電腦攝影機問題)：當想要購買經由網路傳輸到電腦控制的攝影機時，若輸入“攝影機”與“電腦”搜尋出來的結果，出現都是視訊使用的攝影機，而並非所需要的攝影機，必須重新設定關鍵字。

案例 H(電腦共享問題)：當想要一台電腦可以兩個人使用時，又不知道如何這樣軟體程式名稱，一開始嘗試著關鍵字輸入“電腦”、“共享”，搜尋結果會看到一些是區域網路分享的設定，後面的搜尋結果看到電腦分享器的商品，電腦分享器比較符合資訊需求。

案例 I(黃色塑膠繩問題)：當想要買封箱的黃色塑膠繩來時，根本不知道那個專有名詞是甚麼，也不知道要去哪邊可以找到相關物品，就以外型當作關鍵字輸入“塑膠繩”，搜尋出來的結果全部都是另外一種軟的塑膠繩，所以關鍵字完全不能滿足資訊需求。

將這些不同的情況分成以下四個狀況：

情形 1、知道資訊需求的內容，也知道多個或一個資訊需求的關鍵字。

- 閱覽搜尋結果符合大量資訊需求內容，結束搜尋過程。
- 閱覽搜尋結果出現資料歧異龐雜，操作關鍵字(keywords refine)回到情形 1。例如案例 A 的菜市場名字問題。
- 閱覽搜尋結果找到新的資訊需求，但仍不滿足大量需求回到情形 1。例如案例 B 的筆電筆比較問題。
- 閱覽搜尋結果發現關鍵字方向錯誤，重新進入情形 1。例如案例 C 的價錢與位元問題。

情形 2、知道資訊需求的內容，但不知道資訊需求的關鍵字。

- 搜尋部分資訊，找到資訊需求的關鍵字入情形 1。例如案例 D 的消毒劑藥名問題。
- 搜尋部分資訊，發現搜尋結果有誤進入情形 2。例如案例 E 的跳電開機問題。

情形 3、不知道資訊需求的內容，但知道多個或一個資訊需求的關鍵字。

- 閱覽搜尋結果已經滿足少量資訊需求，結束搜尋過程。
- 閱覽搜尋結果找到資訊需求內容入情形 1。例如案例 F 的本體論報告問題。
- 閱覽搜尋結果發現關鍵字方向錯誤，操作關鍵字進入情形 3。例如案例 G 的電腦攝影機問題。

情形 4、不知道資訊需求的內容，也不知道資訊需求的關鍵字

- 搜尋部分資訊，找到資訊需求的關鍵字入情形 3。例如案例 H 的電腦共享問題。
- 搜尋部分資訊，發現搜尋結果有誤進入情形 4。例如案例 I 的黃色塑膠繩問題。

1.2 Google 難題(Google-hard Problem)

由於以上狀況定義 Google 難題為以人的知識為背景，而超出一般人力最輕鬆的操作模式就成為難題，以 Nielsen Online[13]調查全美使用者，一個網頁閱覽平均 57 秒，和 Google 搜尋結果的關注程度有「金三角」現象[12]，50%使用者只會關注到第七個結果。推算若搜尋過程超過六分半鐘，就會造成使用者額外的煩惱，在 Google 上無法輕鬆得到答案將可視為一個難題。

Google 搜尋研究是由兩位史丹佛大學的理學博士生 Larry Page 和 Sergey Brin 在一九九六年早期建立的，開發一個對網站之間的關係做精確分析的搜尋引擎，此搜尋引擎的精確度勝於當時使用的基本搜尋技術，當時研究被稱作 BackRub，因為系統會檢查反向連結(blacklinks)，紀錄使用者的點擊以做評估站點的重要性。Google 搜尋建議運用使用者的點擊性特性，把符合搜尋詞的熱門長詞建議給使用者。Google 雖然是目前市占率最高的搜尋引擎，不過仍然有 Google 難題存在，搜尋引擎並不能解決所有的搜尋問題，所以還有很多可以改進的空間。

Google 難題最原始的解決方法是運用人的智慧去解決，人力去做反覆地搜尋動作和閱覽搜尋結果，經過一段長時間的搜尋過程找出符合資訊需求的資料。現在也有許多試圖要解決 Google 難題的方法如以下三種：

1.專業項目網站：

永慶房屋[19]—透過網路平台將房屋資訊發布出去，提供使用者搜尋查詢房屋資訊與地理位置。

比價王[18]—由網頁管理者發布 3C 產品資訊，提供各家通訊行業者的價錢比較。

2.互動平台網站：

Yahoo 奇摩知識[20]—透過 Web2.0 的概念結合社群的力量，以網路互動模式，由使用者提出問題，然後由知道解答的會員回答，進而使用者的資訊需求。

ChaCha[15]—模仿圖書館員服務的模式，可以付費請人幫你查詢資料。

3.人工智慧網站：

Quintura[16]—將搜尋詞輸入到 yahoo 搜尋引擎，找出與搜尋詞相鄰近的詞。

搜尋羅盤(Google wonder wheel)[21]—宛如 Google 建議一般，提供與搜尋詞相關熱門的搜尋詞。

在搜尋引擎上輸入關鍵字就會提供大量網頁資訊給使用者，每次使用搜尋引擎就出現上千筆結果等著去閱覽，常常需要閱覽數十頁面甚至百頁之後才發現一些關鍵資訊，或者根本浪費時間在閱覽不相關的內容。人們往往需要耗費許多時間在過濾資料，人力效率不僅低落還有遺漏重要資訊的問題。如果可以快速得知內容的重點便可以避免掉寶貴人力浪費，萃取機制希望幫人們省去上述那些麻煩，幫使用者分析網頁資訊去除不必要的資料，提供關鍵資訊縮短使用者找尋的時間。因此本文擬建立一個萃取機制，將搜尋網路資料做半自動化分析，使得碩大的資料去蕪存菁成為精簡的資訊。

2. 萃取機制架構

現在人們常常利用搜尋引擎找尋資料，輸入關鍵字就會獲得很多搜尋結果，但資訊需求還是需要依靠人力去找尋資訊，因此需要一個萃取機制協助人對資料進行萃取與探勘[1][2][3]。例如：搜尋相機出現的是目前市面上賣的產品，但是可能要找的相機拍照技術，因此需要微調搜尋。建立萃取機制需以下的初步步驟：

1. 蒐集網路資訊，過濾字元產生文本。

2. 運用長辭優先演算法，使文本以詞為單位。

3. 利用 N-gram 演算法建立詞集，詞集加入關聯性比重排名。

4. 使用詞夾子演算法從文本中，找出相同屬性的同位詞。

搜尋引擎已成為網路應用服務中重要的一部分，而 Google 搜尋引擎排名運用反向連結及搜集點擊率，特徵是由大多數人點擊出來的關鍵字排名，所以位於前面的排名網頁內會含有關鍵字，搜尋出擁有關鍵字的網頁是一種精度搜尋結果。大部分的使用者正是要找尋含有關鍵字或相關連的網頁與資訊，萃取機制的目的便是協助使用者找到這些資訊。當想要知道數位相機如何拍出好的照片時，由萃取機制可以毫無遺漏的處理資料，並不會有人為的疏失而遺漏重要資訊。萃取機制會將所有網頁資訊收集起來，先運用長詞優先演算法[4][6]以詞為單位的句子，再利用 N-gram 演算法[8]顯示出加權比重的關鍵詞資訊，不必動用人力去一頁一頁瀏覽搜尋引擎的網頁排行，從詞集中明確發現“光圈”、“快門”、“曝光”，不但省去過去大量過濾的時間，還可以輕易找到想要的詞。

2.1 同位詞

同位詞定義為與樣本詞出現在相同位置的詞，所謂相同位置指的是擁有共同錢文、後文，從這樣學習規則找出來的詞稱為同位詞。例如：樣本詞為“台北縣”，從文本中找到前文“由”和後文“政府”，利用前文和後文當作夾子對文本學習出來“台中市”、“高雄縣”、“花蓮縣”...等，這些訓練出來的詞通稱為同位詞。

當使用者需要找尋相關聯的詞，搜尋引擎並無法提供你名詞相關聯性，這時候可以運用詞夾子[9][11]找出相關的同位詞。並非每個使用者的資訊需求都是一樣的搜尋結果排名，在 Google 以絕大多數人的鏈結排名前提下，最前面排名會是以關鍵字為主題的資訊，這樣子排名並無法滿足所有使用者。在資訊需求是同屬性的關鍵字，以 Google 搜尋引擎的排名是無法找到。同屬性的關鍵字搜尋是一種廣度搜尋非精度搜尋，使用詞夾子演算法正好解決廣度搜尋。在需要做演算法比較或是報告的情況之下，需要大量同位詞的關鍵字，以狂蒐集同位詞的情況正是一個 Google 難題，無法以搜尋引擎的點擊特性去解決，萃取機制產生同位詞正好符合資訊需求。當搜尋“退火演算法”時，希望在文本中進行搜尋其他演算法，運用詞夾子演算法可以找尋同位詞，限定綴詞可以更精確的找出演算法，所以在詞夾子演算法中加入後綴“法”之後，就能文本中找出符合的同位詞，會出現“螞蟻演算法”、“基因演算法”、“遺傳演算法”、“模擬退火演算法”、“禁忌演算法”、“最佳化演算法”、“啟發式演算法”、“粒子群演算法”...等，這樣的狂蒐同位詞以搜尋引擎需要人力去閱覽大量網頁資訊，使用詞夾子演算法就迅速了解目前與“退火演算法”同屬性的演算法。

2.2 整體架構

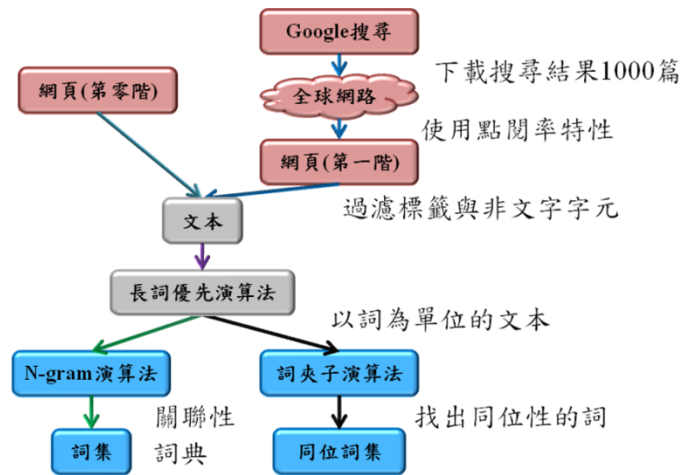


圖 1 網路資訊萃取架構

提出一個萃取網路資訊路機制來處理網頁資料，圖 1 首先在第零階網頁(瀏覽器目前的頁面)過濾掉網頁標籤以及非文字字元，先將文本透過長詞優先法建立以詞為單位的句子，接著利用 N-gram 演算法對句子進行斷詞產生出來詞集，這些詞集可以依照出現頻率與符合詞庫作比重加權，進而可以排列出第零階網頁中重要的詞集列表。使用者可以輕鬆瀏覽詞集以了解網頁重要資訊，點選列表中的詞進行搜尋詞的增加或避免某些搜尋詞，快速找到想要的網頁資訊。萃取機制也可以對第零階網頁使用詞夾子演算法，輸入樣本詞經由機制找出詞夾子，機制再透過詞夾子找出目前在第零階網頁同位詞。

萃取機制提供第一階網頁(使用者搜尋 Google 結果的網頁)的萃取功能，系統可以去下載第一階網頁讓使用者進行萃取資訊的動作。下載大量的網頁資料可以增加文本的字數，在較大的文本作搜尋萃取資訊的依據，可以減少資料不夠廣和詞集狹小的問題，詞集會有較多不同的相關聯性，詞夾子也可以找出更多的同位詞。

萃取機制和以往的研究不同，系統架構使用的文本是由網路上網頁資訊。第零階網頁從目前瀏覽器中擷取，第一階網頁是由 Google 搜尋引擎搜尋網頁，Google 搜尋引擎是由使用者查詢次數做排名和搜尋引擎優化(Search Engine Optimization)，以往的機制需要人去對文本選擇和特徵比重，本機制利用 Google 廣大的使用群幫系統做好排名。

2.3 長詞優先法(longest matching method)

長詞優先法是目前被廣泛使用和具代表性的斷詞方法。此方法是從經驗法則上所建立：句子中通常是許多有意義的詞組合，此方法優先斷出最長的有效詞，從句子的一端開始取最長的詞串逐一比對辭典內的詞，若發現符合的詞把它當作斷詞的結果，再把句子中未符合的字串重覆一樣的動作，直到整句都斷詞完畢。

從句子開頭的第一個字開始，從開始到結尾取最長的字串逐一比對辭典中的詞，直到句子的最後一個字結束。以這段話「大學生要做十一件事。」執行長詞優先演算法斷詞。將句子最常有效詞與詞庫作比較，符合最長有效詞的詞切割成一個詞，剩下的為符合的字串重覆動作

與字典作比較，最後以第一字串到最後一個字串找出最少詞數，將這句話斷「大學生」、「要」、「作」、「十一」、「件」、「事」六個詞，最把六個詞收入詞集。

2.4 N-gram 法

N-gram 統計式語言模型是從統計法上所建立：此方法對文件內容字串進行分割，以 n 個字串為一個詞分割出所有的詞，所分割的字詞出現次數若高過於門檻值，認定為一個獨自的詞彙。

當要從網站的文章進行斷詞的話，以這段話「大學生要做十一件事」這句話為例進行 2-gram 的斷詞，將從第一個字元開始以四個字為一個詞，將這句話斷「大學」、「學生」、「生要」、「要做」、「做十」、「十一」、「一件」、「件事」八個詞，最後把超過出現次數門檻值的詞收入詞集。

2.5 混合式語言模組運用

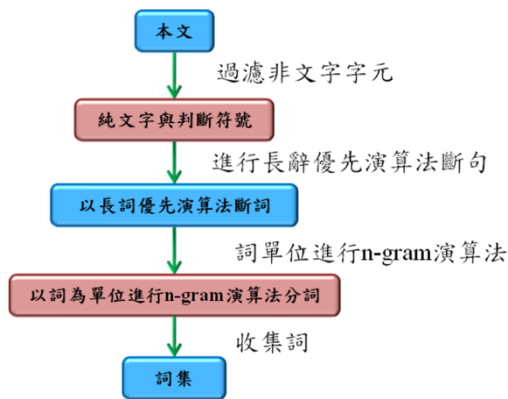


圖 2 語言模組架構

從網路上查詢資訊，第一步收集 Google 網頁的資料，去除網路標籤和過濾非文字字元之後，第二步運用長詞優先演算法將以詞為單位，第三步設定搜尋詞進行 N-gram 演算法，以中文詞庫當作詞範本找出符合的詞擁有較高的權重如圖 2，加上其他比重做為萃取資訊重要性。詞庫分成正詞和負詞兩種詞庫如圖 3，負詞庫中的詞以沒有意涵的詞為區分，像是「類似」無庸置疑是一個詞，在中文的意涵中在「類似」後面接名詞才是重點，詞庫也可以由詞夾子演算中得到新的詞。

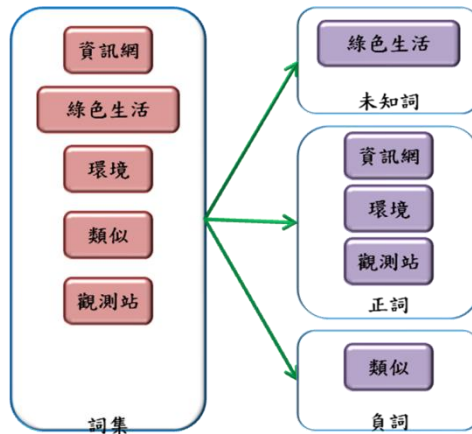


圖 3 詞庫分類

本機制運用在網路文章上找出新的用詞，網路上時常出現新穎的詞彙，這些新的詞彙在一開始中文斷詞會被當成未知詞看待。詞夾子演算法可以找詞的特性，一般專有名詞會使用符號將他區隔開來，因此在網路上常用新詞彙就可以透過學習的模式，將新詞的特性設定若是為專有名詞，符合門檻值就可以將它加入詞庫之中。從 Google 使用者點擊特性的第一階網路訓練過程之中，所形成新的點擊特性詞庫，和以往只有系統操作者及少數人定義詞不同，本機制是有上千億次的搜尋結果當作背景而完成的詞庫。

2.6 詞夾子演算法(Term-Clip algorithm)

中文詞集是不可能收入全部詞的詞集，而分類詞集更是無法收錄所有的分類，所以有各項各樣的研究如何收集詞的方法，詞夾子演算法是將一個搜尋詞分為“前文”、“前綴”、“中綴”、“後綴”、“後文”的區分如圖 4，“前文”與“後文”用來當作夾子在文本中找出候選詞，候選詞裡面可以包含或不包含“前綴”、“中綴”、“後綴”，在訓練選詞過程可限定綴詞，找出同位詞集更加準確。詞夾子演算法是一個不需要詞庫就可以找尋詞集的方法，是透過學習的方式來建立詞集，也可運用詞的特性輔助詞庫的缺詞。

詞夾子演算法是首先會以樣本詞對文本進行比對，找出符合樣本詞的前文與後文，再以前文和後文當作詞夾子到文本中，尋找具有想同屬性的候選詞，超過設定門檻的候選詞可以加入樣本詞之中，再一次訓練時，就會有更多的樣本詞和詞夾子。中文詞的結構[10]中有綴詞的特性，當要搜尋某種特定的詞可以設定樣本詞中綴詞，便可以將樣本詞方向鎖定在特定區塊。

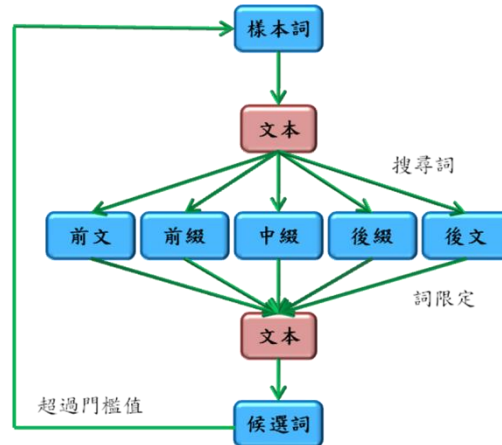


圖 4 網路資訊萃取架構

3. 優缺點

經由長辭優先演算法和 N-gram 演算法混合使用，可以互相彌補各自的缺點。長詞優先演算法的缺點是無法索引出詞庫沒有的詞，而 N-gram 演算法並不會有這的缺點，不過 N-gram 卻會斷出許多長詞中的短詞，這點以詞為單位的句子用長詞優先演算法補強，還有使用學習方式的詞夾子進行文本探勘出同位詞。雖然無法像使用自然語言處理[5]那樣學習精準，但是在可用度上卻有速度上的優勢，不需要耗費大量的學習訓練[7]過程，短時間內就可以產生有效的效果，在網路應用上是一個不錯的解決方案。

4. 結論 (Conclusions)

本篇提出以混合式斷詞法和詞夾子演算法嘗試解決 Google 難題，架構中以長詞優先演算法和 N-gram 演算法混合使用斷詞，詞集加上相關聯性的比重和排除不重要的負詞，用以解決精度搜尋上的問題。另外提出詞夾子演算法找尋相同屬性的同位詞，用以解決狂蒐同位詞的廣度搜尋問題上。以電腦系統做到減少瀏覽的時間和協助搜尋操作，的確節省很多寶貴的人力資訊。

5. 參考文獻 (References)

- [1] Yi-Che Chan., "Google-Based Two-Stage Text Segmentation and Learning Question Type Identification from Wikipedia for a Multilingual QA System.", NATIONAL CHENG-KUNG UNIVERSITY, 2007.
- [2] Zhao, Hai, Huang, Chang-Ning and Li, Mu., "An improved chinese word segmentation system with conditional random field. In Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing.", 2006, pp.162-165.
- [3] Zhang, Y.and Clark, S., "Chinese segmentation with a word-based perceptron algorithm.", In Proceedings of ACL, 2007, pp.840-847
- [4] 方心伶,「中文斷詞與注音」,碩士論文(指導 教授:江永進),清華大學統計學研究所,2008。

- [5] 李志豪「以基因規劃法為基礎的中文斷詞模型」，碩士論文(指導教授：姜正雄)，玄奘大學資訊科學研究所，2008。
- [6] 林千翔「基於特製隱藏式馬可夫模型之中文斷詞研究」，碩士論文(指導教授：張嘉惠)，中央大學資訊工程研究所，2006。
- [7] 王亮超「領域本體論為基之網頁知識擷取機制設計」，碩士論文(指導教授：王昌斌)，南華大學管理研究所，2006。
- [8] 英家慶「使用 N-gram 模型對中文文件自動分類」，碩士論文(指導教授：顏秀珍)，銘傳大學資訊工程研究所，2007，第 78-91 頁。
- [9] 張尚斌「詞夾子演算法在專有名詞辨識上的應用-以歷史文件為例」，碩士論文(指導教授：項潔)，台灣大學資訊工程學研究所，2005。
- [10] 許薰衣「宋詞斷詞與本體論之建置」，碩士論文(指導教授：曾憲雄)，交通大學理學院，2006。
- [11] 謝育平、楊龍廉、趙建宏、黃銘立、古馮文、林郁智「使用詞夾子建立中文典籍分析加值服務」，銘傳大學 2009 國際學術研討會，銘傳大學資訊工程研究所，2009，第 78-91 頁。
- [12] Maggy「用戶搜尋行為研究：google 搜尋結果金三角現象及其意義」，
<http://seo.seo2000.net/%e7%94%a8%e6%88%b6%e6%90%9c%e5%b0%8b%e8%a1%8c%e7%82%ba%e7%a0%94%e7%a9%b6%ef%bc%9a%e6%90%9c%e5%b0%8b%e7%b5%90%e6%9e%9c%e9%87%91%e4%b8%89%e8%a7%92%e7%8f%be%e8%b1%a1%e5%8f%8a%e5%85%b6%e6%84%8f/>，December 2009。
- [13] 那福忠「美國報紙網站 閱讀率低於 1%」，
http://mag.udn.com/mag/digital/storypage.jsp?f_ART_ID=213064，November 2009。
- [14] 張正芊「comScore：7 月全球搜尋市占率 Google 以 68%居冠 Bing 成長 41%」，
<http://news.pchome.com.tw/science/cnyes/20090901/index-12517755281647209005.html>，November 2009。
- [15] 「chacha」，<http://www.chacha.com/>，December 2009。
- [16] 「quintura」，<http://www.quintura.com/>，December 2009。
- [17] 「人造衛星」，
<http://zh.wikipedia.org/zh-tw/%E4%BA%BA%E9%80%A0%E8%A1%9B%E6%98%9F>，November 2009。
- [18] 「比價王」，<http://www.eprice.com.tw/>，December 2009。
- [19] 「永慶房屋」，http://www.yungching.com.tw/CB/CB_Default.asp，December 2009。
- [20] 「奇摩知識」，<http://tw.knowledge.yahoo.com/>，December 2009。
- [21] 「搜尋羅盤」，
http://www.google.com.tw/search?hl=zh-TW&rlz=1W1GGLT_zh-TW&tbo=1&tbs=ww%3A1&q=Google-hard+problem&meta=，November 2009。
- [22] 「隧道技術」，http://www.cec.com.tw/page.php?cat_id=183，November 2009。