知識全文檢索架構的探討－中文歌謠
Frame Of Knowledge Retrieve — Chinese Ballad

洪朝富、陳慧津、邱淑琴、陳怡芬　　吳翠華*
真理大學資訊管理系　　東京大學中文系*
S1787111@email.au.edu.tw

摘要

隨著資訊科技的急速發展，以及寬頻網路的逐漸實現，促使大家使用電腦上網檢索的機會愈來愈多。因此我們希望建構國內外第一個歌謠相關的研究網站，提供研究者上網檢索歌謠，透過我們的知識分類模式歸類，解決過去受到藏書空間與圖書保存及無法充份供給社會大眾利用等問題。

由於今日中文檢索系統大多採用缺乏知識分析模式的自動斷詞與利用關鍵詞的比對模式，所以本研究嘗試將朱介凡歌謠分類模式配合認知科學的『 frame + 基模 』建構歌謠的知識結構，幫助我們做知識推理分析。期望此一查詢模式能幫助學者專家達到初步的分析，同時因本系統仍然保有中文檢索的基本型態，也能供一般人士作歌謠的查詢與簡介。

關鍵詞：斷詞、框架、基模。

ABSTRACT

Because of the information technology developed at high speed and WWW's general popularization, the opportunity of everybody retrieve in the web more and more. We hope establish the first retrieval website related to the ballad in domestic and foreign to provide scholar to retrieve. By our knowledge classify model to sort out to settle problem which must have space to collection books and books preserved can't provide everyone to use fully in the past.

Today's Chinese retrieved system almost adopt automatic word segmentation and the keyword retrieval model which lack of knowledge analyze model , so the research try establish knowledge structure that combine the jie-fan zhu's ballad classify model and the "frame + script" of cognitive science. To help us do knowledge inference analyze. We hope this retrieve model can do expert first analyze service, in the meanwhile this system still have basic Chinese retrieve model that can provide people in general retrieve and brief introduction of ballad.

Key Words： Word Segmentation、 Frame、 Schema.

一、緒論

資訊科技的發達，促使大家上網檢索的機會愈來愈多，傳統的檢索技術利用關鍵字比對，幫助使用者從大量缺乏結構化資料中檢索出所需的文件，但往往檢索出的文件量太多，還需使用者逐一過濾，無法切確符合使用者所需。且此種傳統檢索技術只能檢索出符合查詢條件的文件，無法歸納及推衍隱含在大量文件中的知識，所以我們希望針對此種問題做深入探討。

期望藉由本系統的開發，將非結構化的歌謠予以結構化，使關鍵詞檢索模式提升為概念檢索模式。藉由歌謠依人、事、時、地、物斷詞，可得到各篇歌謠的關鍵詞彙，表現出歌謠的內含，進而可歸納及推衍隱含在歌謠中的知識供後續探勘使用。

在全文檢索中，關鍵詞搜尋無法切確使用者所需，本系統透過檢出率和精確率印證系統效能的確優於關鍵詞檢索模式。

本系統的設計主要針對專家學者，專家學者在做一項研究時，除了要花時間搜集資料外，還得花精力將資料吸收為知識，所以希望藉由本系統讓專家學者易於找資料，將節省出的時間、精力再作其他用途。此外，專家學者有相當的歌謠文學背景，易於印證出系統可行性。本系統也供一般檢索人士查詢使用。

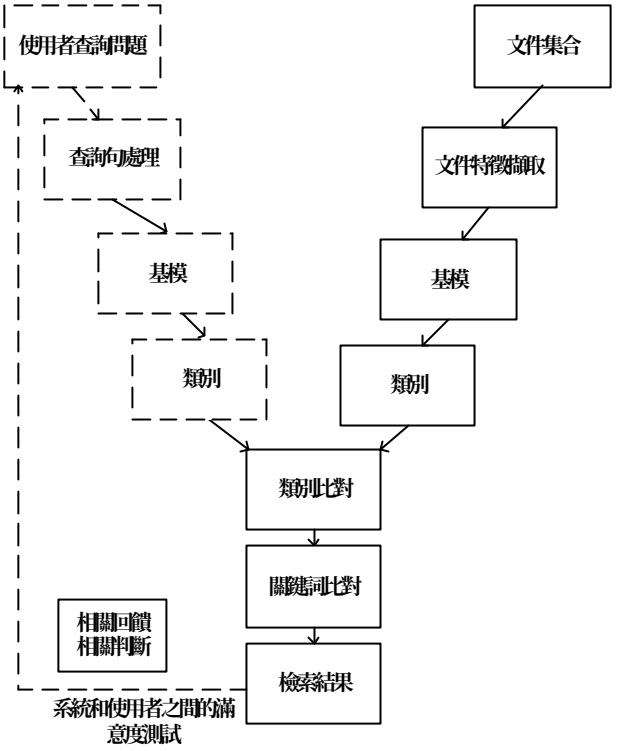
二、研究方法

2.1 全文資訊檢索

從資訊檢索概念（如圖一）可簡單區分為四個主要的研究領域，包括查詢介面、檢索引擎、自動索引與相關評量。本文研究範疇界定於自動索引理論（圖中實線部分）。

所謂索引就是在於分析文件內容、決定文件特徵，並且將文件以特徵形式代表的整個過程。索引的目的是希望系統能夠提供使用者查詢到正確相關的文件。因此，自動索引乃研究各種索引方法，建立良好的索引形式，將文件相關的內容以有效的索引形式表達於內部，以提升資訊檢索的檢出率與精確率。[黃雲龍，1998]

概念是認知後抽象信息的表達，我們所使用的語言、文字，皆為概念的一種表達形式。人類透由思考、認知、概念到語言文字構成資訊檢索形式與內容關係的關鍵變數[黃雲龍，1998]。使用者藉由本身認知抽取其關鍵詞檢索所需資訊，並構成使用者本身的知識結構，而系統方面，根據文件的特徵擷取其屬性、基模、形成類別，架成一個知識體系，將兩者間的知識結構做比對，找出使用者的目的為何，確定類別後再依關鍵詞檢索，即可有效找出使用者要的資訊。

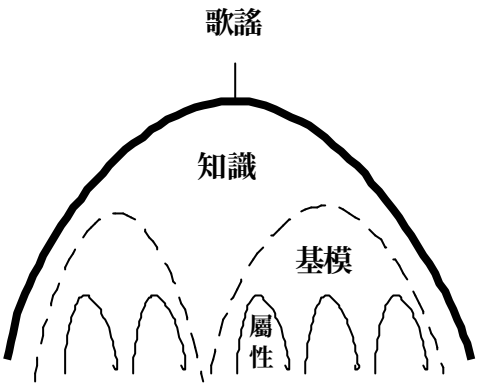


圖一：全文資訊檢索概念圖

2.2 知識表徵

本研究已將傳統的關鍵詞檢索提升為概念式檢索，因傳統檢索系統並沒有將資料做分類，也就是沒有本系統的知識架構部分，而是依關鍵詞比對後的資訊即傳給使用者，但往往檢索出的文件量仍太大。本系統的不同在於先將歌謠建立知識架構，透過使用者鍵入的關鍵詞預測使用者欲查詢的類別，再針對此類別做檢索，已將傳統關鍵字檢索提昇為概念式檢索，相信這樣更可檢索出使用者要的資料。

然而，無限狀態的研究難以達到有效之概念搜尋，所以必需清楚的一大重點是本研究已將知識侷限於有限狀態下（朱介凡歌謠分類模式），在這有限狀態之下所做的系統結合才不致於缺乏定性，且在這有限狀態下的固定模組，經由策略之間的選擇所決定出的知識樹也才不失其彈性。



圖二：知識表徵圖

2.3 知識模式的分類

中文資訊大多存在著閱讀時間、記憶和文字理解等問題較難掌握快速掌握材料，所以要花費相當的時間去吸收和轉換這些材料後，才有可能做出更深入的分析並且建構出更完整的體系。如果要有相當顯著的成果，就要花費積年累月的工夫，這種因為資訊中文化所存在的種種問題，使得我們想將中文資訊和斷詞整理在一起，同時也藉由中文資訊和斷詞的合併觀念，引發了我們想對歌謠領域做深入的知識分析探討與研究。

綜觀以上問題，以我們所探討的歌謠為例做出下列說明：

1. 斷詞：利用人類的觀感斷詞，製造詞庫，然而為使詞庫正確表現，有和專家學者做互動的必要。
2. 歌謠的分類：首先利用歌謠的知識分類，將歌謠依朱介凡的分類(如：童話世界兒歌、兒童遊戲歌...等)分式分類，在每大類之下再分出其細類，細類底下可能還有細類，然後在最細類根據人、事、時、地、物，定義出其特徵詞。
3. 記錄相關資訊：定義歌謠的基本資訊，如作者、年代、地點、情境(特殊人、事、物)，並予以記錄。
4. 統計、綜合分析：統計出一首歌謠中出現的詞與次數，然後比對各分類細項中常用字詞的數量，依照詞出現的量研判出歸屬的分類，再做特徵比對完成情境的描述。

2.4 知識架構

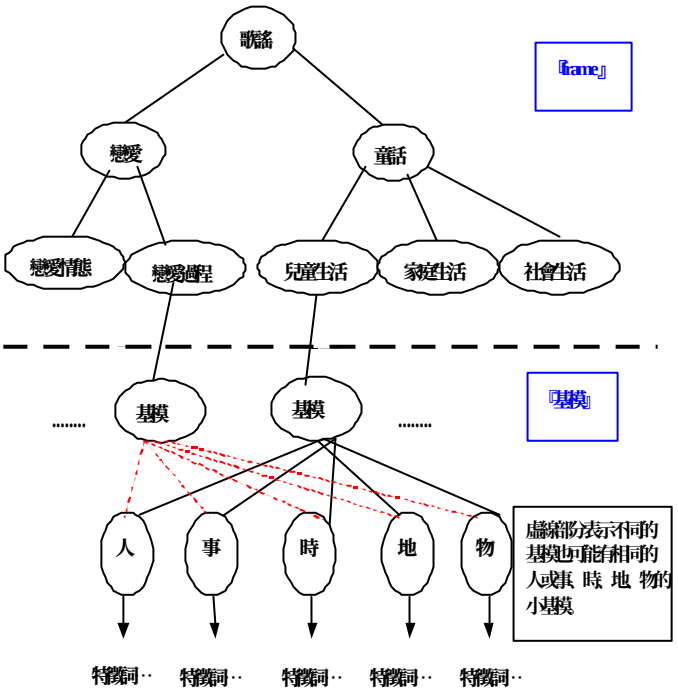
歌謠是由詞彙組成，由於中文是個抽象的概念，所以我們希望用科學統計方法來分析歌謠，藉由統計出的數據看出歌謠的特徵，再導入「frame+基模」的模式，建立一套標準解釋模式。從搜尋面來看，傳統搜尋系統以統計方式利用關鍵詞比對找出相關的文章，但截取出的文章量往往太大，還需使用者逐一的過濾；由於中文的不結構化，所以我們改以「frame+基模」來建其架構，使其 frame 與基模間的關係定義清楚，再透過我們的篩選方式，使檢索出的量大為精簡，真正符合使用者所需。

利用認知科學的角度來看我們將此研究用『frame』+『基模』來做推理解釋。

1.frame：適用於描述固定的物體、事件和動作。包含了它所描述的情況或物體的多方面訊息，也包含了物體必須具有的屬性，其所描述的是它們所代表的概念性典型特徵情況。

2.基模：是敘述性知識的整合單位。所以用到「基模」這個詞的共通元素指的便是有組織的知識架構。

本研究把 frame 當是我們類別表現方示，直接採用朱介凡先生的分類模式，每個類別下的歌謠再運用基模的特性、結構將資料儲存，再依人、事、時、地、物五個事件做詞的分類，我們的做法是將網站的歌謠首先將歌謠鍵入資料庫中，經過下列方式將歌謠依我們提出的知識分析模式作歸類：



圖三：知識架構圖

傳統檢索系統僅是針對「詞的統計頻率」，並沒有將資料做分類，也就是沒有本系統的知識架構部份。只是單純將整個資料庫依關鍵詞比對，找到的資料就傳回給使用者，但往往檢索出的文件量仍太大，本系統的不同在於先將歌謠做分類整理，透過使用者鍵入的關

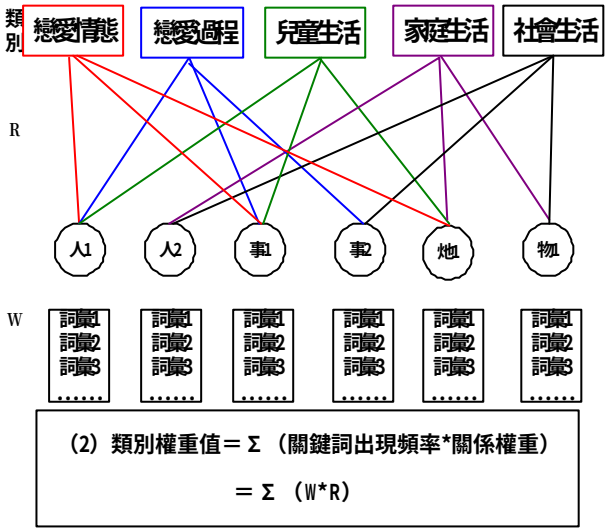
鍵詞預測使用者欲查詢的類別，再針對此類別做檢索，相信這樣更可檢索出使用者要的資料。

2.5 知識推論原理

基於索引的目的，如何選擇合適的索引詞彙來代表文件內容？直覺的想法是該詞彙有沒有出現在文件內，所以須計算出詞彙頻率，如下列（A）式，f_i^k 是索引詞 k 在第 I 個文件的出現頻率，F^k 是索引詞 k 在文件集合內的總次數。

$$(1)\quad F^k = \sum_{i=1}^{n} f_i^k$$

關鍵詞和類別之間的關係，藉由圖四得知，先統計所有詞彙在資料庫出現的次數，以 W 表示詞出現的頻率，小基模和類別間的關係權重以 R 表示，我們已將每個類別下的關係權重做了正規化，所以每個類別下的權重值加總等於 1。



（2）類別權重值＝Σ（關鍵詞出現頻率*關係權重）

＝Σ（W*R）

圖四：知識推論圖

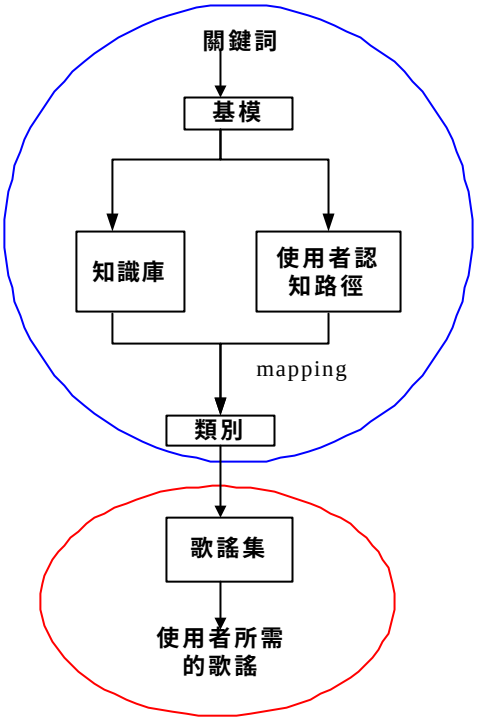
2.6 檢索方法

為使本系統達成有效之相關性全文檢索，所以本研究將使用者輸入之關鍵字檢索做了兩次的篩選動作，我們的檢索方法可分為二個階段：

階段一：1.利用關鍵詞做權重的比對找出其所屬的幾個基模。

2.藉由基模的排列組合做 rule 的決定，選取出所屬類別的知識樹，經過這次的篩選動作，資訊量可大量縮減，並將範籌定義於此一知識樹區塊中，此時知識樹的呈現為一？述性表徵。

階段二：1.利用階段一所選取的類別區塊和輸入的關鍵詞做比對，經由這個二次篩選的動作，又可將資訊量二次縮減，找出使用者所需之真正相關性資訊，此時的歌謠知識樹所做的是程序性表徵之呈現。



圖五：檢索方法圖

2.7 評估準則

參考過去西方的研究報告可知檢出率及精確率是主要的評量指標。但是傳統上以這兩個指標各為縱軸與橫軸，展現系統或模型在兩個指標上的消長關係。由實驗得知：檢出率隨著群集索引構面增加而降低。精確率隨者群集索引構面增加而上升。所以須同時兼顧檢出率與精確率的成對關係，把不同的群集索引構面下檢出率與精確率的特性能夠呈現。 [黃雲龍，1998]

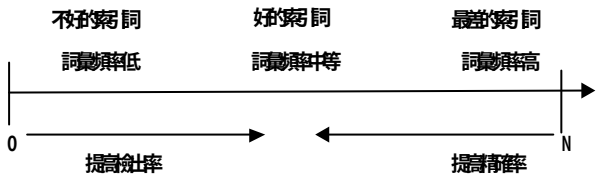
故本系統評估程序的步驟有：系統自動擷取、系統正確擷取、資料庫正確資料量，分別說明如下：

1. 資料庫正確擷取：資料庫中正確資訊量（A）
2. 系統自動擷取：系統擷取到之總歌謠篇數（B）。
3. 系統正確擷取：專家判定系統擷取到之總歌謠篇數中正確的歌謠篇數（C）。
4. 計算檢出率、精確率：利用系統擷取檔，

比對其正確的篇數，並計算其檢出率、精確率：

（3）檢出率 = C / A
（4）精確率 = C / B

如何選出合適的關鍵詞來代表文件內容，其中大有學問，因如果某個詞彙出現次數很高，它也許是一個很重要的識別因子。但是如果該詞彙同時出現在很多文件內，相對的其價值可能就不如出現在較少文件內的詞彙。所以詞彙頻率的高低也攸關關鍵詞取決與否，以圖六[黃雲龍，1998]來看更為清楚。



圖六：索引詞特性與文件頻率關係圖

可推出下列幾點：

1. 詞彙頻率低，運算過程中可能在計算權重、選取類別時會將使用者目的類別刪去，系統顯示出找不到使用者要的文件，所以這時須提高系統的檢出率，本研究的解決方法為使用同義詞庫，讓類別計算權重時把同義的詞彙一併計算進去，類別確定後再用關鍵詞搜尋，即可解決這問題。
2. 詞彙頻率中等則為好的關鍵詞，表此詞彙集中度高、廣度也高，可使文件分離，出現在特定類別中，彼此分離的情況將有利於檢索。
3. 詞彙頻率高，表此詞彙集中度低，不是出現在特定類別中，有無此詞彙對文件糾纏或分離狀況必無影響，無法正確推出使用者的意圖，檢索出給使用者的資訊可能過多，無法達到精簡。
4. 檢出率與精確率存在消長關係，檢出率上升，精確率則下降。若且則若，這中間兩者清長數字是矛盾的，而其最佳狀態是兩組數字都處於最大值。所以使用者若想搜尋出更大範圍資訊量時，須將門檻值降低，則選取出的類別相對多，找出的資訊量也大，檢出率提高，如此將導致精確率跟著下降；反之，使用者想得到較佳精確率，則將門檻值提高。

三、系統運作

3.1 人工斷詞

採用朱介凡先生分類中的「童話」和「戀愛」二類較具代表性的歌謠做為歌謠來源，將這兩類中之各篇歌謠採人工方式斷詞，並鍵入資料庫中，茲以下篇歌謠為例說明人工斷詞的步驟：

「星星，月亮，叭叭狗兒上炕，不想爹，不想娘，單想給媳婦兒抓癢癢。」

1.斷詞

首先斷詞得到：星星、月亮，叭叭狗兒、上炕、不想 2、爹、娘、單想、給、媳婦兒、抓癢癢。

2.若是有不必要的詞，則去除之，以得到有意義的詞：本篇中並無不必要詞，故不去除任何詞彙。

3.依「人、事、時、地、物」將所斷出來的詞分類

人：爹、娘、媳婦兒。
事：不想 2、單想、抓癢癢、給、上炕。
時：
地：
物：星星、月亮、叭叭狗兒。

3.2 同義詞庫

由於不同地區、年代，同樣一個意思可能用法有所不同，所以我們建立一同義詞庫，先作詞的同義轉換再去找歌謠，一開始先把使用者要找的歌謠範圍擴大，有相同意思的詞所屬類別都考慮進來，免除使用者為了找符合歌謠的關鍵詞還得一一費神的猜關鍵詞正確用法。

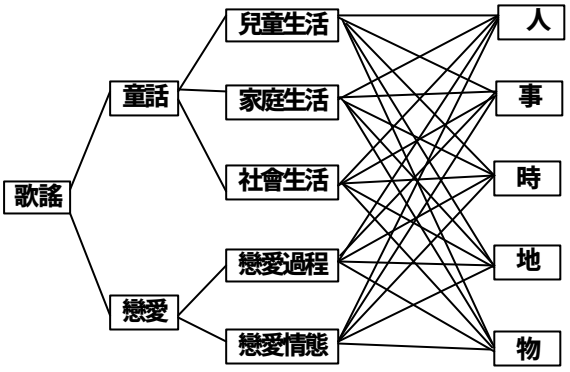
例如：「爸爸」、「爹」、「親爺」分別出自於不同篇的歌謠，但意思都相同，都是「爸爸」的意思，故我們會將這三個詞列為同義詞，建入同義詞庫中。倘若使用者欲找尋有關爸爸的歌謠，本系統會自同義詞庫中檢索出相同解釋的詞彙，做同義詞的轉換，找出這三個詞的歌謠做篩選。

3.3 分類模式

歌謠主要分為兩大類，分別是「戀愛類」和「童話類」，戀愛類別之下再細分為「戀愛生活的過程」、「戀愛生活的情態」，童話類別之下再細分為「兒童生活」、「家庭生活」、「社會生活」。

將各個子類別之下的歌謠再依人工方式加以斷詞，所斷出的詞彙以基模的方式依「人、事、時、地、物」的方式展開。由於歌謠的篇章頗多，詞數也相對的多，為了看出各類別的特徵，我們又將人、事、時、地、物做了更細部的分類，以利系統做群集性分析。

系統架構如下：



人：男主角、男生、女主角、女生、年幼、代名詞、其它

事：口、感官、手、足、身、心、行為、動物、自然、東西

時：早、晚、日子、時間經過

地：地名、特定地、方向、家

物：人、動物、植物、自然、用品、食物、人行為、其它

圖七：歌謠分類模式

3.4 **關係權重**

我們將每個詞出現的頻率一一做統計，再將每個類別內基模的詞數做一合計，結果如下：

表一：基模出現於子類的次數表

大類		童話			戀愛	
Fj \ Fi		兒童生活 F1	家庭生活 F2	社會生活 F3	戀愛過程 F4	戀愛情態 F5
人	男主角	1	2	0	15	23
	男生	0	14	10	25	4
	女主角		14	3	20	23
	女生	2	19	5	1	5
	年幼	11	4	4	1	0
	代名詞	16	22	4	28	26
	其它	2	1	7	2	7
事	口	7	14	24	13	25
	感官	1	7	4	5	14
	手	19	18	16	35	42
	足	17	14	7	21	23
	身	10	7	4	6	11
	心	10	6	3	50	33
	行為	20	33	22	29	43
	動物	0	2	1	3	2
	自然	1	6	3	7	2
	東西	5	5	7	6	3
時	早	0	1	2	1	5
	晚	0	0	4	3	3
	日子	0	1	2	2	3
	經過	0	0	0	6	0
地	地名	10	2	5	4	4
	特定地	6	5	13	24	15
	方向	4	3	4	1	3
	家	4	9	4	3	9
物	人	4	4	6	23	32
	動物	9	11	7	3	11
	植物	3	9	3	13	9
	自然	4	5	4	15	19
	用品	19	18	27	14	27
	食物	13	15	20	14	23
	人行為	3	1	0	15	13
	其它	6	2	3	3	4
合計		Σ F1j =212	Σ F2j =274	Σ F3j =228	Σ F4j =411	Σ F5j =466

將同一類別的基模做加總，並做正規化，使得合計為 1，權重運算式正規化過程如下：

（3） $W1j = F1j / \Sigma F1j$

（4） $W2j = F2j / \Sigma F2j$

（5） $W3j = F3j / \Sigma F3j$

（6） $W4j = F4j / \Sigma F4j$

（7） $W5j = F5j / \Sigma F5j$

表二：基模出現於子類的權重表

大類		童話			戀愛	
Fj \ Fi		兒童生活 F1	家庭生活 F2	社會生活 F3	戀愛過程 F4	戀愛情態 F5
人	男主角	0.00472	0.00730	0	0.03650	0.04936
	男生	0	0.05109	0.04386	0.06083	0.00858
	女主角	0.02358	0.05109	0.01316	0.04866	0.04936
	女生	0.00943	0.06934	0.02193	0.00243	0.01073
	年幼	0.05189	0.01460	0.01754	0.00243	0
	代名詞	0.07547	0.08029	0.01754	0.06813	0.05579
	其它	0.00943	0.00365	0.03070	0.00487	0.01502
事	口	0.03302	0.05109	0.10526	0.03163	0.05365
	感官	0.00472	0.02555	0.01754	0.01217	0.03004
	手	0.08962	0.06569	0.07018	0.08516	0.09013
	足	0.08019	0.05109	0.03070	0.05109	0.04936
	身	0.04717	0.02555	0.01754	0.01460	0.02361
	心	0.04717	0.02190	0.01316	0.12165	0.07082
	行為	0.09434	0.12044	0.09649	0.07056	0.09227
	動物	0	0.00730	0.00439	0.00730	0.00429
	自然	0.00472	0.02190	0.01316	0.01703	0.00429
	東西	0.02358	0.01825	0.03070	0.01460	0.00644
時	早	0	0.00365	0.00877	0.00243	0.01073
	晚	0	0	0.01754	0.00730	0.00644
	日子	0	0.00365	0.00877	0.00487	0.00644
	經過	0	0	0	0.01460	0
地	地名	0.04717	0.00730	0.02193	0.00973	0.00858
	特定地	0.02830	0.01825	0.05702	0.05839	0.03219
	方向	0.01887	0.01095	0.01754	0.00243	0.00644
	家	0.01887	0.03285	0.01754	0.00730	0.01931
物	人	0.01887	0.01460	0.02632	0.05596	0.06867
	動物	0.04245	0.04015	0.03070	0.00730	0.02361
	植物	0.01415	0.03285	0.01316	0.03163	0.01931
	自然	0.01887	0.01825	0.01754	0.03650	0.04077
	用品	0.08962	0.06569	0.11842	0.03406	0.05794
	食物	0.06132	0.05474	0.08772	0.03406	0.04936
	人行為	0.01415	0.00365	0.00000	0.03650	0.02790
	其它	0.02830	0.00730	0.01316	0.00730	0.00858

合計	Σ W1 j =1	Σ W2j =1	Σ W3j =1	Σ W4j =1	Σ W5j =1

四、系統評估

針對研究方法提出的架構、理論開發系統，再由評估準則做系統評估，分別輸入關鍵詞-娘（表三），關鍵詞-爹（表四）和二個關鍵詞-爹、娘（表五）三組做對照，實驗結果如下：

表三：查詢句相關評量表（娘）

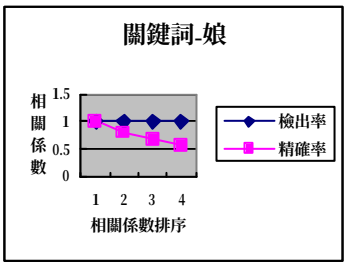
相關係數排序	相關係數	相關與否	檢出率	精確率
1	0.851139	相關	100%（4/4）	100%（4/4）
2	0.073389		100%（4/4）	80%（4/5）
3	0.043903		**100%（4/4）**	**67%（4/6）**
4	0.031569		100%（4/4）	57%（4/7）

表四：查詢句相關評量表（爹）

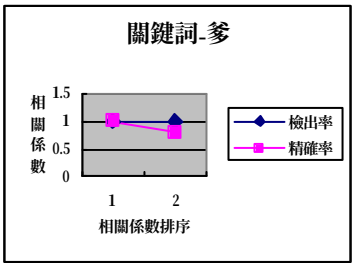
相關係數排序	相關係數	相關與否	檢出率	精確率
1	0.732616	相關	100%（4/4）	100%（4/4）
2	0.20753		100%（4/4）	80%（4/5）

表五：查詢句相關評量表（爹、娘）

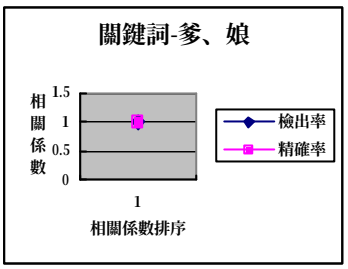
相關係數排序	相關係數	相關與否	檢出率	精確率
1	0.807892	相關	100%（4/4）	100%（4/4）



圖八：查詢句相關評量圖（娘）



圖九：查詢句相關評量圖（爹）



圖十：查詢句相關評量圖（爹、娘）

傳統關鍵詞檢索，檢索出資料量太大導致精確率無法提升，我們希望藉由本系統來改善傳統檢索的不足。由圖八看出：傳統關鍵詞檢索輸出給使用者七篇歌謠，其中四篇才是使用者所需，精確率為 57%；而本系統的比對模式做了兩次篩選，篩類別時，考慮人的不一致性，選出前三名權重較大的類別後再依關鍵詞檢索，使得精確率達到 67%，而關鍵詞檢索因未做類別的篩選，資訊連同雜訊一同呈現給使用者，精確率卻僅止 57%。由這兩個系統對照可看出本系統的精確率有明顯提升，相信若資料量更大時使用本系統效果更為顯著。

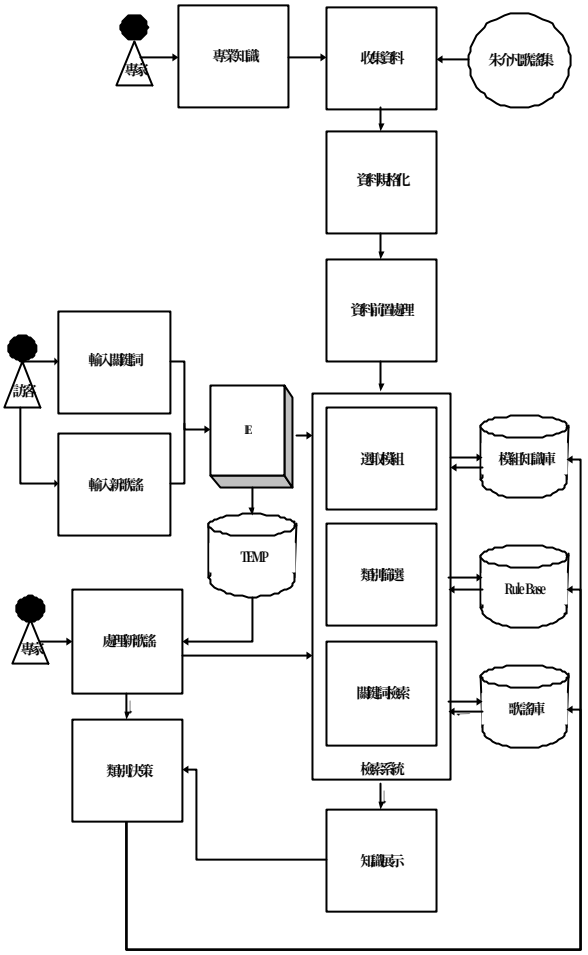
由圖八、圖九對照：圖九精確率、檢出率的差距小於圖八，表示關鍵詞-爹在其類別內的關係強度大於關鍵詞-娘在其類別內的關係強度。

由圖八、圖十對照：二個關鍵詞的精確率和檢出率的差距顯然地較一個關鍵詞來的小，也印證著當使用者輸入較多關鍵詞時，未知資料減少，目的較明確，系統可正確鎖定類別，檢索出的資訊較為精簡，精確率自然提高。

由此實驗可看出，由認知科學的「frame+基模」方式為架構的檢索系統優於傳統檢索系統，更能查詢出符合使用者所需的資訊。

五、系統架構

由於近年來網際網路的盛行，如果只墨守在自己的資訊系統中，不與外部的資訊連結，會如鴕鳥躲避敵人的行為相同，只將自己的眼睛躲起來，最後為敵人所吞食，因此我們將在網際網路上建構此一歌謠發展系統，期望藉由使用者與系統之間的互動，使本系統更為完善。



圖十一：系統架構圖

六、結論及未來研究方向

6.1 結論

本研究之歌謠檢索系統不僅僅有管理和搜尋的功能，其對於推廣、發揚，甚至教學或研究，都存在著極大的幫助，因為透過此全文檢索系統，已把歌謠做有規劃的整理，如此將可輕易的找到所需的資訊，不但省時、省力，並且因資訊的完整性，使得找到的是真正較符合使用者需求的資訊。

本研究針對歌謠的存取及呈現方式，可歸納成以下的結論：

1. 斷詞方面採用混合式斷詞法，利用「frame+基模」的方式來做搜尋可以結合詞庫式斷詞法及統計式斷詞法的優點，加上以長詞優先的規則可以獲得較大意義的詞彙。
2. 利用人、事、時、地、物的進一步分析，將可更清楚明瞭地將歌謠的內容做更細部的分析。
3. 有別於傳統檢索系統僅做簡單關鍵詞比對，本研究所建構出的歌謠知識體系，由關鍵詞✍基模✍類別✍歌謠選取，做了二階段的篩選過程，透過此系統更能找出符

合使用者所需之資訊。

6.2 未來研究

本系統藉由已經分類好的歌謠，透過斷詞加上人事時地物對用詞做分類，統計出各類別中人事時地物所有的詞，完成我們的知識分類模式。但目前為止，本系統所做到的僅止於查詢部份而已。

我們期望未來可以完成新增歌謠的功能，當有新的文章讀入時，即可利用斷詞與同義字的合併後，再和統計好的詞庫做比對，計算各分類的權重，評估得分最高的類為這篇新文章的歸屬類別。藉由歌謠全文檢索系統架設於網際網路上，使得系統能不斷地擴充，進而提供全球的專家學者使用，增進中文歌謠的交流。

在未來，期望能對專家做滿意度測試，經由專家給予的回饋進行權重的調整，使系統的權重運算更為精準。另外，還能便利專家作資料的排比分析，以具體的數據說明歌謠的種種特徵。

七、參考文獻

[1] 王三慶，“漢籍數位典藏之建置、利用與分析---以《全宋詞》為例”，中央研究院第三屆國際漢學會議論文，2000。

[2] 王三慶，“從文化傳統看明月清風與大江東去”，2000。

[3] 朱介凡，“中國兒歌”，純文學叢書，1978。

[4] 侯永昌，楊雪花，“以模糊理論和遺傳演算法為基礎的中文文件自動分類之研究”，模糊系統學刊，第四卷，第一期，p45-57，1998。

[5] 施良方“學習理論”，麗文教育叢書，p171-255，1996。

[6] 許勝傑，許中川，“整合中文處理、資料檢索及資訊擷取技術於文件資料探勘”，1999 科際整合管理研討會，p119-132，1999。

[7] 許勝傑，許中川，“中文新聞文件斷詞”，第十屆國際資訊管理學術研討會論文集，p968-9，1999

[8] 陳克建，陳正佳，林隆基，”中文語句的研究－斷詞與構詞”，技術報告，TR-86-006，中央研究院，南港，1986。

[9] 黃雲龍，“中文全文文件群集索引理論研究與實證“，圖書與資訊學刊，p44-68，1998。

[10] 黃燕萍、許中川，"中文社會新聞文件資訊擷取"，第十屆國際資訊管理學術研討會論文集，p975-982，1999。

[11] 黃譯瑩，“課程統整之意義探究與模式建構”，國家科學委員會研究彙刊：人文及社會科學，第八卷，第四期，p616-633，1998。

[12] Bob McKercher, “A chaos approach to tourism”, Tourism Management 20, p425-434,1999.

[13] Chen, A., J. He, L. Xu, F. C. Gey, and J. Meggs, “Chinese Text Retrieval Without Using a Dictionary”, Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval , p42 – 49,1997.

[14] Chien, K. –J. and S. –H. Liu, “Word Identification for Mandarin Chinese Sentences”, Proceeding of CoLING-92, 14th International Conference on Computational Linguistics, p101-107,1992.

[15] Feldman, R. and I. Dagan, “Knowledge Discovery in Textual Databases (KDT) ”, Proceedings of First International Conference on Knowledge Discovery & Data Mining, p112-117,1995.

[16] Ming-Syan Chen, Jong Soo Park and Philip S. Yu, “Data Mining for Path Traversal Patterns in a web Environment”, proceedings of the 16th ICDCS, p385-392,1996.

[17] Nic, J., M. Briscbois and X. Ren, “On Chinese Text Retrieval”, Conf. Proc. of SIGIR, p225-233,1996.

[18] Peca, K., ”Chaos theory: A scientific basic for alternative research methods in educational administration.”, paper presented at the Annual Meeting of the American Educational Research Association , p20-24,1992.

[19] Yeh, C. L. and H. J. Lee, “Rule-Based Word Identification for Mandarin Chinese Sentences – A Unification Approach”, Computer Processing of Chinese and Oriental Languages, Vol. 5, No. 2, p97-11,1991