

# 行政院國家科學委員會專題研究計畫 成果報告

## 以意圖為基之自然語言網路搜尋支援系統 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 99-2410-H-343-023-  
執行期間：99年08月01日至100年07月31日  
執行單位：南華大學電子商務管理學系

計畫主持人：王昌斌  
共同主持人：陳宗義、楊惠媚  
計畫參與人員：碩士班研究生-兼任助理人員：龔純慧  
碩士班研究生-兼任助理人員：謝雯雀  
碩士班研究生-兼任助理人員：熊偉傑  
碩士班研究生-兼任助理人員：王宏遠

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中華民國 100 年 10 月 28 日

# 以本體論結合意圖為基之智慧型搜尋引擎之研究

喻英雲  
南華大學  
資訊管理學研究所  
violet.yu@gmail.com

王昌斌  
南華大學  
資訊管理學研究所  
cbwang@mail.nhu.edu.tw

張大薇  
大同技術學院  
企業管理系  
jfk8470@ms2.ttc.edu.tw

## 摘要

目前一般人在網路上找資料尋求協助的方式大致有以下二種，一為至搜尋引擎輸入關鍵字，雖然回應速度很快，但往往查得大批非相關資料，需一一檢視，花費許多時間。二則於特定地方如網路論壇等尋求協助，雖然可避免上述問題而找到較合適的解答，但此種人工回覆方式，必須付出比上述方法更多的等待時間。針對上述問題，本研究希望能找出一種方式，以自然語言輸入問題描述之後，自動且快速地於網路上搜尋到相關資料，並且可以獲得正確的知識。

本研究以中文問句做分析，包含三個步驟，首先以中研院 CKIP系統做斷詞及詞性標記，且以 5W1H系統化的歸納方法來搭配有限自動機(Finite Automata)的運作方式以比對方式找出搜尋的意圖型態及關鍵字，接著以關鍵字搭配領域本體論 (Ontology)擴展為語意網 (Semantic Net)，最後將對應 5W1H資料庫所得出之搜尋意圖對應字與語意網做結合，於網路搜尋引擎中做搜尋處理。本研究實驗結果顯示我們所提出的加入 5W1H對應字方法確可改善其搜尋之準確率，以幫助查詢者於網路中搜尋出符合本身意圖之答案。

關鍵詞：5W1H、意圖、本體論、有限自動機

# 以本體論結合意圖為基之智慧型搜尋引擎之研究

摘要腦應用已相當普及 (行政院主計處電子處理資料中

目前一般人在網路上找資料尋求協助的方式大致有以下二種，一為至搜尋引擎輸入關鍵字，雖然回應速度很快，但往往查得大批非相關資料，需一一檢視，花費許多時間。二則於特定地方如網路論壇等尋求協助，雖然可避免上述問題而找到較合適的解答，但此種人工回覆方式，必須付出比上述方法更多的等待時間。針對上述問題，本研究希望能找出一種方式，以自然語言輸入問題描述之後，自動且快速地於網路上搜尋到相關資料，並且可以獲得正確的知識。

本研究以中文問句做分析，包含三個步驟，首先以中研院 CKIP系統做斷詞及詞性標記，且以 5W1H系統化的歸納方法來搭配有限自動機 (Finite Automata)的運作方式以比對方式找出搜尋的意圖型態及關鍵字，接著以關鍵字搭配領域本體論 (Ontology)擴展為語意網 (Semantic Net)，最後將對應 5W1H資料庫所得出之搜尋意圖對應字與語意網做結合，於網路搜尋引擎中做搜尋處理。本研究實驗結果顯示我們所提出的加入 5W1H對應字方法確可改善其搜尋之準確率，以幫助查詢者於網路中搜尋出符合本身意圖之答案。關鍵詞：5W1H、意圖、本體論、有限自動機

## 1. 前言

台灣的資訊產業在世界上佔有一席之地，資訊業的蓬勃發展以及政府的重視，使得電腦設備的普及率相當高，截至 95年12月底止，我國個人電腦設置數達1,041.5萬台，另外我國政府機關設置暨應用電腦的人機比：行政機關為 0.93(電腦338,639台)、公營事業機構為 0.66(電腦139,039台)、公立學校為26(電腦602,799台)、公立研究機構為 1.12(電腦12,208台)，顯示政府機關積極推動資訊化作業，電

心，2006)。

在電腦的各項應用之中，以網路應用最為廣泛，影響也最為深遠，根據調查，我國經常上網人口數已達 990百萬人，連網主機數高居世界第 6 (Internet Systems Consortium, 2007)，雖然網路廣為大眾應用，不過是這十幾年的事情，卻顛覆了傳統的學習，提供了另一追求知識的管道，拜網路蓬勃發展之賜，生活中食、衣、住、行、育、樂各方面所需的知識，幾乎均可於網路上搜尋而得。

有鑑於此，教育部「挑戰二〇〇八~e世代人才培育計畫」將「建構全民網路學習系統」列為計畫之一，其中「開發中小學網路學習內容計畫」提到「在資訊化時代的社會，國民具有資訊應用知能與完善的數位化學習環境是國家發展的基礎條件」，目前各級學校之硬體設備均已建置完成，惟網路學習之內涵尚有待加強，希望能透過網路經驗分享並推廣。在網路教學資源共創、共享的環境中，使網路學習內容得以普及應用並發展。(教育部全球資訊網)

就目前發展現況來看，網路資訊的內容相當豐富，但這也產生了一個問題，如何在數量龐大的網路文件中找到我們想要的資料。以目前來說，關鍵字查詢是最普遍、也是一般成認最快的作法，但是這時候，網路原本資料豐富的優點反而成了一個很大的問題，動輒百萬筆的資料，以不到 1秒的時間迅速列出，可是使用者卻需花上不少時間作篩選的動作。

網路論壇也是網路上相當蓬勃的一項發展，許多知識性的網站均設置此一功能，希望藉由大家的經驗分享，增加彼此的交流，以期增長知能，此舉固然毋須像關鍵字查詢般一一過濾資料，但是必需被動等待答覆，通常並非即時可得，甚至遇到無人答覆之困境。

如何擷取二者之長以轉化成最有用的搜尋方

式是當前重要課題，既要快速求解，又能增加搜尋結果之準確度。因此本研究針對問題詢答機制為主，做更深入的探討。對使用者所下的自然語言問句做解析，以 SWIH 有系統化的歸納出問句型態模型[16]，快速解析所屬之意圖，且配合所對應之意圖資料庫並做領域本體論 (Ontology) [9,15]之語意網擴展，經意圖與關鍵字結合轉換後的字串於網頁整合型搜尋引擎機制中搜尋，期望讓使用者有效率地依其意圖獲得所需答案。

## 2. 文獻探討

### 2.1 SWIH

張鐘尹[7]認為，對於一個自然語言問句，除了關鍵詞部分之外，仍有其他可做為分辨問句間差異性的部分，本研究於表 1 分類出 SWIH 的意圖。

表 1 SWIH 意圖類型

意圖	問句	關鍵字	意圖
What	什麼是四則運算？	四則運算	四則運算
What	何謂學習障礙？	學習障礙	學習障礙
What	什麼是數學？	數學	數學
Who	誰知道輕度障礙？	輕度障礙	人名
Who	誰會數學除法？	數學除法	人名
Who	誰了解數學教學？	數學教學	人名
Why	為什麼會有學習障礙？	學習障礙	原因
Why	為什麼不會加法？	加法	原因
Why	你為何會發生車禍？	發生車禍	原因
When	何時開始學數學？	學數學	時間
When	他離開是什麼時候？	離開	時間
When	他畢業是什麼時間？	畢業	時間
Where	台北是在哪？	台北	地點
Where	嘉義是什麼地方？	嘉義	地點
Where	阿里山的地點？	阿里山	地點
How	要如果取得駕照？	取得駕照	方法

### 2.2 語意網

語意網是一種以網路結構為基礎之知識表示法，主要利用節點 (Node)與弧(Arc)進行知識架構組成，當中，節點可以表示物件、概念或特定領域中之情境，而弧則表示節點間之關聯。因此，Ruan et al. [12]利用語意網之概念，描述醫學專有名詞與其相關資訊來源間之關係，以建構線上醫學資料辭典。

Smolentsev[14]提出一種稱為動態語意網 (Dynamic Semantic Network, DSN)之知識表示法。其以語意網形式描述資料之概念、物件和關係的階層屬性，具有 (1)整合宣告式知識與程序式知識、(2)可平行查詢語意網之各元件及 (3)根據累積之知識發展網絡關係等特性。此外，該表示法的各個知識節點包含運算程序 (即解釋所收集資訊之資訊處理演算法，用以發展問題答案、知識節點間聯繫方式與某特定時間之行為)，因此具有自我學習之特性。

### 2.3 自然語言處理技術

在李坤霖[3]的研究中，透過語意文法及停用字的篩選，將詢問問句分為意圖部分及關鍵詞部分，意圖部分是找出使用者的意圖目的。關鍵詞部分則利用事先建好的關鍵詞樹作比對，而比對過程中會分別求出詞間相似度及節點間的機率，進而找出最佳路徑，這類利用關鍵詞比對來處理語意分析的方式為目前國內自然語言學者較常使用。

### 2.4 詢答系統 (Q&A System)

根據 QA TREC Specifications [11]之定義，詢答系統應可依據使用者的問題來回答其所關心的資訊，而不只是輸出相關文件。

由此可知，詢答系統基本上主要由資訊檢索系統及資訊擷取系統所組成，資訊檢索系統先依據使用者的“問題”篩選出所有可能的文件，然後再交給資訊擷取系統擷取出使用者所關心的資訊。

一般的詢答系統在前端均有問題處理器 (Question Processor)先行處理使用者之問題，其主要目的是希望確認出『問題的種類 (Information of asking point)』以及『建立關鍵字』 [13]。

### 3. 圖形化語意網轉換機制

如圖 1 所示，本研究以語意分析機制 (Semantic Analysis Mechanism)、語意網轉換機制 (Semantic Net Transformation Mechanism) 及意圖轉換機制 (Intention Transformation Mechanism) 為核心，這三種機制主要是幫助我們產出問題意圖 (Intention)、關鍵字及對應字結合。

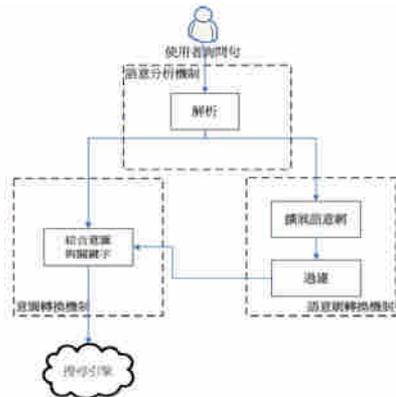


圖 1 圖形化語意網轉換機制架構圖

#### 3.1 語意分析機制

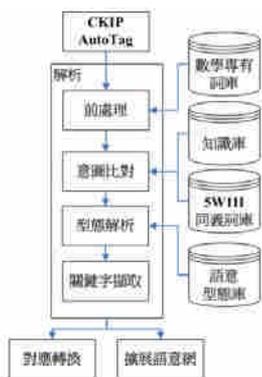


圖 2 語意分析機制

本機制如圖 2 所示，主要包含五步驟，各步驟之主要功能如下：

##### 3.1.1 CKIP AutoTag

此步驟將使用者的問題做斷詞處理與詞性標記，並將句中關鍵字標記出來，以「如何學習除法」這個句子為例，利用 CKIP[4] 進行斷詞及標記後，

將分為三部分「如何 (D) 學習 (VC) 除法 (Na)」，當中 D、VC、Na 分別表示「副詞」、「及物動詞」、「普通名詞」。

##### 3.1.2 前處理 (Pre-Processing)

將步驟 1 斷詞完成卻出現錯誤的詞做結合，如「四則運算」此一專有名詞被斷成「四 (Neu) 則 (Nf) 運算 (VC)」三部分，Neu、Nf 分別表示「數詞定詞」、「量詞」，透過此步驟與數學專有詞庫做比對，可將「四」「則」「運算」三個詞合併為單一專有名詞「四則運算 (Na)」。

##### 3.1.3 意圖比對 (Intention Matching)

將問句字串配合 5WH 同義詞庫比對出所屬問句的意圖類型，其詞庫是由牛維娟及楊宸彥 [2,6] 之研究論文改良後歸納所得，如表 2 所示。

表 2 5WH 同義詞資料表

what	什麼	何謂	什麼是
how	如何做	怎樣做	要如何
why	為什麼	怎麼會這樣	何故
who	誰	何人	什麼人
when	何時	什麼時候	在什麼時間

##### 3.1.4 型態解析 (Type Parsing)

本研究搜集網路上之使用者問句，歸納出 5WH 之六種型態模型：How、What、Who、Where、What、Why，如圖 3 至圖 8 所示，接著以圖 9 所示之 Knuth-Morris-Pratt (KMP) [8] 演算法做型態解析，以建立有限自動機 (Finite Automata) 與語意型態庫做快速掃描，進而判斷出型態類型，如圖 10。

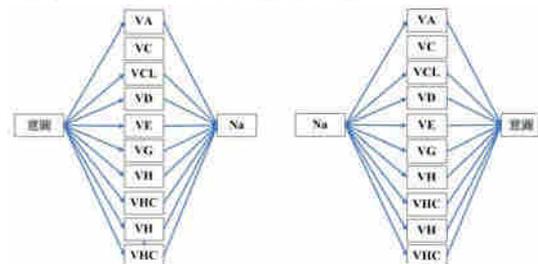


圖 3 型態一：How



圖4 型態二：What

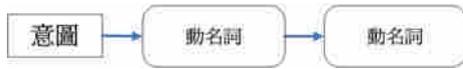


圖5 型態三：Who

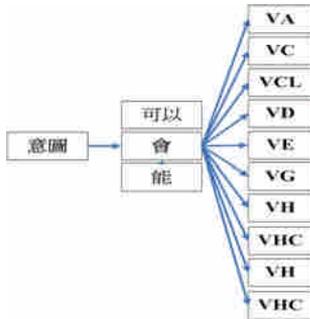


圖6 型態四：Why

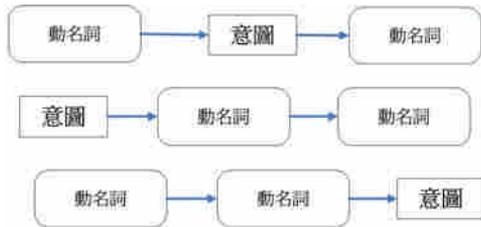


圖7 型態五：When



圖8 型態六：Where

### 3.1.5 關鍵字擷取 (Keyword Extract)

經上述步驟，問句字串中包含名詞 (N) 的字串被擷取出來作為關鍵字，如 Na、Nb、... 等，可供下一步驟做擴展語意網之用。

**Kunth-Morris-Pratt演算法：**  
 輸入：P及T，字樣及本文字串； f 陣列(失敗鏈陣列)如果T的長度事前不可預知，在演算法使用它的部分可以用一個 end-of-string 測試來取代。 P的長度應該在設定 f 陣列時已求出。  
 輸出：T字串中 P的一份拷貝之開始註標。如果找不到可與 P 匹配者，註標值將為 TL+1。

```
function KMPmatch ( P, T : String; f : IndexArray ) : index;
var
  j, k : index; { j indexes text
  characters; k indexes the pattern and
  f array. }
begin
  j := 1; k := 1; while j ≤ T.L and k ≤
  P.L do if k = 0 or j k t = p then j :=
  j + 1; k := k + 1 else { follow fail
  arrow } k := f[k]

  end { if } end { while }; if L k ≤ P then L KMPmatch := j
  - P { match found } else KMPmatch := j { := +1 L jT, no
  match } end { if }
```

end { KMPmatch }

圖9 Knuth-Morris-Pratt 演算法

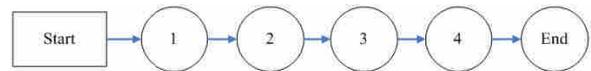


圖10 有限自動機流程圖

### 3.2 語意網轉換機制本機制包括三

步驟，如圖 11所示。

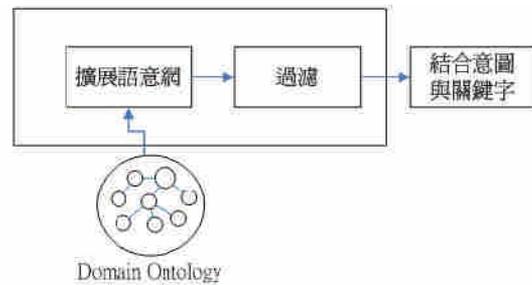


圖11語意網轉換機制

### 3.2.1 擴展語意網 (Extended Semantic Net)

被擷取出的關鍵字，利用領域本體論 (Domain Ontology) 延伸出相關概念而形成語意網，如圖 12 所示。

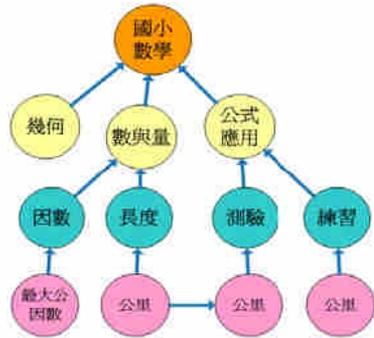


圖12 國小數學領域之 ontology架構

### 3.2.2 過濾(Filtering)

根據本體論技術，在系統上設立一參數門檻值，可依使用者之需要做增減。如圖 13。

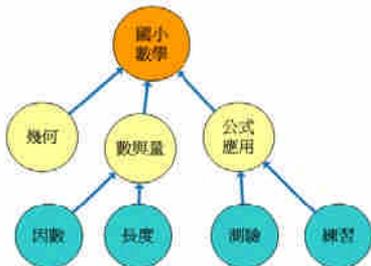


圖13 國小數學領域之 ontology架構

### 3.3 意圖轉換機制

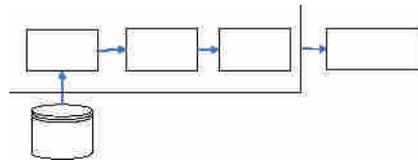


圖14 意圖轉換機制

如圖 14所示，本機制目的在將意圖與關鍵字結合之後，自 5W1H對應資料庫將「動、名詞」對應出來，接著轉換成一般句型並過濾詞性後置於搜尋引擎進行資料搜尋。

#### 3.3.1 對應比對 (Corresponsively) 透過

5W1H對應資料庫如表3的比對，將所對應的「動、名詞」找出來。

根據Han and Kamber[10]所提之Apriori演算法，可自大量交易資料中找尋出隱藏之有用資訊，並以此做為關係字而搜尋出隱藏的 5W1H對應字。例如：顧客購買印表機時，是否會一併購買報表紙?或者購買IBM PC主機時，是否傾向於搭配

ViewSonic螢幕。表3 5W1H對應資料

意圖 類型	類型	Domian Keyword	對應 名詞	對應 動詞
		數學知識	來源	發生 發現 過程
	等腰 三角形 面積		理由 原因 ...	
數學 類	四則 運算		..	
	數學的 教學模 式			
	平行 四邊 形			
		數學知識	日期	發生
		等腰三角形面積	起源	發現
wh e n	數學 類	四則運算	時間	發源
		數學的教學模式	.....	發跡
	平行 四邊 形			
		數學知識		介紹
	等腰 三角 形面 積			
數學 類	四則 運算			
	數學 學模 式			
	平行 四邊			

本研究假設方式將已知關鍵字『四則運算』

及意圖『How』，透過搜尋引擎做文件搜尋，找出100篇與國小數學相關之文件，作為研究樣本。以此樣本進行 Apriori演算法計算，將關鍵字及隱藏之 5W1H對應字的支持個數及信心水準彙整出來。圖15主要將搜尋後之文件做關鍵字及對應字支持個數的彙整，以幫助本研究做關鍵字及 5W1H對應字之相關程度驗證。



圖15 彙整關鍵字及對應字次數

首先找出一頻繁項目組，透過 Apriori\_gen產生候選項目組，由 Lk-1產生Ck。對產生的候選項目，我們將利用 has\_infrequent\_subset演算法來做修剪，以節省搜尋時間。

### 3.3.2 排序轉換 (Ranking Transformation)

經上一步驟，從資料表中對應出的「動、名詞」配合關鍵字做排序轉換，本研究經過文件訓練後，得出「四則運算」+「方法」=「名詞」+「對應字」之結果，透過王亭雅 [1]所提之方式可直接進行結合轉換。

以圖16為例：

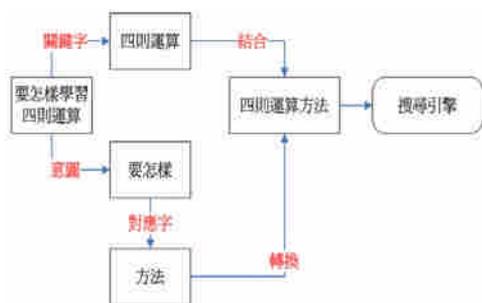


圖16 轉換示意圖

### 3.3.3 詞性過濾 (Filtering)

將轉換過後的字串做詞性過濾，讓下一步驟整合型網頁搜尋引擎機制做搜尋處理，以解決使用者之問題。

## 4. 系統開發與實作

### 4.1 資料來源

本研究之樣本收集，主要透過搜尋引擎做資料搜尋，假設以『等腰三角形面積』、『平行四邊形』及『智能障礙』三個關鍵字，配合『How』、『who』及『why』三種意圖型態所對應之『動、名詞』對應字做網頁搜尋，將有關國小數學網頁之前三十篇有效樣本作為本研究之測試樣本，並以黃耀民 [5]所提之準確率 (Precision ratio)公式進行比較評估，公式如下：

$$\text{precision} = \frac{\{relevant\} \cap \{retrieved\}}{\{retrieved\}}$$

### 4.2 實驗結果

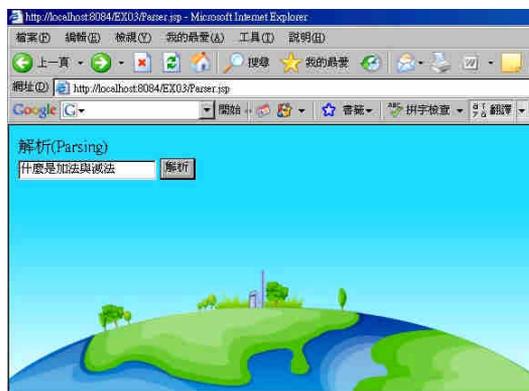


圖17 操作介面

圖 17為使用者以自然語言輸入一問句後，經系統解析可進入下一步驟。

圖 18顯示經解析後之問句，可比對出關鍵字及意圖，並分析出使用者問句之型態。

系統將意圖及關鍵字與 5W1H意圖對應資料庫比對後。選擇出所需之意圖關鍵字，可置於搜尋引擎中做搜尋。

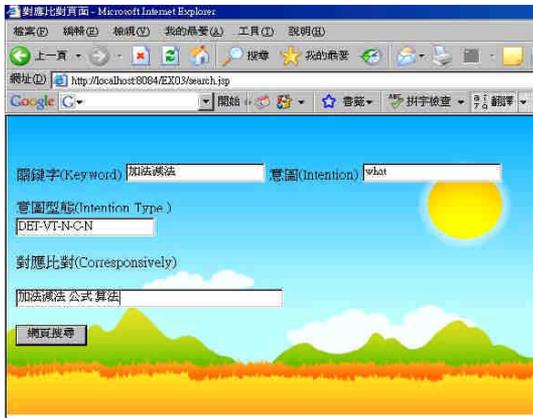


圖18 驗證對應頁面



圖19 網路搜尋頁面

圖 19主要是將關鍵字結合系統分析之對應字於網路上做搜尋，期望能利用此一機制解決使用者的問題。

表4 於Google搜尋資料之準確率統計表

	how	why	who
單一關鍵字	0.433	0.133	0.000
兩個關鍵字	0.533	0.500	0.133
三個關鍵字	0.567	0.567	0.167
四個關鍵字	0.567	0.600	0.167

表5 於Yahoo搜尋資料之準確率統計表

	how	why	who
單一關鍵字	0.300	0.1	0.000
兩個關鍵字	0.367	0.333	0.1
三個關鍵字	0.533	0.367	0.167
四個關鍵字	0.533	0.433	0.167

表4及表5表示，本研究將關鍵字及對應字於

搜尋引擎 Google及 Yahoo做搜尋後，以 Precision 公式來評估之準確率。

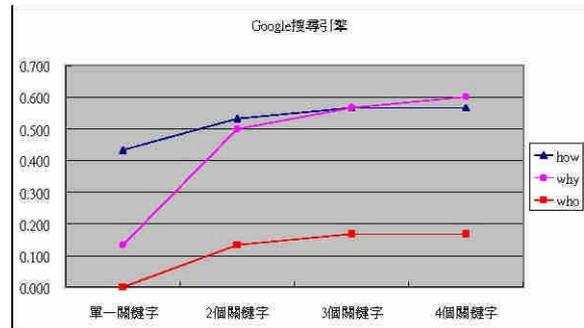


圖20 於Google搜尋資料之準確率比較圖

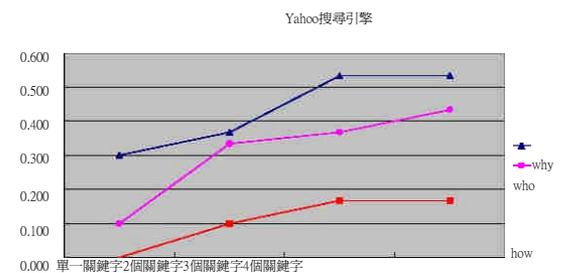


圖21 於Yahoo搜尋資料之準確率比較圖

圖 20及圖 21表示，本研究加入對應字後，於 google及 yahoo兩搜尋引擎做資料搜尋，經計算後約可提高 30%之準確率。

## 5. 結論與未來展望

本研究主要是在探討自然語言的處理技術、意圖與關鍵字的擷取、擴展語意網、型態解析及問題轉換。作法為針對幾項問題進行分類，自使用者問句中擷取出意圖類型及關鍵字，從而進行語意擴展及結合轉換，最後透過搜尋引擎尋找更適合的資料以回覆使用者，進而達到本研究之目的。

### 5.1 研究結

(1) 論：透過本系統，使用者更容易了解其敘述問句之真正涵意。

(2) 圖形化語意網可應用於其他相關領域，例如知識地圖等，而透過語意網之擴展，使用者能挖掘出豐沛的知識。

- (3) 本研究歸納出 5W1H 之問句類型，可提高資料搜尋之準確性及速度。
- (4) 自然語言經意圖及關鍵字之轉換後，可大幅提高搜尋之準確率。
- (5) 有限自動機的建立，可改善比對的效率及準確率。

## 5.2 未來展望

在 5W1H 對應庫中，意圖類型的對應字眾多，可透過權重庫來設定權重值，利用使用者每一筆處理過之資料做訓練，形成一回饋機制，未來希望能在針對此一問題做深入探討，達到本研究之最終目的。

## 參考文獻

- [1] 王亭雅(2003)。網頁搜尋引擎問答。未出版之碩士論文，國立中正大學資訊工程研究所，嘉義縣。
- [2] 牛維娟(2003)。應用於 USENET 之 Q&A 系統之研究與設計。未出版之碩士論文，元智大學資訊管理研究所，桃園縣。
- [3] 李坤霖(2000)。網際網路 FAQ 檢索中意圖萃取及語意比對之研究。未出版之碩士論文，國立成功大學資訊工程研究所，台南市。
- [4] 中央研究院中文斷詞系統，URL：<http://ckipsvr.iis.sinica.edu.tw/>
- [5] 黃耀民(2005)。以字句擷取為基礎並應用於文件分類之自動摘要之研究。未出版之碩士論文，國立台灣師範大學資訊工程研究所，台北市。
- [6] 楊宸彥(2002)。運用剖析概念圖進行中文詢答之研究。未出版之碩士論文，國立台灣大學資訊工程學研究所，台北市。
- [7] 張鐘尹(1997)。漢語對話中的疑問句。未出版之碩士論文，國立台灣大學語言學研究所，台北市。
- [8] Baase, S., & Van Gelder, A. (2000). *Computer Algorithms: Introduction to Design and Analysis*. 3<sup>rd</sup> Ed. Addison Wesley.
- [9] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- [10] Han, J., & Kamber, M. (2001). *Data Mining : Concepts and Techniques*, 5th Ed. (pp. 230-235). San Francisco: Morgan Kaufmann.
- [11] QA Track Specifications, URL : <http://www.research.att.com/~singhal/qa-track-sepc.txt>
- [12] Ruan, W., Buerkle, T., & Dudeck, J. (2000). Object-oriented design for automated navigation of semantic networks inside a medical data dictionary. *Artificial Intelligence in Medicine*, 18(1), 83-103.
- [13] Srihari, R., & Li, W. (2000). A Question Answering System Supported by Information Extraction. *Meeting of the North American Chapter of the Association for Computational Linguistics* (pp.166-172).
- [14] Smolentsev, S. V. (2002). The Identification and Self-Organization Problems in Dynamic Semantic Networks. *IEEE International Conference on Artificial Intelligence Systems* (pp. 35-39).
- [15] Smith, B., & Welty, C. (2001). Ontology: Towards a New Synthesis, In *Proceedings of the International Conference on Formal Ontology in Information Systems* (pp. 3-9), Ogunquit, Maine.
- [16] Tsai, W. D. (2000). The Hows of Why and the Whys of How. 台灣語言學的創造力學術研討會。

# 出席國際學術會議心得報告

計畫編號	NSC 99-2410-H-343-023-
計畫名稱	以意圖為基之自然語言網路搜尋支援系統
出國人員姓名 服務機關及職稱	王昌斌 南華大學教授
會議時間地點	民國 100 年 7 月 18 日至民國 100 年 7 月 21 日
會議名稱	The 2011 International Conference on Image Processing, Computing Vision & Patterns Recognition
發表論文題目	Applying Support Vector Regression to Predict Structure in Image Completion

## 一、前言

此次出國主要目的是到美國拉斯維加斯參加 IPCV2011 會議發表論文，該會議屬於 WORLDCOMP 2011 的一個分會，WORLDCOMP 2011 共包括 12 個分會，是一個非常大的會議，共有 88 個國家超過一千人參加。會議時間 7 月 18 日至 7 月 21 日共四天，開會場地是 Monte Carlo 飯店。

此 7 月 18 日早上 8:30 在 Monte Carlo 飯店的戲院開幕，接下來是一系列的演講及論文發表，7 月 18 日晚上 9:10 大會晚宴，晚宴期間認識了一些與會學者，7/19~7/20 挑選了幾篇有興趣的論文聽聽順便跟一些與會學者談談，發現美國一些學校也有招生壓力，也發現美國電腦相關科系招生不好，多數美國人不念電腦相關科系，值得審思。

此次會議中，有來自世界各地的學者與專家，其中不乏且具有豐富實務經驗的專家，但參與人員仍然以學術界的學者為主。所發表之上千篇論文中其議題與內容極具實用性與開創性，並在學術上有極高之價值。

## 二、參加會議經過

在本次研討會的四天的議程中，第一天行程中安排各領域的知名學者進行專題演講，將各人目前研究成果與與會來賓分享。隨後針對各個不同的議題進行論文發表，本次會議涵蓋範圍如下：

- Software tools for imaging
- Image-based modeling and algorithms
- Illumination and reflectance modeling
- Motion and tracking algorithms and applications
- Biometric authentication
- Event recognition techniques in image sequences
- Medical image processing and analysis
- Image geometry and multi-view geometry

- Segmentation techniques
- Geometric modeling and fractals
- Scene and object modeling
- Image data structures and databases
- Image compression, coding, and encryption
- Image display techniques
- Digital imaging for film and television
- Image formation techniques
- Image generation, acquisition, and processing
- Image feature extraction
- Novel document image understanding techniques + OCR
- Enhancement techniques
- Novel noise reduction algorithms
- Mathematical morphology
- 3D imaging
- Watermarking methods and protection
- Wavelet methods
- Mosaic, image registration and fusion methods
- Color and texture
- Image restoration
- Printing technologies
- Interpolation techniques
- Shape representation
- Video analysis
- Indexing and retrieval of images (image databases)
- Signal and speech processing
- Object recognition
- Multi-resolution vision techniques
- Face and gesture
- Stereo vision
- Soft computing methods in image processing and vision
- Machine learning technologies for vision
- Performance analysis and evaluation (real-time vision)
- Camera networks and vision
- Sensors and early vision
- Active and robot vision
- Cognitive and biologically inspired vision
- Fuzzy and neural techniques in vision
- Graph theory in image processing and vision
- Special-purpose machine architectures for vision
- Dimensionality reduction methods in pattern recognition
- Classification and clustering techniques
- Symbolic learning
- Statistical pattern recognition
- Invariance in pattern recognition
- Knowledge-based recognition
- Structural and syntactic pattern recognition

- Applications including: security, medicine, robotic, GIS, remote sensing, industrial inspection, nondestructive evaluation (or NDE), ...
- Case studies
- Emerging technologies
- Workshop on Multimedia Systems and Applications:

-

所提之論文「Applying Support Vector Regression to Predict Structure in Image Completion」經大會審核，獲准於 19 July 2010 下午 3:40 在 Session 9「Novel Algorithms and Applications」中發表。在報告結束之後，有多位專家學者提出他們的疑問與看法，經我們進一步的解說後並與之討論，獲益良多。

### 三、與會心得

- (1) 本次的研討會是世界性的學術研討會，藉由參加本次的會議報告與討論的過程，不只達到學術交流的目的，更能吸取到最進許多國際知名學者的研究成果令人受益匪淺。要成為一位優秀的國際性學者，除了不斷地充實自己各方面的能力避免閉門造車之外，更需不斷地吸取國際舞台的經驗，甚至參加國際性的研究計畫。

# 國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/26

國科會補助計畫	計畫名稱: 以意圖為基之自然語言網路搜尋支援系統
	計畫主持人: 王昌斌
	計畫編號: 99-2410-H-343-023- 學門領域: 資訊管理
無研發成果推廣資料	

99 年度專題研究計畫研究成果彙整表

計畫主持人：王昌斌		計畫編號：99-2410-H-343-023-				計畫名稱：以意圖為基之自然語言網路搜尋支援系統	
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	0	1	40%	篇	
		研究報告/技術報告	0	1	20%		
		研討會論文	0	1	40%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （本國籍）	碩士生	4	4	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	50%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	1	50%		
		專書	0	0	100%		章/本
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （外國籍）	碩士生	4	4	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	<p>無</p>
--	----------

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

# 國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表  未發表之文稿  撰寫中  無

專利： 已獲得  申請中  無

技轉： 已技轉  洽談中  無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究計畫之效益與貢獻為下列幾點：

（1）中文自然語言搜尋技術創新

本計畫將歸納 5W1H 之一般化模型，使之能更快速判斷出問句類型及意圖，利用 5W1H 六種意圖擴展語意網取出對應字串再與 CKIP 中文斷詞系統取出之關鍵字結合做搜尋，較傳統單一搜尋方法提高準確率。為此一搜尋方法實為中文自然語言搜尋技術之創新。（實驗驗證結果）

（2）提昇搜尋準確度與節省搜尋成本

利用本計畫所建立之智慧型搜尋引擎，使用者僅需輸入自然語言，系統即會自動解析與搜尋。相較於傳統一般關鍵字搜尋，使用者需求之資料能較早被尋得，因而達到節省搜尋時間之目標。

（3）應用領域將可擴大適用（依不同領域建構後，再建立上層分類機制）