

本體論支援之研究學者網頁蒐集器

楊勝源
聖約翰科技大學
電通系
ysy@mail.sju.edu.tw

陳鼎安
聖約翰科技大學
電通系
93412048@student.sju.edu.tw

吳奇峰
聖約翰科技大學
電通系
93412018@student.sju.edu.tw

摘要

網際網路蓬勃發展的時代，浩瀚網際中滾滾的資訊洪流，常令使用者丈二金剛摸不著邊。如何讓使用者搜尋到有用的資訊，早已成為使用者本身或搜尋系統業者的重要需求。因此，主動式網頁搜尋技術便成為當代資訊系統中不可或缺的重要工具。本論文提出一本體論支援之網頁蒐集器 OntoCrawler，使用者只要輸入關鍵字，系統在領域本體論的支援下，主動提供關鍵字判斷比對，藉以提升網頁蒐集的精準度。本技術已實際架構在 Google 及 Yahoo 兩搜尋引擎上主動蒐集相關研究學者網頁，並將結果存入資料庫提供後端系統做進一步的處理。初步的實驗結果驗證本技術確實能高度提升網頁蒐集的精準度。

關鍵詞：網頁蒐集器、本體論、Protégé。

Abstract

In this most developed and quickly shifting era of Internet, huge amount of information makes users ambiguous to adapt what information is necessary. How to search advantage information has become essential part in users or even the search engine manufacturers themselves. Therefore, the active webpage searching technique plays an important role in the contemporary information system. In this paper, an OntoCrawler based on the ontology-supported technique for webpage searching was proposed, in which only user entered some keywords would the system supported by domain ontology actively provide comparison and verification for those keywords so as to up-rise precision rate of webpage searching. This technique has practically been installed in Google and Yahoo search engines and furthermore searched some related scholars' personal webpages and accordingly stored the results into a database to let the backend systems to do advanced processes. The preliminary experiment outcomes proved the OntoCrawler

based on ontology-supported techniques proposed in this paper could really up-rise the precision rate of webpage searching.

Keywords: Webpage Crawler, Ontology, Protégé.

1. 緒論

早期在網際網路尚未出現前，為了查詢散佈在各個分散主機中的龐雜文件，曾有過 Archie、Gopher 等簡單的文字搜尋工具[1]，但其查詢的結果仍不盡如人意。這就是 Yahoo 目錄查詢出現的最大動機，透過事先分類索引及預先蒐集好的網路資訊，提供使用者更便捷快速的資訊查詢。隨著資訊量和收錄連結數的增長，搜尋服務系統也開始提供簡單的資料庫查詢。1998 年 10 月 Google 搜尋引擎正式誕生，它是當代最令使用者滿意的搜索引擎之一。Google 以其優越的搜尋演算法，不被競價排名所籠罩的口碑成為全世界最為人所知的搜索引擎。

Google 採用關鍵字查詢方式，提供含有超過 80 億個索引網址[8]。因此僅使用少量的關鍵字查詢，卻往往傳回令人窒息的龐大查詢結果；以查詢「CPU」為例，就有大約 171,000,000 查詢結果。如此冗長的查詢結果及排列方式，不僅讓使用者必須費時費工的仔細瀏覽後，才能挑選出有用的資訊外；更顯示出搜尋系統本身並無法完整且精準地了解使用者的查詢意圖。造成這樣結果的主要原因：或許是使用者輸入的關鍵字不夠完整與不能清楚表示出使用者的查詢需求；再加上所輸入關鍵字在各種不同領域間都有諸多同名異義字，當系統無法完整且精準地判斷使用者的查詢需求，以及沒針對特定領域分別處理的情況下，終究導致系統蒐集過多橫跨多個領域的龐雜查詢結果[9]。這樣的結果往往讓多數的使用者難以從中挑選出真正的資訊需求，因此，如何讓網頁蒐集技術能夠精準、明確地搜尋到使用者所需要的資料，早已成為使用者和資訊系統間必須面對的重要議題。

隨著網際網路的快速普及使用搜尋引擎

的相關技術早已成為人們生活上不可或缺的科技能力。人們要在網際網路上搜尋想要的資訊，就必需自行前往各個獨立的搜尋引擎，輸入關鍵字來達成資訊蒐集的目的。基於一般搜尋引擎的搜尋範圍相當廣泛；再加上使用者輸入的關鍵字不夠完整清楚，致使傳回的資料量過於龐雜無章，使用者更無法快速準確地找到需要的資訊。因此，我們希望設計出一個具主動擴展查詢語句之主題式蒐集器 (Focused-Crawler)，不僅能輔助與降低使用者查詢作業上的負擔，更能用以支援網頁搜尋系統的核心。

綜言之，本研究之研究主題在於應用本體論 (或知識本體，Ontology) 技術設計出相關學者網頁之知識本體；並藉由本體論建構工具 Protégé 建構與分析研究學者網頁的知識本體類別；再配合 MS SQL Server 資料庫建立出一研究學者網頁相關關鍵字知識本體分享平台；最後，利用 Java 語言[12]建構這個網頁主題式蒐集器 (OntoCrawler: Ontology-supported Focused-Crawler)，引進學者知識本體提供關鍵字判斷比對，不僅排除因同名異義字所額外取回的網頁，更提升針對研究學者網頁蒐集的精準度，達成兼顧支援使用者查詢作業與相關搜尋系統核心之的。

2. 背景知識與相關開發技術

2.1 本體論

本體論原本是哲學領域中的論點，主要探討生命或現實事物的知識本質；在人工智慧領域方面，主要是用來定義一個領域的知識內容，表達知識，解決溝通、共用分享等問題；資訊科學領域中，對於電子商務 (E-commerce) 及知識管理 (Knowledge Management) 等研究發展有很大的幫助[7,9,10]。

本體論能提供完整的語意模型，意指在特定領域中，有關於該領域相關的各式物件、物件屬性與物件彼此間統一架構下的基礎知識，具有共用與重複使用的特性，可用來解決彼此間溝通和共用分享的問題，透過本體論來描述知識內容的架構，可以完整地呈現一個特定領域的知識核心，自動地瞭解相關領域資訊、溝通及存取，甚或更進一步推論出新的知識與結果，對於資訊系統的建立與維護，是個非常有力的工具[9,10]。根據[9]定義，認為本體論描述特定領域的概念 (Concept)；概念的描述包含元件 (Entities)、特徵 (Attributes)、程序 (Processes) 以及渠等之間的關係，可由

不同的方式來表現，通常包含字彙 (Vocabulary)、規格 (Specification) 等。圖 1 就是一個 PCDIY 知識領域的本體論，主要是定義電腦相關的基本知識，顯示部分概念的階層關係及其相關特徵。

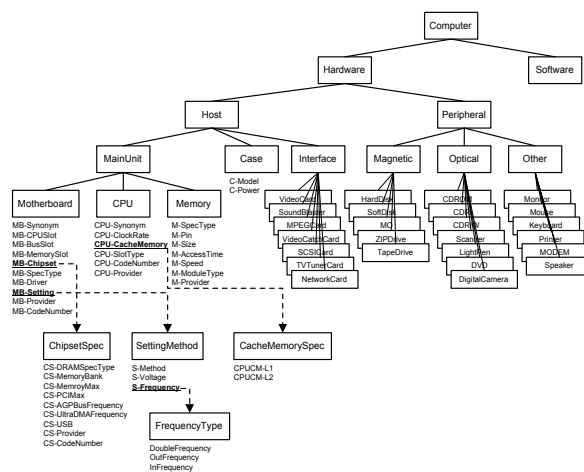


圖 1. PCDIY 本體論

2.2 正規表示法

所謂正規表示法 (regular expression) 就是一份用來描述某種特定順序的字元序列。這種俗稱樣式 (pattern) 的描述方式可以拿來在其他字元序列當中，搜尋符合的樣式結果。正規表示法可以使用萬用字元、字元集合，以及各種量詞 (quantifier) 來進行指定[3]。支援正規表示法的類別有兩個：Pattern 和 Matcher。我們會使用 Pattern 去定義一個正規表示法。對於要和另一個序列進行匹配的 pattern，則會使用 Matcher。圖 2 即是本系統採用正規表示法的程式片段。

```
Pattern BodyLabel=Pattern.compile("(?s).*?>", Pattern.CASE_INSENSITIVE);
Matcher RM=BodyLabel.matcher(s3);
Content=RM.replaceAll("");
Pattern Space=Pattern.compile("(?s)&.*?;", Pattern.CASE_INSENSITIVE);
Matcher CM=Space.matcher(Content);
Content=CM.replaceAll("");
```

圖 2. 本系統採用正規表示法的程式片段

2.3 相關開發技術

本系統採用 MS SQL Server 來做為後端本體知識資料庫分享平台。MS SQL Server 是目前最常被使用的一種關聯式資料庫管理系統[5]，SQL (Structured Query Language) 則是一種用在關聯式資料庫的查詢語言，用來取得資料庫中的資料。MS SQL Server 具備了以下特點[1,2,4]：

- (1) 不論是大企業或是個人需求，都可以提供適合的解決方案；
- (2) 具有預儲程序功能，在資料預儲程序效能極佳；
- (3) 在視覺操作方面，操作簡單；
- (4) 與微軟其它開發工具整合度高；
- (5) 管理帳戶系統簡單、方便易用；
- (6) 安全性、容錯性較高；
- (7) 資料庫支援較大。

圖 3 為 MS SQL Server 管理介面。畫面左方區塊為物件總管，可進行建立資料庫，設定權限、安全性、維護、備份等工作。

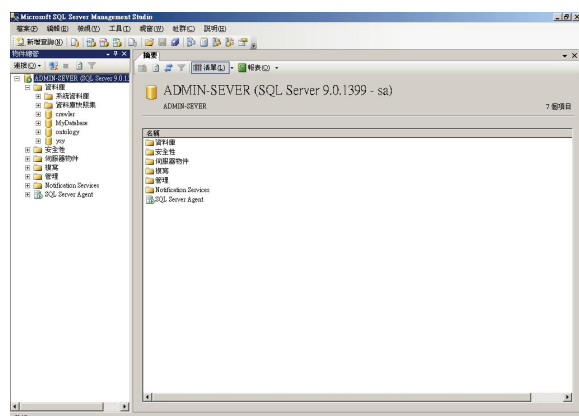


圖 3. MS SQL Server 管理介面

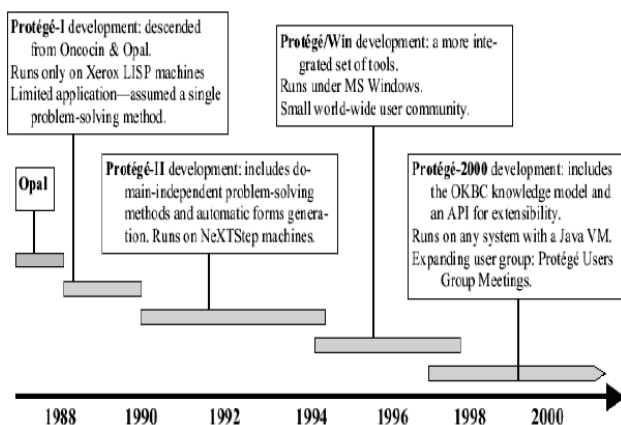


圖 4. Protégé 的發展歷程[1]

本體論建構工具 Protégé 是由美國史丹佛大學 SMI 中心 (Stanford Medical Informatics) 所研究開發的知識本體自由軟體。Protégé 不但是目前在世界使用來建置本體論最重要的平台之一，也是全球支援本體論的平台中最廣為使用的一個[4,6]。圖 4 即為 Protégé 的發展歷程，本研究使用版本為 Protégé 3.3.1。Protégé 是一套開放式、免費與制定完整的本體論建置工具，其最大特色是應用本體論觀念設定。此工具透過 Protégé-OWL Class、Protégé-Properties、Protégé-Forms、

Protégé-Individuals 與 Protégé-OWLViz 多類元件進行知識本體的編輯與製作。讓知識工作者可以建構一個以本體論為基礎的知識管理系統，且建置完成之本體論不但可轉換成各種不同的本體論格式，諸如 RDF(S)、OWL、XML 等；也可直接匯入資料庫系統，譬如 MySQL 及 MS SQL Server，支援的功能較其它本體論建置工具來的完整[11]。茲將其優點分述如下[4,6]：

- (1) 客製化的使用者介面：提供使用者快速簡便地建立特定領域的專有介面；
- (2) 彈性的知識溝通能力：Protégé 提供使用者存取廣泛格式的知識；
- (3) 強大的知識再用性：Protégé 提供自定使用者介面的功能，讓使用者可設計一個介面，並應用到不同的知識庫中；
- (4) 強大快速的知識嵌入能力：Protégé 可提供各種不同的元件支援，方便快速將知識嵌入到其他應用程式。
- (5) 歷經科學驗證：Protégé 為開放式軟體，並將所有 Protégé 的架構 (圖 5) 完全開放，可提供大眾自行新增或移除 Protégé 中的元件。

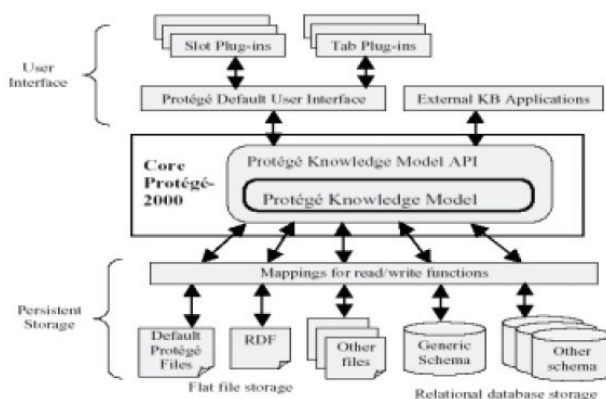


圖 5. Protégé 的系統架構圖[1]

3. 系統架構

3.1 系統運作與技術

圖 6 描繪出 OntoCrawler 運作的系統架構，茲將其各部分功能及相關技術細述如后。

- (1) Action：執行網頁蒐集的前置工作。將輸出區域清空，並把輸入的文字轉換成 URI 碼嵌入 Google 的搜尋網址，例如：
<http://www.google.com.tw/search?hl=zh-TW&q=%E5%AD%B8%E8%80%85&meta=>；最後再檢查輸入欄位是否為空的，如果

是的話將提醒使用者輸入相關查詢關鍵字 [10]。

- (2) LinkToGoogle：宣告一個 URL 元件並把 Action 方法中轉換好的 URI 碼加上 Google 搜尋網址；接著使用一個 BufferedReader 去讀取，並使用 while 迴圈逐行加入 String 元件 line，最後，把 line 輸出成 txt 檔，作為最後分析時的參考，檔案內容即為該頁面的 html 原始檔。

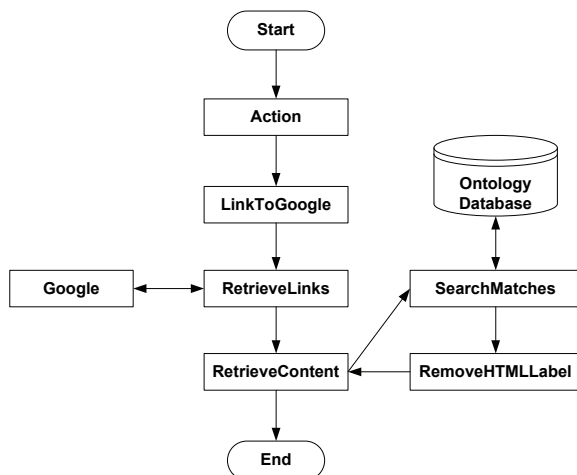


圖 6. OntoCrawler 的系統架構圖

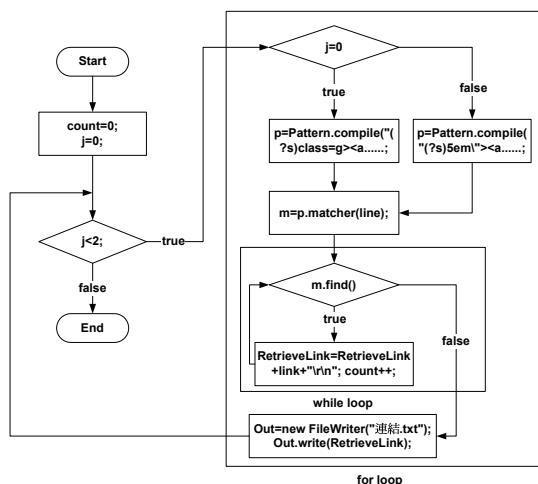


圖 7. RetrieveLinks 演算法

- (3) RetrieveLinks：會先將一個 int 元件 count 設為 0，方便爾後做為搜尋進度的顯示。使用前述的正規表示法從 line 中尋找是否有符合的 URL，因為 Google 網頁排版的內縮關係無法一次把所有連結取出，所以使用一個 for 迴圈並跑兩次。兩次所下的正規表示法的語法也稍微不同，為的就是方便將符合條件的相關連結全數取出；最後將所有連結加入 String 元件 RetrieveLink，同時也輸出成 txt 檔提供系

統做進一步的處理，相關處理流程示如圖 7。

- (4) RetrieveContent：由於 RetrieveContent 處理上較為費時，而且執行順序問題會讓系統介面看似完全停擺，所以採用執行緒 (thread) 來設計。從 Swing 的事件執行緒中擺脫束縛，允許在網頁蒐集過程中，適當的對介面進行變更；接著，把一個 Boolean 元件 crawling 設為 true 代表正在執行網頁蒐集，並將蒐集進度表的進度條設為 0%，介面按鈕標籤設為「停止」，輸入方塊亦改為不可編輯；然後用一個 BufferedReader 讀入 RetrieveLink，使用 while 迴圈逐行讀入，也就是每次檢視一個 URL 連結，並且實際的去連結該 URL，在判定其為何種編碼方式的網頁後，再以正確的編碼方式讀入該網頁的 html 原始檔，並加到 String 元件 s3 輸出成 txt 檔方便系統做進一步的處理。完成上述步驟後，使用 SearchMatches() 方法來判定該網頁是否為我們所設定的網頁蒐集範圍；如果是，就執行 RemoveHTMLLabel() 方法將 html 原始檔之中的 html 標籤移除，只剩下純文字內容，方便系統做進一步的後續處理與分析。最後，把蒐集過的網頁數除以總網頁數，該值即為目前的蒐集總進度，再將 RetrieveLink 清空，防止下一次的蒐集過程有誤；最後，將輸入方塊還原為可編輯，按鈕標籤還原為「開始」，crawling 改為 false，結束整個網頁蒐集程序，圖 8 顯示出 RetrieveContent 整個運作流程。

- (5) SearchMatches：支援 RetrieveContent 內部呼叫，來判定該網頁是否為我們所設定的網頁蒐集範圍。它連結我們的本體論資料庫（詳見 3.2 節），並取出裡面的內容與呼叫此方法時的 s3 內容做比對，假如有符合我們所設定的門檻值（本系統將門檻值設為 6，表示符合百分 50 的網頁蒐集條件），就會傳回一個內容為 true 的 Boolean 元件 matches，表示為符合條件的網頁；反之則傳回 false，表示該網頁不符合蒐集條件。
- (6) RemoveHTMLLabel：跟 SearchMatches 方法一樣是支援 RetrieveContent 內部呼叫，將 html 原始檔之中的 html 標籤移除。本系統使用了一連串的正規表示法來移除呼叫此方法時的 s3 中的 html 標籤，得到一個乾淨的純網頁內容文字檔。

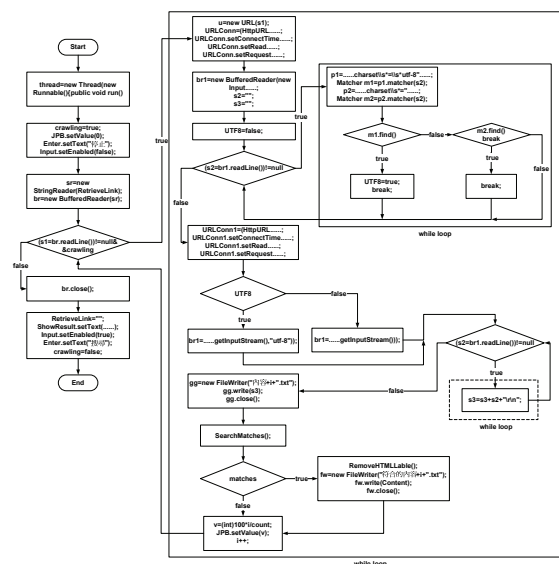


圖 8. RetrieveContent 演算法

3.2 建置本體論資料庫

現今本體論的研究大致可分為兩個方向：一種是針對特定領域建置大量的本體論，透過建立特定領域的本體論，支援該領域的知識剖析；另一種為研究如何建構與表示本體論[11]。本系統的本體論建置屬第一種，亦即利用建置好的研究學者本體論資料庫，支援OntoCrawler進行相關學者網頁的蒐集。本研究學者本體論資料庫的建置包括：研究學者相關概念統計與分析及本體論資料庫的建立兩階段，茲細述如后。

	A	B	C	D	E	F	G	H	I	J	K	L	M
	professor	office	Research	project	publication	Laboratory	link	courses	activities	service	education	resume	
1	郭大雄	○	○		○		○				○		
2	劉朝陽	○	○				○						
3	李文雄	○	○	○									
4	李珣山	○	○	○									
5	吳家鵬	○				○						○	
6	項楷	○	○	○	○					○			
7	陳文遠	○	○	○	○			○	○			○	
8	陳信希	○	○	○	○	○						○	
9	鄧曉明	○	○	○	○	○							
10	傅耀添	○	○				○					○	
11	林智仁	○			○			○					
12	林永豪	□	□							○	□		□
13	李德材	○	○	○	○			○					

圖 9 台灣大學研究學者網頁調查表

首先，我們針對研究學者網頁的首頁進行統計調查，藉以擷取學者首頁會出現的相關概念及其同義字。圖 9 例示出部分我們搜尋並記錄 50 筆台灣大學研究學者網頁中出現相關概念的統計分析，圖中符號「○」代表這個概念明確出現在首頁；符號「□」則代表這個概念出現在學者首頁的下一個超連結網頁。圖 10 就是我們參照[10]中所提 PCDIY 本體論建置的研究學者本體論架構。

圖 11 為本體論建置的第二階段：研究學者本體論資料庫的建置。此部份主要是將建置於 Protégé 中的本體論轉換成 MS SQL 資料

庫，步驟如下：

- (1) 使用 Protégé 定義圖 10 所分析出來的研究學者本體論架構，方便與其他研究學者分享我們的本體論。
- (2) 將 Protégé 知識庫所建置好的檔案匯出成 XML 檔再匯入 Excel 進行修正。這正是 Protégé 的強大知識再用性與快速知識嵌入能力的明證。
- (3) 最後再將 Excel 匯入 MS SQL Sever，完成本系統本體論資料的建置。

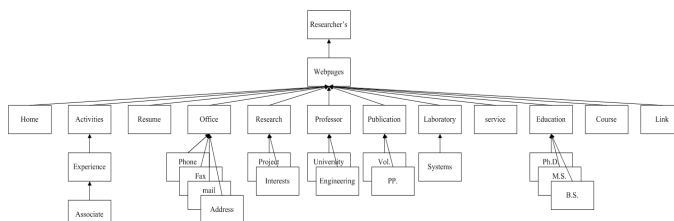


圖 10. 研究學者本體論架構

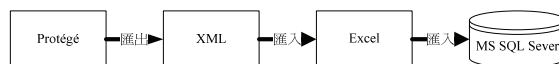


圖 11. 研究學者本體論資料庫轉換流程

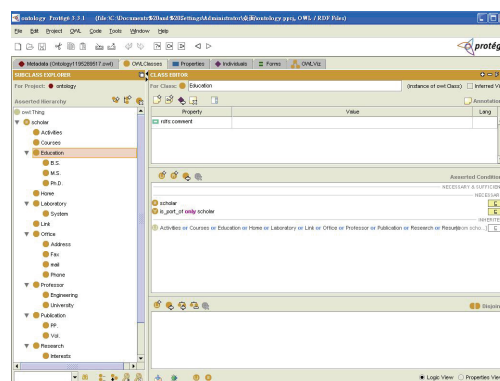


圖 12. Protégé 建置領域本體論的畫面

圖 12 顯示出利用 Protégé 開發領域本體論的畫面。畫面左邊區塊為研究學者網頁出現的相關概念的名稱。我們利用在左邊區塊上方 OWL Classes 裡的 Create Subclass 功能建立研究學者網頁相關概念名稱。在中間上方的 Properties 區塊可定義每一相關概念的屬性及其限制。圖 13 為研究學者領域本體論架構圖。以畫面中央框為例，學歷 (Education) 這個概念，後方連接相關概念為碩士 (M.S)、博士 (PH.D)、學士 (B.S)。應用上，我們將其設為相關概念，代表學歷不外乎就是這些相關概念的組合，方便 OntoCrawler 網頁蒐集器引用來與蒐集回來的網頁內容進行比對。如果比對到上述四項中的一項，我們就假設 Education 這個概念成立，藉以判斷網頁是否符合蒐集條件的依據。圖 14 為 Education 概念的 XML 表

示法。

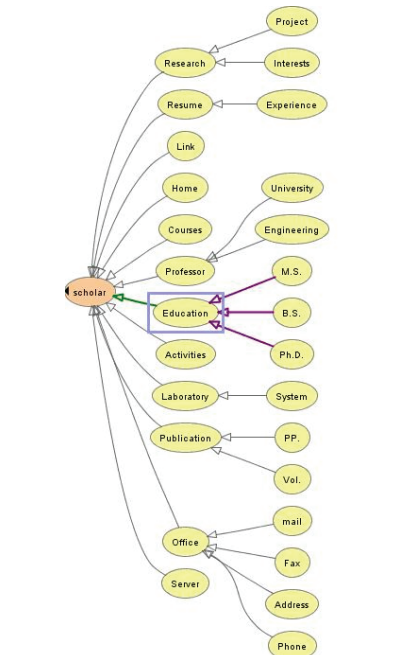


圖 13. 研究學者本體論架構圖

```

- <class>
  <name>Education</name>
  <type>owl:Class</type>
  <own_slot_value>
    <slot_reference>rdfs:subClassOf</slot_reference>
    <value value_type="class">@_:A36</value>
    <value value_type="class">scholar</value>
  </own_slot_value>
  <own_slot_value>
    <slot_reference>rdfs:type</slot_reference>
    <value value_type="class">owl:Class</value>
  </own_slot_value>
  <superclass>@_:A36</superclass>
  <superclass>scholar</superclass>
  <template_facet_value>
    <slot_reference>is_port_of</slot_reference>
    <facet_reference>:VALUE-TYPE</facet_reference>
    <value value_type="string">Instance</value>
    <value value_type="class">scholar</value>
  </template_facet_value>
</class>
- <class>
  <name>B.S.</name>
  <type>owl:Class</type>
  <own_slot_value>
    <slot_reference>rdfs:subClassOf</slot_reference>
    <value value_type="class">@_:A0</value>
    <value value_type="class">Education</value>
  </own_slot_value>
  <own_slot_value>
    <slot_reference>rdfs:type</slot_reference>
    <value value_type="class">owl:Class</value>
  </own_slot_value>
  <superclass>@_:A0</superclass>
  <superclass>Education</superclass>
  <template_facet_value>
    <slot_reference>is_port_of</slot_reference>
    <facet_reference>:VALUE-TYPE</facet_reference>
    <value value_type="string">Instance</value>
    <value value_type="class">Education</value>
  </template_facet_value>
</class>

```

圖 14. 研究學者 Education 概念的 XML 表示法

4. 系統驗證及比較

本實驗將以最為人使用的搜尋引擎 Google 及 Yahoo 當做比較對象。我們舉台灣大學資訊工程學系教授「郭大維」為關鍵字進行研究學者網頁蒐集標的為例，並說明如后。

首先，我們對 Google 及 Yahoo 鍵入相同的關鍵字「郭大維」，查詢回傳部分頁面如圖 15、16 所示，得到的比較結果示如表 1。其中， NW_T 表示所有的回傳網頁數 (Number of Totally Retrieval Webpages); NW_C 為正確的回傳網頁數 (Number of Correct Webpages); NW_R 則為相關回傳網頁數 (Number of Related Webpages)，但未必是正確的目標網頁。我們以公式(1)與(2)分別定義網頁準確率 (Precision Rate, R_P) 及回取率 (Recall Rate, R_R)。經逐一比對回傳頁面後，得到 Google 的準確率及回取率分別為 1%與 2%; Yahoo 則分別為 1%與 3%。

$$R_P = \frac{NW_C}{NW_T} \quad (1)$$

$$R_R = \frac{NW_R + NW_C}{NW_T} \quad (2)$$



圖 15. Google 查詢結果部分畫面



圖 16. Yahoo 查詢結果部分畫面

表 1. Google 與 Yahoo 前 100 筆搜尋結果比較

	NW_C	NW_R	NW_T	R_P	R_R
Google	1 筆	1 筆	100 筆	1%	2%
Yahoo	1 筆	2 筆	100 筆	1%	3%

我們以 Google 與 Yahoo 為網頁蒐集的基底，在 OntoCrawler 鍵入相同的關鍵字「郭大維」，查詢回傳部分頁面如圖 17 所示，得到的比較結果示如表 2。



圖 17. OntoCrawler 以 Google 為蒐集基底回傳畫面

表 2. 以 Google 與 Yahoo 為基底前 100 筆搜尋結果比較

	NW_C	NW_R	NW_T	R_P	R_R
OntoCrawler (Google)	1 筆	1 筆	5 筆	20%	40%
OntoCrawler (Yahoo)	1 筆	2 筆	4 筆	25%	75%

比較整理表 1 與表 2 的結果後，透過 OntoCrawler 技術的協助針對 Google 的準確率與回取率均提升 95% ($\frac{20\% - 1\%}{20\%}$, $\frac{40\% - 2\%}{40\%}$)；針對 Yahoo 的準確率與回取率更提升到 96% ($\frac{25\% - 1\%}{25\%}$, $\frac{75\% - 3\%}{75\%}$)。從上述比較中，除了顯現 OntoCrawler 確實在研究學者相關網頁蒐集上比搜尋引擎 Google 及 Yahoo 有較高的準確率及回取率；更展現及驗證我們所提出技術的可行性。

5. 結論

本論文已提出一本體論支援之網頁蒐集器 OntoCrawler，使用者只要輸入關鍵字，系統在領域本體論的支援下，主動提供關鍵字判斷比對，不僅排除因同名異義字所額外取回的網頁，更提升針對研究學者網頁蒐集的精準度，達成兼顧支援使用者查詢作業與相關搜尋系統核心之的。本技術已實際架構在 Google 及 Yahoo 兩搜尋引擎上主動蒐集相關研究學者網頁，並將結果存入資料庫提供後端系統做進一步的處理。初步的實驗結果驗證本技術確實能高度提升網頁蒐集的準確率與回取率。

本 OntoCrawler 更可輕易地與後端系統作結合，因為我們不僅採用開放源碼的設計哲

學，更利用諸多業界常見的開發工具，為的就是希望本技術能廣為相關學者採用。至於系統擴充性，只要套入其他領域的本體論就能夠將其設為 OntoCrawler 的搜尋主題，針對結合其他搜尋引擎的部分，只要瞭解其相關查詢語句的意涵，即可輕易轉換查詢基底。持續改善本系統的執行效能、擴充本體論資料庫與相關聯結界面及持續發展與後端系統結合的中介程式等都是我們未來的研究工作。

致謝

這篇論文是國科會計劃 (NSC-95-2221-E-129-019) 的部份研究成果，我們由衷地感謝國科會的經費支援。

參考文獻

- [1] J.H. Gennari et al., "The evolution of Protégé: and environment for knowledge-based systems development," *International Journal of Human-computer studies*, 58, 2003, 89-123.
- [2] S.Y. Yang and C.S. Ho, "FAQ-master: An Intelligent Web Information System for PCDIY," *Proc. of 2004 National Workshop on Internet and Communication Technology*, Taipei, Taiwan, 2004, 143-148.
- [3] S.Y. Yang, "A Website Model-Supported Focused Crawler for Search Agents," *Proc. of the 9th Joint Conference on Information Sciences*, Kaohsiung, Taiwan, 2006, 755-758.
- [4] W.E. Grosso, H. Eriksson, R.W. Ferguson, J.H. Gennari, S.W. Tu, and M.A. Musen, "Knowledge Modeling at the Millennium: the Design and Evolution of Protege-2000," SMI Technical Report, SMI-1999-0801, 1999.
- [5] 邱俊諧、陳均銘、趙銘，多媒體資料庫檢索系統，逢甲大學資訊工程學系專題報告，民91。
- [6] 許蒞彥、吳旻雲、張成儀，"應用OWL建構Ontology-based自由軟體知識分享Web平台," *ICOS 2005 開放源碼研討會暨「國科會自由軟體研發專案」成果發表會*，民94。
- [7] 許薰尹、曾憲雄，宋詞斷詞與本體論之建置，國立交通大學理學院（網路學習）學程碩士論文，民95。
- [8] 陳麴合、蔡玉娟，超連結與關鍵字頻分析之搜尋引擎研究，國立屏東科技大學資訊

- 管理學系碩士論文，民91。
- [9] 楊勝源、朱毓君、何正信，“以本體論強化網路FAQ系統之解答整合能力,” 第六屆人工智慧與應用研討會論文集，高雄，台灣，2001, 52-57.
- [10] 楊勝源、吳志偉、何正信，“結合規則挖掘與行為預測來強化網路資訊查詢系統,” 第六屆人工智慧與應用研討會論文集，高雄，台灣，2001, 574-579.
- [11] 簡宏傑、蘇東興，以語意網為基礎之特殊教育學習教材庫搜尋引擎之研究與實作，明新科技大學資訊管理學系碩士論文，民95。
- [12] 羅友志譯，大師談JAVA，上奇科技股份有限公司，民92，6-1~6-66。