

南 華 大 學

資訊管理學系

碩士論文

以本體論結合意圖為基之智慧型搜尋引擎之研究

An Intelligent Search Engine

Based on Ontology and Intention Comparison

研 究 生：喻英雲

指 導 教 授：王昌斌 博士

中 華 民 國 九 十 七 年 六 月

南 華 大 學

資訊管理學研究所

碩 士 學 位 論 文

以本體論結合意圖為基之智慧型搜尋引擎之研究

研究生：喻英雲

經考試合格特此證明

口試委員：

王易斌
邱宏林
鍾國貴
阮金萍

指導教授：王易斌

系主任(所長)：鍾國貴

口試日期：中華民國 97 年 6 月 18 日

誌 謝

時光荏苒，兩年時間匆匆而過，畢業多年之後重拾學生身分，並順利完成學業，值此畢業之際，心中充滿著感激與愉悅，這裡有著誨人不倦的師長，面對聽課時常常一頭霧水的學生，老師仍不厭其煩般指導；還有一群熱心可愛，互相加油打氣的同學，何其有幸能在這樣的環境中學習。

在撰寫本篇論文的過程中，著實遇到不少困難與瓶頸，甚至一度有了放棄而另尋題目的想法，多虧指導教授 王昌斌老師的包容與耐心，給予諸多建議，且犧牲下班時間與同學們進行討論，以及育銘學長的毫不藏私鼎力相助，終於能夠完成這篇論文。而撰寫以及口試期間，本所 鍾國貴所長、 邱宏彬老師、 陳仁義老師以及中正大學 阮金聲老師給予的寶貴建議，都讓論文的內容更加嚴謹，在此致上萬分謝意。

最後要感謝家人及同事間的體諒，方能使我無後顧之憂專心向學，感激之情無法言喻。

以本體論結合意圖為基之智慧型搜尋引擎之研究

學生：喻英雲

指導教授：王昌斌 博士

南 華 大 學 資 訊 管 理 學 系 碩 士 班

摘 要

教育部挑戰二〇〇八～e 世代人才培育計畫中，將「建構全民網路學習系統」列為其三大重點領域之一，目前已在各級學校建置完成計畫所需之硬體設備，期望藉由這些設備幫助使用者在網路上共創、共享教學資源，以促進網路學習之應用與發展。

雖然已有完善之設備，但在網路之利用上仍有未及之處，就現況而言，一般人在網路上尋找資料或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，但往往查得大批無關資料，尚需一一檢視，花費許多時間才能得到想要之資料。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，但此種方式所需之等待時間較長。針對上述問題，本研究提出一種新的以本體論結合意圖為基之搜尋模式，使用者以自然語言輸入問題描述之後，搜尋引擎即可自動分析自然語言之語意，進而尋得

符合使用者意圖之相關資料。

本研究搜尋資料之模式分為五個步驟進行，分別為建構數學領域本體、建構 5W1H 意圖句型庫、問句型態分析、意圖轉換、關鍵字擷取與擴展，最後將關鍵字組置於現有之搜尋引擎進行搜尋。

實驗結果顯示，本研究所提出之結合數學領域本體與 5W1H 意圖進行搜尋之方式，確實可提高搜尋資料之準確率並縮短查詢時間。

根據政府所做的一項調查，數學是許多學生懼怕，甚至討厭的學科，因此本研究以國小數學領域為主，希望在數學的教學與學習上均能幫助使用者獲得協助。

關鍵詞：5W1H、意圖、本體論、有限自動機、模糊理論

An intelligent search engine based on ontology and intention comparison

Student : Ying-Yun Yu

Advisor : Dr. Chin-Bin Wang.

Department of Information Management
The M.I.M. Program
Nan-Hua University

ABSTRACT

“Establish an Environment that supports Life-long Learning” is one of the three focus areas in the “Challenge 2008: E-Generation Manpower Cultivation Plan” proposed by the Ministry of Education. The project has implemented the required computer hardware in schools of all levels to facilitate the sharing and creation of online teaching resources, and develop and promote the application of online learning.

Currently, the way most users search information online can be categorized into two types: 1. Entering keywords into search engines. Normally users receive responses from search engines in a short time, but much of the retrieved information is irrelevant. Thus they need to spend much time in reviewing the search results item by item to obtain the information they really need. 2. Seeking information or helps from certain places online, such as internet forums. Information obtained this way is more likely to be what the users wanted but it takes longer to receive responses. It seems that each of these two search methods has its advantages and disadvantages. Therefore, we propose a new way to search information on the internet, namely the natural language search method.

Users seeking information can enter their problem description in natural language format into a search engine. The search engine automatically analyzes the nature of the problem to use this information to speed up the search process. Thus it will be able to quickly search the database and return appropriate information to users.

The five main steps of this research are: developing ontology for mathematics, structuring intention-sentence pattern database, transforming intention, identifying and expanding keywords, and finally including keyword sets in the search engine to improve the searching process.

Computational results show that combining mathematics ontology with 5W1H-intention can improve the accuracy of the search and shorten the search time. Therefore, it can help users obtain the information they desired on the internet efficiently and effectively.

According to a government's survey, mathematics is the subject that most students fear or even dislike. For this reason, our research focuses on the applications of the proposed methods to the mathematics curriculum of elementary schools. We hope this method can help users in the teaching and learning of mathematics.

Keywords: 5W1H, Intention, Ontology, Finite Automata, Fuzzy

目 錄

論文口試合格證明	II
誌謝	III
中文摘要	IV
英文摘要	VI
第一章 緒論	1
第一節 研究背景	1
第二節 研究動機	3
第三節 研究目的	4
第四節 論文架構	5
第五節 研究範圍	6
第二章 文獻探討	8
第一節 本體論(Ontology)	8
壹、本體論之定義	8
貳、本體論之組成	9
參、本體論之應用	10
肆、建立知識本體的步驟	11
第二節 自然語言處理系統	16
壹、處理流程	16
貳、斷詞處理	17
參、斷詞規則	20
第三節 中文語法	21
壹、語法的意義	21
貳、語詞排列之影響	23
第四節 模糊理論(Fuzzy Theory)	23
壹、模糊集合	24
貳、模糊數	26
參、模糊集合運算	28
肆、解模糊化	29
第三章 研究步驟	31
第一節 本體論架構	33

壹、	決定研究領域與範圍	33
貳、	列舉知識本體中的重要詞彙	33
參、	定義類別及其階層	36
肆、	建立實例	36
第二節	建構句型庫	37
壹、	中文語法與 5W1H 意圖之結合	37
貳、	中文句型依 5W1H 意圖之分類	38
第三節	句型分析	41
壹、	斷詞處理	41
貳、	詞性重設	43
參、	句型比對	44
第四節	意圖轉換	46
第五節	關鍵字擷取	47
壹、	關鍵字過濾	48
貳、	關鍵字擴展	49
第四章	實驗驗證	52
第一節	網頁擷取內容設計	52
壹、	實驗環境	52
貳、	實驗設計	53
參、	系統設計	54
第二節	實驗結果	56
第三節	資料分析	60
第五章	結論與未來方向	65
第一節	結論	65
第二節	未來方向	66
參考文獻	68
附錄一	中研院平衡語料庫詞類標記集	71
附錄二	線上斷詞服務採用之精簡詞類列表	72
附錄三	本研究與 Yahoo 搜尋引擎搜尋結果比較 1	73
附錄四	本研究與 Yahoo 搜尋引擎搜尋結果比較 2	74

表 目 錄

表 1	各種斷詞法之比較	19
表 2	語詞排列相異之語意列表	23
表 3	高的歸屬函數的程度	25
表 4	明確集合與模糊集合區分表	26
表 5	句型庫種類與功能	31
表 6	知識本體問題需求答問表	33
表 7	數學領域重要詞彙	34
表 8	「分數」之擴展詞彙	35
表 9	「通分」之擴展詞彙	35
表 10	句型與 5W1H 意圖之結合列表	38
表 11	5W1H 意圖對應表	46
表 12	候選關鍵字詞性標記列表	48
表 13	取樣文件記錄示例	57
表 14	本研究與 Yahoo 搜尋引擎搜尋結果準確率之比較	58
表 15	搜尋準確率之成對母體平均數差異檢定	60
表 16	關鍵字個數不同之問句準確率列表	61
表 17	關鍵字個數不同之 t 檢定結果	62
表 18	符合需求樣本於搜尋結果中出現之順序	63
表 19	關鍵字組排列相異之排序對應表	64

圖 目 錄

圖 1	個人電腦設置數及電腦用戶的人機比例.....	1
圖 2	研究流程圖.....	7
圖 3	「九年一貫課程」知識本體論架構.....	10
圖 4	由上往下定義類別及其階層.....	13
圖 5	由下往上定義類別及其階層.....	13
圖 6	數學領域不同層級.....	14
圖 7	TOVE 本體論工程建置步驟.....	15
圖 8	自然語言處理架構.....	16
圖 9	布林邏輯與模糊邏輯的範圍.....	24
圖 10	三角模糊數.....	27
圖 11	梯形模糊數.....	28
圖 12	圖形化介面語意轉換架構圖.....	32
圖 13	數學領域類別階層.....	36
圖 14	數學領域本體之 Classes 與 Individuals.....	37
圖 15	句型歸類方法.....	39
圖 16	中文句型庫資料表.....	40
圖 17	句型分析流程圖.....	41
圖 18	中研院斷詞系統處理流程圖.....	42
圖 19	「最簡分數」之斷詞結果.....	43
圖 20	數學專有名詞詞庫.....	43
圖 21	Knuth-Morris-Pratt 演算法.....	44
圖 22	有限自動機流程圖.....	45
圖 23	句型分析流程圖.....	45
圖 24	5W1H 意圖對應詞庫.....	47
圖 25	關鍵字過濾流程圖.....	49
圖 26	數學領域知識本體資料庫.....	50
圖 27	分數與 what 結合之擴展關鍵字.....	50
圖 28	「通分要怎麼做」之查詢流程.....	51

圖 29	實驗進行流程圖	53
圖 30	本研究設計之搜尋介面	54
圖 31	輸入自然語言問句	55
圖 32	系統比對出意圖及關鍵字	55
圖 33	Yahoo 搜尋引擎之搜尋畫面	56
圖 34	關鍵字個數不同之搜尋結果準確率比較圖	59
圖 35	關鍵字個數與搜尋準確率之關係圖	61

第一章 緒論

第一節 研究背景

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，根據行政院主計處電子處理資料中心(民96)統計，截至95年12月底止，我國個人電腦設置數達1,041.5萬台，其中家庭部門為614.9萬台；企業機關學校為426.6萬台，與94年底比較，共增加66.2萬台，而我國政府機關設置暨應用電腦的人機比亦逐年攀升，由圖1可看出其逐年遞增之趨勢。

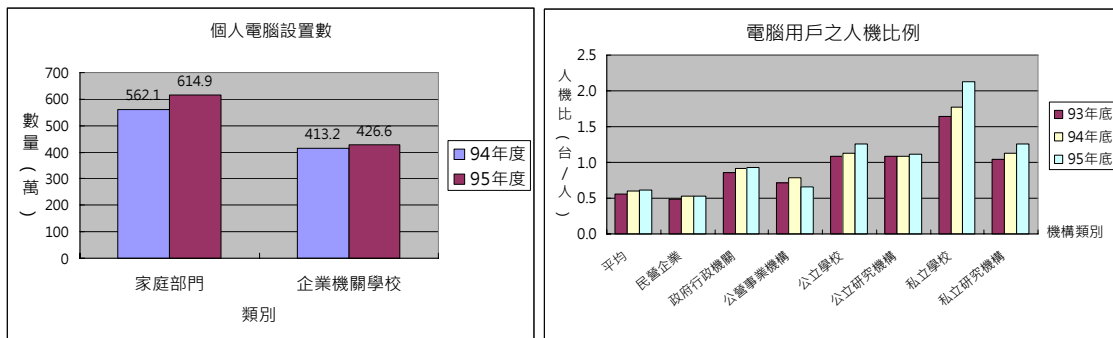


圖 1 個人電腦設置數及電腦用戶的人機比例

資料來源：行政院主計處電子處理資料中心(民96)

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，根據資策會電子商務研究所(民97)調查，至97年1月，我國網路使用人口已達1,476萬人，個人連網普及率為64.4%，至於觀測網際網路硬體建置規模重要指標之一的連網主機數，至去年底已達511萬部，較96年1月增加69萬部，成長率達16%，排名為世界第13名、亞太第4名。由以上數據顯示，我國之資訊國力可謂不容小覷。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(民93)訂定2005~2008年之教育施政主軸為「創意臺灣、全球布局~培育各盡其才新國民」，其中，「建構數位化學習環境」被列為培育現代國民的行動方案之一。

此外教育部挑戰二〇〇八~e世代人才培育計畫(民94)中，亦將「建構全民網路學習系統」列為子計畫之一，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，此計畫目前仍持續進行中，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。

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，如何在浩瀚網海中找到我們想要的資訊？目前，關鍵字查詢是最普遍、也是一般咸認最快的作法，但是採用這種方式，原本網路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，並非筆筆皆為使用者所需，可用者往往隱身其中，須一一檢視，如此一來，雖然搜尋引擎搜尋資料所需時間極為短暫，但是使用者進行資料篩選的工作時間卻可能達數倍之多。

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，由使用者提問，希望藉由大家的經驗分享，增加彼此交流，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，但卻必需被動等待答覆，通常並非即時可得，甚至遇到無人答覆之困境。

因此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，亦即既要快速得到搜尋結果，又能增加搜尋結果之準確度。針對此需求，本研究認為輸入之語詞與搜尋方式乃重要關鍵。以人類語言學習歷程來說，兒童一歲左右可以說出字或詞，三歲時可以習得主要句法。在兒童牙牙學語時期，他人僅能由隻字片語判斷甚至猜測其所欲表達之意義，此時或有猜測錯誤之時。待其漸長，習得自然語言之後，方能與他人溝通無礙，因此我們是否亦可據此推論，自然語言其實是網路搜尋之最佳語言，惟有完整之描述，方能正確判斷使用者欲獲取之知識。

第二節 研究動機

目前以自然語言查詢之研究正方興未艾，若能有一詢答系統，當使用者輸入問題描述之後，即可自動且快速地於網路上搜尋相關資料，亦即毋需使用人力，即可提供即時且正確的協助，則目前所存在的問題，不論在時間或是資料準確度這兩個面向，均能獲得解決，應可對知識的傳遞有極大的助益，這也是目前許多學者專家努力的目標。

當一個問句被提出時，其中可能蘊含時間、地點、原因…等因素，因為使用者的目的不同，希望得到的資料也不同。比方說使用者輸入「火車」這個關鍵字，也許是問開車時間，但也有可能是想知道在哪裡搭車，這些不同的目的，我們把它稱為「意圖」，在不同的意圖下，所找到的資料也應不相同，因此本研究希望能整理出一個蘊含意圖之中文句型庫。然而中文句型甚為複雜，單純利用句型庫恐仍不敷所需，因此本研究將另外建置5W1H意圖對應詞庫進行比對，以為輔助。

不同領域的使用者，看到相同的關鍵字各有不同的解釋，比方說「鹽水」，在地理專業領域上屬於「地名」，但在一般人看，就是加了鹽的水，如果在找資料的時候，事先考慮意圖與領域的不同，給予

不同的關鍵字，那麼對使用者進行資料查詢時應有所幫助。因此，在研究智慧型搜尋引擎時，如何將意圖與領域本體的概念加入關鍵字之中，進而提升效能，成為筆者的研究動機。

以現有課程來看，數學是許多人懼怕甚至討厭的學科，根據教育部統計處於八十七年十月份完成的「中小學生對課程不喜歡比率」之調查報告(楊新芳、葉明利，民90)顯示，國小、國中及高中學生最喜歡的科目都是數學。該調查同時發現，小學生不喜歡數學的比率隨年級增加而更加提高，由一年級的16.11%，一路提升到六年級的46.05%，也就是說在六年級的學童中，平均約每兩人就有一人討厭數學。所以對關心數學學習的師長來說，如何教學，使得學生能夠接受而不產生懼怕的感覺，甚且主動學習數學，是一項很重要的課題，這些都需要經驗的傳承與再學習。傳統的研習，將所有人集中授課，對象僅為特定少數人，且為單純的「施」，對求知若渴且背景知識不同的大眾來說，這種作法已不敷所需。透過網路學習系統，提升專業知能的學習模式正是學習的一項新主流。

因此，如何能更有效率地利用網路來協助使用者，不論是師長或者是莘莘學子，均能找出數學方面問題的解答，亦成為動機之一。

第三節 研究目的

基於上述問題背景及研究動機，本研究之目的為發展一種智慧型搜尋引擎，希望能建構一查詢系統，此系統能達到下列目標：

- 壹、進行查詢時，使用者僅需輸入之自然語言，系統即可自動分析該問句，產生合適之關鍵字進行查詢，使得所需資料能依其意圖做排序。
- 貳、透過圖形化介面，以智慧型搜尋引擎進行數學領域方面知識查詢

時，所得資料之準確率能優於一般關鍵字搜尋方式。

參、透過圖形化介面，以智慧型搜尋引擎進行數學領域方面知識查詢時，符合使用者需求之資料，相較於一般關鍵字搜尋方式，能較早被尋得而節省搜尋時間。

肆、能建立可用之數學領域知識本體，供系統選用合適之關鍵字。

伍、分析影響資料搜尋準確率與排序之因素，進而提高搜尋效果。

陸、驗證將領域本體與 5W1H 意圖結合，以產生領域關鍵字與意圖擴展關鍵字之方式，是否能協助系統正確選用關鍵字。

第四節 論文架構

本論文共分五章：

第一章為緒論，說明研究背景、研究動機、研究目的、研究架構及研究範圍。

第二章為文獻探討，針對研究相關之文獻逐一探討，包括本體論、自然語言處理、中文語法分析、模糊理論等。

第三章為研究步驟，本研究之各個步驟如下：

- 一、 建構數學領域本體論。
- 二、 建構5W1H之意圖句型庫。
- 三、 輸入查詢字串，進行問句型態分析。
- 四、 問句無法明確判別型態時，以模糊理論分析其歸屬程度。
- 五、 結合5W1H句型庫進行意圖轉換。
- 六、 利用數學本體論挑選出領域關鍵字，並結合轉換後之意圖關鍵字，將關鍵字組置於搜尋引擎進行查詢。

第四章為實驗驗證，以本研究發展之搜尋模式設計搜尋介面，並且利用現有之搜尋平台進行實驗。

第五章為結論與未來方向，除根據實驗結果，歸納出本研究之結論外，另提出研究不足之部份，以提供未來研究者參考。

本研究之流程如圖2所示。

第五節 研究範圍

不同領域所建構之領域本體並不相同，本文以數學領域為研究範圍，因此僅建構數學領域本體，並且以國小學程為主，且以中文之搜尋為主要研究目標。

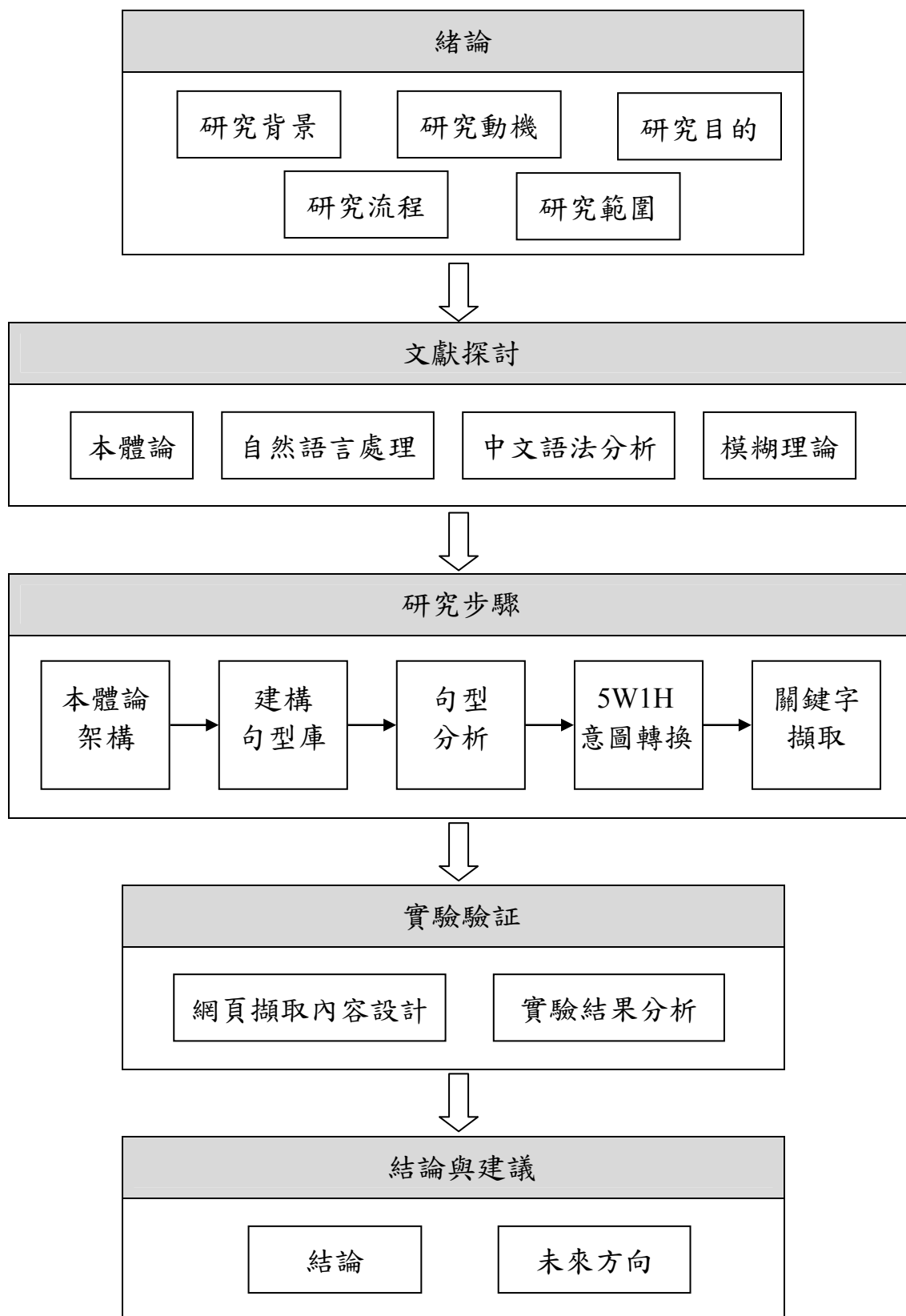


圖 2 研究流程圖

第二章 文獻探討

本研究希望能利用本體論與意圖之結合，增加網路搜尋之準確率與節省時間，而意圖之判斷又與語句的分析相關，在分析語句時，極可能面臨既有資料庫並無收錄以至於無法比對情形。本章針對相關領域加以探討，內容包含本體論、自然語言處理系統、中文語法研究及模糊理論等相關領域，期望在面臨各項問題時，能提供一正確方向。

第一節 本體論(Ontology)

壹、本體論之定義

本體論原本為哲學領域中的一個名詞，主要探討存在本身，但是目前知識工程領域中也開始應用本體論的觀念，不過意義已經不同，以下列出學者對於本體論的看法：

一、Gruber (1993)：

Ontology是概念的明確描述，源自哲學，藉由定義專有名詞的集合來描述本體。

二、Wielinga and Schreiber (1993)：

Ontology是一種理論，探索何種實體能存在知識代理人的思想中心。

三、Gruber (1994, as cited in Uschold & Gruninger, 1996)：

Ontologies 對於共享的概念是一致的，這些共享概念包含塑造領域知識模型的架構、與代理人溝通的特殊協定。

Ontologies 對於特定領域的陳述也是一致的，分享知識內容時，Ontologies 以明確的格式定義代表字彙。

四、Neches, et al. (1991)：

Ontologies 提出建構領域知識的基本規則，以定義基本用語及其關聯的方式構成目標領域的辭彙。

五、Gruber (2008)：

Ontology是為了塑造領域知識的模型或論述而定義的代表元素。

由以上各學者的看法可知，本體論乃是針對領域知識，將其中各種組成元素下一明確的定義，而此定義需有規格化的陳述，如此方能不同代理人之間溝通，並且可進行概念的共享與共創。

貳、本體論之組成

目前不論以何種語言來描述本體，它們的結構都極為類似，多數學者均認為由本體論以下四種元素組成：

一、實例(Individual)：

實例是本體論中最明確、最底層的物件，不同領域會有不同的描述，可能包含具體的物件，如動物、植物、桌、椅...等等，或者是數字、文字等等。

二、概念(Concept)：

或稱類別(Class)，概念是由多個底層物件(實例)所組成的集合，藉由概念將資料表達出來，進而描述出所有領域的資料，概念可能為一個範圍、領域或是主題。概念之下又可分出子概念，以進一步描述本身所處之領域。

三、屬性(Attributes)：

屬性是對物件的描述，字彙與概念都屬於物件，而每一物件本身的特徵、特性均須由屬性來描述，使得物件本身更能被了解。

四、關聯(Relationships)：

建構出Ontology的架構之後，除了以屬性表達物件本身之外，還需定義物件與物件之間的關係。

以下舉實例說明本體論之架構，如圖3所示，「九年一貫課程」是這個本體論中最上層的概念，語文、數學、社會…等領域為隸屬於九年一貫課程之下的學習領域，為其子概念，與九年一貫課程存在一種「is-part-of」的關係，而在「多邊形」這個概念中，三角形、四邊形為其屬性，用來描述多邊形。

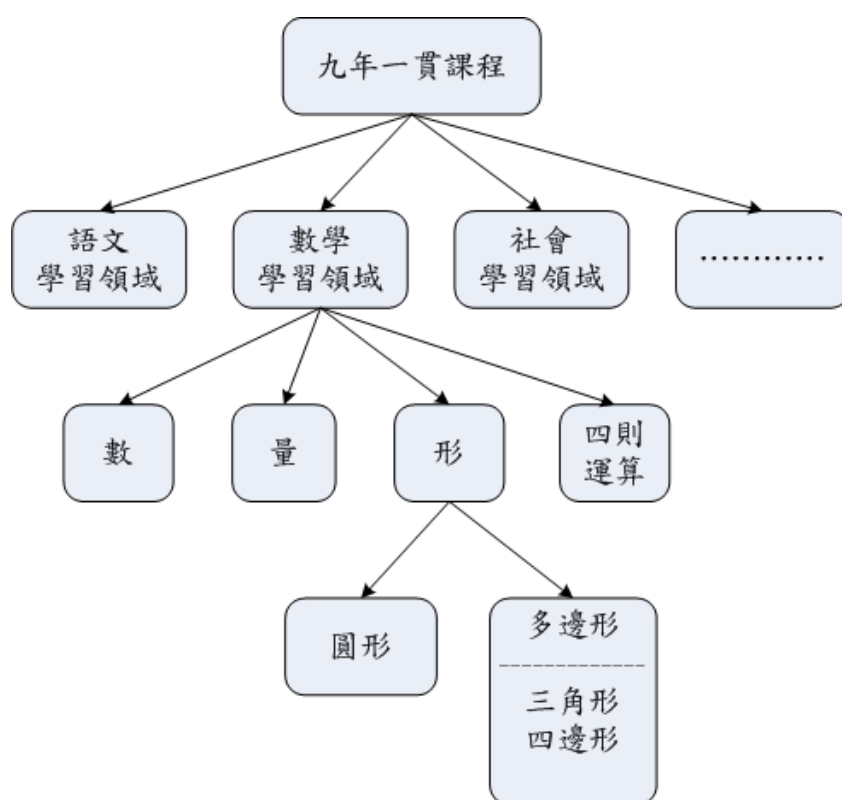


圖 3 「九年一貫課程」知識本體論架構

資料來源：本研究設計

參、本體論之應用

對需要分享領域知識的研究者來說，Ontology 定義了共通的字彙，當中包含可由機器判讀的領域基礎概念之定義與其間的關係 (Noy & McGuinness, 2001)。

通常建置Ontology的原因有下幾種：

- 一、讓不同的人或軟體代理人之間能夠分享彼此的知識。
- 二、使領域知識可重覆使用，研究者可以利用或合併現存的Ontology進行研究，而不需重複建置。
- 三、使得領域假說更加清楚明確，對於電腦專門知識較淺薄的人來說，修改與學習都更加容易。
- 四、從現有的知識中區分出領域知識。
- 五、試圖使用或是擴展領域知識時，Ontology當中合乎格式的名詞分析相當有用。

由以上五點可知，當我們在研究某一領域知識時，可以利用Ontology 規格化的語言來明確表達物件本身與物件之間的關係，使得其中的物件均能代表領域內相關知識，而這些領域知識也能夠以人與電腦皆能了解的方式來表達，並能為不同研究者通用、擴展，甚至不同領域研究者亦能藉由規格化的描述，迅速進入狀況並且加以利用。

目前本體論在資訊上的應用相當廣泛，例如檢索系統、知識管理系統、詢答系統...等，均被普遍利用。此外，在網路發展居重要地位的「語意網」中，本體論亦為一項非常重要的技術，畢竟電腦僅為一冰冷機器，若希望其能理解人類語言，必需藉助於結構化的陳述。

肆、建立知識本體的步驟

Noy and McGuinness 認為建立知識本體時，並沒有所謂「正確」的方法論，他們將建置知識本體的步驟由決定領域與範圍至建立實例方法論，他們將建置知識本體的步驟由決定領域與範圍至建立實例分為七個階段，而Kim, Fox, and Grüninger (1999) 亦提出TOVE (Toronto Virtual Enterprise)本體工程方法，包含由背景與問題描述到本體論評估等四項程序，以下分別就上述兩種方式做更進一步說明：

一、Noy and McGuinness：

1. 決定知識本體的領域與範圍：

開始發展本體時，可以先由下列幾個問題來進行：

- (1) 知識本體涵蓋的領域為何？
- (2) 我們使用知識本體的目的為何？
- (3) 知識本體的資訊應該用來回答哪一類的問題？
- (4) 知識本體為誰所用？誰來維護？

條列出的問題均能利用基礎知識來回答，並能以此問題來決定領域本體的範圍。

2. 考慮利用現存的知識本體：

目前已經有許多可再利用的知識本體存在網路及文獻中，例如Ontolingua ontology library、DAML ontology library，或是公開的商業化知識本體，例如UNSPSC、RosettaNet、DMOZ等等

3. 列舉知識本體中的重要詞彙

把所有的詞彙條列出來，例如數、量、形、小數、分數、三角形、角、邊…等數學領域詞彙，對於向使用者作陳述或是解釋很有用。

4. 定義類別以及類別的階層

定義類別及其階層有三種方式：

(1) 由上而下(top-down)

由較籠統的概念開始，其後接較專業的子類別(subclass)，例如「數學領域」為一較籠統之概念，「數」、「量」、「形」則為較專業之數學名詞，因此定義「數學領域」為父類別(superclass)，「數」、「量」、「形」為其下之子類別，如圖4所示。

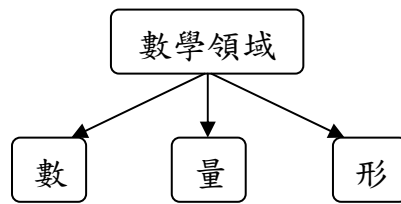


圖 4 由上往下定義類別及其階層

(2) 由下而上(bottom-down)

由一或數個較特殊的類別(階層中的葉節點)開始，隨後接父類別，例如「三角形」等形狀為較特殊之類別，屬於「形」此一概念之下，因此由各形狀開始定義，往下則接「形」，為其父類別，如圖5所示。

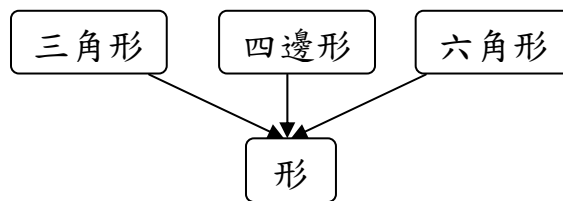


圖 5 由下往上定義類別及其階層

(3) 結合式(combination)

結合top-down以及bottom-down兩種方式，首先定義較顯著的概念，然後把他們做特殊化或一般化的推論。也許從最上層的「數學領域」以及最下層的「正三角形」開始，兩者與中間階層的「三角形」有關，此外，我們要產生所有的「形」的類別，從而產生一些中間階層概念，如圖6。

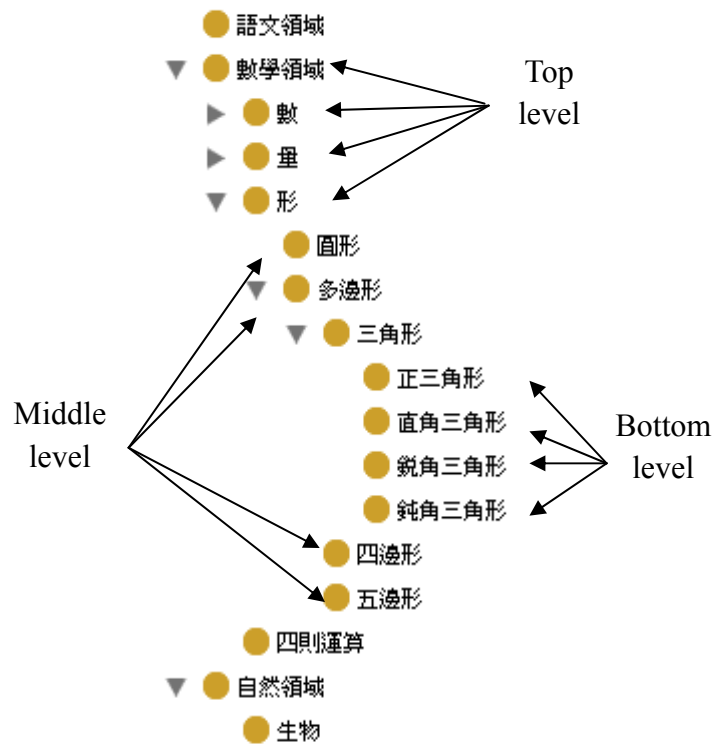


圖 6 數學領域不同層級

5. 定義類別屬性(slot)

僅類別無法回答步驟一提出的問題，我們必須描述概念的內部結構，例如定義正三角形中，邊的數量以及角度值。

6. 定義屬性的限制

屬性中，值的型態可以有不同的限制，例如將學號限制為五位數。

7. 建立實例

首先選擇一個類別，接著建立類別的實例，最後填入屬性值。

二、TOVE本體論工程

TOVE本體論工程建置之步驟可分為四項，其建置過程顯示如圖7。

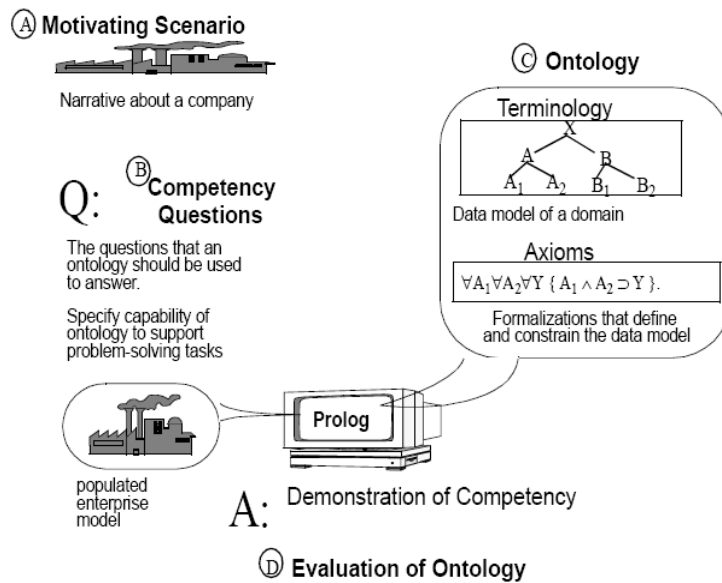


圖 7 TOVE 本體論工程建置步驟

資料來源：Kim, H. M., 2002

1. 應用情境(Motivation Scenario)

設計領域本體通常是因為遇到問題，因此需要在一開始就針對問題以及可能的解決方案作詳細描述，本階段可以使以後的維護人員了解最初設計本體論的用意。

2. 口語化的能力問題(Informal Competency Question)

本階段為使用者的需求，是日後評估本體論是否成功的重要條件之一，利用自然語言來表示。一個理想的Competency Question 應具備階層式的關聯，也就是說後面的Competency Question 會使用到前面Competency Question 所產生的答案。在設計本體論時以能回答Competency Question 為目標來進行。

3. 本體化(Ontology)

本階段又分為四個步驟：

- (1) 定義所有本體論中會用到的專門用語。
- (2) 建立所有詞組與關聯的階層模型。

- (3) 經由建立的詞組將口語化的問題轉換成格式化的問題，並由一階邏輯來表示，本階段中，藉由提供本體論中的原則 ($T_{ontology}$)與實際物件及關聯(T_{ground})，以此推論出 Competency Question (Q) 的答案，即 $T_{ontology} \cup T_{ground} \models Q^1$ 。
- (4) 將前幾個階段無法完全表達複雜語意或限制的定義加以補強。

4. 能力驗證(Demonstration of Competency)

本階段利用 Competency Question 來評估之前所建立的本體論是否符合第一階段的問題描述。也就是將真實資料 (T_{ground}) 輸入之後，評估是否每個 Competency Question 都能被正確地回答，若可以的話，就表示找到一個 Q，能夠滿足 $T_{ontology} \cup T_{ground} \models Q$ 。

第二節 自然語言處理系統

壹、處理流程

使用者輸入文字之後，自然語言處理流程大致經斷詞、語法分析、語意理解、產生查詢代碼及資料查詢等步驟，最後再將結果輸出。Kao, Cercone, and Luk (1988)提出一自然語言處理之架構，如圖8所示。該架構可分為五個主要單元：自然語言剖析器、語意解譯器、查詢產生器、查詢解譯器及資料庫。

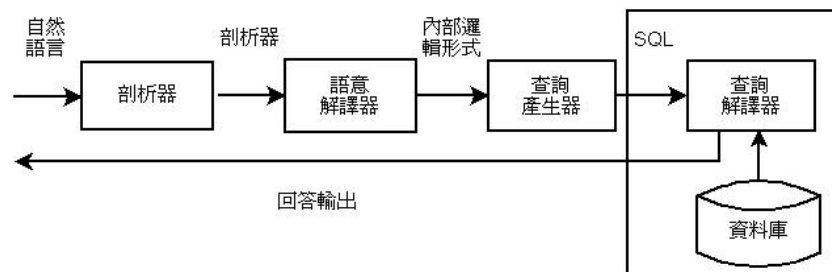


圖 8 自然語言處理架構

¹ $T_{ontology} \cup T_{ground} \models Q$ 中， \models 代表蘊含，亦即在 $T_{ontology} \cup T_{ground}$ 成立的每個模型，在 Q 也成立。

貳、斷詞處理

中文以字為基礎，由字的排列組合形成詞，再由詞的排列造成句。斷詞處理在中文的自然語言處理上是最基礎的工作，每一個斷出來的詞都可能是關鍵字，但是詞的長短並不固定，一字可為一詞、另有雙字、三字甚至多字都可能成為一個詞，而詞與詞之間並沒有明顯的區隔，因此若進行斷詞時，句中斷字的位置不同，即可能造成完全不同的語意甚至產生錯誤。

目前中文斷詞的方法，依黃勝龍(民91)之整理，大致可分為詞庫式斷詞法、規則式斷詞法、統計式斷詞法、混合式斷詞法等四種，分述如下：

一、詞庫式斷詞法

事先以人工方式收集各種語詞並存入詞庫，再將詞庫中的語詞逐一與需要斷詞的語句進行比對，依「長詞優先」與「前詞優先」規則，找出句中可能的中斷點與有意義的詞彙。因詞彙已經過人工過濾篩選，因此可找出適合的詞彙，並去除無意義的詞彙。

二、規則式斷詞法

自然語言當中並不只有詞彙，尚有其他語法規則，因此部分學者認為斷詞僅為自然語言之一部分，仍需強調語言現象，使用前事先建立語法規則作為判斷之依據，並與詞庫資料作比對，方能找出適當的辭彙組合。例如，「就……來說」為規則句型，表示「就」字出現後，通常伴隨著「來說」一詞的出現。規則式斷詞法即藉由這種簡單的規則來進行自然語言處理。

三、統計式斷詞法

統計式斷詞法必需先利用大型的語料庫進行訓練後，方能實際應用，目前有下列幾種方式：

(一) 鬆弛法(范長康，民79)：

將斷詞視做一種對句中各字做「字詞指派」的過程。利用句中字詞間的組成關係做為指派方式的約束條件；並以機率式鬆弛循環建立指派機率的修正模式。在執行鬆弛程序時，這些約束條件將剔除不相容的指派，最後找出正確的斷詞結果。

(二) Sporat-Shih 統計斷詞法：

計算句中兩相鄰文字出現的頻率，找出相鄰機率最大者視為詞，再處理句中其他文字，如此反覆運算直到兩相鄰機率小於固定值，此法僅能找出單字或雙字詞，正確率不高(黃勝龍，民91)。

(三) 周竝弘與張俊盛(民81)使用模組化架構來結合影像模式與語言模式，利用影像模式產生候選字，在語言模式上運用機率式作法進行斷詞層次的分析，進而從影像模式所產生的候選字列中挑出正確字。

(四) 使用一些特徵值，如詞頻，詞長，詞性等，並給予權重，準確率可達到百分九十九以上(Chiang, Chang, Lin & Su, 1992)。

四、混合式斷詞法

前述三種斷詞法各有其優缺點，因此Yeh and Lee (1991) 提出混合式斷詞法，以聯併(Unification)為基礎，首先利用詞庫搜尋可能的斷詞組合，接著利用構詞規則將其簡化，再以一階馬可夫機率模型排列出所有可能的結果，然後依照機率值排列所有可能的組合，最後使用HPSG剖析器(Head-driven Phrase Structure Grammar Parser)逐一過濾，以確認該斷詞組合是否合乎文法(陳稼興、謝佳倫、許芳誠，民89)。

五、各斷詞法的比較

各斷詞法均有其優劣存在，本研究參考王聖中(民83)、陳稼興等(民89)、陳孟華(民91)、黃勝龍(民91)、林信宏(民92)等研究之看法，列出

各斷詞法之特性及優、缺點製成表1，採用時應考慮系統本身，選擇合適之方法。

表 1 各種斷詞法之比較

類別	特性	優點	缺點
詞庫式斷詞法	1 ^d .以人工挑選詞彙收錄於詞庫 2 ^d .依長詞、前詞優先規則進行	1 ^c .選出之關鍵詞較具意義 2 ^e .可建立領域詞彙，提高準確度	1 ^c .人工方式不夠客觀 2 ^c .詞庫須時常更新以因應時代變化 3 ^d .可能產生長詞掩蓋短詞的現象
規則式斷詞法	^a 事先建立語法規則，再與詞庫資料比對，找出詞彙組合	1 ^e .可同時處理斷詞與語意 2 ^e .不受限於資料多寡 3 ^d .與領域無關之詞彙可依特殊語法處理	1 ^b .受詞庫品質影響很大 2 ^d .規則之間可能造成衝突的現象 3 ^d .語法整理較為困難
統計式斷詞法	1 ^b .依機率值決定斷詞的位置。 2 ^c .需先建立各類別之語料庫，以足夠數量之文件作訓練	1 ^c .執行效率高 2 ^e .不受限於詞庫未收錄新詞或語法規則而導致無法正確斷詞	1 ^c .好的語料庫建置不易 2 ^b .詞頻會因語料庫之建構者而異 3 ^b .詞長大於二，效率降低 4.統計模型之大小需適度太大花費時間較長，太短則準確率降低

資料來源：^a王聖中(民83)、^b陳稼興等(民89)、^c陳孟華(民91)、^d黃勝龍(民91)、^e林信宏(民92)

本研究中，斷詞部分之處理係採用中央研究院資訊科學所CKIP中文詞知識庫小組[中研院詞庫小組]開發之中文斷詞系統，該系統斷詞問

題乃依據中央標準局「資訊處理用中文分詞規範」處理(中研院詞庫小組)，其分詞原則如下：

(一) 基本原則：理論性原則、指導原則

1. 語意無法由組合成分直接相加而得到之字串應該合為一分詞單位(合併原則)。
2. 詞類無法由組合成分直接得到，應該合為一分詞單位(合併原則)。

(二) 輔助原則：操作性原則

1. 有明顯分隔標記應該加以切分。(切分原則)
2. 附著語素盡量和前後詞合為一個分詞單位。(合併原則)
3. 使用頻率高或共現率高的字串盡量視為一個分詞單位。(合併原則)
4. 雙音節結構之偏正式動詞盡量視為一個分詞單位。(合併原則)
5. 雙音節加單音節之偏正式名詞盡量視為一個分詞單位。(合併原則)
6. 內部結構複雜之詞盡量加以切分。(切分原則)

參、斷詞規則

將語句斷詞時，每一個語詞都是字本身或是相鄰字的組合，有各種組合方式，也就是說同一語句可能產生不同的斷詞組合，進而產生不同的意義。斷詞規則主要就是在解決斷詞時出現的歧義問題，在陳永德(民86)的研究中，發現長詞優先、詞頻對比與前詞優先三種規則確實存在，並且前詞優先規則比詞頻對比規則更有效率。

一、長詞優先

中文詞的長短並不固定且詞與詞之間並無明顯區隔，除了由字與

字組成詞，還可能由字與詞、或者是詞與詞來組成另一個詞。例如「嘉義車站」，可以是「嘉義」與「車站」兩個詞彙，或者是「嘉義車站」一個詞彙，依長詞優先的原則，我們會斷為「嘉義車站」一個詞，而「嘉義車站」這個詞的涵義要比「嘉義」或者「車站」來得準確。

二、 前詞優先

一句中有多種斷詞組合，若兩相鄰詞組中有部份字重疊，例如「...體育場所以爲了...」，可以斷成「體育場」「所以」「爲了」，或是「體育」「場所」「以爲」，在前詞優先原則下，我們以「體育場」為前詞，斷成「...體育場-所以-爲了...」。

三、 詞間頻率對比

此規則即以詞彙出現的頻率來決定，例如「畢氏定理」，可能有「畢氏」、「定理」或者是「畢氏定理」這三種組合，這時以出現頻率較高的詞彙為最後結果，因此語料庫的建立及訓練非常重要。

第三節 中文語法

壹、語法的意義

依學者何永清(民94)整理，語法又可稱文法，其解釋有以下二類：

一、 各學者的見解

1. 郎峻章：語法包含兩部份，一是詞法，研究詞類、詞的變化、構詞法等。二是造句法，研究組詞造句的規則。
2. 傅銘第：語法是研究語言結構方式的科學。
3. 鄒熾昌：任一民族之語言，其詞類採用、字句排列、均有一特殊而普遍的方式，這是他們習慣的規則，研究這種規則的方法就是文法。
4. 黎錦熙：國語有我們常用而公認的習慣與規則，將這些習慣與

規則歸納出來並加以整理、排列、說明就叫做「國語文法」，「簡稱語法」。

5. 呂叔湘：文法就是語句組織的條理。
6. 陸儉明、馬真：語法就是一種語言組詞造句的規則。
7. 程祥輝、田小林：語法就是專門分析語言結構的，是某種語言組合法則的總體，是使用這種語言的人約定俗成的。
8. 李宇明：語法就是語言的結構規則，即詞語的組合規則，那些詞語能互相組合，哪些不能；哪些詞語能以這種方式組合，哪些詞語能以另一種方式組合；詞語在結合時形式是否發生變化以及如何變化均有一定的規則，這些規則就是語法。

二、辭書的解釋

1. 重編國語辭典修訂本(教育部編)：語詞、語句形成的規律。有時包括口語和書面語，與「文法」一詞相近。
2. 新辭典(三民書局)：研究語言組織法則的學問，即文法。
3. 幼獅少年百科全書：語法是語言的結構方式，例如「黑貓」表示黑色的貓，「黑」字必須放在「貓」字前面，如果放在後面，意思就有差別，這就是語法。
4. 漢語大辭典(羅竹風主編)：語言的結構方式，包括詞的構成和變化、詞組和句子的組織。
5. 學典(三民書局)：語言中文詞、文句的組織規律和結構法則。
6. 新編國語日報辭典(國語日報社)：「文法」是研究語文的詞、語、句的連結、配置、構造、組織等習慣規則的學科。又釋「語法」為文法，研究語言組織法則的學問。
7. 現代漢語知識大辭典(高更生等主編)：語法又稱文法，包括語素、詞、詞組、句子、句組等單位，由詞或詞組組成句子，均

有規律可循。

綜合以上各專家學者所述，我們可以知道，「語法」是針對某種語言之習慣性用法的研究，主要在研究句子的構造，字詞排列的規律性，而規律性，正是在利用機器判讀自然語言時，相當重要的一種性質。

貳、語詞排列之影響

語句中，字詞的排列有一定的順序，位置不同意義就可能有所改變，甚至成為錯誤的句子，另一方面也可能出現組合的位置不同，意義卻相同的情況。以「假分數是什麼」為例，該句經 CKIP 斷詞系統處理後，得出該問句乃是由假分數(Na)、是(SHI)以及什麼(Nep)²所組合而成，若語詞組合位置不同，將造成如表 2 之結果。

由表 2 可以看出，問句 1 與問句 2，雖然詞的排列順序不同，但意義是一樣的，問句 3 與問句 4 則是錯誤用法。

表 2 語詞排列相異之語意列表

編號	語詞組合	句型	語意
1	假分數-是-什麼	Na-SHI-Nep	詢問假分數的意義
2	什麼-是-假分數	Nep-SHI-Na	詢問假分數的意義
3	假分數-什麼-是	Na-Nep-SHI	錯誤句型
4	什麼-假分數-是	Nep-Na -SHI	錯誤句型

第四節 模糊理論(Fuzzy Theory)

日常生活各種現象中包含了確定、隨機、模糊現象等三大類，為解決實務上各種問題，數學家發明了許多不同的方法，遇上確定現象

² Na、SHI、Nep 分別代表普通名詞、是、指代定詞

問題時可以利用幾何、代數、數學分析、微分方程等方法，若是隨機現象則能考慮機率與統計，但面臨模稜兩可的問題時，傳統的數學已經無法解決，這時候就能利用模糊理論加以處理了。

而在語意的分析當中，也可能產生判別上的不確定性，此時即可利用模糊理論產生較合適之結果。

模糊理論是在 1965 年，由美國加州大學柏克萊分校 L. A. Zadeh 教授發表在「資訊與控制」(Information and Control) 學術雜誌上，名為「Fuzzy Sets」，初期並未受到重視，但時至今日，已能應用於農、工、氣象、社會、人文以及家電用品之控制等等，近年來更朝向專家系統方向發展。對於具有模糊性(blur)、含糊性(vagueness)或曖昧性(ambiguity)等不確定性(uncertainty)的資料或語言，均可藉由模糊集理論加以描述分析而獲得較適當且合理的結果(黃昭勝，民 95)。下面就介紹幾個模糊理論的相關概念。

壹、模糊集合

傳統的二值布林邏輯(two-valued Boolean logic)為一明確集合，特徵值僅有 0(假)與 1(真)，模糊邏輯與之不同，屬於多重值(multi-valued)由 0 與 1 之間的數來表示其歸屬於「假」或是「真」的程度，如圖 9 所示。

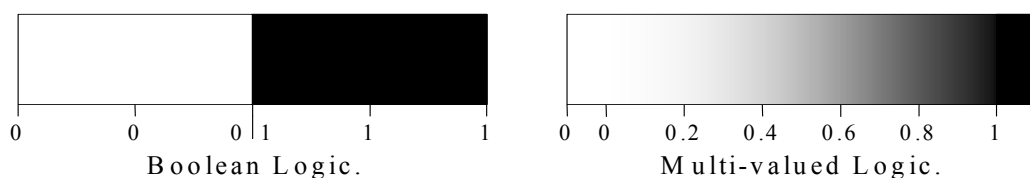


圖 9 布林邏輯與模糊邏輯的範圍

資料來源：Negnevitsky, 2005, p.89

在現實世界中，人們的思維、感覺及語意表達等，很難以數字來

精確表達，例如身高為高或矮的區分，如果以身高超過 180 公分以上稱為「高」，未達 180 公分為「矮」，因此若有一個人身高 180 公分，無庸置疑將被歸類至「高」這個集合，而 179.5 公分則被歸類至集合「矮」，雖然在古典數學上此答案為正解，但在一般人的感覺上，兩者卻無差異，因此以模糊集合來解決此一歸屬問題，即以「179.5 公分偏向高的程度有多少」來表達會更為清楚。

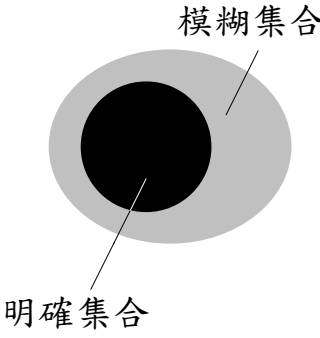
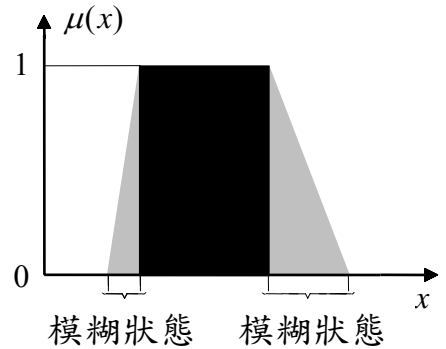
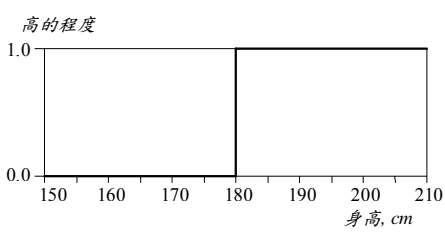
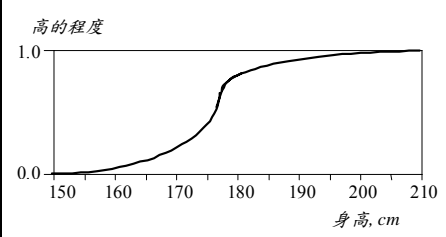
例如表 3 為十個人的身高資料，各有其歸屬函數的程度，以模糊集合來看，身高 180 公分偏向高的程度為 0.82，205 公分的程度為 1，屬於絕對的高，158 公分偏向高的程度則為 0，屬於絕對的矮。

表 3 高的歸屬函數的程度

編號	身高(cm)	歸屬函數的程度	
		明確	模糊
1	210	1	1.00
2	205	1	1.00
3	200	1	0.98
4	180	1	0.82
5	179	0	0.78
6	172	0	0.24
7	165	0	0.15
8	158	0	0.06
9	155	0	0.01
10	150	0	0.00

本研究另整理傳統的明確集合與模糊集合之區分，詳列於表 4。

表 4 明確集合與模糊集合區分表

類別	明確集合	模糊集合
函數屬性	特徵函數	歸屬函數
值域	{0,1}	[0,1]
解釋範圍	明確集合是模糊集合的特例	模糊集合是明確集合的延伸
		
圖形表示		
函數式	$f_A(x) : X \rightarrow \{0, 1\}$, where $f_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$	$\mu_A(x) : X \rightarrow [0, 1]$, where $\mu_A(x) = 1$ if x is totally in A ; $\mu_A(x) = 0$ if x is not in A ; $0 < \mu_A(x) < 1$ if x is partly in A .

資料來源：本研究整理

貳、模糊數

要表示模糊集合，除了繪圖的方式外，通常可以利用下列兩種方法來表示：

1. 離散型：

假設全集 Ω 的元素個數為 $n < \infty$ ，即 $\Omega = \{x_1, x_2, \dots, x_n\}$ ，則模糊集合 A 可表示為 $A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \dots + \mu_A(x_n)/x_n$ ，其中「+」號表示“或”，「/」表示“分隔”， $\mu_A(x_1)/x_1$ 表示全集中的元素 x_1 ，對應到一個歸屬值 $\mu_A(x_1)$ 。

2. 連續型

全集 Ω 的元素有「無限多個」，模糊集合 A 表示為 $A = \int_{\Omega} \mu_A(x)/x$ ，其中「 \int_{Ω} 」表示全集 Ω 的元素是連續的，即有無限多個元素，此時可以模糊數來表示，模糊數為具有下列三種性質的模糊集合：

1. 模糊集合在實數軸上
2. 模糊集合是凸的
3. 模糊集合經正規化

下面列出幾種常用的模糊數(Kaufmann & Gupta, 1991)：

(1) 三角模糊數(Triangular Fuzzy Numbers)：

三角模糊數以 $A(a, b, c)$ 表示， $\forall x, a, b, c \in R$ ，其歸屬函數 $\mu_A(x)$ 之計算及定義如圖10所示。

$$\begin{aligned} \mu_A(x) &= 0, & x \leq a \\ &= (x-a)/(b-a), & a \leq x \leq b \\ &= (c-x)/(c-b), & b \leq x \leq c \\ &= 0, & x \geq c \end{aligned}$$

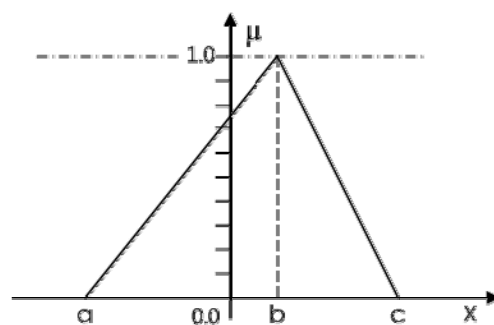


圖 10 三角模糊數

(2) 梯形模糊數(Trapezoidal Fuzzy Numbers)：

梯形模糊數以 $A(a,b,c,d)$ 表示， $\forall x, a, b, c, d \in R$ ， $\mu_A(x)$ 之計算及定義

如圖11：

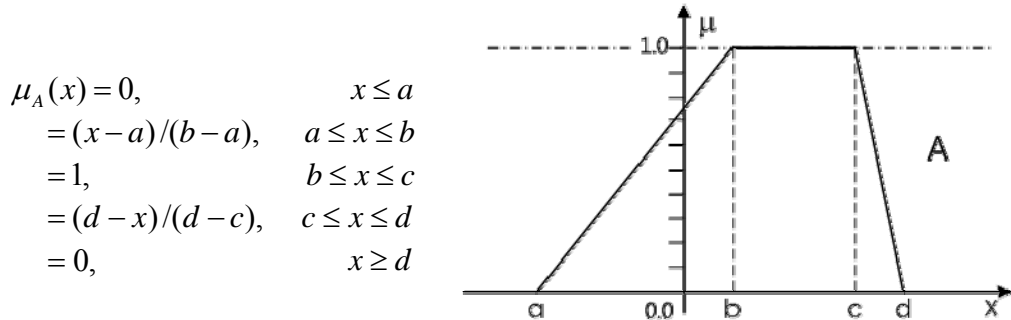


圖 11 梯形模糊數

參、模糊集合運算

以身高舉例說明如下：

一、 補集(Complement)：元素不屬於集合的程度。

$$\bar{A} = X \setminus A \Leftrightarrow \mu_{\bar{A}} = 1 - \mu_A(x) \quad (1)$$

例如：高=(0/180,0.25/182,0.5/185,0.75/187,1/190)

NOT高=(1/180,0.75/182,0.5/185,0.25/187,0/190)

二、 包含(Containment)：子集合的元素歸屬於集合的程度較小。

$$B \subseteq A \Leftrightarrow \forall x, \mu_B(x) \leq \mu_A(x) \quad (2)$$

例如：高=(0/180,0.25/182,0.5/185,0.75/187,1/190)

非常高=(0/180,0.06/182,0.25/185,0.56/187,1/190)

三、 交集(Intersection)：同時屬於兩集合的元素有多少。

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x), \text{ where } x \in X \quad (3)$$

例如：高=(0/165,0/175,0.0/180,0.25/182.5,0.5/185,1/190)

中等=(0/165,1/175,0.5/180,0.25/182.5,0.0/185,0/190)

則高 \cap 中等=(0/165,0/175,0/180,0.25/182.5,0/185,0/190)

四、 聯集(Union)：元素在每一集合中的最大歸屬值。

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x), \text{ where } x \in X \quad (4)$$

例如：高=(0/165,0/175,0.0/180,0.25/182.5,0.5/185,1/190)

中等=(0/165,1/175,0.5/180,0.25/182.5,0.0/185,0/190)

則高 \cup 中等=(0/165,1/175,0.5/180,0.25/182.5,0.5/185,1/190)

肆、解模糊化

模糊集合僅由範圍表示，因此轉換為一明確數值，可做為模糊排序過程中所使用的工具(邱閔鴻，民96)，解模糊化有許多方式，以下列出常用的三種(吳金照，民90)：

一、 重心法

Yager (1981)提出的重心法(Center of Gravity Method)，是目前最常使用的方法，主要是求取模糊集合的幾何中心，以中心值來代表整個模糊集合，計算公式如公式(5)-(7)所示：

假設 $f_i(x)$ 為三角形模糊函數 $T = (l, m, u)$ 之歸屬函數， $R(T)$ 為模糊集合 T 之重心值。

$$R(T) = \frac{\int_l^m x f_i(x) dx}{\int_l^m f_i(x) dx} \quad (5)$$

x 表示對歸屬度之一個重要性測量權數。

令 $T_i = (l_i, m_i, u_i)$ ， $i = 1, 2, 3, \dots, n$ ，為 n 個三角形模糊函數，則

$$\int_{l_i}^{u_i} f_{ii}(x)dx = \frac{1}{2}(l_i - u_i) \quad (6)$$

$$\int_{l_i}^{u_i} f_{ii}(x)dx = \frac{1}{6}(l_i + m_i + u_i)(u_i - l_i) \quad (7)$$

所以 T_i 的重心值為 $R(T_i) = \frac{1}{3}(l_i + m_i + u_i)$ ，由此可計算出 n 個模糊函數

的重心值

二、距離測量法

假設 $\tilde{T}_1 = (l_1, m_1, u_1)$ 及 $\tilde{T}_2 = (l_2, m_2, u_2)$ 分別為正三角模糊函數，其距離運算式如公式8，可作為轉換函數進行語意變數的解模糊化，如公式9：

$$d(\tilde{T}_1, \tilde{T}_2) = \sqrt{\frac{1}{3}[(l_1 - l_2)^2, (m_1 - m_2)^2, (u_1 - u_2)^2]} \quad (8)$$

$$R = \frac{d^-}{d^+ + d^*} \quad (9)$$

其中， $d^* = d(\tilde{T}, \tilde{T}^*)$ ， $d^- = d(\tilde{T}, \tilde{T}^-)$ ， \tilde{T}^* 為模糊最佳值(1,1,1)， \tilde{T}^- 為模糊最差值(0,0,0)。

三、最大歸屬度法 (Meam of Maximum Method)

以歸屬函數中歸屬度值最高的元素值作為該模糊集合的解模糊化值，若此元素值不只一個，則取所有相對應元素值的平均值，以其代表解模糊化值。

第三章 研究步驟

本研究希望能發展一智慧型搜尋引擎，利用圖形化介面，供使用者輸入查詢字串，並傳回搜尋後之結果，而此模式能協助使用者達到快速且準確之目標。本章將發展流程分為資料庫建置以及語意轉換二個主要階段。

一、資料庫建置：

此階段建置之資料庫，包含中文句型庫、5W1H意圖對應詞庫、數學領域知識本體資料庫以及數學專有名詞詞庫，各資料庫功能如表5。

表 5 句型庫種類與功能

名稱	功能
中文句型庫	將斷詞處理後所得之句型與中文句型庫進行比對，可判斷使用者所提問句為5W1H中何種類型問句。
5W1H意圖對應詞庫	將斷詞後各語詞與本意圖對應詞庫比對，可判斷使用者所提問句為5W1H中何種類型問句。
數學領域知識本體資料庫	將斷詞後各語詞與本資料庫比對，可擷取出數學領域相關詞彙，做為領域關鍵字。
數學專有名詞詞庫	僅收錄數學領域專有名詞，供斷詞後比對領域名詞，若有需要則進行詞性重設。

資料來源：本研究建置

二、語意轉換：

本階段利用前一階段所建之各資料庫分別進行句型分析，5W1H意圖轉換以及關鍵字擷取，最後利用既有搜尋引擎進行搜尋，語意轉換架構如圖12所示。

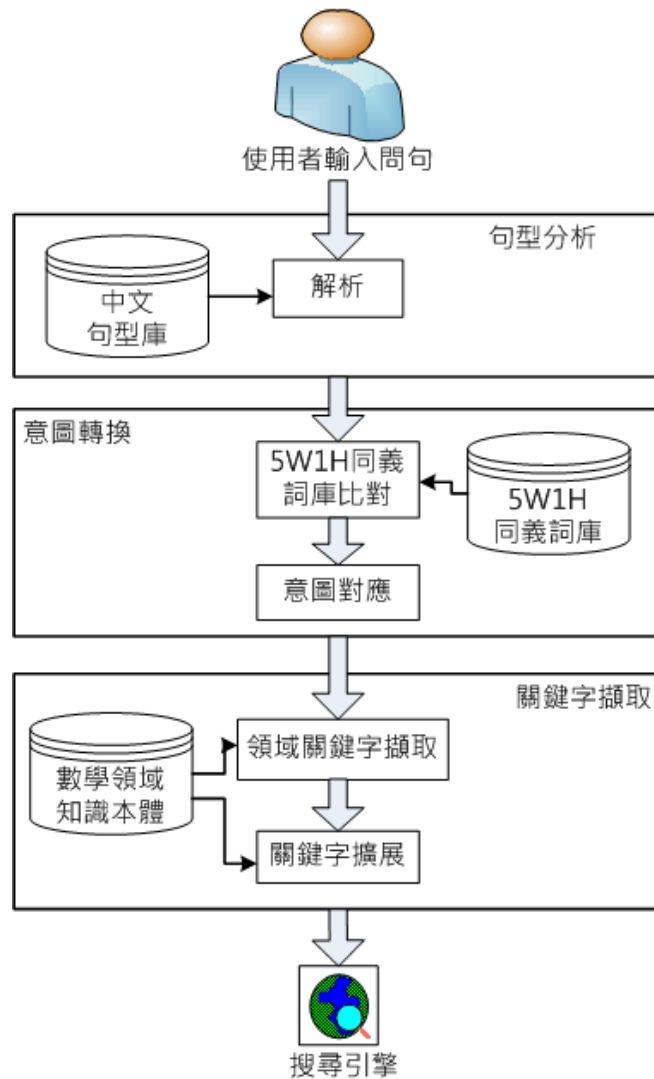


圖 12 圖形化介面語意轉換架構圖

以下將如何建構數學領域知識本體及進行語意轉換之各步驟運作方式詳細說明如下：

第一節 本體論架構

參考各學者所提建構知識本體步驟，本研究建立數學領域本體時分為四個步驟：

壹、決定研究領域與範圍

建置知識本體之初，可以先由下面幾個問題來導出實際需求，詳列如表6：

表 6 知識本體問題需求答問表

項次	問題	實際需求
1	知識本體涵蓋的領域及範圍為何？	國小數學領域
2	是否有現存領域本體可供利用？	尚未發現
3	知識本體的資訊應該用來回答哪一類的問題？	從事數學教學以及學習數學時可能遇到之問題
4	知識本體的使用者？	數學教學者與數學學習者
5	建置知識本體的目的？	建置資料庫供查詢之用

在TOVE的方法中，Fox and Gruninger (1998)認為建置領域本體時，可以利用「詢答集」來進行，此詢答集採階層方式，高層級為問題，回答之答案則為低層級。本研究依此「問題－答案」模式進行本體之建置，例如「幾年級開始學分數－二年級」、「甚麼是真分數－小於1」，藉由問題與答案的連結逐步形成本體架構。

貳、列舉知識本體中的重要詞彙

數學領域詞彙來源有二方面，一方面由九年一貫課程綱要之數學領域能力指標分年細目(教育部，民92)當中，依各學習階段尋找詞彙，整理如表7。

表 7 數學領域重要詞彙

階段	數與量	幾何	代數	統計與機率
一 (一—三年級)	數、計算、量、聲音、圖像、數字、加法、減法、乘法、除法、位數、累加、倍、積、單位、分、概數、等分、真分數、分母、合成、分解、離散量、連續量、和、被減數、交換律、括號	形體、分類、角、方位、相對位置、鉛垂線、水平線、水平面、直線、平面、輪廓線、周界、立體堆疊、面積、體積	算式	資料、分類、直接對應表格、一維表格
二 (四—五年級)	位值、計算機、直式、四捨五入、進位、捨去、估算、單位分數、等值分數、假分數、帶分數、時間、時刻、24 時制、長度、容量、重量、角度、面積、體積、單位，實測、估測、量感、千米、毫米、公升、毫公升、時、分、秒、等號、對稱性、遞移性、結合律、分配律、驗算、運算、數列、速率、距離、線段、數線。	幾何、性質、異同、東、西、南、北、位置、方向、鉛垂直線、水平直線、平行、張開程度、旋轉程度、平面圖形、線對稱。		長條圖、報讀、交叉對應表格、二維表格、統計圖
三 (六—七年級)	非負整數、億、兆、異分母、有限小數、近似值、誤差、千公斤、公噸、千公升、公乘、百平方米、公畝、千平方米、公頃、面積、公式、半徑、容量、容積、比、比例、正比、反比、比值、率、百分率、p. p. m.、因數、倍數、公因數、公倍數、面積公式、最大公因數、最小公倍數、質數、合數，質因數分解	直角、方位、位置、垂直、對稱、比例、三角形、四邊形、圓、梯形、長方形、平行四邊形	正、負、未知量、變量、	平均數、中位數、折線圖、圓形圖、機會、比值、百分率

資料來源：本研究整理

另一方面則以領域詞彙為基礎，5W1H問句類型為其子類別，分別依各問句類型假設題目並推想解答，此解答為實例，如表8以及表9，由這兩方面整理領域詞彙，以供本體建置之用。

表 8 「分數」之擴展詞彙

分數					
why	where	who	what	when	how
平均	遊戲	九章算術	平分	國小	整數
應用	埃及	劉徽	用途	二年級	有理數
分	起源	最早	應用	小學	分母
	網站	阿默斯	分母	課程標準	分子
	數學史		意義	埃及	分線
	數學		定義	考工記	小數
	中國			周髀算經	步驟

資料來源：本研究整理

表 9 「通分」之擴展詞彙

通分					
why	where	who	what	when	how
異分母	遊戲	教材教法	意義	國小	步驟
比較	網站	教學設計	介紹	五年級	作法
分數計算	練習	老師	同分母	小學	擴分
分數加法		困難	約分		約分
分數減法		研究	擴分		教法
			分母		
			分子		

資料來源：本研究整理

參、定義類別及其階層

Noy and McGuinness曾提出三種定義類別及其階層的方式，本研究採取由上而下方式進行，由範圍最廣之「數學」領域開始，其下接子類別「數」「量」「形」等概念，「數」之下又可分「整數」「分數」「小數」，為「數」之子類別，以此類推逐步建立類別及階層，如圖13所示：

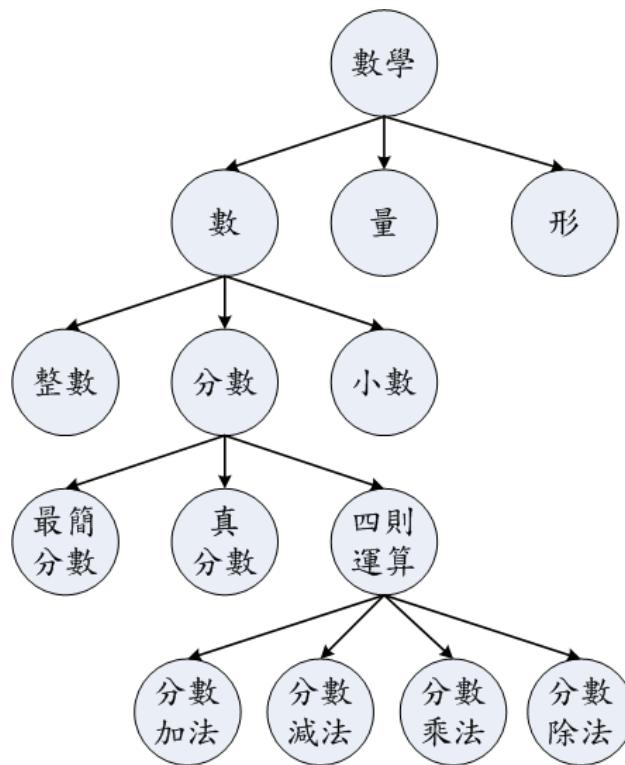


圖 13 數學領域類別階層

肆、建立實例

本研究利用Stanford大學發展之protégé軟體建立知識本體，該軟體使用介面如圖14所示，左側為數學各類別階層，右側為各類別之實例。

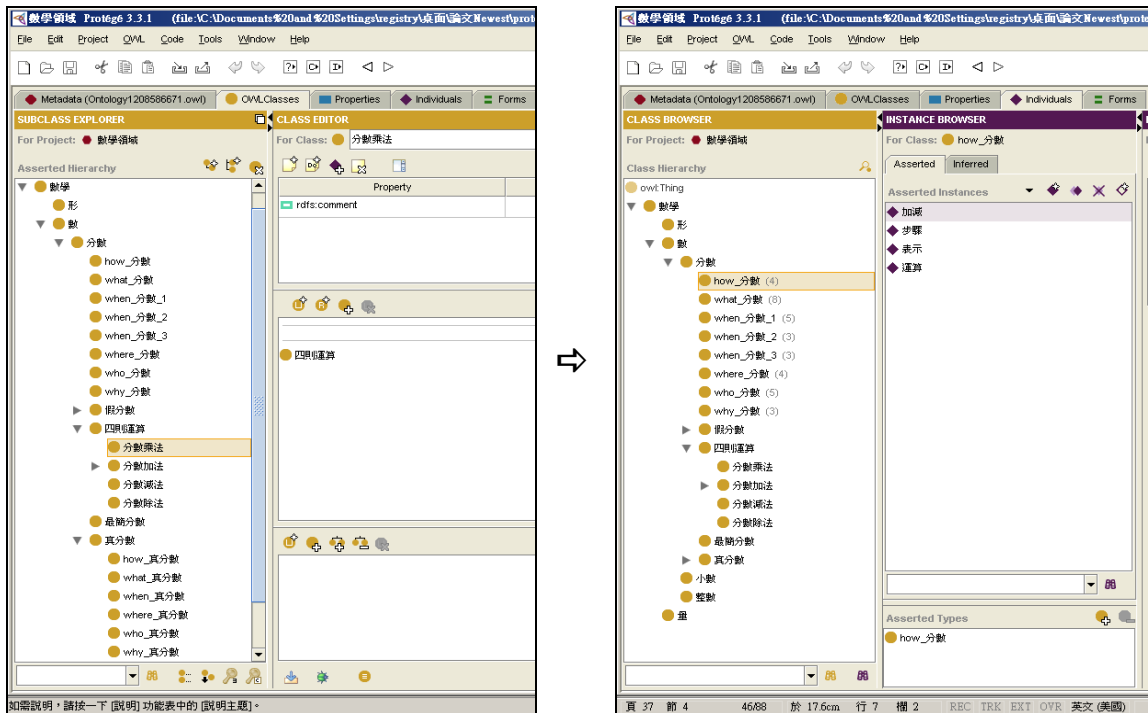


圖 14 數學領域本體之 Classes 與 Individuals

第二節 建構句型庫

壹、中文語法與 5W1H 意圖之結合

一個詢問句當中，可協助檢索的分為兩部分，一個是「關鍵詞部分」，另外則是「意圖部分」，而中文疑問詞問句與英文的WH 問句相當接近(李坤霖，民89)。因此本研究將問句類型區分為5W1H六種意圖，分別為why、where、who、what、when、how，利用問句分解後各語詞的詞性排列，將句型與意圖結合而歸類如表10。

表 10 句型與 5W1H 意圖之結合列表

意圖	問句	句型	關鍵字
why	為什麼(D)要(D)學(VC)分數(Na)	D-D-VC-Na	分數
why	學(VC)分數(Na)是(SHI)為什麼(D)	VC-Na-SHI-D	分數
why	為何(D)搭(VC)汽車(Na)	D-VC-Na	汽車
where	車站(Nc)在(P)哪裡(Ncd)	Nc-P-Ncd	車站
where	哪裡(Ncd)有(V_2)熊貓(Na)	Ncd-V_2-Na	熊貓
where	書(Na)放(VC)在(P)什麼(Nep)地方(Na)	Na-VC-P-Nep-Na	書
who	誰(Nh)會(D)教(VC)數學(Na)	Nh-D-VC-Na	數學
who	誰(Nh)是(SHI)班長(Na)	Nh-SHI-Na	班長
who	牛頓(Nb)是(SHI)什麼(Nep)人(Na)	Nb-SHI-Nep-Na	牛頓
what	假分數(Na)是(SHI)什麼(Nep)	Na-SHI-Nep	假分數
what	什麼(Nep)叫做(VG)假分數(Na)	Nep-VG-Na	假分數
what	何謂(VG)分數(Na)	VG-Na	分數
when	何時(Nd)開始(VL)學(VC)英文(Na)	Nd-VL-VC-Na	英文
when	什麼(Nep)時候(Na)離開(VC)	Nep-Na-VC	離開
when	幾(Neu)歲(Nf)讀(VC)幼稚園(Nc)	Neu-Nf-VC-Nc	幼稚園
how	分數(Na)要(D)怎麼(D)計算(VC)	Na-D-D-VC	分數
how	如何(D)學好(VC)英文(Na)	D-VC-Na	英文
how	畫(VC)梯形(Na)的(DE)方法(Na)	VC-Na-DE-Na	梯形

資料來源：本研究整理

貳、中文句型依 5W1H 意圖之分類

建構句型庫之目的在將句型依 5W1H 六種型態進行分類，本研究搜

集使用者問句，利用其斷詞後之結果進行整理，分類後之句型列於下方。每一問句可區分為數個部分，影響意圖判斷的部分，將各語詞之詞性列出，其餘無關的部分則合併為「Sentence」。以「為什麼要學分數」為例，斷詞結果為「為什麼(D) 要(D) 學(VC) 分數(Na)」，「為什麼(D)要(D)」可看出意圖為屬於why類型之問句，因此將詞性列出，而「學(VC) 分數(Na)」與意圖無關，因此僅以「Sentence」表示，方式如圖15所示，斷詞後各詞性代表之意義詳列於附錄一。

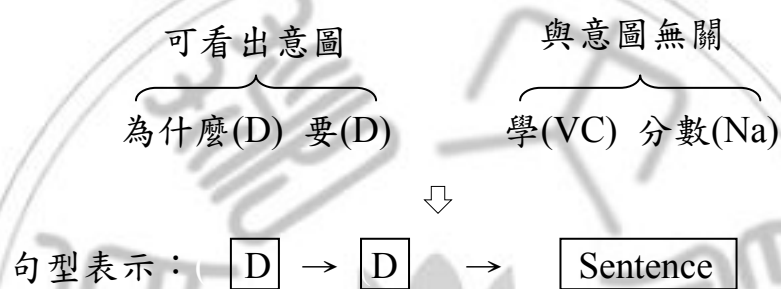


圖 15 句型歸類方法

一、 Why

[D] → [D] → [Sentence]

[Na] → [D] → [Sentence]

[Sentence] → [V] → [Na]

(V include VA、VB、VC、VD、VE、VF、VH、VJ)

二、 Where

[Ncd] → [V] → [Sentence]

(V include V_2、VA、VB、VC、VD、VE、VF、VH、VJ)

[Nep] → [Na] → [Sentence]

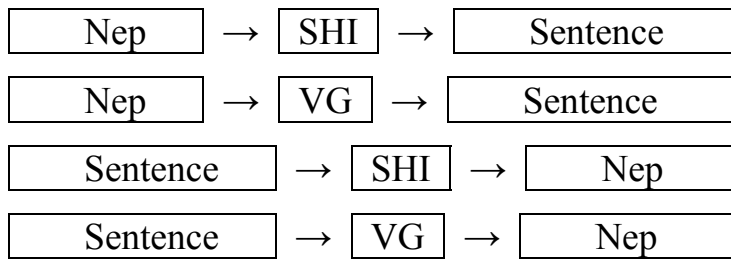
[Sentence] → [P] → [Ncd]

三、 Who

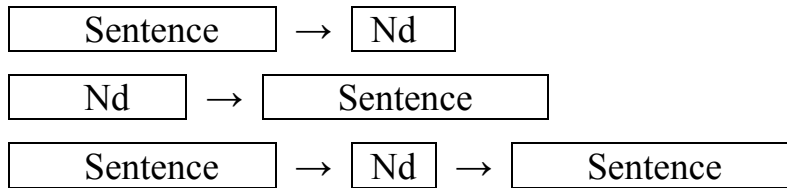
[Sentence] → [SHI] → [Nh]

[Nh] → [Sentence]

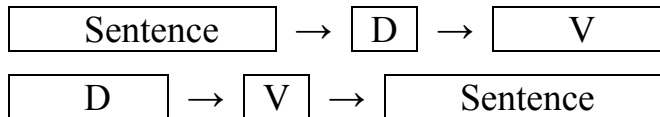
四、 What



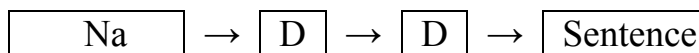
五、 When



六、 How



(V include VA、VB、VC、VD、VE、VF、VH、VJ)



將中文句型整理為上述六類之後，接著依據此六種型態建構中文句型庫，如圖16所示。

ID	Intention	Type
1	Nep-SHI-N	what
2	Nep-P-N	what
3	Nep-VG-N	what
4	N-VT-DET	what
5	N-P-DET	what
6	N-VG-DET	what
7	VT-SHI-N	what
8	VT-P-N	what
9	VT-VT-N	what
10	DET-VT-N	what
11	DET-VT-N-C-N	what
12	DET-VT-DET-M-NW	what
13	ADV-ADV-VT-N	how
14	DET-DET-M-N-P...	who
15	ADV-ADV-N	why
16	N-VT-DET-VT	why
17	N-ADV-VT-P-N	why
18	ADV-N-N-N-ADV-VI	why

圖 16 中文句型庫資料表

第三節 句型分析

本研究進行句型分析時分為三個步驟，分別為斷詞處理、詞性重設及句型比對，圖17為進行分析之流程圖，下面就各步驟詳細說明：

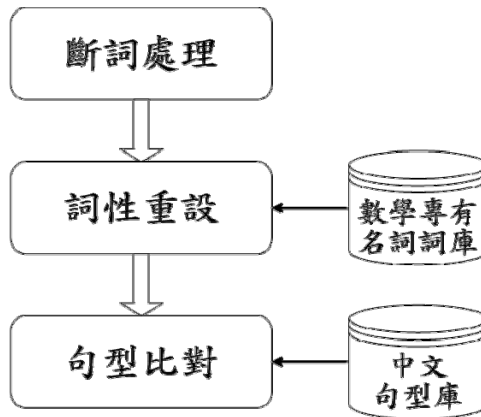


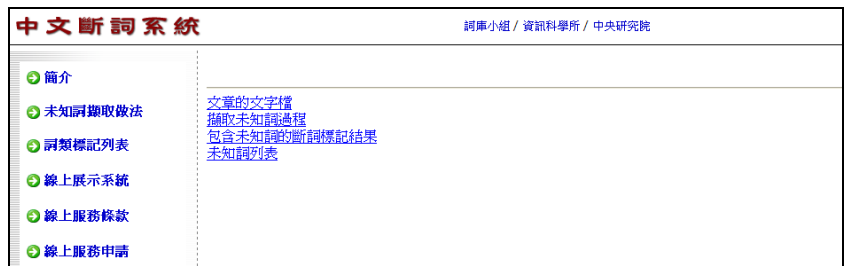
圖 17 句型分析流程圖

壹、斷詞處理

使用者輸入問句之後，利用中研院開發之「中文斷詞系統」進行處理，以問句「什麼是假分數」為例，經斷詞處理並解析後得結果為「什麼(Nep) 是(SHI) 假分數(Na)」，其進行步驟如圖18所示。



於展示系統中輸入「什麼是假分數」



選擇選項 3：包含未知詞的斷詞標記結果



輸出斷詞標記結果：「什麼(Nep) 是(SHI) 假分數(Na)」

圖 18 中研院斷詞系統處理流程圖

貳、詞性重設

詞性重設並非必定被執行之步驟，因中文斷詞系統非專為數學領域而設計，因此部分領域名詞可能無法被正確解析，若遇此狀況，則需進行語詞合併以及重設詞類標記。以專有名詞「最簡分數」為例，斷詞處理後，詞類標記原為「最(Dfa³) 簡分數(Na)」，如圖19，與數學專有名詞詞庫如圖20比對後，可將兩組語詞合併，成為「最簡分數」一詞，詞性則重設為普通名詞Na。

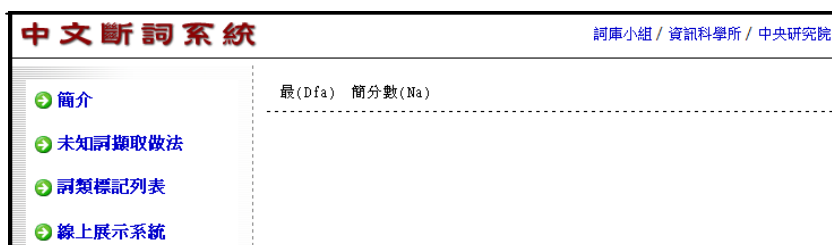


圖 19 「最簡分數」之斷詞結果

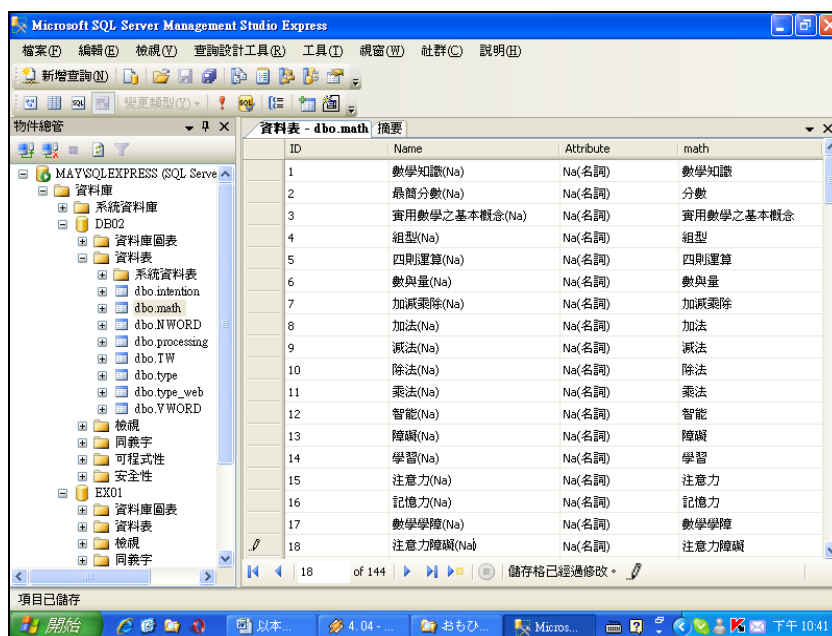


圖 20 數學專有名詞詞庫

³ Dfa 為程度副詞

參、句型比對

本研究以圖21所示之Knuth-Morris-Pratt(KMP)演算法做型態解析，以建立有限自動機(Finite Automata)，並利用中文句型庫做快速掃描，進而判斷出型態類型，如圖22所示。圖23顯示不同類型之六個問句，其實際比對並進行意圖轉換之流程，由輸入→斷詞→建立問句句型→意圖比對結果。

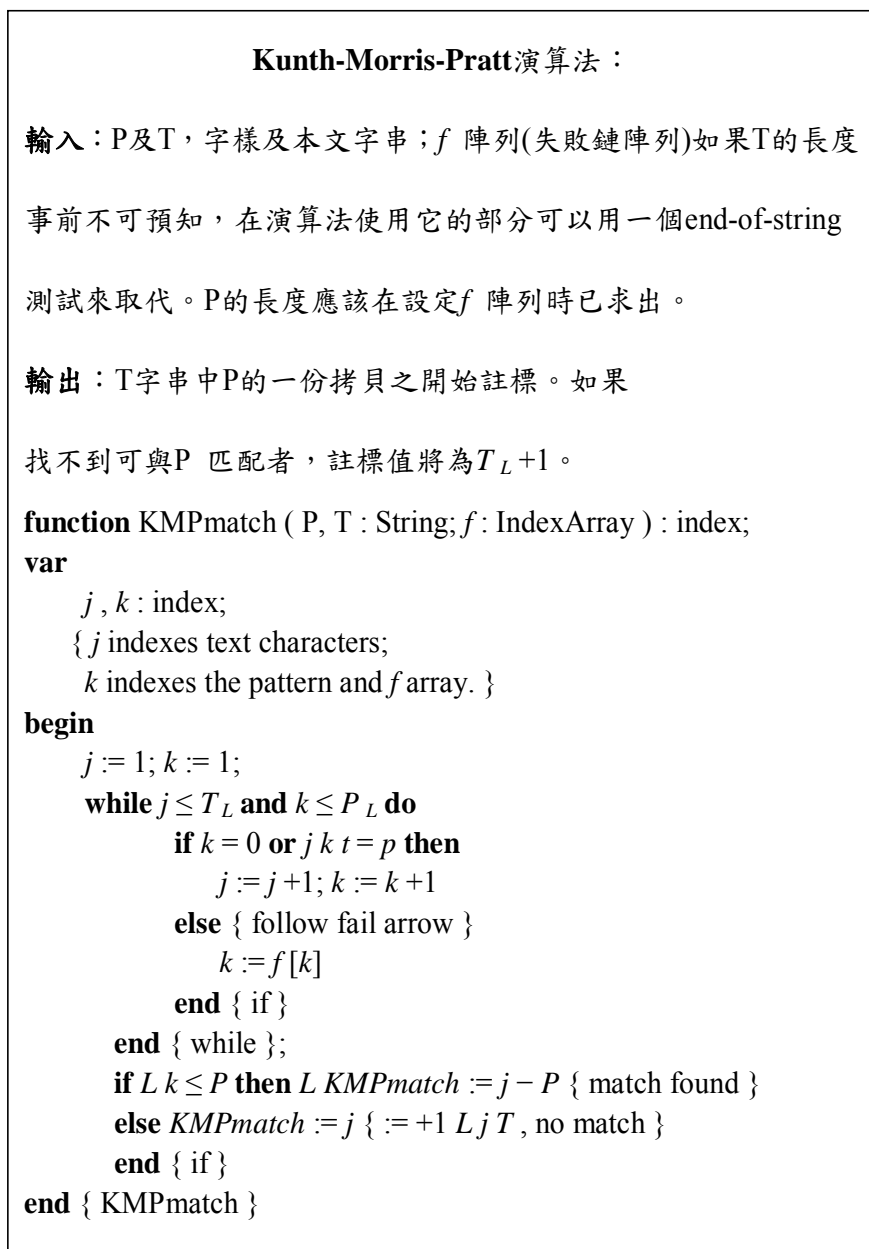


圖 21 Knuth-Morris-Pratt 演算法

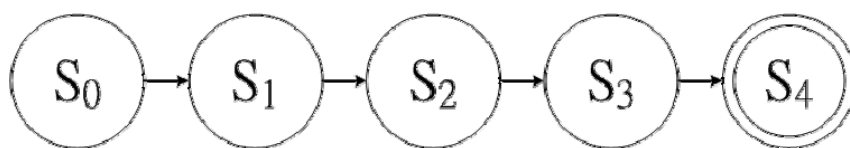


圖 22 有限自動機流程圖

輸入問句	CKIP 斷詞處理	斷詞結果
為什麼要約分→	為什麼(D)要(D)約分(VB)
哪裡可以學數學→	哪裡(Ncd)可以(D)學(VC)數學(Na)
誰發明分數→	誰(Nh)發明(VC)分數(Na)
什麼是假分數→	什麼(Nep)是(SHI)假分數(Na)
幾年級開始學分數→	幾年級(Nd)開始(VL)學(VC)分數(Na)
分數要怎麼分類→	分數(Na)要(D)怎麼(D)計算(VC)

↓ 輸出

比對出意圖	與句型庫 比對	問句句型
why	←.....	D-D-VB
where	←.....	Ncd-D-VC-Na
who	←.....	Nh-VC-N
what	←.....	Nep-SHI-Na
when	←.....	Nd-VL-VC-Na
how	←.....	Na-D-D-VB

圖 23 句型分析流程圖

第四節 意圖轉換

本章第二節雖已將問句型態做出六種類型區分，但中文語言相當複雜，此基本歸類仍嫌不足，尚須與5W1H意圖對應詞庫比對，以達到正確之意圖判斷，如此方能進行意圖轉換，方法如下。

將使用者輸入之問句斷詞，所得各詞組經詞性重設後，與5W1H意圖對應詞庫比對，對應出使用者之意圖類型，本研究使用之5W1H意圖對應詞庫如表11與圖24所示。

表 11 5W1H 意圖對應表

意圖	對應字
why	為什麼、怎麼這樣、因何、何因、何故、為何會、怎麼會、什麼原因
where	何地、何處、哪裡、那兒、哪裡有、什麼地方、什麼地點、在哪邊、何方、在哪、在哪邊、在哪
who	誰、何人、什麼人、哪一個人
what	什麼、何謂、什麼是、是什麼、什麼叫做、叫做什麼、意義、何意
when	何時、什麼時候、幾時、什麼時間、什麼階段、幾年級、幾歲
how	如何做、怎麼做、要怎樣、要怎麼、什麼方法、什麼方式、如何解、作法、做法、怎樣做、怎麼辦

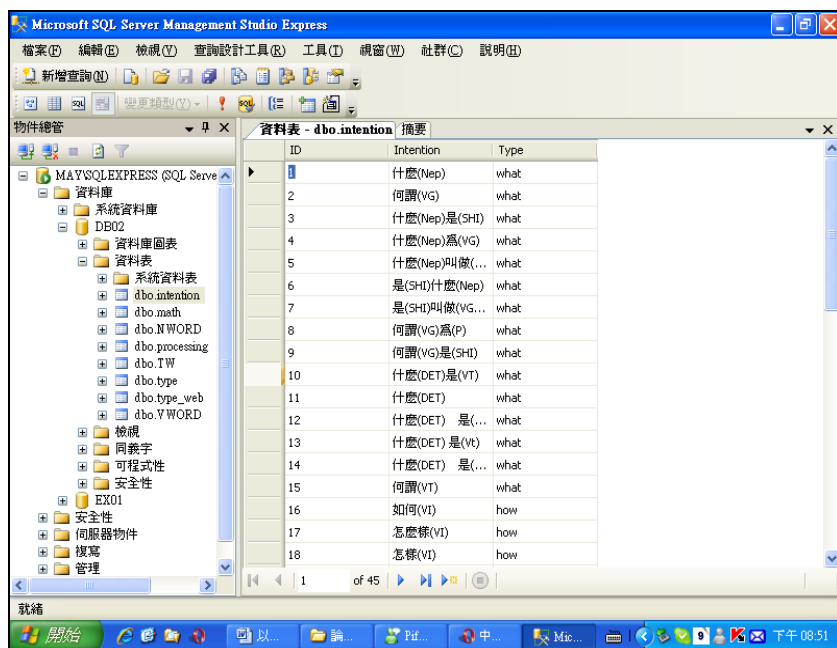


圖 24 5W1H 意圖對應詞庫

以問句「約分的作法」為例，經斷詞系統解析後得「約分(VB)的(DE)作法(Na)」，由表11得知「作法」為「how」的對應字，因此判斷本問句之意圖屬於「how」的類型。

第五節 關鍵字擷取

本研究擷取名詞、動詞做為關鍵字，由中研院詞庫小組技術報告 93-05發表之「中文詞類分析」中，參考楊宸彥(民91)之研究，取出一般問句中可能出現之關鍵詞類，比對詞類標記為名詞如Na、Nb、Nc、Nd以及動詞如VA、VB、VC...等之語詞，將其列為候選關鍵字，整理如表12。

斷詞系統雖解析出語詞，但僅僅標示詞性而無該語詞本身所代表之意義，因此將語詞與資料庫進行比對，並採用過濾及擴展兩種方式擷取關鍵字。

表 12 候選關鍵字詞性標記列表

詞類標記	屬性	語詞範例
Na	普通名詞	數學、分數、數學家、原因
Nb	專有名稱	牛頓、畢達哥拉斯
Nc	地方詞	台灣、嘉義、阿拉伯
Nd	時間詞	端午節、中秋節
VA	動作不及物動詞	教學、上課、游泳
VB	動作類及物動詞	約分、分類、除名
VC	動作及物動詞	發明、練習、搜尋、解決
VD	雙賓動詞	索取、寄、送
VE	動作句賓動詞	詢問、提示、討論
VF	動作謂賓動詞	勸告、企圖、打算
VH	狀態不及物動詞	高大、急促、稀少
VJ	狀態及物動詞	剩下、長達

壹、關鍵字過濾

問句經斷詞處理後，將產生多組語詞，首先去除不可能之詞彙，留下候選關鍵字，接著結合數學領域本體架構，選擇領域相關詞彙，即為數學領域關鍵字，此種由多組語詞中篩選出領域關鍵字之過程稱為關鍵字過濾。

以「分數在實際生活中的應用為何」為例，斷詞處理結果得「分數(Na) 在(P) 實際(VH) 生活(Na) 中(NG) 的(DE) 應用(Na) 為何(D)⁴」，參考表12可看出，其中候選關鍵字有四個，分別為詞性屬於Na的「分數、生活、應用」等三個語詞，以及詞性為VH的語詞「實際」，

⁴ P、NG、DE、D 分別代表介詞、後置詞、副詞、「的、之、得、地」

將這四個候選關鍵字與數學專有名詞詞庫比對，得出數學領域方面的關鍵字應為「分數」，圖25為進行關鍵字過濾之流程。



圖 25 關鍵字過濾流程圖

貳、關鍵字擴展

以領域關鍵字為基礎，根據該問句之意圖以及領域本體，擴展出更多相關詞彙，此種由一個關鍵字增加為數個關鍵字的過程稱為關鍵字擴展。

以「分數在實際生活中的應用為何」為例，過濾後之領域關鍵字為「分數」，且意圖對應後得出本句屬於「what」類型之問句，將「分數」與「what」結合，利用意圖對應詞庫以及如圖26所示之領域知識本體資料庫進行比對後，得出數個語詞，如圖27所示。這些語詞為往後進行搜尋時將選用之關鍵字，稱為擴展關鍵字。

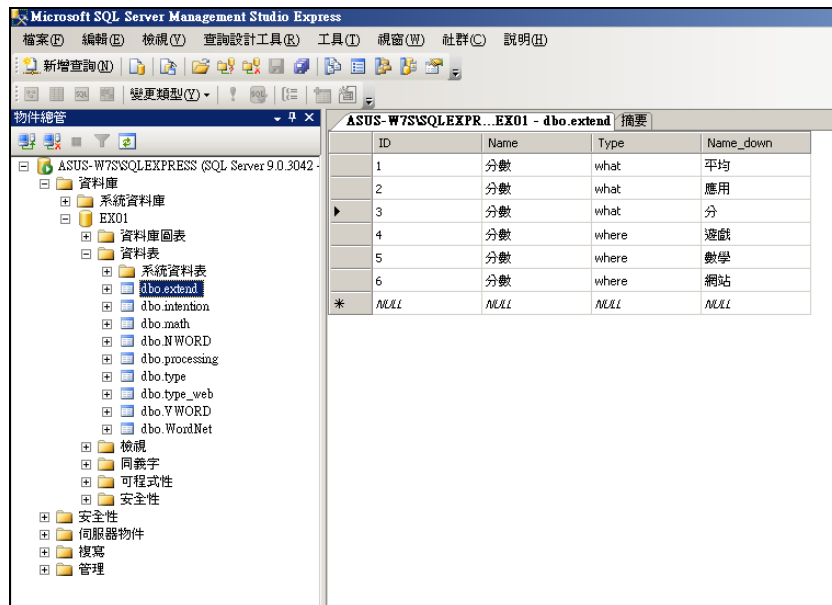


圖 26 數學領域知識本體資料庫

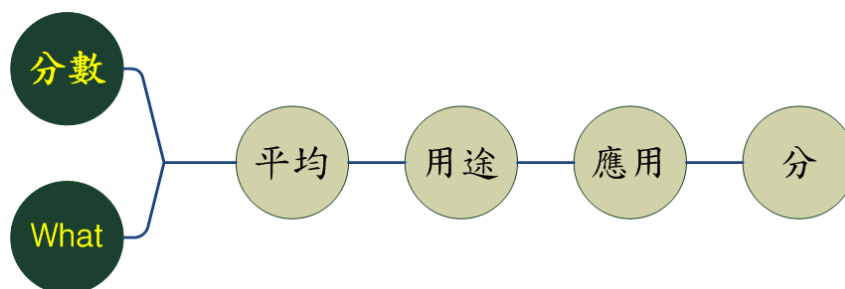


圖 27 分數與 what 結合之擴展關鍵字

將本章所述之系統運作過程，自使用者輸入問句開始至輸出關鍵字置於搜尋引擎查詢為止的整體流程，大致上可分為七個步驟進行，試舉詢問句「通分要怎麼做」為例，圖28可清楚表示其流程，並說明如下：

1. 輸入問句：通分要怎麼做。
2. 斷詞解析：「通分(Nd) 要(D) 怎麼(D) 做(VC)」。
3. 意圖轉換：得問句類型為「how」。

4. 關鍵字過濾：得領域關鍵字「通分」。
5. 關鍵字擴展：由資料庫擴展出如表9所示之意圖關鍵字「步驟、作法、擴分」。
6. 關鍵字結合：將領域關鍵字與意圖關鍵字結合，得「通分、步驟、作法、擴分」。
7. 網路搜尋：將關鍵字組置於網路搜尋引擎進行搜尋。

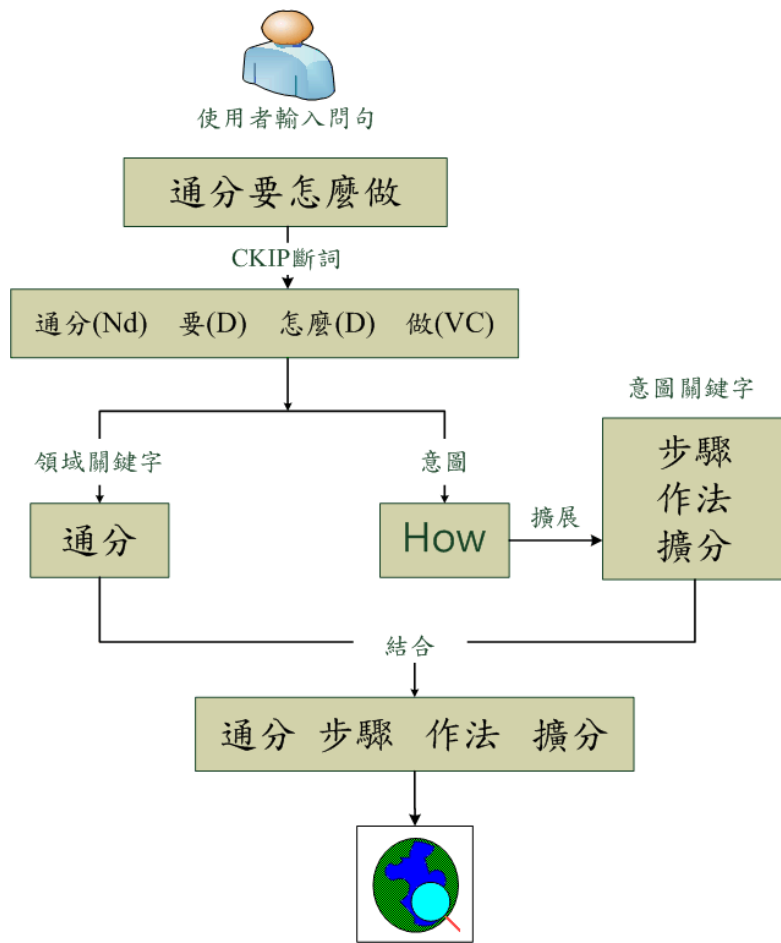


圖 28 「通分要怎麼做」之查詢流程

第四章 實驗驗證

本研究設計一智慧型搜尋引擎，以驗證第三章所提之方法是否可行。藉由該介面進行實驗資料蒐集，並以下列四個問題進行分析，以驗證本研究所提之查詢模式是否能達到快速且準確之目標：

一、搜尋結果之前三十筆文件中，與使用者需求相關之文件數量為何？依此計算其準確率。

二、系統給予之關鍵字數量愈多，是否愈能提高準確率？

三、符合使用者需求之資料，在前三十筆搜尋結果中出現的順序是否往前提升？

四、關鍵字組之排列組合與資料之排序之相關性如何？

本章首先介紹驗證系統，接著提出實驗結果，最後根據所得資料進行分析。

第一節 網頁擷取內容設計

本研究建置SQL Server及WEB Server進行實驗驗證，所用之軟硬體設備如下：

壹、實驗環境

一、硬體：桌上型電腦

1. CPU：Intel Core2 Quad CPU Q6600 @2.40GHz

2. RAM：2.00GB

二、軟體

1. 作業系統：Microsoft Windows XP Professional Service Pack 2。

2. 資料庫系統：Microsoft SQL Server 2005 Express Edition。

3. 程式開發平台：JDK 1.6.0、SQL JDBC 1.1.1501.101、
NetBeans IDE 5.0。

4. 程式語言：Java。

貳、實驗設計

本研究所使用之資料樣本，皆透過網路搜尋而得，考慮經由Yahoo搜尋引擎搜尋時，可一併搜尋出論壇型態之Yahoo知識資料，因此本研究以Yahoo搜尋為主。自搜尋所得之前三十筆資料中，逐筆檢視該筆資料是否與使用者需求相關。

實驗資料分兩組進行搜集，如圖29：

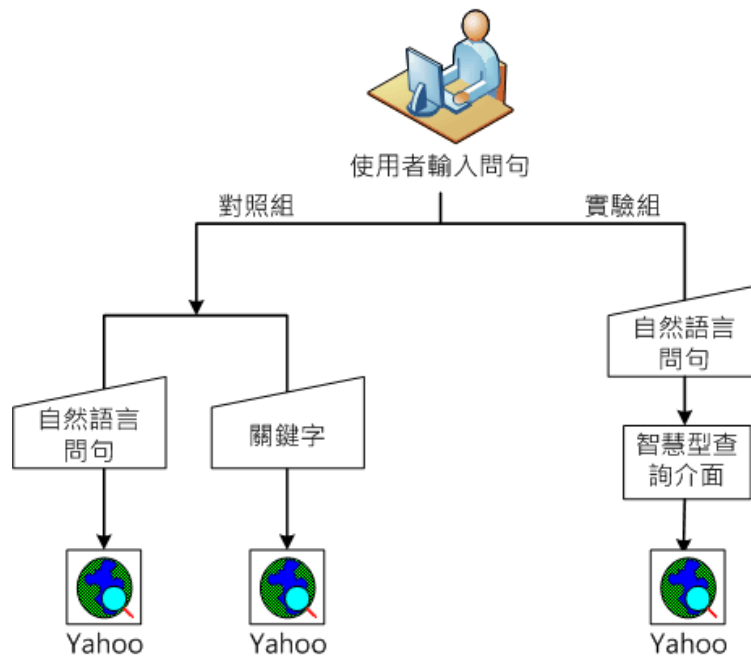


圖 29 實驗進行流程圖

一、對照組：

直接利用Yahoo搜尋引擎進行搜尋，語詞輸入分為兩類：

1. 全句搜尋：

輸入與數學領域相關之自然語言問句，如「幾年級開始學分數」。

2. 由字面拆解出可能關鍵字進行搜尋，如「幾年級、分數、學分數、開始」。

二、實驗組：

透過本研究設計之界面，輸入詢問句，由系統自動篩選關鍵字，並置於Yahoo搜尋引擎進行搜尋，問句之產生係任選數學領域關鍵字，並結合意圖類型「why、where、who、what、when、how」而得。

參、系統設計

一、網頁介面：

圖30為本研究設計之搜尋介面，供使用者輸入自然語言進行搜尋。



圖 30 本研究設計之搜尋介面

二、查詢步驟：

以問句「哪裡有分數遊戲」為例，搜尋步驟如下：

步驟一：使用者於介面中輸入問句，如圖31。

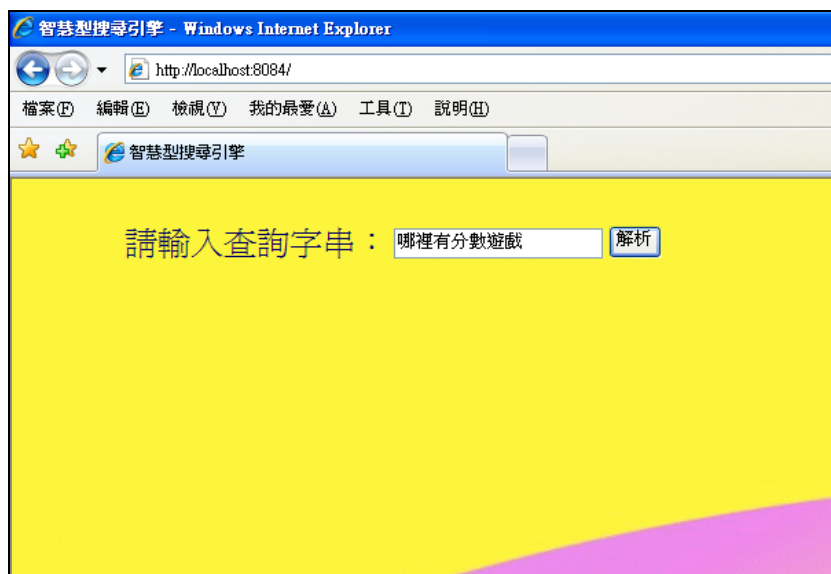


圖 31 輸入自然語言問句

步驟二：系統處理後，得意圖為where，並列出四個關鍵字：分數、遊戲、數學、網站，如圖32。

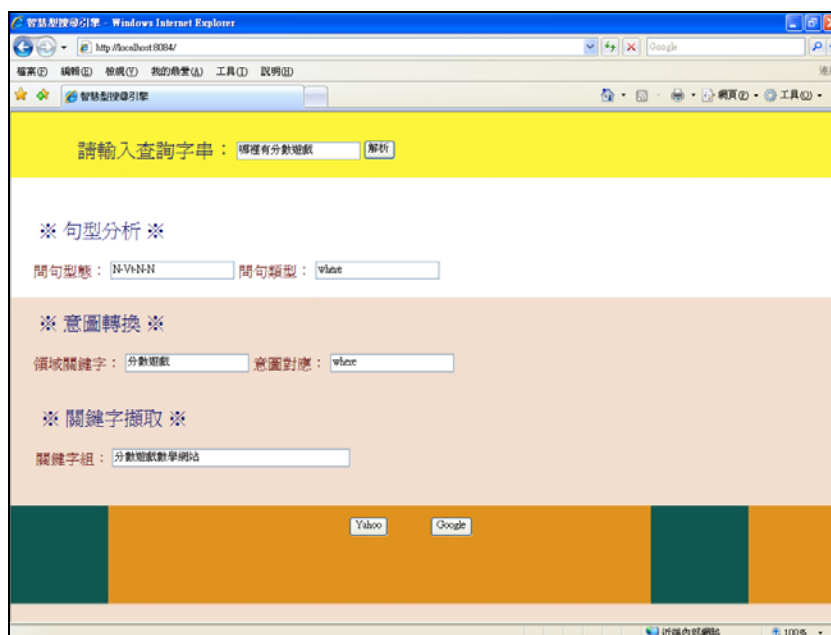


圖 32 系統比對出意圖及關鍵字

步驟三：將步驟二所得之關鍵字組傳送至Yahoo搜尋引擎進行搜尋，結果如圖33。



圖 33 Yahoo 搜尋引擎之搜尋畫面

第二節 實驗結果

本研究利用31個問句進行測試，以Chakrabarti (2003)所提之公式進行準確率(Precision ratio)計算，如公式10。

$$precision(k) = \frac{1}{k} \sum_{1 \leq i \leq k} r_i \quad (10)$$

其中 k 代表取樣文件數，在本研究中取前三十筆資料檢視，因此 $k = 30$ ， r_1, r_2, \dots, r_k 代表第一至三十筆資料，若第 k 筆為符合⁵使用者需求之資料，則 $r_k = 1$ ，不符合者 $r_k = 0$ ，紀錄方式如表13。

⁵ 符合之判定方式為檢視資料時，若該筆資料內容可解答使用者之提問，則將該筆資料視為「符合」

表 13 取樣文件記錄示例

k	r_k	k	r_k	k	r_k
1	1	11		21	
2		12		22	
3		13	1	23	
4		14		24	
5	1	15		25	
6		16		26	
7		17		27	1
8	1	18		28	
9		19		29	
10		20		30	

註：空格代表0，省略不記

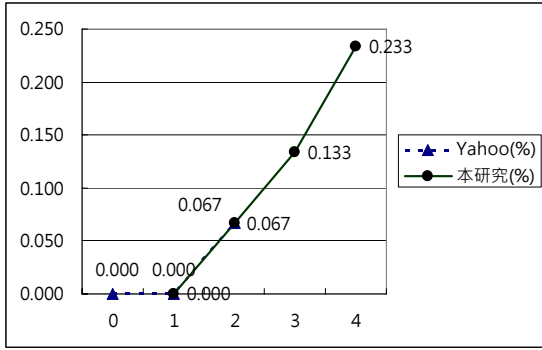
系統除領域關鍵字之外，另加入意圖關鍵字合併搜尋，表14列出進行實驗的問句、類型、使用之關鍵字，以及分別利用本研究設計及Yahoo搜尋引擎進行搜尋之結果，結果以準確率差距表示。

除兩種搜尋模式之準確率差距外，本研究還希望了解關鍵字數與準確率的相關性，因此在前面8筆問句中，另外記錄關鍵字數由一個逐漸增加至四個時，搜尋結果之準確率，並以此繪製圖34，圖中縱軸為準確率，橫軸為關鍵字個數。Yahoo搜尋之記錄中關鍵字數為0表示直接輸入問句，另外依問句不同，採用之關鍵字由一至三個不等，本研究採用之智慧型搜尋，關鍵字數則均為一至四個。

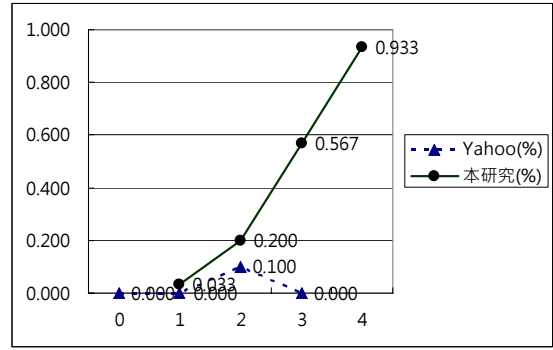
表 14 本研究與Yahoo搜尋引擎搜尋結果準確率之比較

編號	實驗問句	類型	準確率 差距(%)	關鍵字
1	幾年級開始學分數	when	0.167	分數幾年級課程標準國小
2	分數是甚麼時候出現的	when	0.833	分數埃及考工記周髀算經
3	最早研究分數的國家	where	-0.100	分數數學史埃及中國
4	哪裡有分數遊戲	where	0.367	分數遊戲數學網站
5	什麼是分數	what	0.033	分數平分意義分母
6	分數加減要怎麼做	how	0.100	分數加減運算步驟
7	甚麼是真分數	what	0.167	真分數名詞解釋定義分子小
8	為什麼 0 不能當分母	why	-0.100	分母 0 理由意義
9	通分要怎麼做	how	0.133	通分步驟分母最小公倍數
10	四邊形怎麼分	how	0.033	四邊形種類定義
11	四邊形有幾種	what	0.133	四邊形分類角度邊
12	三角形的面積怎麼算	how	-0.033	三角形面積公式算法
13	三角形分成幾類	what	0.200	三角形分類角度
14	怎麼進行分數教學	how	-0.033	分數教學設計方法
15	甚麼是最簡分數	what	0.267	最簡分數定義約分
16	圓周率是甚麼時候發明的	when	0.100	圓周率數學史最早
17	甚麼是圓周率	what	0.200	圓周率定義近似值
18	圓周率要怎麼算	how	-0.100	圓周率方法步驟
19	誰發明了圓周率	who	0.133	圓周率數學史最早
20	想要分數的題目	where	0.000	分數數學題目試卷
21	如何求最小公倍數	how	-0.067	步驟算法公式
22	甚麼是最小公倍數	what	-0.033	名詞解釋定義
23	如何求近似值	how	0.067	取法步驟
24	甚麼是近似值	what	0.100	名詞解釋定義
25	甚麼是概數	what	0.133	名詞解釋定義
26	請解釋等值分數	what	0.167	名詞解釋定義
27	要怎麼算等值分數	how	0.000	步驟算法公式
28	甚麼是鉛錘直線	what	0.000	名詞解釋定義
29	誰發明阿拉伯數字	who	0.067	0 數學史最早
30	甚麼是 24 時制	what	-0.033	名詞解釋定義
31	24 時制是誰規定的	who	0.067	制定

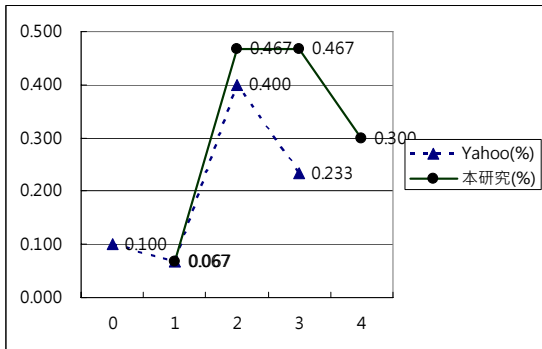
註：準確率差距 = 本研究搜尋準確率 - Yahoo 搜尋準確率



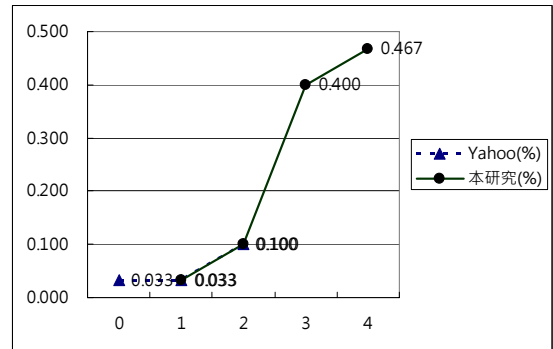
問句1



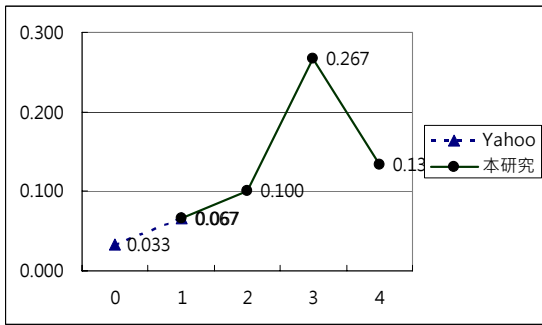
問句2



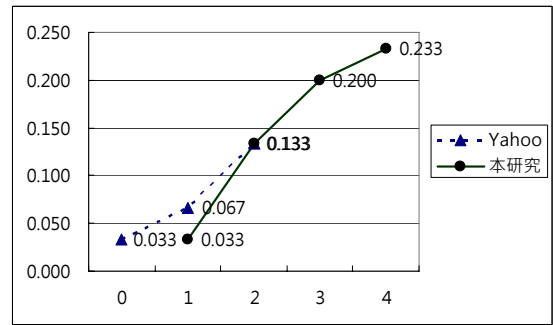
問句3



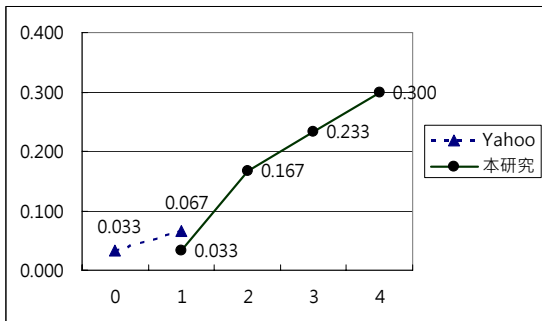
問句4



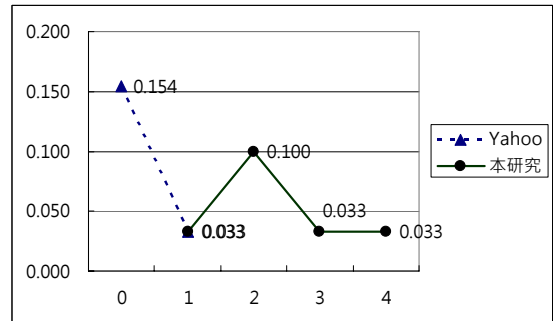
問句5



問句6



問句7



問句8

圖 34 關鍵字個數不同之搜尋結果準確率比較圖

第三節 資料分析

依據實驗結果，本節就前述問題進行分析，驗證是否達成本研究之目的，茲分述如下：

一、搜尋資料準確率是否提高之分析

本研究針對Yahoo與本研究設計界面之搜尋準確率進行分析，採用相依樣本t檢定法。

設 μ_D 代表準確率差距的平均數，若 $\mu_D = 0$ 表示兩者並無差異(H_0)，若 $\mu_D > 0$ 則表示本系統確實能提高資料搜尋的準確率(H_1)，分析結果如表15。

表 15 搜尋準確率之成對母體平均數差異檢定

	本研究	Yahoo
平均數	0.216	0.120
變異數	0.035	0.012
觀察值個數	31	31
皮耳森相關係數	0.386	
假設的均數差	0	
自由度	30	
t 統計	3.018	
P($T \leq t$) 單尾	0.003	
臨界值：單尾	1.697	

由於P值 < 0.01 ，顯示本研究設計之搜尋系統的確能提高資料搜尋之準確率。

二、關鍵字數量與搜尋準確率是否相關

關鍵字數之數量與問句查詢之準確率是否有其關聯性存在？觀察表16以及圖35，當關鍵字數小於或等於3時，關鍵字數量與搜尋準確率

多呈現正相關，但關鍵字數大於3時，則出現兩筆負相關，因此本研究針對關鍵字數為3及4時，進行相依樣本t檢定。

表 16 關鍵字個數不同之問句準確率列表

關鍵字數	資料別							
	1	2	3	4	5	6	7	8
1	0.000	0.033	0.067	0.033	0.067	0.033	0.033	0.033
2	0.067	0.200	0.467	0.100	0.133	0.133	0.167	0.100
3	0.133	0.567	0.467	0.400	0.267	0.200	0.233	0.033
4	0.233	0.933	0.300	0.467	0.167	0.233	0.300	0.033

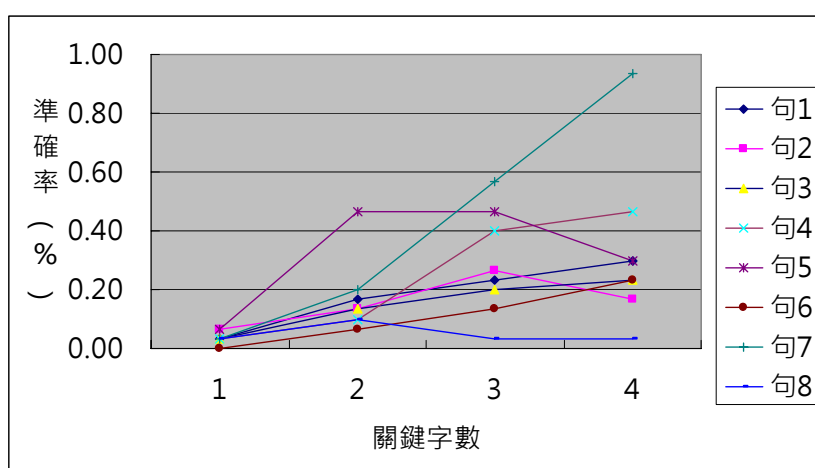


圖 35 關鍵字個數與搜尋準確率之關係圖

設 μ_D 代表準確率差距的平均數， $\mu_D = \mu_4 - \mu_3$ 。

設 $H_0: \mu_D = 0$ ，兩者並無差距。

$H_1: \mu_D > 0$ ，利用4組關鍵字進行查詢，結果優於3組。

t 檢定分析結果如表 17，P 值為 0.220，大於 0.05 顯著水準，表示利用 4 個關鍵字或 3 個關鍵字進行查詢時，兩者之準確率並無顯著差異。

表 17 關鍵字個數不同之 t 檢定結果

	4 字	3 字
平均數	0.333	0.288
變異數	0.074	0.032
觀察值個數	8	8
皮耳森相關係數	0.832	
假設的均數差	0	
自由度	7	
t 統計	0.819	
$P(T \leq t)$ 單尾	0.220	
臨界值：單尾	1.895	

三、符合使用者需求之資料，在樣本中出現的排序是否提升？

本研究取三個問句，將樣本中符合意圖資料的所在位置繪製成表 18，以觀察其資料排序狀況。第一列之 1~30 分別代表文件的排序編號，另外關鍵字數為 0,1,2,3,4 時，分別以紫、藍、紅、綠、黃五色標記符合使用者意圖之文件出現的位置。以第二列為例，紫色方格所在位置為 0 與 21 交會處，此代表關鍵字數為 0 時，檢視至第 21 筆文件方能查詢出相關資料。

由表 18 可看出，利用本研究之搜尋介面進行查詢時，問句 1 與問句 2 相關資料之排序確實有往前提升之趨勢，而問句 3 雖無較優之結果，但也能保持與直接利用 Yahoo 進行搜尋時之同等排序狀況。

表 18 符合需求樣本於搜尋結果中出現之順序

搜尋介面	關鍵字數	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	結果	
Yahoo	0																																1
	1			藍		藍																											2
	2		紅										紅							紅							紅						4
本研究	1			藍		藍																										2	
	2		紅										紅							紅							紅					4	
	3	綠			綠								綠					綠	綠					綠								6	
	4	黃	黃									黃			黃			黃									黃			黃		7	
Yahoo	0			紫	紫									紫																		4	
	1		藍																													1	
	2			紅																												1	
本研究	1		藍																													1	
	2	紅				紅	紅				紅			紅																		5	
	3	綠	綠																綠									綠	綠	綠	綠	7	
	4	黃		黃		黃	黃									黃		黃	黃					黃			黃					9	
Yahoo	0												紫																			1	
	1			藍	藍										藍																	3	
	2																															0	
本研究	1			藍	藍																											2	
	2												紅							紅												2	
	3											綠		綠													綠					3	
	4														黃															黃		3	

四、關鍵字在詞組中之排列是否影響搜尋之準確率與資料排序？

以問句「什麼是分數」為例，分別以第一組關鍵字組「分數-意義-平分-分母」與第二組關鍵字組「分數-平分-意義-分母」進行查詢，此二組雖採用相同關鍵字，但其排序不同，準確率分別為 0.133% 與 0.167%，以第二組較高。

將搜尋樣本之前十五筆資料排序記錄於表19，發現其中第9、10、15三筆資料差距均達到10以上，資料排序前後差距甚大，但此三筆資料均為非相關資料，符合意圖之資料4與資料7則差異不大。

表 19 關鍵字組排列相異之排序對應表

「分數-意義-平分-分母」 資料序	「分數-平分-意義-分母」 資料序	差距
1	1	0
2	2	0
3	3	0
4*	6	2
5	5	0
6	4*	2
7*	9	2
8	7*	1
9**	22	13
10**	>30	
11	13	2
12	10	2
13	16	3
14	19	5
15**	30	15

註：*表示該筆資料符合使用者意圖，**表示該筆資料差距>10

第五章 結論與未來方向

本研究之目的，主要是希望能發展出一種智慧型搜尋引擎，使用者經由此引擎進行資料查詢時，能提高搜尋之準確率並減少搜尋時間，因此提出建構領域本體，並結合中文語法與5W1H意圖擴展關鍵字之方法進行搜尋。搜尋時仍使用現有之搜尋引擎為工具，但中間需先經過本研究設計之介面處理關鍵字擷取部分，經實際驗證及資料分析，顯示確實可達到這兩個目的，足證本研究提出之方法確實可行。

本章分為兩部分，第一節針對研究時之發現及結果分析提出結論，第二節將本研究不足部分提出，供後續研究參考。

第一節 結論

本研究提出之方法，經實際設計搜尋系統，搜集資料分析後，得出以下結論：

- 一、利用智慧型搜尋引擎進行查詢時，使用者僅需輸入自然語言，系統即自動產生多個關鍵字，已達到自動化之目標。
- 二、以智慧型搜尋引擎進行數學領域知識查詢時，與一般關鍵字搜尋方式進行準確率之 t 檢定，其 P 值小於顯著水準0.01，由此可知此搜尋模式確實可以提高資料查詢之準確率。
- 三、以智慧型搜尋引擎進行數學領域知識查詢時，與利用一般關鍵字搜尋方式相比較，符合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。
- 四、本研究利用「問題－答案」模式進行數學領域本體之建置，所建構之領域本體在系統擷取關鍵字時佔一關鍵地位，由實驗結果觀

之，此本體有其正確性及可用性存在。

五、本研究希望能找出影響資料搜尋準確率與時間的因素，針對部分可能原因進行實驗，得出以下結論：

1. 本研究擷取相同關鍵字組，更動其中語詞之排列進行實驗，發現此更動不只造成相關資料之排序異動，準確率亦有影響，可見關鍵字組之間的排列組合與準確率及搜尋時間均有關聯存在，表示即使關鍵字組為相同之語詞，其間之排列組合仍須不斷修正，方能得到最佳效果。
2. 系統給予之關鍵字數在3字以下時，字數與準確率大致上成正相關，但超過3字以後開始有準確率下降的情形，應為關鍵字選取之因素，因此資料庫能否完善建置相形重要。

六、本研究所提出之結合領域本體與5W1H意圖產生領域關鍵字與意圖擴展關鍵字之方式，能達到正確選用關鍵字目的，提高搜尋準確率。

此外本研究尚發現，以智慧型搜尋引擎進行查詢時，可找出較多數學領域相關文件，對數學知識的獲得有幫助。而系統進行意圖對應時，首先由中文句型資料庫產生，應再輔以5W1H意圖對應詞庫，如此可產生較正確之意圖，並藉由此擴展出意圖關鍵字。

第二節 未來方向

本研究設計之智慧型搜尋引擎，雖經驗證所提之方法確實可行，但仍有未及之處，在此提出以供後續研究之參考：

- 一、 關鍵字之排列組合影響搜尋時間，因此應考慮加入語詞重要度之計算，以提升搜尋引擎之效率。
- 二、 目前資料庫之建置採人工建立方式，費時耗力，若能建立回饋機

- 制，使用者進行搜尋時同時回傳最佳關鍵字組，可促進系統之自動化。
- 三、 網路使用者常有其約定俗成之句法及用字，這個部分尚未納入資料庫，未來可考慮引入，充實資料庫，對資料查詢有所幫助。
 - 四、 使用者輸入錯誤將導致系統無法正確擷取關鍵字，目前未進行處理，未來若能運用模糊理論，計算其歸屬程度，更能增加系統實用性。
 - 五、 本研究範圍僅限於數學領域之國小部分，未來若能就其他學習領域及階段進行建置設計，對網路學習之內涵能有加深加廣作用。
 - 六、 目前僅針對是否能盡快找到較多之有用資料為主，對於資料之正確性尚無法一一核對，事實上在檢視過程中，許多資料內容其實是一樣的，只是作者及所在網址不同，因此均一一列出。資料取得容易為網路發達之優點，卻易成為錯誤傳播的溫床，正是一體兩面。因此確保資料本身的正確性，為一亟需解決的問題。

參 考 文 獻

一、中文部份

中文斷詞系統。<http://ckipsvr.iis.sinica.edu.tw/>

中央研究院資訊科學所CKIP中文詞知識庫小組。現代漢語平衡語料庫。上網日期：97年12月27日，檢自：

<http://rocling.iis.sinica.edu.tw/CKIP/20corpus.htm>

王聖中(民83)。語法式中文斷詞之研究。未出版之碩士論文，淡江大學資訊工程所。

朱怡霖(民91)。中文斷詞及專有名詞辨識之研究。未出版之碩士論文，國立臺灣大學資訊工程學研究所。

行政院主計處電子處理資料中心(民96)。95年電腦應用概況報告。上網日期：96年12月7日，檢自：

<http://www.dgbas.gov.tw/ct.asp?xItem=17966&CtNode=4920>

何永清(民94)。現代漢語語法新探。台北市：台灣商務印書館。

吳金照(民90)。應用模糊理論於火力電廠廠址評選研究。未出版之碩士論文，台北科技大學生產系統工程與管理研究所。

李坤霖(民89)。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，成功大學資訊工程系。

周立弘、張俊盛(民81)。語言模式在中文文字辨識上的應用。第五屆計算語言學研討會論文集，147-165。

林信宏(民92)。結合語意分析輸入修正之研究—以交通查詢為例。未出版之碩士論文，世新大學資訊管理學研究所。

邱閔鴻(民96)。運用模糊層級分析法於平衡計分卡權重分配之研究。未出版之碩士論文，南華大學資訊管理學研究所。

范長康(民79)。以鬆弛法做中文斷詞及其應用[摘要]。未出版之博士論文，國立交通大學資訊工程研究所。

教育部全球資訊網(93年10月21日)。教育部未來四年施政主軸架構。上網日期：96年12月7日，檢自：

http://www.edu.tw/content.aspx?site_content_sn=13600

教育部國民教育司全球資訊網(92年11月14日)。92年國民中小學課程綱要。上網日期：97年1月7日，檢自：

http://www.edu.tw/eje/content.aspx?site_content_sn=4420

教育部電子計算機中心(94年3月29日)。教育部電子報第120期---重大政策。

上網日期：96年12月7日，檢自：
<http://140.111.34.116/old/120/important.htm>

- 陳永德(民86)。中文斷詞中長詞優先、詞頻對比與前詞優先規則之使用[摘要]。未出版之博士論文，國立臺灣大學心理學研究所。
- 陳孟華(民91)。以模糊理論與關聯法則為基礎之二階中文辭彙分割法。未出版之碩士論文，淡江大學資訊工程學系。
- 陳稼興、謝佳倫、許芳誠(民89，7月)。以遺傳演算法為基礎的中文斷詞研究。*資訊管理研究*，2(2)，27-44。
- 黃居仁、陳克健、陳鳳儀、魏文真、張麗麗(民86)。「資訊處理用中文分詞規範」設計理念及規範內容。*語言文字應用*，1997年(1)，92-100。
- 黃昭勝(民95)。結構半主動模糊控制基因演算法之研究。未出版之碩士論文，國立成功大學土木工程研究所。
- 黃勝龍(民91)。智慧型客服代理人之研究—以自然語言應用在交通查詢為例。未出版之碩士論文，世新大學資訊管理學系。
- 楊宸彥(民91)。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系。
- 楊新芳、葉明利(民90)。電腦融入數學學科教學之設計。2001資訊素養與終身學習社會國際研討會。逢甲大學。上網日期：97年1月25日。檢自：
<http://dspace.lib.fcu.edu.tw/handle/2377/4277>
- 資策會電子商務研究所(民97)。How Many。上網日期：97年2月10日，檢自：
<http://www.find.org.tw/find/home.aspx?page=many&p=1>

二、西文部份

- Chakrabarti, S. (2003). *Mining the web: Discovering knowledge for hypertext data*. San Francisco, CA: Morgan Kaufmann Publishers.
- Chiang, T. H., Chang, J. S., Lin, M. U., & Su, K. Y. (1992). Statistical models for word segmentation and unknown word resolution. *Proceedings of COLING-92*, 101-107.
- Fox, M. S., & Gruninger, M. (1998). Enterprise modeling, *AI Magazine*, 19 (3), 109-121.
- Gruber, T. (2008). Ontology. In Liu, L., & Özsu, M.T. (Eds.), *Encyclopedia of Database Systems*. Springer-Verlag.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Gruber, T., SRKB mailing list. (1994), as cited in Uschold, M., & Gruninger, M.

- (1996). Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*, 11(2), 93-155.
- Kao, M., Cercone, N., & Luk, W.-S. (July, 1988). Providing quality responses with natural language interfaces: The null value problem. *IEEE Transactions on Software Engineering*, 14(7), 959-984.
- Kaufmann, A., & Gupta, M. M. (1991). *Introduction to fuzzy arithmetic: Theory and applications*. New York: Van Nostrand Reinhold.
- Kim, H. M. (2002). XMLhoo!: A prototype application for intelligent query of XML documents using domain-specific ontologies. *Proceedings of the 35th Hawaii International Conference on Systems Science*. USA.
- Kim, H. M., Fox, M. S., & Grüninger, M. (1999). An ontology for quality management: Enabling quality problem identification and tracing. *BT Technology Journal*, 17(4), 131-140.
- Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & et al. (1991). Enabling technology for knowledge sharing. *AI Magazine*, 12(3), 36-56.
- Negnevitsky, M. (2005). *Artificial intelligence: A guide to intelligent systems* (2nd ed.). New York: Addison Wesley.
- Noy, N. F. and McGuinness, D. L. (2001, March). *Ontology development 101: A guide to creating your first ontology* (Tech. Rep. KSL-01-05). Stanford University, Knowledge Systems Laboratory.
- Sporat, R., & Shih, C. (1990). A statistical method for finding word boundaries in Chinese text. *Computer Processing of Chinese and Oriental Languages*, 4(4), 336-351.
- Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*, 11(2), 93-155.
- Wielinga, B. J., & Schreiber, A. T. (1993), Reusable and shareable knowledge bases: A European perspective. *Proceedings International Conference on Building and Sharing of Very Large-Scaled Knowledge Bases'93* (pp. 103-115). Tokyo: Japan Information Processing Development Center.
- Yager, R. R. (1981). A procedure for ordering fuzzy subsets of the unit interval. *Information Sciences*, 24, 143-161.

附錄一 中研院平衡語料庫詞類標記集

簡化標記	對應的CKIP詞類標記 ⁶	
A	A	/*非謂形容詞*/
Caa	Caa	/*對等連接詞，如：和、跟*/
Cab	Cab	/*連接詞，如：等等*/
Cba	Cbab	/*連接詞，如：的話*/
Cbb	Cbaa, Cbba, Cbbb, Cbca, Cbcb	/*關聯連接詞*/
Da	Daa	/*數量副詞*/
Dfa	Dfa	/*動詞前程度副詞*/
Dfb	Dfb	/*動詞後程度副詞*/
Di	Di	/*時態標記*/
Dk	Dk	/*句副詞*/
D	Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	/*副詞*/
Na	Naa, Nab, Nac, Nad, Naea, Naeb	/*普通名詞*/
Nb	Nba, Nbc	/*專有名稱*/
Nc	Nca, Ncb, Ncc, Nce	/*地方詞*/
Ncd	Ncda, Ncdb	/*位置詞*/
Nd	Ndaa, Ndab, Ndc, Ndd	/*時間詞*/
Neu	Neu	/*數詞定詞*/
Nes	Nes	/*特指定詞*/
Nep	Nep	/*指代定詞*/
Neqa	Neqa	/*數量定詞*/
Neqb	Neqb	/*後置數量定詞*/
Nf	Nfa, Nfb, Nfc, Nfd, Nfe, Nfg, Nfh, Nfi	/*量詞*/
Ng	Ng	/*後置詞*/
Nh	Nhaa, Nhab, Nhac, Nhb, Nhc	/*代名詞*/
I	I	/*感嘆詞*/
P	P*	/*介詞*/
T	Ta, Tb, Tc, Td	/*語助詞*/
VA	VA11,12,13,VA3,VA4	/*動作不及物動詞*/
VAC	VA2	/*動作使動動詞*/
VB	VB11,12,VB2	/*動作類及物動詞*/
VC	VC2, VC31,32,33	/*動作及物動詞*/
VCL	VC1	/*動作接地方賓語動詞*/
VD	VD1, VD2	/*雙賓動詞*/
VE	VE11, VE12, VE2	/*動作句賓動詞*/
VF	VF1, VF2	/*動作謂賓動詞*/
VG	VG1, VG2	/*分類動詞*/
VH	VH11,12,13,14,15,17,VH21	/*狀態不及物動詞*/
VHC	VH16, VH22	/*狀態使動動詞*/
VI	VII,2,3	/*狀態類及物動詞*/
VJ	VJ1,2,3	/*狀態及物動詞*/
VK	VK1,2	/*狀態句賓動詞*/
VL	VL1,2,3,4	/*狀態謂賓動詞*/
V_2	V_2	/*有*/
DE	/*的, 之, 得, 地*/	
SHI	/*是*/	
FW	/*外文標記*/	

資料來源：中央研究院資訊科學所詞庫小組編列

⁶ 斜體詞類，表示在技術報告#93-05中沒有定義，即後來增列的。

附錄二 線上斷詞服務採用之精簡詞類列表

簡化詞類	精簡詞類	簡化詞類	精簡詞類
A	A	Nes	DET
Caa	C	Neu	DET
Cab	POST	Nf	M
Cba	POST	Ng	POST
Cbb	C	Nh	N
D	ADV	SHI	Vt
DE	T	T	T
Da	ADV	VA	Vi
Dfa	ADV	VAC	Vt
Dfb	ADV	VB	Vi
Di	ASP	VC	Vt
Dk	ADV	VCL	Vt
FW	FW	VD	Vt
I	T	VE	Vt
NAV	NAV	VF	Vt
Na	N	VG	Vt
Nb	N	VH	Vi
Nc	N	VHC	Vt
Ncd	N	VI	Vi
Nd	N	VJ	Vt
Nep	DET	VK	Vt
Neqa	DET	VL	Vt
Neqb	POST	V_2	Vt

資料來源：中央研究院資訊科學所詞庫小組

附錄三 本研究與 Yahoo 搜尋引擎搜尋結果比較 1

編號	實驗問句	類型	相關資料量		關鍵字
			Yahoo	本研究	
1	幾年級開始學分數	when	0	7	分數幾年級課程標準國小
2	分數是甚麼時候出現的	when	1	28	分數埃及考工記周髀算經
3	最早研究分數的國家	where	2	9	分數數學史埃及中國
4	哪裡有分數遊戲	where	1	14	分數遊戲數學網站
5	什麼是分數	what	2	5	分數平分意義分母
6	分數加減要怎麼做	how	4	7	分數加減運算步驟
7	甚麼是真分數	what	1	9	分數名詞解釋定義分子小
8	為什麼 0 不能當分母	why	1	1	分母 0 理由意義
9	通分要怎麼做	how	1	5	通分步驟分母最小公倍數
10	四邊形怎麼分	how	2	3	四邊形種類定義
11	四邊形有幾種	what	1	5	四邊形分類角度邊
12	三角形的面積怎麼算	how	9	8	三角形面積公式算法
13	三角形分成幾類	what	2	8	三角形分類角度
14	怎麼進行分數教學	how	6	5	分數教學設計方法
15	甚麼是最簡分數	what	3	11	最簡分數定義約分
16	圓周率是甚麼時候發明的	when	7	10	圓周率數學史最早
17	甚麼是圓周率	what	4	10	圓周率定義近似值
18	圓周率要怎麼算	how	10	7	圓周率方法步驟
19	誰發明了圓周率	who	9	13	圓周率數學史最早
20	想要分數的題目	where	0	0	分數數學題目試卷
21	如何求最小公倍數	how	3	1	步驟算法公式
22	甚麼是最小公倍數	what	6	5	名詞解釋定義
23	如何求近似值	how	1	3	取法步驟
24	甚麼是近似值	what	1	4	名詞解釋定義
25	甚麼是概數	what	1	5	名詞解釋定義
26	請解釋等值分數	what	1	6	名詞解釋定義
27	要怎麼算等值分數	how	0	0	步驟算法公式
28	甚麼是鉛錘直線	what	0	0	名詞解釋定義
29	誰發明阿拉伯數字	who	9	11	0 數學史最早
30	甚麼是 24 時制	what	1	0	名詞解釋定義
31	24 時制是誰規定的	who	0	2	制定

附錄四 本研究與 Yahoo 搜尋引擎搜尋結果比較 2

附表 4-1 幾年級開始學分數(when)

Yahoo 搜尋引擎						
關鍵字數	關鍵字	符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相關 資料量	
0	無(整句查詢)	0	0.000	16,172	0	
1	分數	0	0.000	34,286	0	
2	幾年級 分數	2	0.067	665,000	10	
智慧型搜尋引擎						
關鍵字數	領域 關鍵字	擴展關鍵 字	符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相關 資料量
1	分數	無	0	0.000	40,469,999	5
2	分數	幾年級	2	0.067	665,000	10
3	分數	課程標準	4	0.133	147,300	16
4	分數	國小	7	0.233	46,900	24

附表 4-2 分數是甚麼時候出現的(when)

Yahoo 搜尋引擎						
關鍵字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
0	無		0	0.000	1,061,350	1
1	分數		0	0.000	40,469,999	5
2	分數 出現		3	0.100	5,036,000	6
3	分數 甚麼時候 出現		0	0.000	1,061,350	0

智慧型搜尋引擎						
關鍵字數	領域 關鍵字	擴展關鍵字	符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
1	分數		1	0.033	40,469,999	5
2	分數	埃及	6	0.200	188,599	28
3	分數	考工記	17	0.567	73	21
4	分數	周髀算經	28	0.933	48	28

附表 4-3 最早研究分數的國家(where)

Yahoo 搜尋引擎						
關鍵字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相關 資料量
0	無		3	0.100	117,500	4
1	分數		2	0.067	40,469,999	5
2	分數 最早		12	0.400	444,999	16
3	分數 最早 國家		7	0.233	185,000	9
智慧型搜尋引擎						
關鍵字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相關 資料量
1	分數		2	0.067	40,469,999	5
2	分數	數學史	14	0.467	11,860	30
3	分數	埃及	14	0.467	881	30
4	分數	中國	9	0.300	327	30

附表 4- 4 哪裡有分數遊戲(when)

Yahoo 搜尋引擎						
關鍵 字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
0	無		1	0.033	573,100	1
1	分數		1	0.033	40,469,999	13
2	分數遊戲		3	0.100	14,428,000	4
智慧型搜尋引擎						
關鍵 字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
1	分數		1	0.033	40,469,999	13
2	分數	遊戲	3	0.100	14,428,000	4
3	分數	數學	12	0.400	725,000	30
4	分數	網站	14	0.467	146,400	29

附表 4- 5 什麼是分數(what)

Yahoo 搜尋引擎						
關鍵 字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
0	無		1	0.033	30,690	18
1	分數		2	0.067	39,209,999	13
智慧型搜尋引擎						
關鍵 字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
1	分數		2	0.067	39,209,999	13
2	分數	意義	3	0.100	2,000,000	19
3	分數	平分	8	0.267	13,890	29
4	分數	分母	4	0.133	1624	30
2	分數	定義	2	0.067	890,000	10
3	分數	平分	3	0.100	61,060	23
4	分數	分母	3	0.100	2,191	30
2	分數	平分	4	0.133	60,680	22
3	分數	意義	8	0.267	13,980	30
4	分數	分母	5	0.167	1644	30

附表 4-6 分數加減要怎麼做(how)

Yahoo 搜尋引擎						
關鍵字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
0	無		1	0.033	30,690	18
1	分數		2	0.067	39,209,999	13
2	分數 加減		4	0.133	212,400	22
智慧型搜尋引擎						
關鍵字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果資 料量	數學領域相 關資料量
1	分數		1	0.033	39,209,999	13
2	分數	加減	4	0.133	212,400	22
3	分數	運算	6	0.200	36,150	28
4	分數	步驟	7	0.233	5,735	29

附表 4-7 什麼是真分數(what)

Yahoo 搜尋引擎						
關鍵字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
0	無		4	0.033	2,849	19
1	真分數		1	0.067	20,190	19
智慧型搜尋引擎						
關鍵字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
1	真分數		1	0.033	20,190	19
2	真分數	意義	1	0.033	7,113	30
1	真分數		1	0.033	20,190	19
2	真分數	定義	4	0.133	1078.000	26
3	真分數	分子小	5	0.167	394.000	28
4	真分數	名詞解釋	9	0.300	27.000	23
1	真分數			0.033	20,190	19
2	真分數	名詞解釋	5	0.167	460	24
3	真分數	定義	7	0.233	116	26
4	真分數	分子小	9	0.300	27	23

附表 4- 8 為什麼 0 不能當分母(why)

Yahoo 搜尋引擎						
關鍵 字數	關鍵字		符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
0	無		4	0.154	26	13
1	分母		1	0.033	2,900,000	28

智慧型搜尋引擎						
關鍵 字數	領域 關鍵字	擴展 關鍵字	符合結果 資料量	準確率 (%)	搜尋結果 資料量	數學領域相 關資料量
1	分母		1	0.033	2,900,000	28
2		0	3	0.100	813.000	8
3		理由	1	0.033	1010.000	5
4		意義	1	0.033	62.000	4