

行政院國家科學委員會專題研究計畫結案報告

計畫名稱：故宮文物之美系列 II - 數位博物館關鍵技

術研製 - 子計畫(五)

執行期間：90年7月1日 ~ 91年6月30日

計畫編號：NSC90-2750-H-002-035

計畫主持人：歐陽彥正 教授

共同主持人：陳信希 教授

執行單位：國立台灣大學資訊工程學系

壹、計畫概述

一、計畫目標

在資訊化的社會中，數位博物館扮演著資訊及知識提供者的角色。就資訊及知識提供者這個角色而言，首要的目標便是能提供使用者一個高效率的檢索工具，以迅速而準確的協助使用者找到有興趣的資訊及知識。就這個議題而言，現有的數位博物館大多提供全文檢索及 metadata 欄位檢索。這也是本子計畫在第一年執行期間主要的成果。然而嚴格來講，全文檢索及 metadata 檢索均尚未達到知識處理的層次。同時以目前針對故宮單一數位博物館所設計的中文檢索軟體亦不能符合未來故宮與國際上重量級博物館進行交流整合以及提供國際人士使用的需求。因此，在本年度的計畫中，我們規劃了以下兩個主題：

(1)將知識庫系統引進數位博物館之建構中

(2)整合跨語言檢索功能於多數位博物館中。

綜觀目前國內外著名之數位博物館均尚未引進知識庫系統，同時亦均未提供使用者跨語言整合性檢索。因此本計畫的成果將使故宮數位博物館成為國際上的標竿，同時亦為未來與國際重量級博物館進行交流整合奠定基礎。

二、計畫內容

本子計畫的兩個主要研究重點為：

(1)將知識庫系統引進數位博物館的建構中

(2)完成跨語言整合性檢索功能。

將知識庫系統引進數位博物館的建構中將產生幾項主要影響：

(1)使用者可以近乎自然語言的方式查詢所需的資訊及知識；

(2)系統可透過自動推論的機制發覺隱藏性的資訊及知識；

(3)由於自動推論機制的引進，數位博物館的資訊及知識內容管理對系統維護者而言將更為方便及有效率，同時也大幅將低人為錯誤的機率。

由於數位博物館在資訊化社會中將扮演相當重要的資訊及知識提供者的角色。因此將知識庫系統引進數位博物館的建構中，所產生的影響將是相當巨大的。本計畫的執行，將在這方面作出先驅性的貢獻。

至於在跨語言整合性檢索功能方面，我們知道故宮在世界博物館界占有舉足輕重的地位。因此，未來故宮數位博物館必然需要與國際上重量級博物館進行交流，以達到資源共享之目標。因此，跨語言整合性檢索的功能乃未來故宮數位博物館之推廣與國際化所需之關鍵技術。

由上述的說明可知，本子計畫的研究主題對未來故宮數位博物館的發展有著關鍵性的影響。同時本子計畫所研發的技術更能奠定故宮數位博物館在數位博物館系統建構中的領導地位。

貳、計畫成果報告(Part 1: 知識管理系統)

一、知識管理簡介：

談論知識管理之前，首先要討論資料、資訊及知識彼此間的差異性。知識不是資料，當然也不是資訊。所謂的資料是指單純的信號或信息。舉例來說，身體的血液在動脈中流動的壓力就是一種信號。資訊則是指可被解讀和了解的資料。也就是說，將蒐集到的資料經過整理、統計或分類等資料處理流程，產出有意義的結果。舉例來說：經

由血壓計來測量血壓所得到的一組血壓值就是資訊。而知識是指對人類的生活或企業的經營能創造出行動力與價值的資訊，也就是有價值或有潛在意義的資訊。舉例來說：知道正常血壓值是多少、知道如何預防高血壓就是知識。

針對知識存在的形式，我們可以將知識分成兩類。第一類是指具有結構化，可訴諸於文字編輯，並客觀地傳授給別人的知識，稱作「外顯知識」(Explicit Knowledge)。至於有些無法輕易用言語或文字描述的經驗、判斷和直覺就稱作「內隱知識」(Tacit Knowledge)。資訊科技對外顯知識的儲存和流通極具有實質上的效益，對於知識管理的提升有相當大的助益。因此，為了使內隱知識也可以適當地流通與分享，我們必須將內隱知識轉移至外顯知識，也就是將內隱知識外部化。

「知識單元 (Knowledge Unit)」是知識庫的基本結構成分，意即一個小單位的知識內容。分析與定義知識單元的目的是為了使知識管理系統能夠決定儲存資訊的方式，及建立關聯、編製索引與設計檢索的機制，還有分類的準則。知識單元的粗細將會影響到知識管理系統是否能適當的提供個人或企業有意義的內容。舉例來說：當您的知識管理系統探討的是全球性的議題時，您可設定知識單元為國家。但是當您探討的只是某一個國家的議題時，知識單元或許可設定為省

或州。

知識管理界首屈一指的理論大師 Karl M. Wiig 是全球首位提出「知識管理 (Knowledge Management)」概念的專家。自從 Wiig 於一九九三年提出知識管理一詞之後，許多專家也開始紛紛提出他們個人對於知識管理的定義與見解。

以下列出幾點知識管理的定義：

(一)「有關知識的清點、評估、監督、規劃、取得、學習、流通、整合、保護、創新的活動，並將知識視同資產進行管理，凡是能有效增進知識資產價值的活動，均屬於知識管理的內容。結合個體與團體，將個體知識團體化，將內隱知識外顯化；結合組織內部與外部，將外部知識內部化，將組織知識產品化，則屬於知識管理的過程」。

(二)「在適當的時間，將正確的資訊傳遞給需要的人員，並協助其分享與創造價值，以達到提昇組織競爭優勢之策略」。

(三)「透過資訊科技的應用，將組織內成員所創造的經驗及部門所累積的資訊互相結合，並在知識分享的組織文化下，讓企業累積知識的程度達到乘數效果，進而達到知識創新與永續經營的目的」。

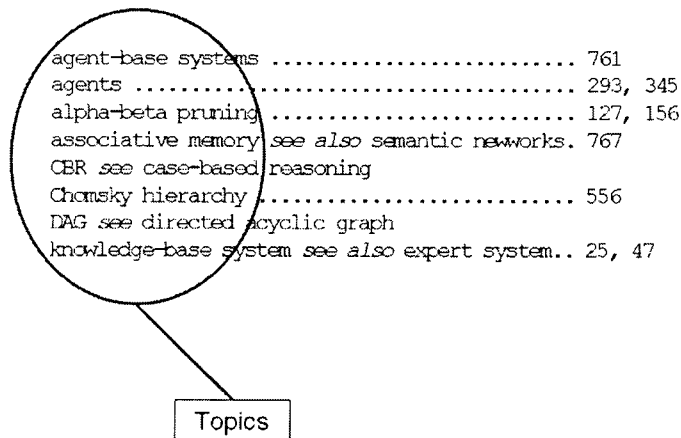
故宮博物院書畫處所整理出來的書畫小辭典，是經過書畫處的專員，經過專業判斷與分析之後，所整理出來的重要詞彙。書畫小辭典的內容包括有中國書畫家的介紹與中國書畫名詞術語的解釋。這些詞彙是用來建構故宮博物院書畫知識庫的主要資訊來源。我們將利用資訊科技，將這些經由專家所建議的詞彙，儲存至知識庫系統。使用者可經有此專業知識庫系統的輔助，能夠容易的來學習並獲取有關中國書畫方面的專業知識，甚至能經由此專業知識庫系統內部機制的運做，來發現有關中國書畫方面的新知識。

二、主題地圖簡介：

綜觀所有知識的內容，皆可歸納至人、事、時、地、物這五個主題。所以任何「領域知識 (Domain Knowledge)」應該只由與此領域相關的人、事、時、地、物這五個層面的主題，及存在於主題之間彼此的關聯性，以及每一主題所擁有的資訊資源共同形成。我們將把一知識領域裡，所有有意義的主題組織起來，形成一知識導航的主題地圖。

主題地圖就好比是書本後面所附的關鍵詞索引。一本好書除了內容要豐富之外，更重要的是它必須要有一份很完備的關鍵詞索引。一份完備的關鍵詞索引，除了完整地列出所有出現於該書本內的重要關

鍵詞之外，更重要的是組織這些關鍵詞之間的關聯。也就是建立關鍵詞之間「見 (See)」及「參見 (See Also)」之參照關係，將具有關聯性之關鍵詞加以連結。讀者即可按圖索驥，輕鬆找到需要的正確資訊。圖一至六是一本人工智慧書籍的索引目錄，我們以此當範例說明關鍵詞索引與主題地圖之間的關係。



圖一、主題與關鍵詞

agent-base systems	761
agents	293, 345
alpha-beta pruning	127, 156
associative memory <i>see also</i> semantic networks.	767
CBR <i>see</i> case-based reasoning	
Chomsky hierarchy	556
DAG <i>see</i> directed acyclic graph	
knowledge-base system <i>see also</i> expert system..	25, 47

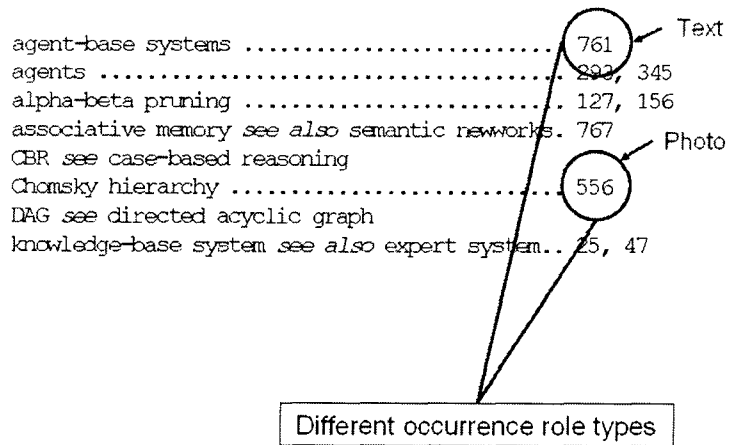
Occurrence

圖二、關鍵詞的呈現頁數

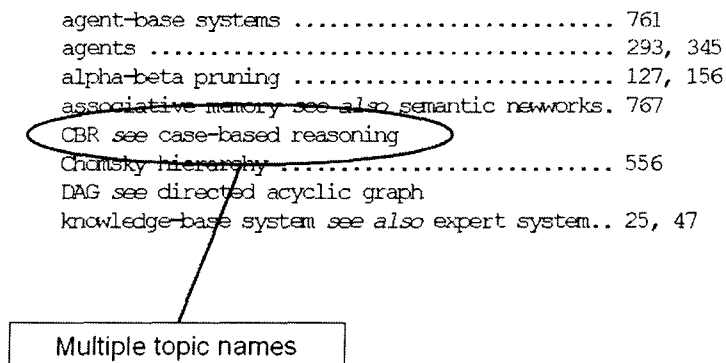
algorithm	agent-base systems	761
	agents	293, 345
	alpha-beta pruning	127, 156
	associative memory <i>see also</i> semantic networks.	767
	CBR <i>see</i> case-based reasoning	
	Chomsky hierarchy	556
	DAG <i>see</i> directed acyclic graph	
graph	knowledge-base system <i>see also</i> expert system..	25, 47

Different topic types

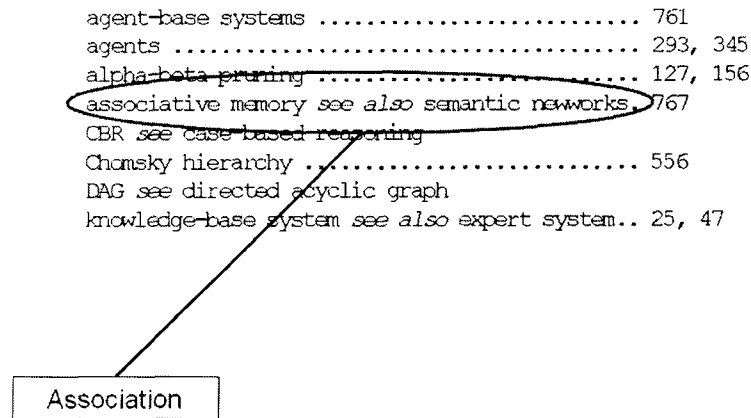
圖三、不同型態的關鍵詞



圖四、不同型態的關鍵詞呈現



圖五、關鍵詞之同義詞



圖五、關鍵詞之相關詞

主題地圖標準規範 (ISO/IEC 13250:2000 "Topic Maps" standard, 簡稱 ISO 13250) 建立的背景是緣自於近年來全球資訊網的爆炸性成長, 刺激了人們對於新世代知識管理技術的需求, 而被提出來的。主題地圖是一個功能強大的全球資訊網導航新機制。伴隨著主題地圖標準規範的制定, TopicMaps.Org 提出主題地圖之可擴充標示語言 (XML Topic Maps), 讓主題地圖得以在網際網路上使用, 並且可應用於各領域知識的知識組織, 方便使用者能有一致性的知識管理策略。

主題地圖主要是由一群主題所 (Topics) 組成, 通常編寫成 XTM 的形式加以保存。除了主題之外, 還包括有關聯 (Associations) 以及

呈現 (Occurrences)。以下以「故宮博物院書畫小辭典」說明構成主題地圖的三個主要元素：

(一) 主題：任何人、事、時、地、物等，凡是能引起使用者討論的對象都可以是一個主題，圖七是表示「故宮博物院書畫小辭典」的各類主題。每一個主題通常可被賦予一個名稱，以及一個統一資源識別符 (Uniform Resource Identifier, URI)。譬如：「蘇軾」是一個主題，「赤壁」也是一個主題，而「前赤壁賦」也是一個主題。我們知道蘇軾陳經擔任過禮部尚書，所以我們也可以定義一個主題叫做「禮部尚書」，只不過「禮部尚書」是另一個主題的型態，也就是有關主題的主題 (Metatopic)。此外對於一個主題而言，時常會因為使用者的不同觀點，而有不同的解釋或看法。譬如：因為蘇軾在書畫藝術分別都有獨特的風格，所以他除了是一位「書法家」之外，也是一位「畫家」。另一方面他在文學方面的造詣更是被眾人所推崇，而名列唐宋八大家之一，所以蘇軾也是一位「文學家」。因此之故，對於一個主題通常會有不同面的分析 (Facet Analysis)。

(二) 關聯：石鐘山在今西湖口縣鄱陽湖畔。蘇軾由黃州前往汝州余中，經遊石鐘山，寫下了石鐘山記。所以「蘇軾」與「石鐘山記」這兩個主題之的關係是一種「著作關係」，其中「蘇軾」扮演的角色是「作者」；「石鐘山記」扮演的角色則是「作品」。又例如說，我們

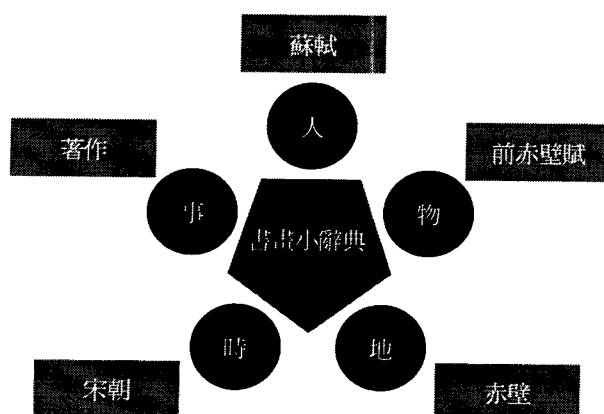
知道「三蘇」是指蘇洵和他兒子蘇軾、蘇轍的合稱，在這當中「蘇洵」扮演的是「父親」的角色，而「蘇軾」及「蘇轍」扮演的則是「兒子」的角色。我們稱這種介於不同主題之間的關係為關聯，關聯的形式可以是一對一、一對多或多對多。關聯是主題地圖的主要功能。因為，將存在於主題之間各類關係，透過關聯的組織與連結後，將形成一領域知識的「知識網 (Knowledge Network)」。如果能進一步的將知識網結合推論引擎，即可提供創新知識與更新知識的能力。譬如：在上例中，我們可經由「蘇軾」及「蘇轍」具有相同的父親，進而得到他們兩個人彼此是具有兄弟關係。

(三) 呈現：一個主題可連結至一個或多個，在某種層面上被視為與該主題切題的資訊資源，這樣的資源稱為該主題的呈現。譬如：「蘇軾」這個主題可出現在許多資源。而每一項資源又可以有不同的作用，譬如：蘇軾的傳記、蘇軾的著作或提及蘇軾的典故等。通常呈現是指儲存於全球資訊網裡任何形式的資源，意即可經由統一資源定位器 (URL) 存取到的資源。由於網路資源來於自四面八方，為了能達到各取所需的目的，對於主題的呈現通常會進行範疇分析 (Scope Analysis)。譬如：依據語文別的不同，對於每一項資源可指定其所屬語系。

總而言之，「主題」、「關聯」及「呈現」就像是三度空間裡的「點」、

「線」、「面」，其目的是用來將一特定領域知識具體化。所以主題地圖也是一種知識呈現（Knowledge Representation）的工具。透過主題地圖，我們可將抽象的知識內容組織成一個有座標概念的知識地圖。每一位使用者皆可依據個人對該領域的認知與了解程度，從他所熟悉的或有興趣的主題方向出發，經由最佳化的自主定位導航輔助，得到他所需要的知識。

由於「故宮博物院書畫小辭典」的內容皆可歸納至人、事、時、地、物等主題，所以非常適用於主題地圖的建置。所以，我們將採用主題地圖為知識管理工具，以主題為知識管理的基本單位，連結主題之間的關聯，組織大量異質性資訊資源，提供使用者在學習故宮博物院書畫知識時，最佳化的自主定位導航系統，以達到知識分享與知識創造的最終目的。



圖七、故宮博物院書畫辭典之主題類型

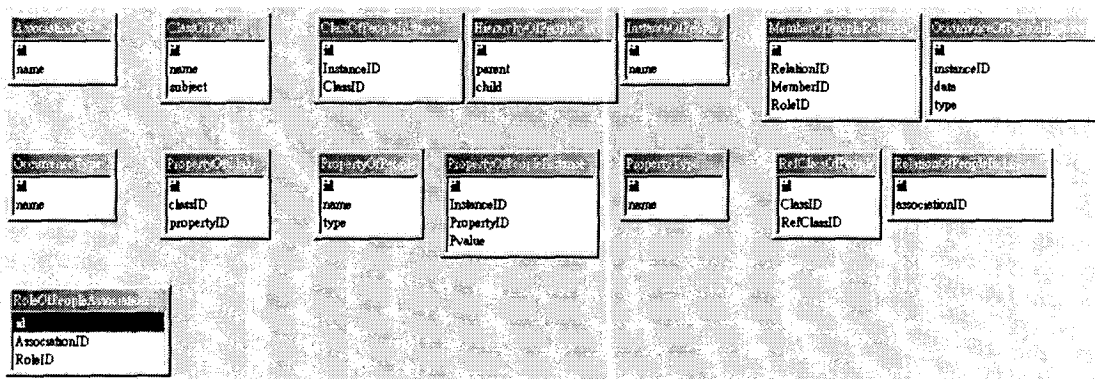
三、系統分析簡介：

就像主題地圖是由許多主題組織而成的，人腦則是由許多神經細胞組織而成的。由於神經細胞是一種有機體，隨著人類的知識成長，大腦所儲存的訊息將會增加，而神經細胞的數量也會隨之增加。所以，當主題地圖所儲存的知識領域擴大時，主題的個數必定會增加。製作一個符合 ISO13250 標準規範的主題地圖，必需事先進行主題及其相關聯之定義，無法隨意更改或增加。尤其如果又是使用 XTM 的形式來記錄主題地圖，則不論是在主題的新增，或是異動等維護作業，都是相當困難的。

為此我們必須採用一種較為彈性的主題維護方式，以方便主題地圖的管理者進行知識管理。資料庫觀念的運用可說是最好的解決方式，因為資料庫是一種專門用來管理（儲存、修改、刪除、產生）企業內資料的系統。資料庫的好處之一就是「資料獨立（Data Independence）」，通俗的說，就是程式中並不包含與資料儲存格式有關的細節，當資料庫的資料結構或存取方式有任何的改變時，都不會影響到整個程式的邏輯。除此之外，使用資料庫還有許多優點列舉如下：（一）資料與資訊可共用共享。（二）可去除資料重覆的現象。（三）系統內資料確保一致性。（四）增加資料保密與安全性。（五）資料與

程式保持獨立性。

我們將依據 TopicMaps.Org 提出的主題地圖之可擴充標示語言當中所制定的文件型態宣告，首先將它轉換成實體關係圖，再依據此實體關係圖來設計關聯式資料庫綱目，以儲存主題地圖的資訊。圖八列出構成整個知識庫的資料庫綱目。



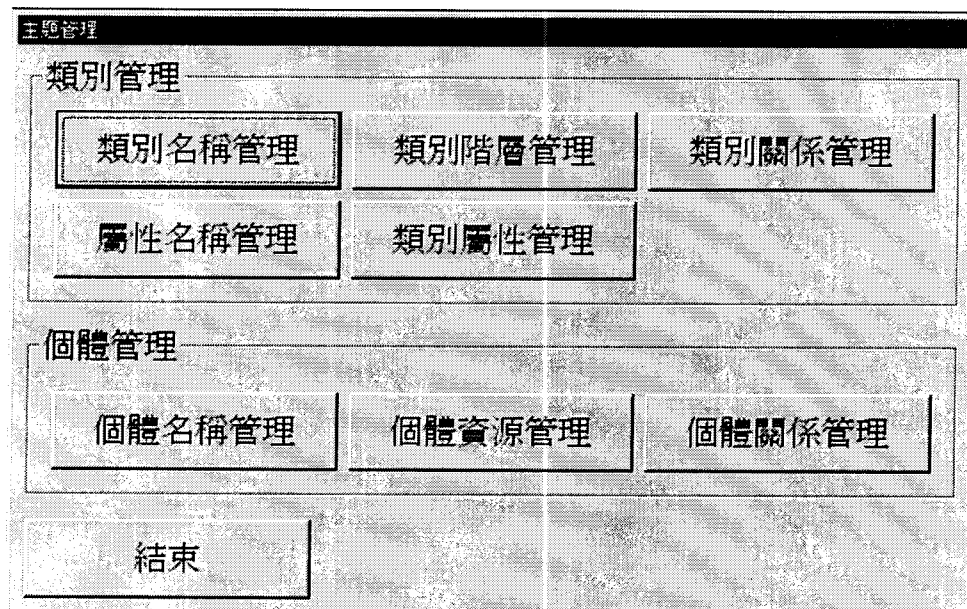
圖八、知識庫的資料庫綱目

四、系統管理介面簡介：

整個主題管理系統的畫面如圖九，我們將系統依據所管理知識的形式分成兩個主要功能，分別為「類別管理」與「個體管理」。所謂的類別是指一個抽象的概念，泛指某一類的個體。而個體則是存在於真實世界中的個別物體，是從某一特定的類別衍生而來的。當我們在描述一個個體時，會針對其特徵加以說明，一個個體所具有的特徵即是個體的屬性。「類別管理」與「個體管理」主要是用來建置知識庫

所要存放的內容，經由此一視窗介面的管理程式，管理者將可快速的與有效的來學習知識庫的管理。

此外，我們也針對使用者設計一個架構於微軟網際網路資訊服務（Internet Information Services）平台上的個人全球資訊網伺服器（Personal Web Server），採用動態的伺服器網頁（ActiveX Server Page）來設計知識庫的查詢瀏覽介面，在這個介面底下，使用者可任一點選他所感興趣的主題為出發點，之後系統會依據主題之間的關聯性，進一步地引導使用者閱讀更多相關性的主題資訊。



圖九、主題管理系統主畫面。

I. 類別管理：

類別管理主要是在建立知識庫的基本知識，也就是本體論（ontologies）的管理，在類別管理功能之下，我們又細分出五項子功能，列舉如下：

（一） 類別名稱管理：

類別名稱管理之維護管理介面如圖十。其主要功能是讓管理者可以維護構成知識庫之基本語詞。舉例來說：在故宮博物院書畫辭典裡，我們依據人、事、時、地、物的原則，分別記錄了幾個重要的語詞，分別是有關地理的「郡」、「縣」、「州」、「書院」等；有關人物的「書法家」、「畫家」、「文學家」、「帝王」；有關事物的「詩詞」、「書法」、「畫」及有關時間的「年代」、「朝代」等。

（二） 類別階層管理：

類別階層管理之維護管理介面如圖十一。其主要功能是讓管理者可以維護類別之間的樹狀架構。因為，在類別名稱管理功能當中，只提供了個別的類別名稱維護，對於類別之間的上下關係，也就是並未提供父子關係的管理。為了能方便將來對於類別名稱進一步的利用。因此之故，我們提供了類別階層管理，用來設定類別之間的父子關係。舉例來說：「人物」類別可在分成「書法家」、「畫家」、「文學家」等；「詩」類別可在分成「田園詩」、「諷諭詩」、「古體詩」及「抒情

詩」等。

（三） 類別關係管理：

類別關係管理之維護管理介面如圖十二。其主要功能是讓管理者可以建立並維護類別之間所存在的關係。它不同於類別階層管理，只是提供樹狀架構的建立；它主要是建立一具方向性的圖像關係。在建立一個類別關係的過程中，必須記錄構成此關係的所有角色，這些角色分必須事先在類別名稱管理當中有所定義。舉例來說：構成「父子關係」的主要角色有「父親」與「兒子」兩個類別；構成「著作關係」的主要角色有「作者」與作品，其中我們將作品分為「詩」、「詞」、「畫」三類。

（四） 屬性名稱管理：

屬性名稱管理之維護管理介面如圖十三。其主要功能是讓管理者可以維護在此一主題地圖底下，所有類別之特徵項目，也就是所謂的屬性。我們必須簡單的紀錄每一屬性的名稱及其所屬型態，以方便未來在個體管理時，能夠具體的來描述一個個體。目前我們所紀錄的屬性包含有有關人物的「出生地」、「出生日」、「字」及「號」等，及畫作的「寬」、「長」等欄位屬性。

（五） 類別屬性管理：

類別屬性管理之維護管理介面如圖十四。其主要功能是讓管理者用來維護每一類別所具備的特徵。在屬性名稱管理功能當中，主要是維護整個主題地圖裡所有可能會用到的屬性。而類別屬性管理則是用紀錄某一類別底下，所擁有的屬性。因此，在建立一個類別屬性之前，必須事先建立屬性名稱。舉例來說：用來紀錄「人物」這個類別的屬性有「出生地」、「出生日」，「字」、「號」等。

類別管理-類別清單(15/15)	
類別名稱	書法家
類別名稱	
畫家	
師父	
徒弟	
父親	
兒子	
年代	
朝代	
書院	
詩	
帝王	
作者	
詞	
郡	
縣	
▶ 書法家	

新增

更新

寫入

取消

查詢

查詢下一筆

刪除

第一筆

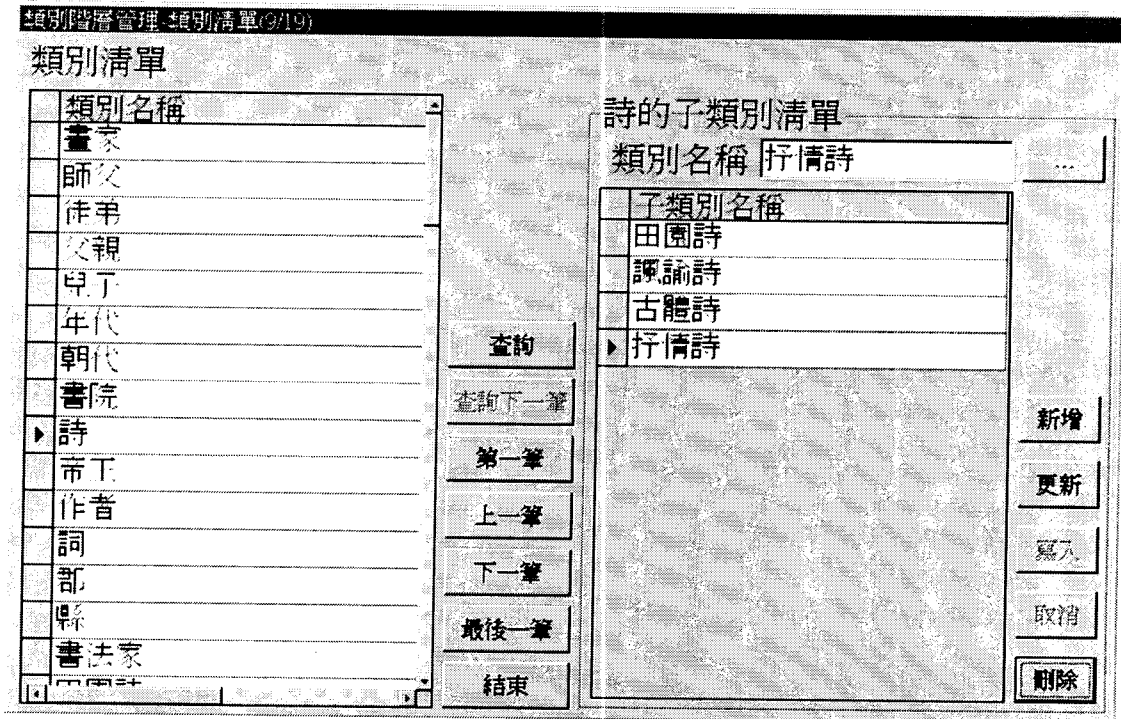
上一筆

下一筆

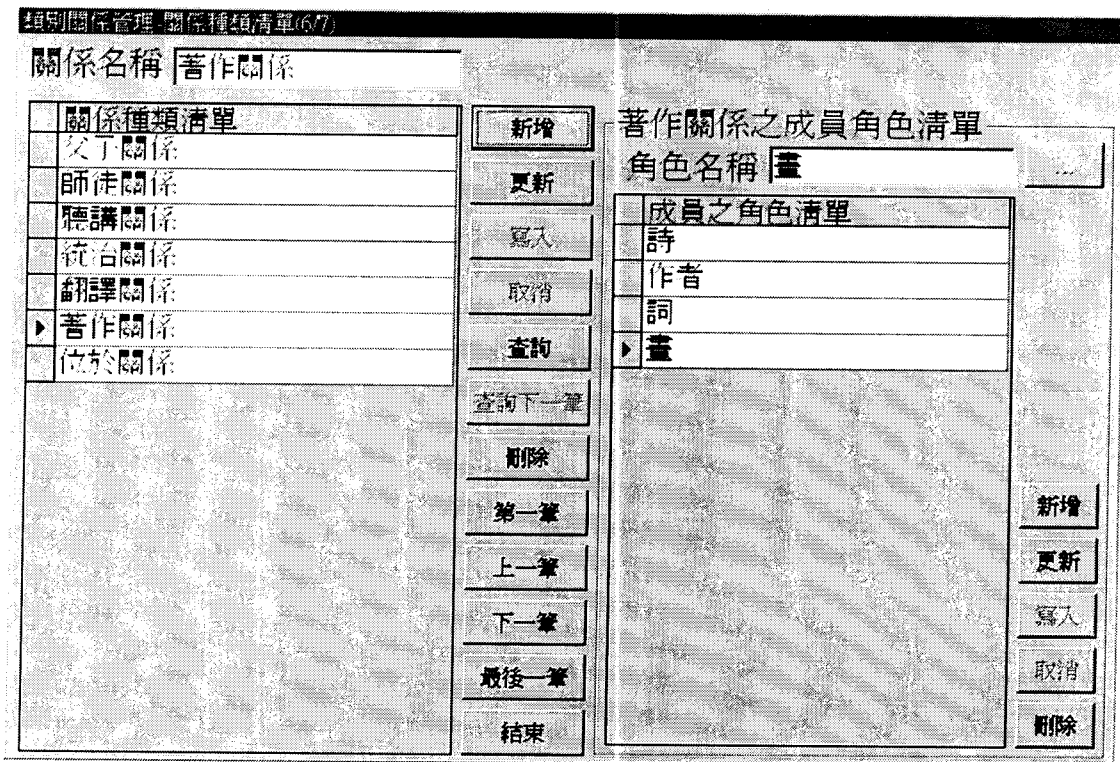
最後一筆

結束

圖十、類別名稱管理畫面。



圖十一、類別階層管理畫面。



圖十二、類別關係管理畫面。

屬性管理-屬性清單(11/11)

屬性名稱

屬性型態

屬性名稱	屬性型態
逝世日	數值
出生日	數值
出生地	文字
字	文字
號	文字
別名	文字
本名	文字
尊稱	文字
長	數值
寬	數值
▶ 職稱	數值

新增
更新
寫入
取消
查詢
查詢下一筆
刪除
第一筆
上一筆
下一筆
最後一筆
結束

圖十三、屬性名稱管理畫面。

類別屬型管理-類別清單(21/21)

類別清單

類別名稱
朝代
書院
詩
帝王
作者
詞
部
縣
書法家
田園詩
諷諭詩
古體詩
抒情詩
畫
▶ 人物

查詢
查詢下一筆
第一筆
上一筆
下一筆
最後一筆
結束

人物類別的屬性清單

屬性名稱

屬性名稱	屬性型態
逝世日	數值
出生日	數值
出生地	文字
字	文字
號	文字
別名	文字
本名	文字
▶ 尊稱	文字

新增
更新
寫入
取消
刪除

圖十四、類別屬性管理畫面。

II. 個體管理：

個體管理主要是在建立構成知識庫的基本事實資訊，也就是可供查詢的具體資訊。在個體管理功能之下，我們可再細分出三項子功能，列舉如下：

(一) 個體名稱管理：

個體名稱管理之維護管理介面如圖十五。其主要功能是讓管理者可以維護構成知識庫之具體且可供查詢的資訊。舉例來說：在故宮博物院書畫小辭典裡，我們依據人、事、時、地、物的原則，分別記錄了幾個重要的個體，分別是有關地理的「黃州」、「赤壁」等；有關人物的「蘇軾」、「宋神宗」、「王安石」、「蔡襄」、「黃庭堅」、「米芾」等；有關事物的「前赤壁賦」、「祭黃幾道文」、「黃州寒食詩帖」、「枯木怪石圖」及有關時間的「元豐初年」、「唐朝」等。

在建立個體名稱的同時，我們必須同時去設定該個體所屬的類別，舉例來說：在「蘇軾」這個個體所屬的類別就有「書法家」、「畫家」、「文學家」、「兒子」及「作者」。這些類別附加在蘇軾身上表示蘇軾是一位書法家、畫家、文學家，蘇軾也可以是一位兒子，另外蘇軾同時是一位作者。

為了進一步的賦予個體具體的特徵，也就是描述個體的狀態，我

們必須紀錄一個體的屬性資訊。舉例來說：用來描述「蘇軾」這個個體的屬性資訊有蘇軾的誕辰日是 1036 年，逝世日是 1101 年，出生地是四川省眉山縣，字子瞻，號東坡居士。

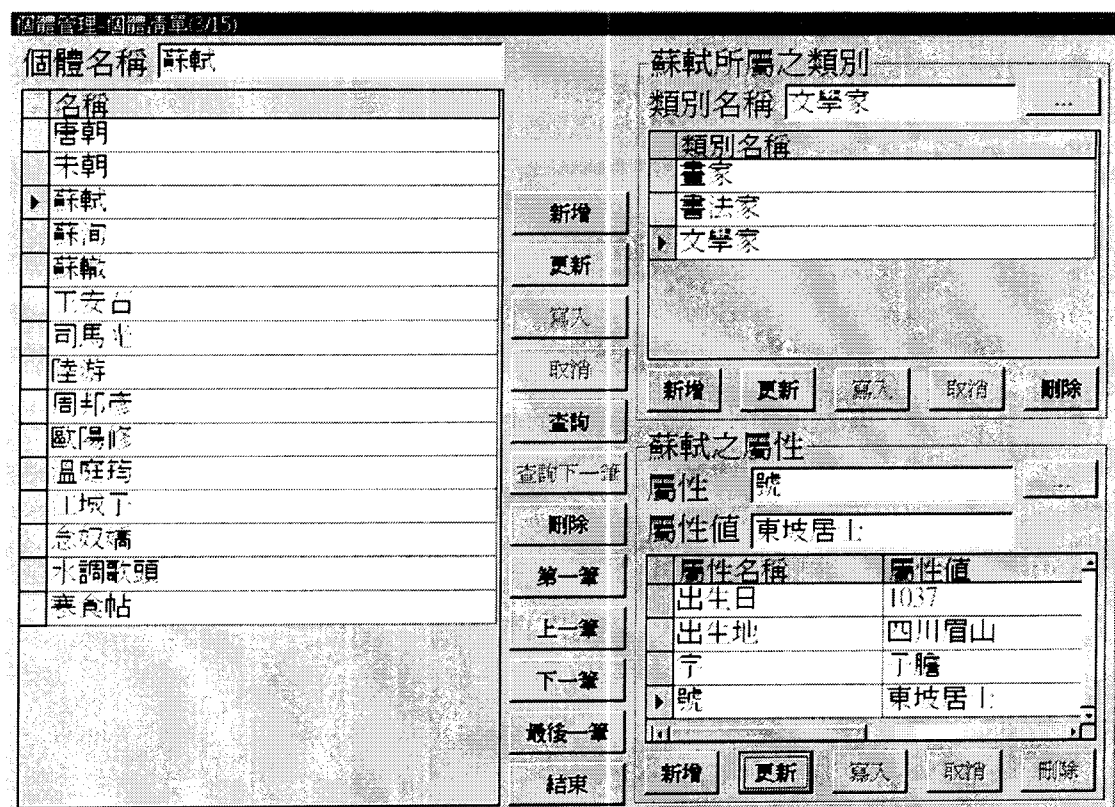
（二） 個體資源管理：

個體資源管理之維護管理介面如圖十六。其主要功能是讓管理者可以維護個體的相關資訊資源，這些資訊資源通常是指位於全球資訊網上，各式各樣的多媒體資源。舉例來說：我們收集了有關「蘇軾」的網站、網頁、圖片等，並將收集到的來源，透過統一資源定位器（URL）的方式紀錄下來。如此，我們將可更完整的來建立每一個體主題的知識庫。

（三） 個體關係管理：

個體關係管理之維護管理介面如圖十七、圖十八。其主要功能是讓管理者可以建立個體之間的關係。在類別關係管理當中，我們已經建立了構成一關係的角色有哪些。在個體關係管理裡，我們首先選擇要建立的關係，之後分別指定在此一關係底下的成員有哪些個體，以及每一個體所屬的角色。舉例來說：我們想要紀錄蘇洵與他的兩位兒子蘇軾及蘇洵之間的關係時，首先先增加一「父子關係」，並在此父子關係底下選擇「蘇洵」這個成員，以及設定其角色為「父親」，另外再選擇「蘇軾」、「蘇洵」，同時設定這兩個成員的角色為「兒子」。

如果我們想要描述「元豐五年（1082年）蘇軾游赤壁後寫下《念奴嬌·赤壁懷古》」這件事情，也可經由個體關係的建立，來描述這件事情。首先我們增加一「著作關係」，由於著作關係裡包含有人、時、地、物，所以我們再進一步的選擇「元豐五年」這個時間個體，以及設定其角色為事件發生的「年代」，另外再選擇「蘇軾」、「念奴嬌·赤壁懷古」，分別扮演「作者」及「作品」這兩種角色，及事件發生的地點「赤壁」。



圖十五、個體名稱管理畫面。

個體清單

個體名稱	查詢
唐朝	查詢下一筆
宋朝	第一筆
蘇軾	上一筆
蘇洵	下一筆
▶ 蘇轍	最後一筆
王安石	結束
司馬光	
陸游	
周邦彥	

蘇轍相關資訊資源

資源位址

資源型態

資源位址	資源型態	
http://www.louhau.edu.mo/www/teach/tangsong/so	文字	新增
▶ http://www.contest.edu.tw/87/endshow/1/sue/works	文字	更新
		寫入
		取消
		刪除

圖十六、個體資源管理畫面。

個體關係管理-關係清單(1/3)

關係名稱 父子關係

關係名稱	新增
父子關係	更新
師徒關係	寫入
統治關係	取消
統治關係	查詢
著作關係	查詢下一筆
著作關係	刪除
位於關係	第一筆
位於關係	上一筆
	下一筆
	最後一筆
	結束

父子關係-1之成員及角色

成員名稱 蘇軾

成員角色 兒子

名稱	角色名稱
蘇軾	兒子
蘇洵	父親
蘇轍	兒子

新增
更新
寫入
取消
刪除

圖十七、個體關係管理畫面。

個體關係管理-關係清單(6/8)

關係名稱 著作關係

關係名稱	新增
父子關係	更新
師徒關係	寫入
統治關係	取消
統治關係	查詢
著作關係	查詢下一筆
著作關係	刪除
位於關係	第一筆
位於關係	上一筆
	下一筆
	最後一筆
	結束

著作關係-6之成員及角色

成員名稱 寒食帖

成員角色 書法

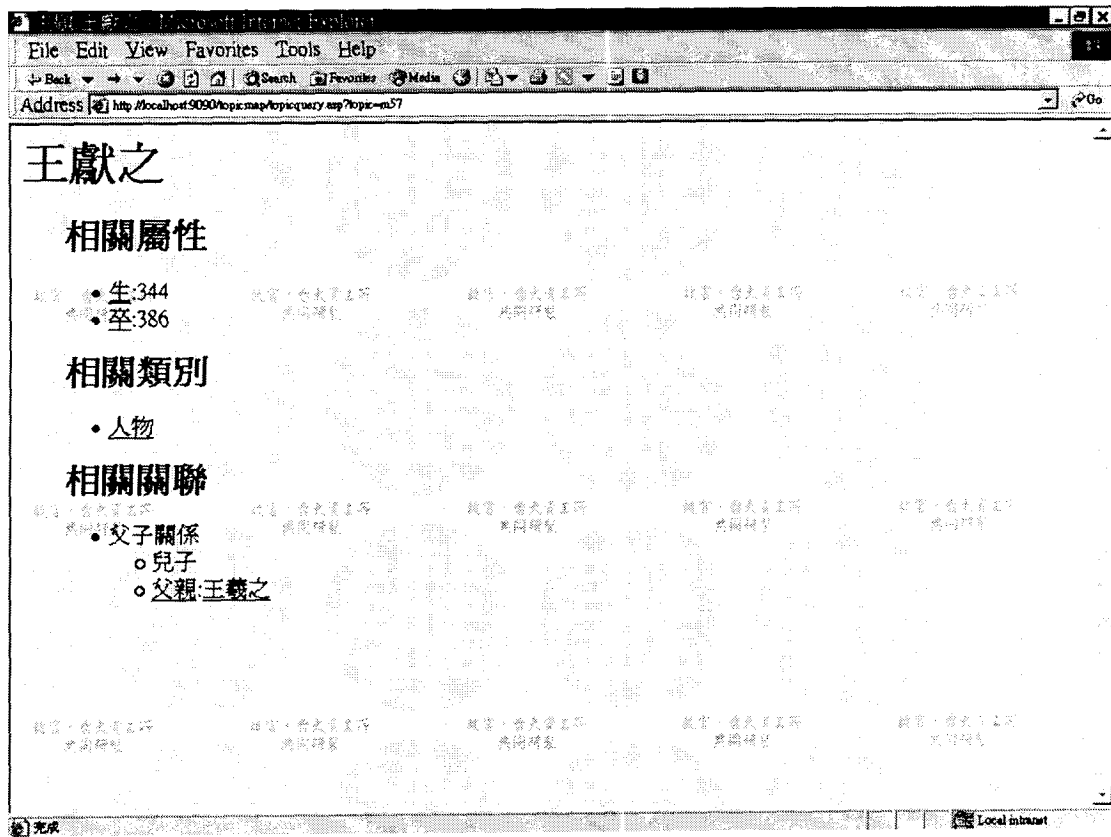
名稱	角色名稱
蘇軾	作者
江城子	詞
念奴嬌	詞
水調歌頭	詞
寒食帖	書法

新增
更新
寫入
取消
刪除

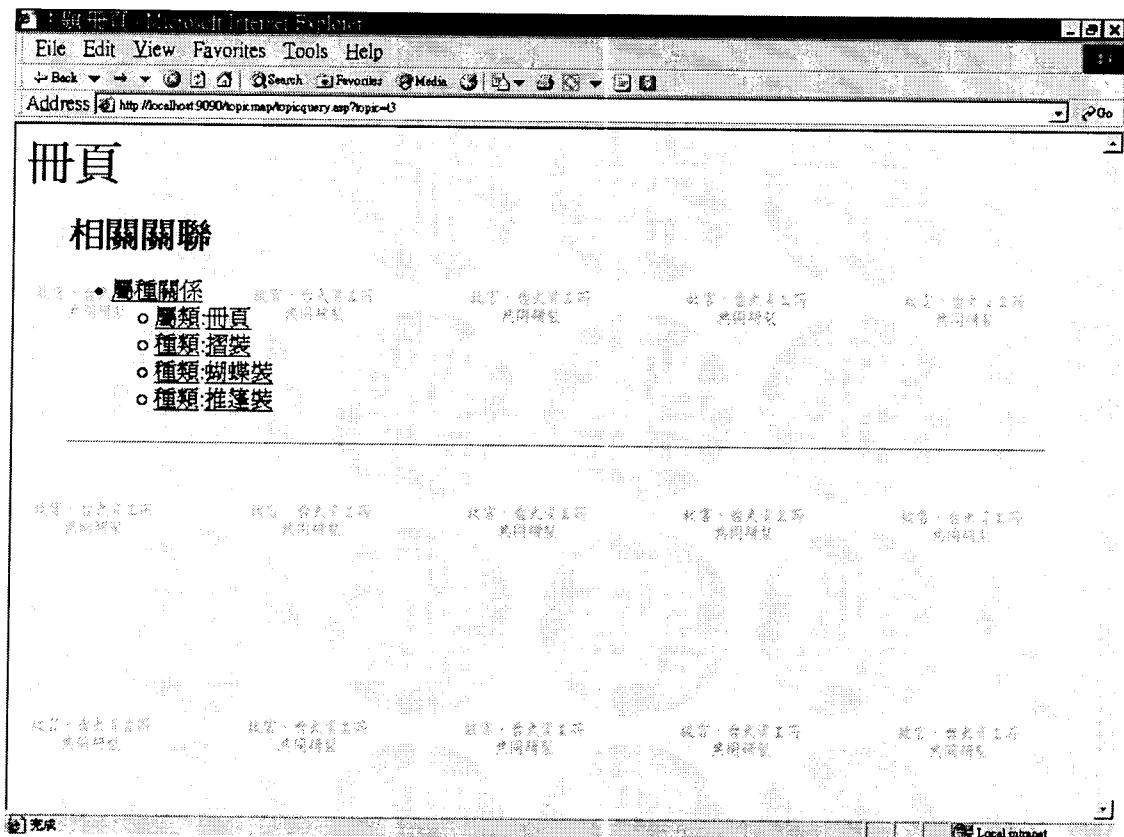
圖十八、個體關係管理畫面

III. 使用者查詢瀏覽：

使用者查詢瀏覽介面如圖十九與圖二十。在圖十九當中使用者可看到某一個體的個別資訊，在圖二十當中使用者則可以看到某一類別之間的關聯性。



圖十九、個體資訊查詢瀏覽畫面



圖二十、個體關係查詢瀏覽畫面

貳、計畫成果報告(Part 2: 跨語言檢索)

一、詢問翻譯系統功能：

在跨語言資訊檢索中，詢問(query)與文件(document)分屬不同語言，因此詢問與文件之間，必須有一個對應，翻譯是重要的運算之一。過去跨語言資訊檢索可能的策略，基本上可以區分出詢問翻譯(query translation)、文件翻譯(document translation)、和不翻譯(no translation)等三類。詢問翻譯是將詢問翻譯成與文件同一種語言，再進行檢索。文件翻譯則相反，是把資料庫的文件翻譯成與詢問同一種語言。如果

以機器翻譯系統來做，馬上會有執行效率的問題。要翻譯大量的文件需要花費大量的時間，這在實際的應用上是不合宜的，所以大部分的系統是採用詢問翻譯的方式。不翻譯的方法是直接將詢問和文件對應到同一個向量空間，在此向量空間中就可以做相似度的計算。Bellcore (Deerwester *et al*, 1990)曾在單語資訊檢索，提出一種稱為 Latent Semantic Indexing (LSI)的方法。Dumais 等人 (1997)進一步把這種觀念引進到跨語資訊檢索，在其訓練過程，英法雙語文件、英語詞彙、法語詞彙都被對應到向量空間。隨後，加入不同語言的文件。沿用 LSI 的基本想法，不管是不是同一種語言，這些文件可以在向量空間呈現出來。因此，這種作法不需翻譯。

為避免龐大的翻譯工作及能快速的應用在其他語言，在此研究中採用了詢問翻譯的方法。由於本系統所要檢索的文件是故宮數位資料，是在限定領域中的文件，所以翻譯後的詢問也要在這一領域中，才能正確的檢索出所要的文件。故宮的文件中有許多的專有名詞、專業用語，一般的字典中並無這些詞，所以要先建立故宮專有名詞的英中對譯辭典。在這部辭典中收集了故宮典藏品品名、人名、專有術詞等。當系統接受到一英文詢問後，從辭典中找出與詢問最接近的詞或詞組，以這個詞或詞組所對應的中文翻譯為詢問的中文翻譯。最後將翻譯後的中文詢問送到檢索系統中，檢索出相關的文件。

當辭典建立好後，要對辭典中的詞做索引，以方便搜尋。製作索引時，先對每個英文字做前處理，包括去停用字、詞幹處理。英文詢問中的字也經過同樣的前處理。除了這兩項前處理外，還要考慮專有名詞音譯的問題。很多的中文專有名詞，特別是人名，是用音譯的方式翻譯成英文。目前音譯所採用的拼音系統有很多套，包括威翟(Wade Giles)、漢語拼音(Pinyin)等，而故宮所採用的是威翟拼音系統。若英文詢問中的專有名詞是用威翟以外的系統，就必須先轉換成相對應的威翟拼音，才能正確的檢索。當做完前處理後，就計算英文詢問與辭典中每一個英文詞組的相似度，找出最相似的詞組。英文詢問與辭典中英文詞組的相似度的計算是根據兩者相同的字在辭典英文詞組中所佔的比率而定，若相同的字越多，佔的比率越高，則相似度越高。

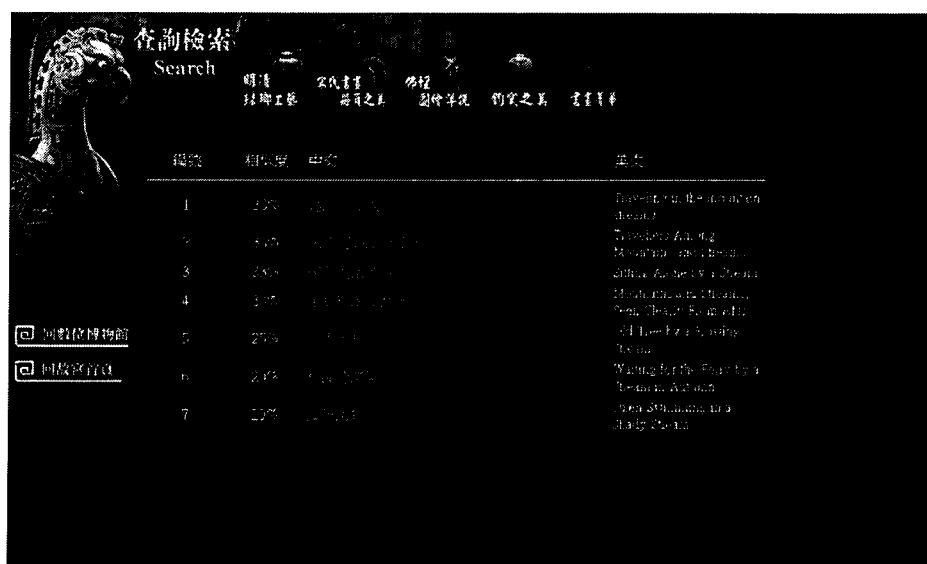
二、使用者介面：

本系統與故宮數位博物館網站整合，透過網頁介面對故宮數位資料做跨語言資訊檢索。一進入系統的介面如圖一，在對話框內輸入英文詢問，按”查詢”鍵後詢問即被傳送到系統處理。當系統將詢問翻譯好後，便將翻譯後的中文詢問回傳給使用者。為提高檢索之精確度，系統不只傳回最相似的翻譯，而是將分數高的翻譯全部傳回給使用者，由使用者來選取最恰當之中文詢問來進行檢索。使用者所看到的查詢翻譯的結果包含了中文及對應的英文翻譯，如圖二所示，使用

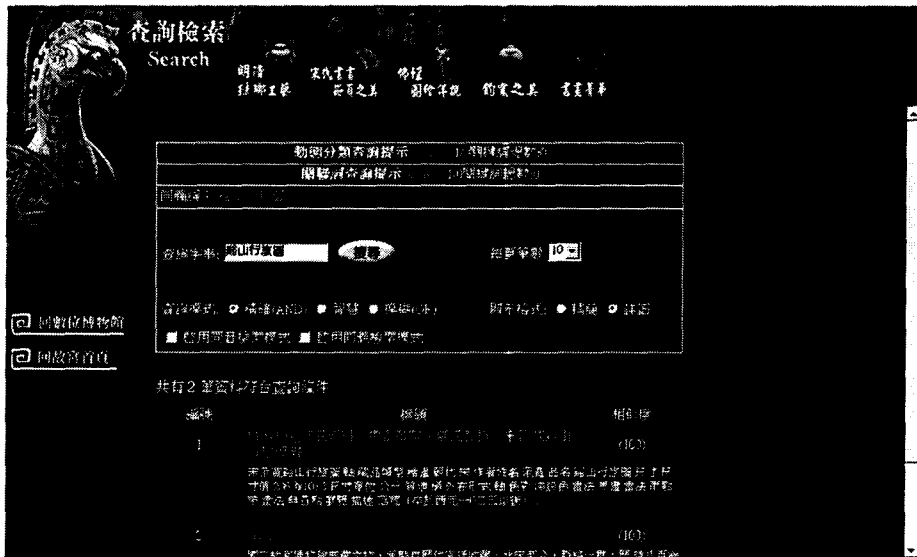
者可以根據英文來判斷哪一個句子最符合他的需求。圖二中每一個中文皆有超連結，按下後即可將中文詢問傳給檢索系統檢索。在圖一的範例中所輸入的詢問是” stream”，系統傳回了七個相似的中文詢問並根據相似度排序。當選擇了第一個翻譯” 谿山行旅圖(Traveling in the mountain streams)” ，系統便將” 谿山行旅圖” 傳給檢索系統檢索。圖三是檢索的結果。使用者可以點選有興趣的文件，檢視文件的內容。圖四是” 谿山行旅圖” 的內容。



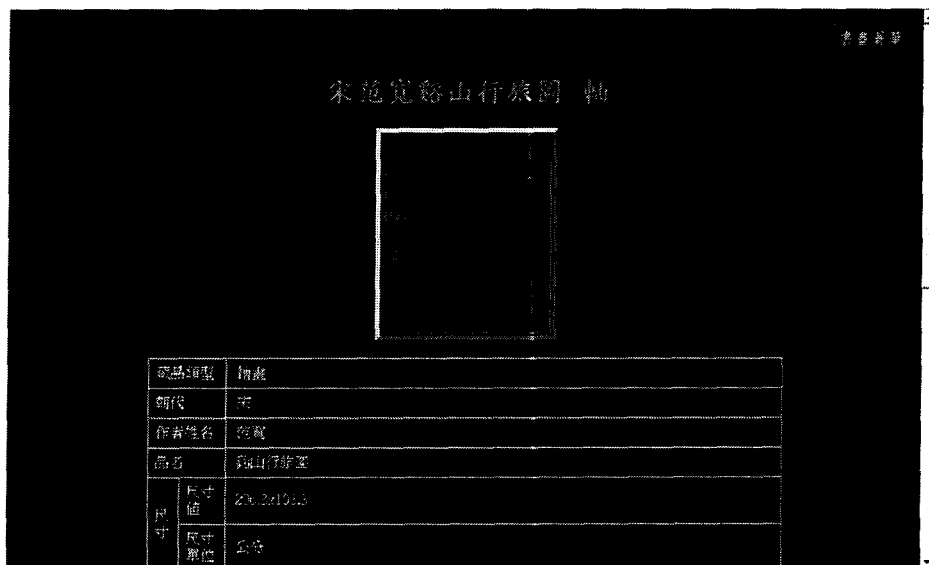
圖一：查詢輸入介面



圖二：查詢翻譯結果



圖三：“谿山行旅圖”檢索結果



圖四：“谿山行旅圖”資料

參、計畫成果檢討

故宮的館藏是豐富且珍貴的，將館藏資料數位化、置於網路上後可以讓更多人來接觸這些珍貴的文化資產。而網路無國界的特性使世界上各地的人都能欣賞故宮的藏品，促進了文化的交流。雖然在網路上資料的傳遞很方便，但是語言障礙的問題讓非中文使用者無法快速的、正確的找尋、使用資料。在此計畫中設計了一套英中跨語言檢索系統以幫助非中文使用者或無中文系統之使用者能正確的找到所要的資料。目前詢問的語言是英文，未來的工作是要將此技術應用在其他的語言上，讓更多的人使用故宮數位資料。

本年度子計畫的兩個主要研究重點為

- (1)將知識庫系統引進數位博物館的建構中；
- (2)完成跨語言整合性檢索功能。

由於數位博物館在資訊化社會中將扮演相當重要的資訊及知識提供者的角色。因此將知識庫系統引進數位博物館的建構中，所產生的影響將是相當巨大的。本計畫的成果，已在這方面作出先驅性的貢獻。

至於在跨語言整合性檢索功能方面，我們知道故宮在世界博物館界占有舉足輕重的地位。因此，未來故宮數位博物館必然需要與國際上重量級博物館進行交流，以達到資源共享之目標。因此，跨語言整

合性檢索的功能乃未來故宮數位博物館之推廣與國際化所需之關鍵技術。

由上述的說明可知，本子計畫的研究主題對未來故宮數位博物館的發展有著關鍵性的影響。同時本子計畫所研發的技術更能奠定故宮數位博物館在數位博物館系統建構中的領導地位。

參考文獻

- Andresen, L. (1997) "Metadata: New Key Concept in Internet Circles," *Bibliotekspresen*, 6, March 1997, 152-153.
- Bearman, D. (1996) "Developments in Metadata Management Frameworks," *Archives and Museum Informatics*, 10(2), 1996, 185-188.
- Bian, K.W. and Chen, H.H. (2000) "Cross Language Information Access to Multilingual Collections on the Internet." *Journal of American Society for Information Science*, Special Issue on Digital Libraries, 51(3), 2000, 281-296.
- Caplan, P. and Guenther, R. (1996) "Metadata for Internet Resources: the Dublin Core Metadata Elements Set and Its Mapping to USMARC," *Cataloging and Classification Quarterly*, 22(3/4), 1996, 43-58.
- Chen, Hsueh-hua, Chen, Chao-chen, & Chen, Kuang-hua (1999) "Metadata Interchange for Chinese Information," *IT and Global Digital Library Development*, Ching-chih Chen (Editor), West Newton: MicroUse Information, 65-74.
- Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and

Harshman, R.A. (1990) "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, 41(6), 391-407.

Dumais, S.T., Letsche, T.A., Littman, M.L. and Landauer, T.K. (1997) "Automatic Cross-Language Retrieval Using Latent Semantic Indexing," *Working Notes of AAAI-97 Spring Symposiums on Cross-Language Text and Speech Retrieval*, 18-24.

Grimes, B.F. (1996) "Ethnologue Language Name Index,"
<http://www.sil.org/ethnologue/names>.

Hershman, T. (1998) "Real-Time Web Language Translators," *Byte*, June 1998, 5-10.