

資工系實務專題研究計畫成果報告

大專院校師資檢索系統

專題編號：097CSIE-S023

執行期限：96 年 1 學期至 97 年 1 學期

指導教授：王正豪

專題計劃參與人員：詹閔淵

壹、摘要

資訊檢索研究的目的是在解決人類對於資訊的需求。然而，隨著不同資訊型態的出現，資訊檢索的技術亦逐漸多樣化，以適應各種資訊檢索型態的呈現需求。

以往在 Google 搜尋關鍵字僅能找出刻板印象的字詞或是因為廣告廠商的付費影響排名，而本系統提供一個介面，透過索引各大專院校師資網站至資料庫中，讓使用者能利用關鍵字，檢索大專院校師資的資訊。利用 AJAX^{註1} 技術提供更方便使用的介面，也提供了點閱排行榜、熱門關鍵字以及顯示歷史查詢資訊作為搜尋關鍵字的參考。

關鍵詞：資訊檢索、web2.0、AJAX、師資、大專院校、中文斷詞系統、資料庫

貳、緣由與目的

隨著網路科技與經濟的快速發展，現今已經進入資訊數位化的科技時代，無論是學界或者是產業界皆利用網際網路來傳遞資訊、交流互動。因此，在資訊的傳遞過程中，資訊傳遞與搜尋的精準度就更受重視，所以可以更明確的確認資訊檢索的重要性。而所謂的資訊檢索係指將資訊與網路結合，使得知識或訊息能以重點或摘要的方式呈現。

根據研究者仔細的分析，目前若利用 Google 等搜尋工具進行資料檢索，如搜尋關鍵字「網路」，Google 會搜尋到較熱門的結果，但研究者發現有關於關鍵詞呈現結果會受限於廣告商贊助提升排名或是熱門點閱率而影響真正的資訊檢索結果。

大專院校的教師提供了產業界的諮詢

與發展的機會，因此提供正確的大專院校師資檢索是相當重要的，但目前並無較完整的系統可供檢索使用，所以本研究欲發展一套師資檢索的系統，提高大專院校師資檢索的準確性與便利性。

參、系統簡介

本系統分成三個部份，第一部份為解析來源網頁，使用斷詞^{註2} 工具提取關鍵字；第二部份為索引網頁，將斷詞後的結果建立索引資料庫；第三部份為使用者介面，提供自動完成與圖形化結果。

以下針對本系統三個部份敘述如下：

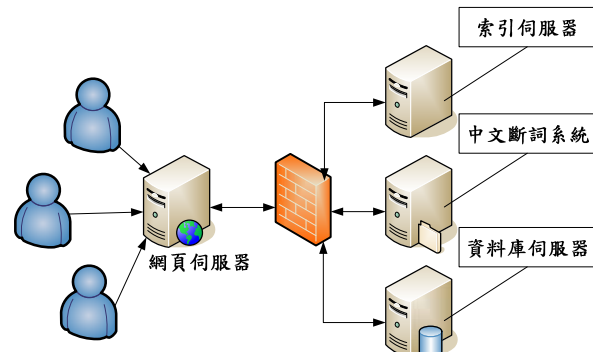


圖 1 系統架構圖

一、解析網頁斷詞：

- 1.自行開發蒐集程式載入 URL 並輸出文字檔供斷詞使用。
- 2.斷詞前將 HTML tag 拿掉，留下純文字的部份，判斷網頁編碼，轉換為 Big5 供斷詞使用（留下 title 及原始 URL 位置）。
- 3.斷詞後拿掉不需要的 tag 並將編碼轉換為 UTF-8，強化中文字的支援避免中文亂碼。
- 4.自動產生網頁檔案，作為索引使用。

^{註1} 建立互動式網頁應用的網頁開發技術

^{註2} 使用中研院詞庫小組開放之 API 將文字分詞

二、索引：

- 1.將斷詞後的網頁檔案進行索引。
- 2.解析斷詞時保留的原始 URL 位址。
- 3.建立資料庫內容供之後統計點閱率。

三、使用者介面：

- 1.熱門查詢以及熱門點閱率。
- 2.AJAX 技術自動完成，使用者輸入第一個詞，即產生該詞相關的關鍵字，以出現的次數做排序，可快速查詢想取得的資料。
- 3.搜尋關鍵字呈現結果後，產生關聯性，讓其他使用者了解查詢過什麼關鍵字，也可以得知相關的資訊。
- 4.圖表呈現關鍵字近期幾天查詢次數。
- 5.圖形化前四名熱門搜尋，排序顯示。

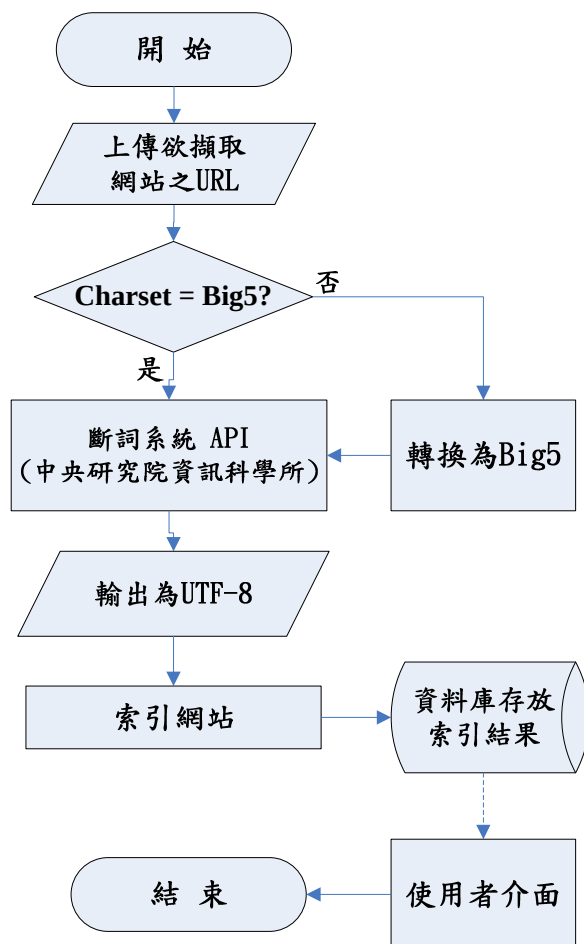


圖 2 系統流程圖

肆、應用技術與工具

本系統修改 Sphider 的 Open Source，系統核心僅支援英文，修改後已經可以支援對中文的搜尋。

網頁技術為 PHP，使用 MySQL 資料庫，應用 AJAX、JavaScript、GD Library

等技術增加網頁 UI 的部份。

實作修改 Bug 時利用 FireFox 的附加元件 FireBug 及 Web Developer 偵錯修改。

伍、結論與建議

本系統使用 PHP + MySQL + AJAX，可以跨平台在 Linux 或是 Windows 上運作，實用性高。本系統在實測驗證後，經評估成效良好，主要改善：

- 1.目前搜尋引擎無法精確的搜尋到使用者想得到的資訊。
- 2.節省篩選時間，使用者想了解大專院校師資，不用到每所學校查詢。

本系統後續發展之應用領域相當廣泛，除了大專院校師資的檢索外，還可應用在其它的檢索方面，例如新聞分類、專業書籍資訊檢索等。據此，研究者臚列下列研究建議，俾供後續研究者參考：

本系統雖有跨平台及支援中文搜尋優點，但如何改進以中文字義為斷詞依據仍須加強。斷詞這部份在研究時，是採用「中央研究院資訊科學所詞庫小組中文斷詞線上服務」的 API，明顯的發現人名或是專有名詞部份，會被分段，而造成索引時的誤判，降低搜尋結果的準確度。

斷詞系統將來要建立一個字典檔，讓索引時能夠辨別，而不再採用別人的 API，如此即可增加搜尋結果的精確度。另外一部分則是增強 UI，比起以往的輸入關鍵字更加方便，朝向圖形介面的方式，漸進式引導使用者所需要搜尋的結果。

最後的部份，則是增加使用者評分，為搜尋出來的結果加上權重，未來有其他使用者再次搜尋這份資料，能快速且精確的顯示結果。

陸、參考文獻

- [1] Sphider,
<http://www.sphider.eu/>
- [2] JpGraph,
<http://www.aditus.nu/jpggraph>
- [3] jQuery,
<http://jquery.com/>
- [4] CSS,
<http://www.w3schools.com/css>
- [5] Suggest Framework,
<http://sourceforge.net/projects/suggest>