

行政院及所屬各機關出國報告
(出國類別：考察)

圖書館資訊服務架構規畫 – 數位圖書館計畫考察報告

服務機關：國立成功大學

出國人 職 稱：講師

姓 名：羅思嘉

出國地區：美國

出國期間：89.11.02-89.11.12

報告日期：90.02.12

圖書館資訊服務架構規畫 – 數位圖書館計畫考察報告

摘要

資訊科技的變化影響了資訊使用者獲取資訊以及面對資訊的態度，同樣地使得圖書館能夠以與過去不同的方式提供各種資訊服務。相對地在資訊時代，圖書館提供的各項資訊服務也面臨新的挑戰。將資料予以數位化，透過網路，以數位圖書館的方式提供讀者所需的資訊服務，為圖書館在資訊時代開展了一個不同服務型態。此次藉由考察美國地區 5 個不同的數位圖書館計畫，透過與工作小組直接的討論以及對運作流程的觀察，建立成大圖書館未來在規畫與推動數位圖書館以及建構資訊服務架構的基礎。訪問單位與計畫包括 Carnegie Mellon 大學的數位化影像資料自動摘要及索引計畫(Informedia)、哥倫比亞大學的個人化多媒體資訊檢索及摘要系統(PERSIVAL)、紐約公共圖書館的數位圖書館計畫、麻省理工學院的學位論文數位圖書館、Tufts 大學的 Perseus。各計畫在其所處理的資料內容、型態以及目標等方面重心各有不同。對成大或國內其他資訊機構，同時擔負研究教學支援、醫療資訊提供以及社區資訊服務等多重功能責任的單位，均能夠提供一些相當具有參考價值的經驗以及資訊。

考察計畫目的

處於資訊時代，資訊科技的變化影響了資訊使用者獲取資訊以及面對資訊的態度，同樣地使得圖書館能夠以與過去不同的方式，提供各種資訊服務。相對地這樣的轉變也使得圖書館服務面臨新的挑戰。透過電腦網路通訊科技的應用，新式的資訊服務改變了傳統圖書館經營的理念，電子資訊服務應運產生，數位化資料、網路化的資訊傳遞成為建構數位圖書館的基本基礎。透過數位資訊、網路化的服務，資訊服務可實施的方式越來越多樣化，這樣的轉變不單使圖書館需要重新審視、調整資訊服務模式及其後支援的架構，同時過去在技術以及成本考量下無法有效處理、提供的資訊服務型態，在資訊技術環境的改變影響下，均能逐漸一一落實。

在國科會的支援下，國內資訊單位、研究機構於 1998 年開始第一階段的數位圖書館/博物館計畫。對國內相關研究影響程度相當高的美國則是在 1994 年開始這方面的計畫，目前已進入第二期的研究階段。數位圖書館計畫所支援的各個

計畫，其研究結果在不同型式及內容主題資料的服務架構上引起了不同討論，對於提供資訊服務的圖書館來說，在資訊服務架構上不論是參與研究或是落實在資訊服務提供上均產生了影響力不一的結果。除了從實作研究落實，如何從各種來源獲取經驗，是知識管理時代相當重要也有效的學習方式。此次考察的主要目的就是期望從了解相關研究計畫的運作狀況及與參與人員面對面的討論，獲取計畫的相關資訊與經驗，做為圖書館參與這方面研究或建構數位化資訊服務的參考。

這次考察的 5 個單位/計畫所建構的數位圖書館在目的、收藏的館藏、服務對象及研究重點各有不同。各計畫建構的數位圖書館以及其如何計畫其服務機制可在不同方面提供有價值的資訊。考察單位包括 Carnegie Mellon 大學，主要是訪問該校的數位化影像資料自動摘要及索引計畫(Informedia)研究小組，Informedia 計畫是以播報新聞資料處理及架構機制研究為主；哥倫比亞大學的個人化多媒體資訊檢索及摘要系統(Personalized Search and Summarization over Multimedia Information, PERSIVAL)，PERSIVAL 計畫是以醫療資訊服務提供研究為主，結合該校醫療中心、電腦科技研究中心以及圖書館三種專業的合作研究計畫；紐約公共圖書館的數位圖書館計畫，該圖書館主要是著重該館典藏社區歷史資料的數位化及資訊服務；麻省理工學院的學位論文數位圖書館主要是建立該校學位論文，同時提供全文資料服務的數位典藏與服務機制；Tufts 大學的 Perseus 是建立一個提供研究古典文獻所需要的原文及參考資料的數位圖書館。這幾個不同計畫在所處理的資料內容主題、資料型態以及目標等各方面重心各有不同。對於成大來說或國內其他資訊機構，同時具有研究教學支援、醫療資訊提供以及社區資訊服務等多重功能的單位，所考察的單位及所進行的計畫均能夠在這些不同任務發展上提供一些相當具有參考價值的經驗以及資訊。

數位圖書館計畫介紹及議題討論

此次考察的第一站是訪問位於美國匹茲堡的 Carnegie Mellon 大學，接著是位於紐約市的哥倫比亞大學及紐約公共圖書館，最後一站則是訪問位於波士頓的麻省理工學院及 Tufts 大學。前後行程共 11 天，各單位平均以一天的時間，與各個計畫執行小組交換意見。行前已透過電子郵件的方式溝通計畫的基本資料以及主要訪問討論的議題，因此討論能夠直接進行所關心的議題。以下分別就各個計畫做一介紹並對相關議題進行探討。

Carnegie Mellon 大學 Infromedia 計畫 - 影像資料自動摘要及索引計畫

Carnegie Mellon 大學數位化影像資料自動摘要及索引計畫始自 1994 年，在第一期的美國數位化圖書館計畫支援下開始。計畫的目的是架構一個影像資料處理自動化的機制，使影像資料的檢索更具效能。據計畫主要主持人 Howard Wactler 博士表示，之所以會以影像資料為計畫發展的目標，主要是因為越來越多的資訊是以影像方式儲存及呈現，而隨著資訊技術環境的改變，資訊檢索者對影像資料的接受度也越來越高，但是這類型資料的典藏、組織、檢索功能卻還相當有限，因此希望在所提出的計畫中可以發展一套自動處理影像資料的技術，以提昇影像資訊的檢索。計畫預定發展的技術功能包括內容自動截取記錄、索引摘要自動化、影像檢索功能 ... 等，並進一步就各個不同條件從記錄的影像資料中分析、列比不同主題的相關資訊。



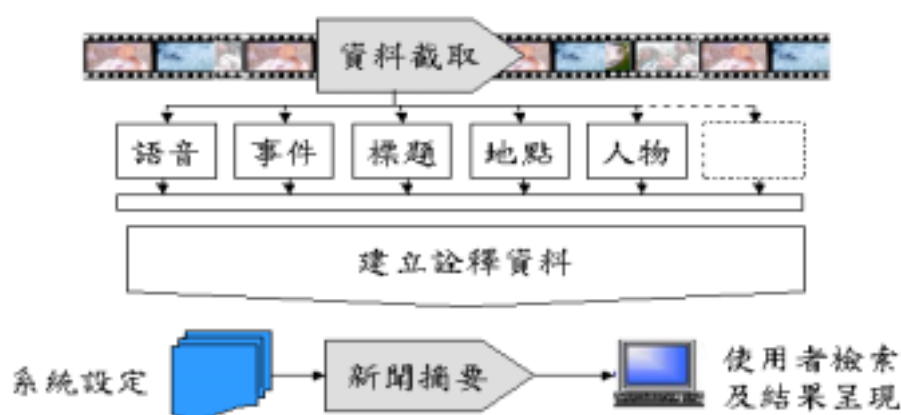
【圖 1：Carnegie Mellon 大學 Infromedia 計畫首頁¹】

Infromedia 先導計畫處理的資訊是以新聞資料為主，目前是以美國有線新聞電視網(Cable News Network, CNN)的新聞為研究測試對象，約有 45,000 段的新聞播報資料，包括 1997 年到目前 CNN 所播報的電視新聞資料。計畫預計達到的目標包括新聞語音的辨識及文字化，經由語音辨識技術將記錄的播報新聞由播報

¹ <http://www.infromedia.cs.cmu.edu/>

的語音記錄下來，轉譯新聞內容使其成為可檢索的資料。除了播報新聞語音的轉譯外，對於新聞影像 Informedia 的處理方式包括截取新聞影像上的標題，透過影像字元辨識將資料予以文字化，做為檢索的詞彙，為了能夠提昇資訊檢索的效能，Informedia 計畫的一部份是針對影像中的人物進行處理，以利人物影像的檢索。人物影像處理部份包括人物影像特徵計算以及顏色的處理。

Informedia 計畫目標的另一部份是建立一個不同的新聞檢索模式。除了將播報新聞的語音辨識及索引的建立外，Informedia 將轉譯資料中的地名轉為經緯度，在內建地圖上予以標註，利用標註後的資訊，以地圖方式顯現新聞事件。資料檢索者可依照地區、時間檢索該地在特定時間的相關新聞資料；或是由地區、時間展現新聞事件的發展狀況。



【圖 2：Informedia 計畫預定目標圖示】

新聞資料處理流程

Informedia 計畫目前所發展出的資料流程包括幾個不同任務，第一步是新聞資料的截取儲存；將每日新聞截取存檔後，系統必須針對每一節新聞進行新聞段落的判讀，將每一則新聞獨立出來；判斷每一則新聞後，系統進一步將新聞的播報語音部份透過辨識，將語音予以文字化；配合每則新聞的新聞鏡頭若附有相關的標題，利用系統的影像截取技術將標題影像截取下來再經由字元辨識處理；新聞影像若包括人物顯現，則人物形體特質以及顏色將予以計算及記錄，做為檢索依據；新聞內容及地理資訊分析；詮釋資料的建立；各種記錄、資料的連結。

一、資料截取

由於 Informedia 計畫鎖定的測試資料是電視新聞，為了加強資訊的更新速率，減少資訊傳遞流程，因此計畫首要的工作目標是如何自動的截取每日每時的

電視新聞，儲存提供系統處理。Informedia 的方式是與當地的有線電視業者達成協議，取得資料轉錄測試權，由系統控制，自動截取、儲存由業者送出的新聞訊號。透過這樣的方式，業者部份不需另行處理新聞資訊的儲存，同時作業流程中的資料傳遞需求可降至最低，人力投資部份也可由系統程式取代，資料的更新可與新聞播報同步。

二、段落判讀

Informedia 計畫主要目的是要達到新聞的自動化處理。截取新聞部份僅是截取、錄製、轉換、儲存，對於新聞內容不做分析，且不區隔每一則新聞，因此第二個工作目標是系統要能夠分辨在每一次播報的新聞中每一則不同的新聞。就人力判斷來說，新聞區段的辨識可清楚且相當容易即可達成，但對於系統程式則不然，Informedia 小組必須從所獲得的影像、語音資料建立判讀每一則新聞的機制與依據。Informedia 小組經過幾種不同嘗試，現階段系統的新聞區段辨識功能是以畫面的切換來判定新聞的轉折。Informedia 工作小組曾嘗試以播報新聞語音詞句的詞彙句型的變化加以判定新聞的切換，但在數次嘗試後發現辨識效果不佳，因此目前是以螢幕畫面的變化來判定，再以內容做為後續驗證處理的依據。

三、語音辨識

判讀每一則新聞後，系統進行播報新聞的語音辨識處理，將語音轉換為文字記錄，做為新聞內容描述資料及索引基礎。據計畫主持人 Wactlar 博士表示，目前 Informedia 所達到的語音辨識正確度為 70% 左右。在檢討過程中，小組成員曾經討論是否應進一步取得更高的語音辨識正確率，但測試所獲得的檢索回收率 (Recall) 以及資料精確率 (Precision)，並未顯示強烈的需求。就新聞資料來說，計畫小組成員對於所獲得的正確率還稱滿意，而要提高轉譯正確度需要再投入的成本不足以平衡可獲得的成果，因此在這個部份 Informedia 暫以所獲得的結果為準。Wactlar 博士認為由於是新聞播報資料，其內容結構的特質使得在檢索上可允許某些程度的誤差，語詞的精確性及概念複雜度不高，所以雖然辨識部份的正確度為 70% 左右，檢索上仍可以達到相當不錯的回收率及精確率；但若是進行其他資訊的檢索可能無法獲得相同的結果。同時資料量以及檢索測試仍是以內部測試為主，因此這樣結果的滿意度還需要進一步驗證。

四、螢幕標題及人物辨識

新聞除了播報語音外，影像是另一需處理的資訊。在這個部分，目前 Informedia 的處理分為兩個部份，一是螢幕畫面標題的處理，另一是畫面人物面部影像的辨識。Informedia 將每一畫面予以分割處理，若是畫面部份顯現相關標題，系統會凍結該畫面，截取畫面文字影像，經過 OCR 技術，將播報新聞螢幕的標題予以文字化，做為可檢索的依據。目前這部份由於部份錯誤對檢索造成問題，因此正確率還有待進一步改進以提高檢索的回收率及精確率。

若影像中包括新聞相關人物或者是主播的影像，Informedia 嘗試分析影像的特質，包括人物面部特徵的計算及顏色的註錄，做為相關檢索的依據。這個部份目前僅處理到面部正面的影像，對於側面的處理仍有限。

五、主題標目

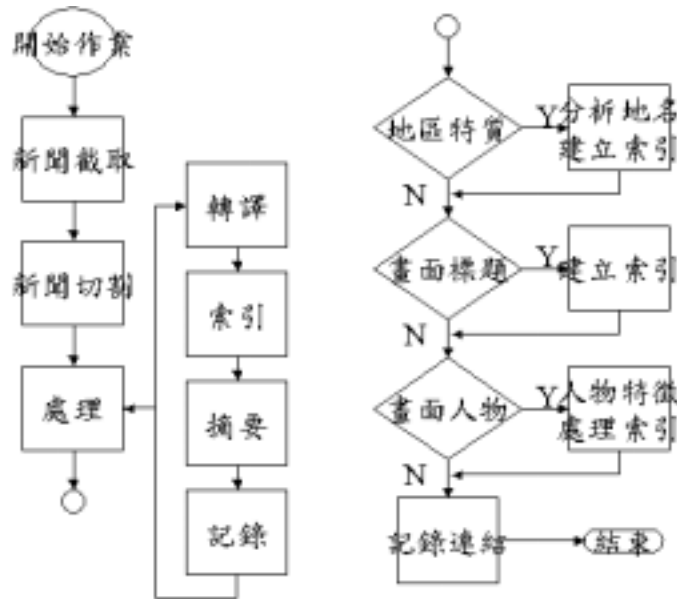
Informedia 在新聞的詮釋資料部份，除了基本資料的註錄以及內容轉錄外，為了加強新聞資料的檢索 Informedia 為每一則新聞加註內容標目。為了加速資料的處理速度、降低資訊處理成本，Informedia 計畫所要達到的目標是建立一套資訊索引及摘要自動化的機制，但同時也必須考慮其正確性。在標題方面 Informedia 所採取的方式是先建立一個指標資料庫，其中包括數則不同的電視新聞資料，這部份資料以專業人工分析的方式，為每一則新聞訂定主題標目。當指標資料庫的主題分析完成後，系統分析這些新聞譯文內容的詞彙，從中找出詞彙出現頻率特質與主題標目間的關係。更新的新聞資料加入資料庫後，系統對新聞譯文進行詞彙分析，從其中找出詞彙出現頻率的特質，再將這些特質與先導資料庫的新聞詞彙互相比對，找出近似值或者是相似者，給予該則新聞相同的主題標目，再以取樣方式驗證。

六、地理資訊

Informedia 對於新聞資料的呈現與解讀除了以每一則播報新聞為主，也希望能夠從時空的角度來分析新聞資訊。就時間來說，每一則新聞的截取與處理可自動加上時間的因子；地理部份，則系統需就內容進一步篩選處理。目前處理的方式是將轉譯資料內容與地理名詞索引典比對，建立地理名詞連結關係。

七、詮釋資料

分析每則新聞的資料元素後，系統以新聞為單元，建立每一則新聞的詮釋資料，包括題名、新聞內容摘要、相關資料連結路徑 ... 等，以該詮釋資料為中心，以物件連結的方式連結與該則新聞有關的各個檔案物件，包括內容語音全文、影像區段、與內容相關的地理 ... 等。



【圖 3：影像資料分析與詮釋資料的建立流程】

Informedia 使用環境

目前 Informedia 資料的檢索是透過由計畫小組所建立的視窗主從式軟體介面進行資料的檢索，未來將發展以超文件通訊協定為基礎的檢索介面。資料的檢索以簡易檢索加上布林邏輯運算為主。

一、檢索介面

Informedia 所提供的資訊檢索介面，主要是以輸入檢索字串為主，系統將索輸入的字串轉換為查詢詞彙，從系統資料庫中查詢相符的新聞譯稿，依詞彙出現的頻率計算各則新聞稿件的相關度，依照相關度顯現找到的資料。



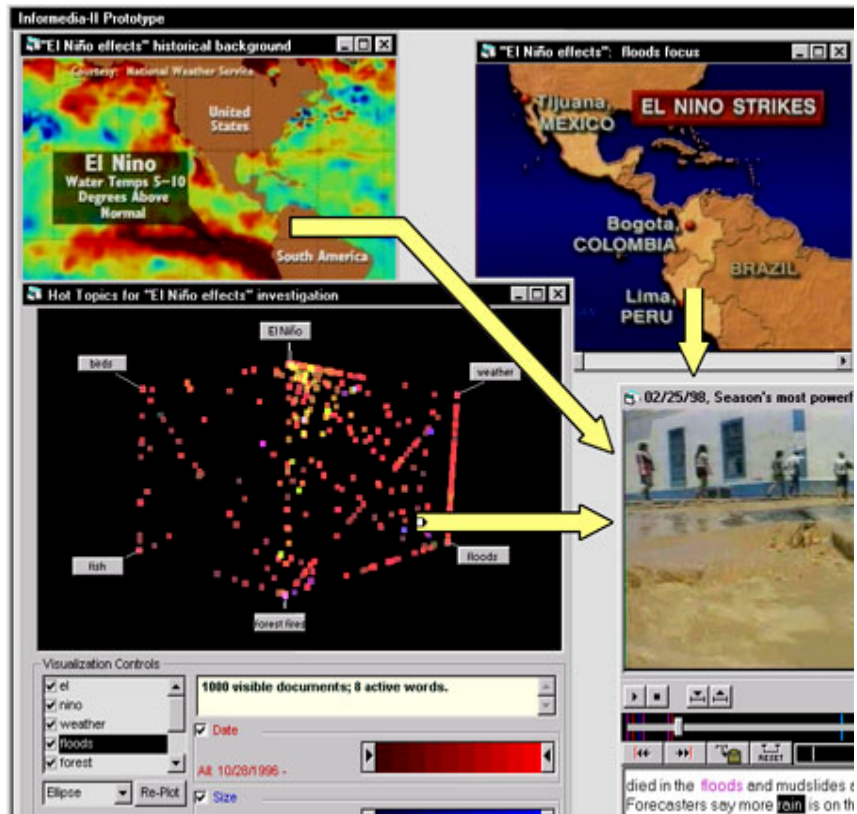
【圖 4：Informedia 新聞檢索及結果顯現】

二、視覺顯現

利用目前資訊技術，為便利資料的利用，Informedia 以圖像方式協助檢索結果的顯現。過去的檢索系統多是文字化的方式處理系統介面，Informedia 系統所處理的為影像資料，在使用介面設計以及輔助工具多也以圖像方式呈現，希望透過視覺化的方式讓系統的利用能夠更具親和力。其中設計包括利用顏色表現資料的閱讀程度；以圖型相對位置，經由遠近方式表示資料內容的相關程度；或是以地圖方式選擇地區的新聞資料。

三、新聞摘要

系統顯現檢索結果是以播報新聞的第一個畫面加上新聞摘要為主，若必要，檢索者可以播放 30 秒的新聞摘要，以確定所找到的是否為所需要的新聞。在資料檢索上以圖像使檢索介面較具親和力。



【圖 5：Informedia 資料檢索結果】

四、同質新聞

系統在處理每一則新聞時會分析其內容地區及時間特質，同時將每一則新聞以自動內容分析方式找出每則新聞的內容主題。利用這些內容特質，系統可以檢索回饋的檢索方式提供類似資料的檢索。

Informedia 計畫的運作目的主要是技術開發及運用，雖然取得新聞資料提供單位的同意可使用其資料，但是所完成的技術與產品尚需某一種商業化的機制，使得產品得以提供一般檢索者利用。目前這正是 Informedia 在進行的工作項目之一。以 Informedia 計畫所完成的技術來看，從每日新聞影像的截取、儲存資料；新聞語音轉譯、全文索引；地區與時間分析；新聞內容摘要 ... 等均可透過系統程式以自動設定方式進行。

Informedia 未來發展方向

目前 Informedia 的發展已獲得相當成果，研究結果的實務化、市場化以及商

業化也在進行中。與工作小組對話過程中，對於下一階段或未來的發展方向有這樣的建議與想法，包括評估模式的建立、新聞事件追蹤、新聞的綜合整理、大規模的使用者測試、大規模的使用者測試等。

一、評估模式的建立

Informedia 計畫所強調的是資訊處理自動化模式的建立，一個引起討論的議題就是這樣的模式在資訊處理以及檢索成效的精準度方面是否經得起各種不同檢索的檢驗。究竟語音辨識、自動索引、內容分析等各個階段的成品，其中可能的誤差對於資訊檢索可能產生什麼樣的影響？這個部份是需要建立一套檢驗評估機制，才不至發生檢索上的誤判。

二、同一事件新聞的追蹤

目前系統對新聞的處理是以每一則新聞為準，雖然在地區、時間以及主題上均進行一定的分析，但是在同一事件的相關新聞上卻未進一步做縱向的連結。因此以目前系統的設計來看，無法將同一事件新聞清楚的連貫，獨立的呈現出來。如此在新聞資料搜尋上，檢索者需要進一步的篩選判斷相關的新聞，同時對於新聞內容也無法由系統就一事件的前後發展做一整體、有效的綜合整理。

三、新聞的綜合整理

如前面所述，Informedia 的每一則新聞是獨立處理。因此單從詞彙的分析，目前 Informedia 在資料的重新組織與綜合整理的功能還相當有限。仍是以新聞檢索為系統主要功能，尚無法綜合的組織整理，以不同的方式重新呈現新聞事件。若 Informedia 未來能夠朝此方向發展，可使得新聞記錄可以不斷的增長，新認知可以寫回資料庫，達到所謂的 Glowing Metadata，為資料加值。

四、大規模的使用者測試

目前 Informedia 為系統運作成效進行了幾種不同的使用者使用行為研究，以了解系統運作的成效及需要修改的部份。由於系統資料版權的問題，Informedia 的資料庫只能夠在一個較為封閉的環境中進行利用，無法開放給各種不同學科背景，具有不同使用目的使用者進行系統測試。現階段系統測試主要是以系統設計者，CMU 校內的師生及部份高中生。目前 Informedia 正在嘗試以 XML 方式設計

系統資料及使用介面，希望能夠在版權的許可下，透過網路提供給更多使用者進行系統的功能測試。

五、不同語文嘗試

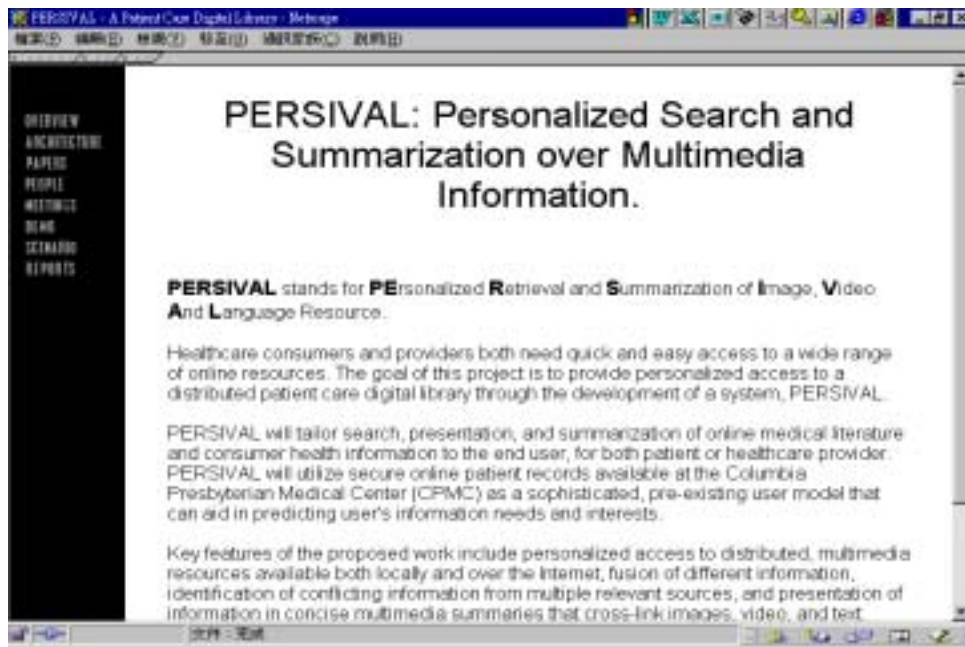
Informedia 計畫現階段已完成英語以及西語的新聞資料轉譯，將再嘗試不同語文的計畫。即將開始合作的計畫包括日文、韓文及賽爾維亞語。Informedia 並不嘗試語文的翻譯，不過在檢索部份可將輸入的英文檢索問題依字詞翻譯為西文，或反之，將英翻譯為西文，再依照字詞進行全文比對，找出可能符合檢索需求的新聞資料。

目前國內新聞資料的處理模式主要是以文字化的方式呈現相關的新聞資訊，對於播報新聞多是以建立相關書目資料的方式檢索錄影訊息，Carnegie Mellon 大學的影像資料自動摘要及索引計畫嘗試以不同方式處理、檢索每日接觸的新聞資料。從對新聞語音資料的轉譯處理、影像資料的分析模式、相關度的視覺化顯現、檢索新聞的摘要處理以及系統訊息的圖像化。Informedia 的研究提供了一個在影像資料處理上可以參考、思考的方向。除了目前 Informedia 的計畫，Carnegie Mellon 大學在影像資料處理方面的相關研究也可做為不同研究領域合作的參考。如目前 CMU 工作小組嘗試以日記式的影像記錄方式記錄每人每天活動，包括行經路線及各種情境的語音反應記錄等，以不同角度進行社會活動分析的研究。希望可與社會學家共同研究、探討不同的田野研究方法，將這樣的研究法運用於社會發展、心理研究上。

哥倫比亞大學個人化多媒體資訊檢索及摘要系統

Personalized Search and Summarization over Multimedia Information

有鑑於醫療服務的接受者及提供者都需要能夠快速及輕鬆的取得各種不同的資源，哥倫比亞大學的個人化多媒體資訊檢索及摘要系統(Personalized Search and Summarization over Multimedia Information, PERSIVAL)的計畫目標是建立一個可以提供個人化的檢索，透過數位化圖書館將醫療資訊傳遞需求者。PERSIVAL計畫在數位圖書館第二期計畫支援下，於1999年開始，預計於2004年完成。計畫目標是希望建立一個可以根據不同需求、對醫學知識可掌握程度的深淺，就檢索者希望查詢的資料，從系統資料庫篩選適合資料的醫學資訊系統。



【圖 6：PERSIVAL 計畫網站首頁²】

PERSIVAL 計畫是以哥倫比亞大學醫學中心(Columbia Presbyterian Medical Center, CPMC)的醫療記錄資料庫為基礎，就患者的過去醫療記錄預測檢索者可能的資訊需求。PERSIVAL 系統希望能夠達到的不單是根據檢索者需求，將資料提供給檢索者，同時還能夠整合網際網路上的資訊，完整的提供給讀者。尤其是典藏資料及網路多媒體資料，包括影像、動畫及文字資訊，依照檢索需求及檢索者

² <http://http://www.cs.columbia.edu/diglib/PERSIVAL/>

重新包裝檢索到的資訊。對醫療者來說，PERSIVAL 希望能整合患者醫療記錄以及系統資料庫所儲存的資料，提供醫生診斷方面的參考資訊、加速醫療效率、降低醫療風險；對患者則是透過系統對資訊的整理，使其對醫療情況能夠有更清楚的了解。

醫療環境中不同的角色對於醫療資訊有不同的需求。患者及其家屬需要醫療資訊能夠幫助其對其狀況有更好的了解；對醫生來說，其需要相關的醫療資訊，協助其對其所照顧的病人提供較好的醫療服務。一方面能夠掌握充分的資訊，避免誤診，另一方面在充分的資訊上可以更好的方法維持健康狀況。PERSIVAL 系統的建立是為了立一個可依據檢索者的背景資料、個人狀況及資訊需求，組織醫療資訊，以不同的型式提供資訊檢索者。PERSIVAL 系統希望建立一套資訊自動判斷機制，從資訊需求者的背景類別、資料型式、內容特質等，組織各種不同的資訊資源，滿足不同樣的資訊需求。為了達到上述的目的研究小組在系統設計上必須建立幾個不同的部份，包括檢索介面、檢索者類型判斷、資料粹取、以及重組機制。

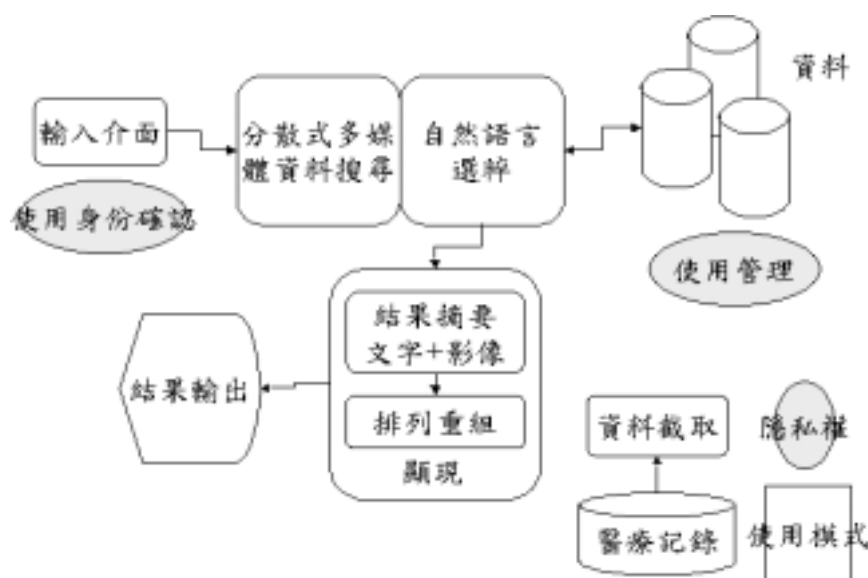
- 檢索部份必須包括一個檢索問題輸入介面，透過介面所獲得與檢索者相關的資訊，從系統醫療資料庫中搜尋檢索者的醫療記錄。
- 以檢索出的醫療記錄，配合輸入檢索指令，檢索資料庫中的相關資訊，比對資料型態、資料內容以及閱讀程度。
- 組織、摘要、呈現所檢運索出的各類型資料，整合重覆資訊，依據檢索者慣用的詞彙，重新整理所檢索到的資訊。
- 發展自動摘要，圖像呈現以及說明檢索結果技術。

PERSIVAL 系統目前是與哥倫比亞長老教會醫療中心(Columbia Presbyterian Medical Center, CPMC)的心臟病醫療部門及糖尿病醫療部門共同合作，同時與 Milstein 圖書館服務的讀者分別進行館內以及在家資訊檢索測試，以進一步了解評估系統設計與執行成果。計畫後期將進行影響評估，了解系統對於醫療者及患者的資訊行為有何影響，同時使用者對系統滿意狀況。

系統運作架構

系統架構分為幾個部份，使用者介面、資料輸出入處理、資料粹取。使用者

介面以及資料截取部份由於涉及資料版權以及醫療記錄隱私權的考慮，必須對使用者權限有所控制，對資料隱密性有所保護。



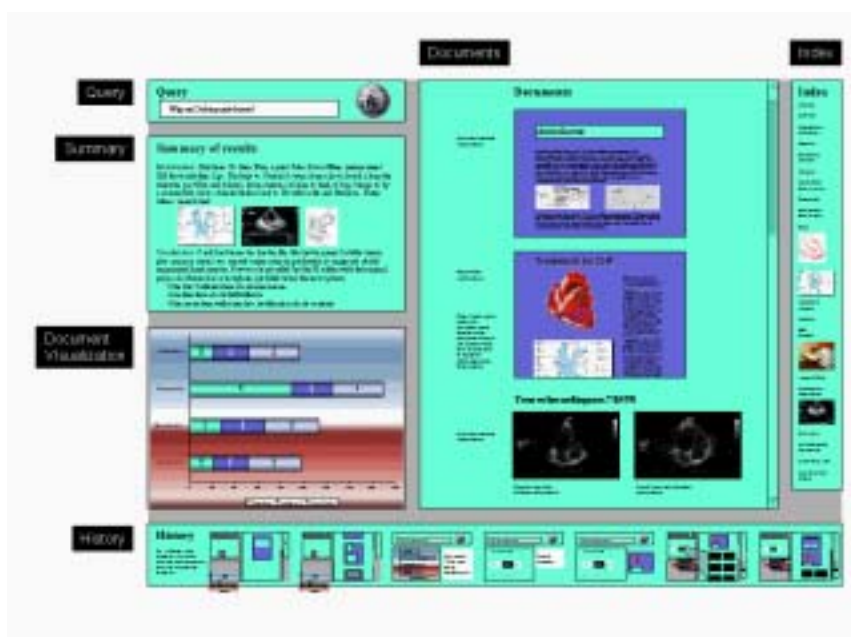
【圖 7：PERSIVAL 系統架構】

系統開始處理檢索者輸入的檢索策略之前，首先要確定的是檢索者的身份，包括其型態類別，醫療行為的提供者或是接受者；對於醫療資訊的掌握程度；預定醫療行為的接受者……。身份及背景資料確定後，系統就資料庫中所收集的文件、網路資源，找出相關的資訊。依前述適用對象資訊的分析，在使用權管理機制的控制下從醫療記錄中找出可能的醫療記錄。綜合所查到的檔案文件以及醫療歷史，按檢索者的適合程度，重新組織、整合、顯現檢索結果。在這樣的架構下，PERSIVAL 提供一個包含檢索結果摘要、相關文件與影像全文、解釋說明、建議檢索詞彙以及檢索歷史的使用者介面。

PERSIVAL 系統特色及運作架構

目前 PERSIVAL 還在建構當中，以下將就第一階段已完成的部份及相關後續計畫做一介紹。PERSIVAL 計畫的工作成員依照 PERSIVAL 計畫需完成的目標設定各個不同工作任務，分成不同的工作小組。主要的領導團隊由 5 位教授組成，其來自不同的研究背景，包括電腦、電機、醫學及資訊組織檢索部門。PERSIVAL 計畫自 1999 年開始，預定 2004 年完成。第一階段計畫目標分為幾個部份，包括分散式的資訊檢索機制、資訊資源的綜合組織、檢索介面與資源呈現、臨床資訊

架構、資料內容與圖書館館藏整合、系統功能成效評估。



【圖 8：PERSIVAL 操作介面】

一、分散式資訊檢索機制

分散式資料檢索機制的建立主要目標是要發展一套搜尋機制檢索一般搜尋無法直接查詢的網路資料庫。越來越多的資料庫是以經由網際網路，以超連結文件為使用介面提供資訊檢索者檢索，但這些資料庫的資料多半無法透過搜尋引擎查詢。PERSIVAL 在資料檢索上第一個工作目標是要能夠分析網路上各個資料庫來源類型，同時透過單一介面系統能夠由當中合適搜尋問題的資料庫查出符合檢索問題語意的資料，並能夠就檢索端的特質進行資料內容的分析與分類。除了文獻類型資料庫，醫療資訊中有不少圖型式的資料。以先導計畫合作的心臟科來說，除了文獻式的資料之外，臨床醫療資訊包括了不少心電圖、器官活動影像等型式的資料，這些資料在臨床醫療資訊是相當重要的。PERSIVAL 的第一個工作小組所要處理的除了文字資料外，也必須發展一套機制能夠處理索引影像資料，甚至進一步解讀這些資料。第一階段工作小組的研究重心是在影像資料區段的分析。

除可就特定資料進行檢索查詢，在結果分析方面，系統在顯現檢索結果前會依照檢索端做較詳盡的結果相關分析，也就是要達到不同的檢索者查詢相同的資料所獲得的資訊結果呈現是依檢索者特質而有所不同。目前完成的資料分析是期刊資料，獲得檢索結果後系統先行比對由醫療記錄中所獲得的資料，進行結果相

關度的計算，重新給予相關排序。

分散式資料檢索的任務中心是要整合各種不同相關資料源，進行跨系統的資訊檢索，透過 PERSIVAL 檢索不同的資訊系統獲得所需要的資訊，正確快速的處理連結心電圖、影像資料；連結文字資料與相關影像資料；整合網路上其他與心臟研究相關的資訊。

二、資訊資源的綜合組織

不同類型的檢索者對醫療相關資訊的解讀程度各有不同，不同的患者族群對於相關的醫療資訊需求也有不同，PERSIVAL 第二個工作任務是建立醫療資訊的整合機制，將所研究文獻與工具書類型的資料做一連結整合，使系統能夠依照檢索者需求提供不同層次的資訊解讀，必要時可以進行詞彙替換，以不同的語彙描述相同的內容。這部份的研究工作小組也是從期刊文獻開始。首要的系統功能是要能夠辨識期刊文獻內容所討論的醫療患者類型，再依照檢索者類型重新組織所檢索到的資訊。PERSIVAL 在第一階段的研究發展出了 DEFINDER (Definition Finder)的運作機制，透過與詞彙控制工具的連結，查詢文獻名詞的定義，必要時可對文獻內容提出解釋 在步驟上要能夠解讀文獻 找出重要概念詞並予以歸類，在與線上工具書連線，整合不同的條件，重新呈現資料。

三、檢索介面

主要是設計一個共通的檢索介面，處理不同資訊資源的查詢，同時建立、分析資訊檢索模式。這個部份的目標是建立使用者介面，提供一個具親和力，且可以提供高效率的檢索操作介面。在第一階段 PERSIVAL 將設計的先導系統與合作醫院的檢索系統整合，測試醫療資訊與參考資料文獻整合處理的狀況。除了以文字方式輸入檢索策略，PERSIVAL 也發展以語音方式輸入查詢指令。

四、醫療資訊架構的建立

醫療資訊工作小組在資訊架構部份的努力是探討如何將醫療記錄運用於醫學資訊的檢索，藉由不同資訊的整合對於資訊檢索有什麼樣的影響及幫助。希望能夠達到的目標是能夠提高資訊系統檢索的效率，使所獲得的資訊能夠更貼切的協助檢索者所面臨的情況。

工作小組首先進行的是文獻的分析，就醫療文獻探討議題的特性分析所使用的資料庫收錄文獻，分析是否可就其運用面向找出文獻的特質，並將文獻予以歸類。這是從用途論來思考，分析文獻內容的相關度。另一部份的工作是從醫療記錄中粹取可以與檢索指令的資訊。醫療記錄與醫學文獻間的語彙連結則是透過資源資訊整合組織的研究成果，透過醫療工具書，取得詞彙的一致性以利重組檢索策略。

五、資訊內容與圖書館

圖書館提供先導測試系統所需要的資訊、分析醫療資訊檢索的特質、蒐集醫療相關資訊資源。先導系統的資料以期刊文獻為主，主題則是以心臟研究相關的期刊為中心。除了建立先導系統所需要的資料之外，圖書館部份亦提供醫療資訊檢索系統的資訊搜尋模式，做為 PERSIVAL 之後系統設計的參考。除了圖書館本身所提供的資訊，網路資源也是 PERSIVAL 系統期望納入資料庫中的資訊來源。圖書館在第一年及後續工作將持續蒐集網路上沒有版權考量、適用資訊整合入系統中。

六、系統評估

系統評估小組著重的就是針對人機介面進行研究探討，同時也深入研究資訊需求，主要是讓系統建立者對檢索者行為有所了解，能夠透過檢索技巧的分析，資訊行為的了解，加強 PERSIVAL 資訊服務的功能。

PERSIVAL 計畫雖離完成還有一段時間上的差距，其中所希望提供資訊功能的技術還有待開發、整合，但從計畫的推動可以看到其中跨學科合作的必要性及互補性對於一個數位圖書館建立的正面影響。藉由跨學科的合作，能夠充分的整合各專業的特長，在發展考量上能夠更完整。除了在專業背景上的整合，資訊資源的整合也是數位圖書館在發展上必須要融入的概念。所提供的資訊從單一面向、獨立的單元，轉換成一組織、有系統，彼此相關連的資訊。而使用者導向的發展也是一個必然的趨勢，其中包含了考慮個人檢索需求的差異性、個人化的服務、以及親和力的使用設計。

紐約公共圖書館 - 數位圖書館館藏

紐約州公共圖書館的各分館所收藏的館藏中或多或少都收錄了一些珍貴、不易得的資料，其中可能包括手稿、唯一版本或是當地獨有的資料。這些資料在使用典藏上可能面臨不同的需求，部份資料可能需要有一套好的典藏機制，部份資料則可能有其使用上特殊的需求，在時空上面臨不同的要求。也有一些資料無法以一般圖書館組織資訊的方式有效地，或在合理成本範圍內予以處理。在各種不同的需求情況下，相當比例的資訊無法獲得充分利用，或是說需求者無法有一個有效的管道獲取這些資訊。資訊科技以及網路技術的運用發展為面臨如此情況的館藏，提供另一種處理機制。在使用及典藏的需求下，紐約公共圖書館展開相關的數位圖書館計畫。計畫主要目標就是將上述資料予以數位化，透過一套具檢索效益的方法組織數位化資訊，經由網路機制提供更多使用者可以從不同地點，以不同方式檢索利用這些在過去使用環境無法容易取得使用的資訊。位於紐約市的紐約公共圖書館於 1996 年開始紐約公共圖書館數位館藏建置的計畫，建立一個數位圖書館，主要收錄圖書館特有，但不易為一般使用者所取得的資料。這個圖書館所收藏的資料與現有圖書館所收藏的資料有別，數位圖書館計畫主要是組織圖書館特藏資料，以有效的方式提供使用。



【圖 9：紐約公共圖書館數位圖書館首頁³】

³ <http://digital.nypl.org>

數位圖書館架構

一、工作小組

為了完整規畫數位圖書館的建置，圖書館特定設立了一個專責小組，Digital Library Program，負責相關業務，由 Michael Alexander 領導。原計畫規畫該小組為 20 人力的工作小組，但至筆者訪問時人力資源 12 位工作成員，未來將在原計畫的支援下增加人員。目前計畫成員背景各異，包括資料內容專家，主要任務是分析資料的主題內容；圖書資訊方面的研究人員，主要是以研究資訊組織、分析使用者資訊行為；系統人員，專責架構系統，完成計畫所希望的目標；影像處理專家，數位館藏將包括相當比例的像片資料，因此有專責影像處理人員。

二、詮釋資料

目前數位圖書館資料分為幾個部份，包括圖書館現藏手稿及檔案的詮釋資料、圖書館所收錄的照片資料影像、作家作品全文記錄等。資料是以標準通用標示語言(Standard General Markup Language, SGML)標記，計畫改為可擴充標示語言(Extensible Markup Language, XML)。使用者均可透過網頁瀏覽器檢索、使用館藏資料。詮釋資料格式部份則目前是以都柏林核心集(Dublin Core Set, DC)為主，未來會擴增納入開放式典藏資訊系統(Open Archival Information System, OAIS)標準。工作小組對資料的著錄處理方式是採以遵循現有且具國際標準發展的格式為主。主要目的考量是未來合作、資料交換的格式一致性的需求；另一方面可架構在過去的經驗，考量特殊需求，擴充現有格式，可降低發展人力需求。在書目詮釋資料部份，所參照的標準為 DC；內容將以 OAIS 所建立的標準為著錄依據。

三、圖書館現有館藏

(一) 特殊館藏詮釋記錄

紐約公共圖書館收藏紐約地區各種人文、藝術表演及當地發展文獻與影音資料。目前數位圖書館所建立的資料為詮釋資料，主要是提供檢索者一個方式檢索圖書館館藏資料。圖書館現藏 7 種不同的館藏詮釋資料，包括 [Humanities Manuscripts & Archives Division Finding Aids](#)、[Billy Rose Theatre Collection Finding Aids ...](#) 等。



【圖 10：數位館藏 – 特藏資料記錄】

(二) 影像資料

影像資料目前是以像片為主，收錄 5 種不同像片館藏資料，主要是紐約地區的歷史、發展圖像記錄。收錄的像片資料包括 Berenice Abbott 的 Changing New York, 1935-1938、Lewis Wickes Hine 的 Construction of the Empire State Building, 1930-1931 ... 等。

(三) 文學作品

全文文學作品目前收錄的為美國 19 世紀非裔女性作家的作品，包括 42 位作家，52 種不同型式的作品。

第一部份資料是依據所訂定的詮釋資料格式，由工作小組逐一就各個館藏內容，內容建立其書目資料，方便使用者檢索。目前透過數位圖書館僅能獲得其詮釋資料，至於完整內容的數位化目前工作小組正在規畫中。影像資料的處理，則是由專責人員先將資料予以攝影，建立數位影像資料。上載後由該專責人員建立其相對應的詮釋記錄，便於資料的控管以及檢索。文學作品資料是以 SGML 標準儲存，資料則是依 Text Encoding Initiative (TEI) 格式標註。

數位圖書館資料的檢索主要是以瀏覽為基礎，每一類館藏資料表列館藏內容，另部份館藏資料提供檢索功能。現階段的發展重心是以資料的數位化及典藏為主。使用上，偏重於提供資料的影像記錄；進一步的內容分析、列比功能則尚未納入圖書館功能中。

麻省理工學院學位論文數位圖書館 - M.I.T. Theses Digital Library

麻省理工學院的學位論文數位圖書館是在該校圖書館與資訊系統工作小組共同合作建立維護的數位圖書館計劃。計劃的前身是圖書館文獻服務部門的數位典藏計劃，之後在與該校資訊系統工作小組合作下將圖書館內部資料典藏計劃擴展為線上數位資料傳遞服務。據其負責人表示，在計劃之初並非以建立一個數位圖書館為導向，僅是希望利用既有的技術以及資料架構，提供搜尋學位論文使用者一個更好的資料獲取方式。麻省理工學院對學位論文的處理向來與其他美國主要大學不同。美國大部分學校的學位論文在學生繳交給學校後，若需要取得資料的全文多是透過 UMI 公司取得，而麻省理工學院研究所的畢業學位論文則是直接向該校購買，不透過第三者。由於該校的研究對學術、業界發展有一定的影響力，因此該校的畢業論文一直以來也是被高利用的資料之一，不單是校內師生的利用，校外對於學位論文也有一定的需求量。在這樣的環境下，麻省理學院一直以來均有一個對外提供論文資料傳遞的機制，而負責這個工作的是該校圖書館的文獻傳遞服務部門。為了有效典藏資料，圖書館將所收錄的論文逐一以微縮技術處理，每一篇學位論文透過照相縮影技術，製作成微縮片。若有讀者需要這樣資料時，透過該校目錄或是 UMI 資料庫的查詢，可從圖書館的文獻服務部門取得學位論文資料。圖書館的文獻傳遞部門在獲得索取要求後，將資料由微縮片在印製成一般可讀的文件。當資訊技術有所改進之後，在文件生產過程中，圖書館在印製論文過程中改以掃描方式將微縮片以掃描方式處理，產生 600dpi 的影像資料，再由印表機印出。

隨資訊技術的改變，在操作上為了便利內部作業，該部門將掃描後的數位資料予以存檔，以方便後續的再利用。隨著外在對資料的需求，圖書館逐漸建立一批館內作業用的數位館藏資料。這個數位館藏在建立之初主要是便利館內在作業上的便利，降低重複掃描的工作。由於用途的單純性，所建立的資料純粹是以便利資料再生為目的，在資料的檢索、規模上以簡單為主。在這樣的作業機制背景下，加上網路技術的發展，該部門開始思考將原有的數位館藏計劃擴大，建立一個收錄麻省理工學院學位論文的數位圖書館。這個圖書館除了收錄過去該校畢業學生所寫的論文外，也計畫提供一個可由線上取得論文全文資料的機制，並與圖書館的館藏目錄做一整合，充分掌握圖書館可提供的各種型式的論文館藏。為了

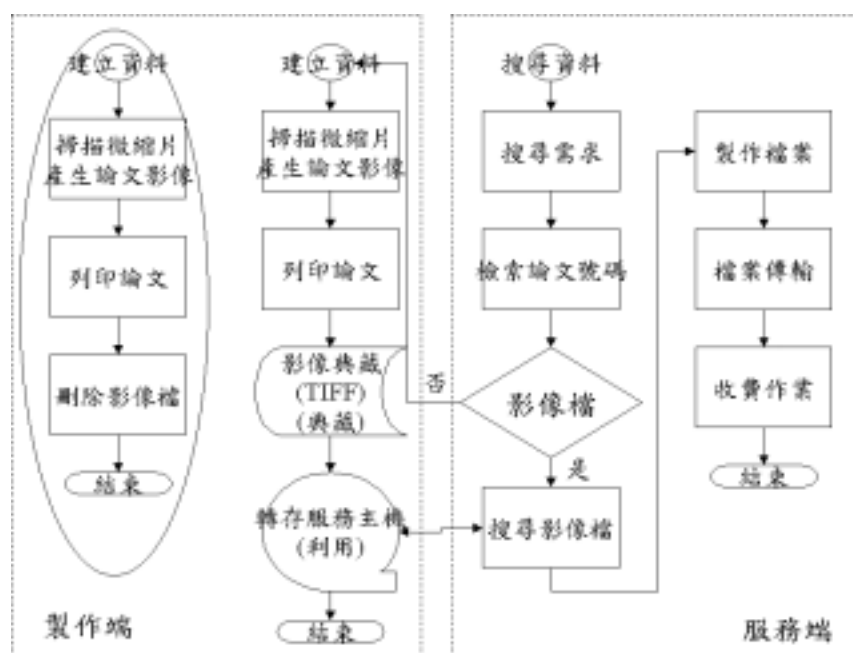
便利資料收集，第二階段的發展，該數位圖書館將提供一個學生線上上載論文資料的機制，在畢業前直接將論文資料加入該數位圖書館，完成畢業程序，同時圖書館也可典藏該資料。

論文數位圖書館

麻省理工學院圖書館的論文服務原先的作法是在印製紙本論文後，將電子檔案刪除；開始檔案典藏作業後，作業流程將原應被刪除的檔案轉變為典藏的資料。以下就數位圖書館作業流程及架構做一介紹。

一、作業流程

系統資料的建立是依需求而定(Create on Request)。MIT 學位論文資料可以回溯至 1880 年，但資料的利用集中於 1980 年以後所出版的論文資料。因此在成本考量以及無典藏保存之虞，MIT 論文數位圖書館的進行方式是收錄每年新增的畢業論文，回溯資料則是依需求決定新增與否。使用介面部份僅可查詢到目前數位圖書館典藏的資料，若是尚未建檔的資訊，使用者透過線上訂購的方式索取所需的資料。工作小組在獲得需求後查詢相關記錄，調閱存檔的微縮資料，進行數位化工作，完成後提供資料。一旦資料進入數位圖書館館藏後，資料的處理由系統執行。

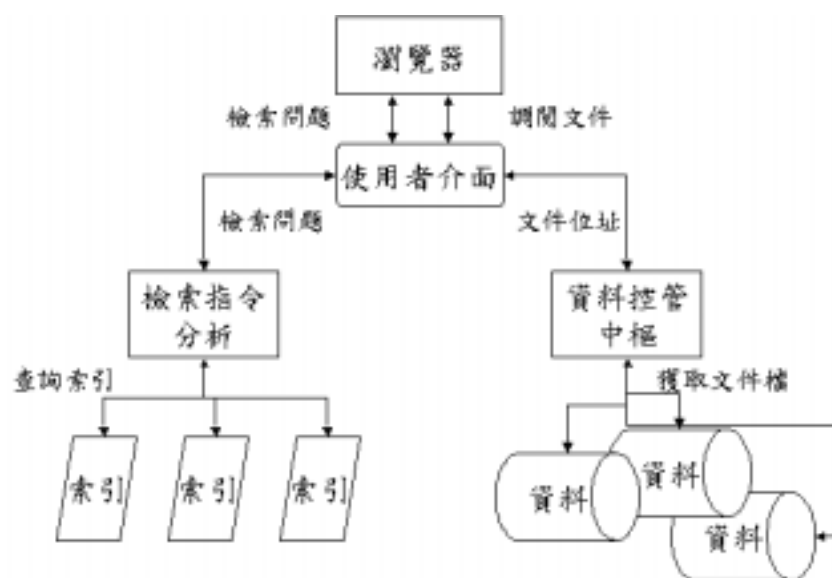


【圖 11：論文資料回溯建檔與資料搜尋作業流程】

為取得與圖書館館藏目錄一致的記錄，每一年圖書館收到新的論文資料第一步工作是建立一個館藏書目記錄，同時上載資料到 OCLC 資料庫取得 OCLC 號碼。論文資料另以兩種不同型式處理，一為製成微縮資料，主要是做為永久典藏用。另一則是以影像方式處理，成為數位圖書館的館藏資料之一。若是回溯資料。則將微縮資料經由掃描製成影像檔，列印與存檔。

二、技術與詮釋資料

工作小組在規畫發展藍圖時，已經具有參與網路化電腦技術參考圖書館 (Networked Computer Science Technical Reference Library, NCSTRL) 計畫，處理技術報告的經驗，在評估成本、效率、使用者熟悉度 ... 等因素下，MIT 學位論文數位圖書館決定以參與 NCSTRL 計畫的技術建立論文數位圖書館，包括硬體、作業系統以及應用介面，使用的作業軟體為 DIENST。影像部份則是透過由 MIT 所發展的軟體 DSpace 處理。



【圖 12：系統架構圖】

在論文數位化處理初期，資料的處理是依 OCLC 號碼儲存，是唯一的資料檢索值。若僅是內部典藏資料排序用，以這個方式處理資料不致有太多使用上的問題。但涉及資料的開放檢索需求以及資料交換，MIT 電子論文計畫的工作小組的工作任務之一是必須為每一篇論文建立詮釋資料，用以說明論文背景資料及作為檢索依據。由於已有處理技術報告的經驗，加上圖書館本身有建立論文詮釋資料

的經驗，因此在格式的訂定以及資料的建立上工作小組不需要從零開始，以 RFC1807 格式為基礎建立論文的詮釋資料。同時圖書館館藏目錄已針對每一篇學位論文建立了機讀編目格式，計畫可以從書目資料庫轉出所需要的欄位資料。工作小組的方式是以技術報告所使用的詮釋資料格式為基礎，考量論文資料特性，以及圖書館對論文資料的處理模式，訂定數位圖書館資料的格式。為便利資料的著錄，工作小組也訂定了著錄指引、資料來源，以及部份欄位是採固定欄位值的方式處理。

【表 1：論文資訊詮釋資料格式】

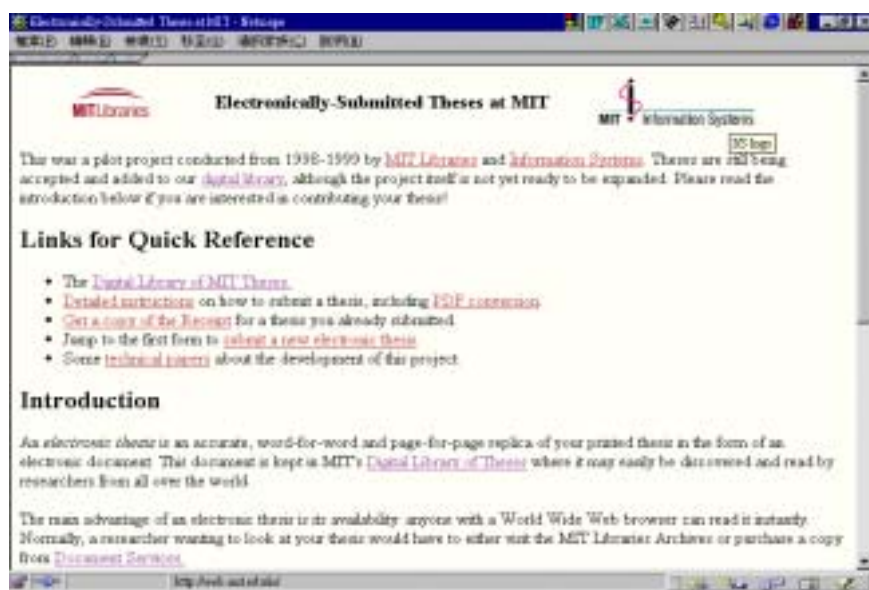
欄位名稱	說名	必備	註
Title	論文名稱	Y	
Author	論文作者	Y	
Date(R)	數位圖書館收錄時間		
Degree	學位類別，M.S.、Sc.D.、Ph.D		
Document Type	論文類別	Y	
Institution	所屬單位，目前僅有一單位，未來考慮擴增	Y	
Department	系所	Y	
Supervisor	指導教授	Y	
Keywords	關鍵字		
Abstract	摘要		
Availability	開放控制碼	Y	
Page count	頁數		
Notes	其他相關訊息		
Grant Information	支援經費		
Date of upload	上載時間	Y	E ⁴
Email	作者電子郵件地址		E
Authentication	資料上載者	Y	E
ID/URN	系統文件編號以及資料位址(URN/URL/handle)	Y	E

論文繳交作業電子化

除了以原有的館藏資料建立數位論文圖書館，MIT 自 1998 年在圖書館及 MIT 資訊系統人員的支援下開始電子版論文繳交研究計畫。計畫初期是在選擇的科系中，小規模的進行先導計畫，主要是以原有作業流程為基礎，分析評估電子傳遞模式的可行性與進行模式。計畫目標是建立一個網頁，透過網頁所提供的資訊及

⁴ 數位圖書館資料專用。

軟體，學生可將論文以可攜式文件格式(Portable Document Format , PDF)將論文資料繳交給學校，同時納入數位圖書館館藏中。這個先導計畫有兩個先決條件是計畫人員必須要達到的，一是在資料準備的步驟方面必須達到簡易、方便的操作要求；另一是在校務上需要取得學校研究所政策委員會及各系所的同意與認可。參與先導計畫的系所包括 MIT 的認知心理學、化工以及電機電腦等三個研究所，而參加的學位論文以博士論文為主，估計有 35 篇左右。先導計畫在 1999 年完成。



【圖 13：論文繳交作業說明】

MIT 學位論文數位圖書館目前收錄 4,000 筆論文全文資料，論文出版年代主要是以集中近 20 年，每周約新增 15 篇論文，網站的連結次數每週為 70,000 次，資料的輸出量每週平均 600MB。系統提供以作者、論文題名查詢，也可依作者及年代瀏覽圖書館所典藏的資料。從先導計畫的成果來看，論文繳交作業數位化的推行問題不在於技術問題，而是政策考量。由於 MIT 學校環境的基本條件，學生對於資訊技術的運用並不陌生。雖然在使用軟體上，有部份資料需透過 PDF 以外的方式處理，但是在作業流程上這些問題是可以被克服的。政策上的爭議則是 MIT 在計畫後續的發展方面需要謹慎考量規畫的，一是在學校是否能夠完全接受數位典藏的模式，另一是授權確認機制的建立，使得資料具有完全的效力。再者許多論文內容是具有專利以及未來發展潛能，如何在開放使用，又不損及學生在專利、著作權益的原則下，建立，資訊傳輸機制。這些是先導計畫完成後未來 MIT 在學位論文數位計畫推行將遭遇到且需要面對解決的課題。

Tufts 大學的 Perseus 計畫



【圖 14：Perseus Digital Library 首頁⁵】

Tufts 大學的 Perseus 計畫是以建立可以提供研究古代文獻所需資訊的數位圖書館為目標。Perseus 計畫之初是以建立數位化古典希臘文學文獻為主，主要是建立一個檢索介面提供一個容易操作的環境，取得研究古希臘文所需要的文獻資訊以及參考資料。計畫主持人 Gregory Crane 從 1982 年開始著手進行以電腦技術輔助學術研究的計畫，當時 Crane 教授取得一些與其研究相關的電子資料，在獲得這些電子化資料後，Crane 教授著手建立一個可以全文檢索方式查詢資料的線上系統，而這個線上系統的發展成就了現在的 Perseus 計畫。1985 年 Crane 教授取得經費可以將原有的全文檢索系統擴大為 10 年計畫，建立一個以收錄古希臘及古羅馬文獻的數位圖書館，以提供研究資訊為主。在計畫之初，Perseus 所建立的數位圖書館是以 CD-ROM 為儲存及傳輸媒體的數位圖書館，到網路以及超文件技術發展成熟後，Perseus 在 1994 年從 CD-ROM Based 轉為網路 Web Based 的數位圖書館。目前 Perseus 計畫主要的經費來源是由美國科學基金會以及人文學會支助，Tufts 大學及 Crane 教授所屬的文學院則是在硬體設備以及部份日常運作所需的資源。Perseus 目前除了原計畫中的古希臘及羅馬文獻方面的資料，目前也擴及其他資料，例如英國文學方面的資料以及研究相關的地理資料，同時 Perseus 研究小

⁵ <http://www.perseus.tufts.edu/>

組也與其他單位合作進行不同以語文需求為基礎的研究計畫。合作的方式是由 Crane 教授的研究群提供需要的經費，合作單位則是提供典藏物件及經費。

Perseus 計畫任務架構

Perseus 計畫包括資料的數位化、相關資訊的連結、使用介面的設計、以及使用績效評量等幾個部份。目前計畫主要重心是在前三者的工作推動。

一、資料的數位化

Perseus 資料的數位化若本文為英語，則可在影像化後透過光學字元辨識軟體即可處理資料的數位化。其中需要進一步修正的部份並不多，尤其是本身狀況不錯的文件，多半不需要後續處理。產生的檔案資料在透過 XML/SGML 編輯軟體標註，如題名、地名及人名，以利後續的檢索作業。目前 Perseus 並未針對數位化後的資料逐一查驗，而是以抽樣方式查核處理。同時仰賴資料的使用者將可能發生的錯誤狀況反應給系統管理者予以修正。從目前系統運作的情況來看，數位化的結果尚稱滿意。Perseus 所處理的資料中有不少資料原件是書籍型式，對於這一部份資料的處理，在考慮成本效益以及資料狀態的維護問題，Perseus 是以委外方式將此批資料交由專業的裝訂商資料予以拆裝，掃描處理。

Perseus 計畫在資料數位化方面所遭遇較大的問題是要處理非英文的資料。Perseus 主要是以建立古希臘及羅馬文研究所需資料為主，因此除了英文資料外，有一大部份的資料為古希臘文及拉丁文。這個部份資料的影像化沒有太多的問題，但是為了滿足其後的檢索功能，而需建立的文字型式資料則無法如同英文資料一般，可以透過 OCR 的辨識功能完成。前者無適當的軟體可利用，後者則雖有軟體，但是辨識效果不佳，需投入相當的後處理成本。因此評估後，Perseus 研究小組決定將這部份的數位化工作轉由外包的方式處理，同時將轉換為文字型式的資料予以標示註記，產生結構化的資料型式。

Perseus 在資料數位化方面所遭遇的主要問題是必須要建立一套可以辨識古希臘文及羅馬文的機制，同時所建立的資料在不同的使用介面上能夠被辨識，以利資料的流通。Perseus 是以 Beta 內碼處理文字資料，在 CD-ROM 及目前的 Web 階段為顧及使用者使用介面，除了資料建立及使用介面設計，系統需將資料內碼所需要的相關軟體包含在系統功能中。未來 Perseus 計畫考慮將資料儲存內碼改為 Unicode，提高通用性。資料架構是以 TEI 為詮釋資料格式，再視需要於以擴

充。目前 Perseus 可將 TEI 格式或非 TEI 格式資料轉換至系統中。

二、研究參考文獻的建立

Perseus 計畫目標是建立以提供研究資料為主軸的數位圖書館，因此除了古典文獻本身的數位資料，相關研究所需的參考工具書也是 Perseus 計畫數位資料相當重要的一部份。在考量、處理版權問題後，參考資料先經過數位化的處理，再依照前述文獻標註的結果，連結重要詞彙、概念的解釋資訊。

三、架構資料查詢介面

Perseus 資料查詢是以文獻時間軸為主，在每一個部份提供全文式的關鍵字檢索及資料類型及語文的設定；同時在操作過程中，可在時間軸上進行檢索資料的切換。對於文中內容部份的解釋，則視參考文獻的完整性，透過連結可獲取參考文獻對於概念的解釋資訊，或是館藏資料的切換。

四、研究小組及決策程序

Perseus 研究小組除了計畫主持人 Gregory Crane 教授外，還包括 8 位全職人員，主要是以具文學研究學科背景的學者為主，同時包括系統設計人員以及圖像資料處理專才。Crane 教授提到，理想上同時具有系統設計以及學科背景的研究成員是對計畫發展最有利的，但是若在兩者不可得兼情況下，系統人員以及學科專家的平衡是相當重要的。Crane 教授表示，計畫雖涉及相當程度的程式及系統設計，但在主題分析及系統發展上，具主學科背景的成員有助於系統的發展以及有效的呈現及運用文獻資料。

計畫推動、運作過程的決策是以小組及負責人 Crane 教授為主，決策項目包括館藏的選擇、處理方式、運用技術 ... 等。若涉及與其他單位的合作計畫，則視合作方式及權責分配共同處理。

現有典藏資料

Perseus 計畫所收錄的數位館藏除了一開始的古典文獻，隨計畫的發展目前 Perseus 數位館藏還包括文藝復興時期的作品，以 16 世紀劇作家 Christopher Marlowe 完整作品為主，另外 Perseus 也開始建立莎士比亞作品的數位資料。除了

文學資料，Perseus 對應文學的發展，收集當時期的歷史資料，目前以英國倫敦為主。至於美國部份則是包括波士頓地區的歷史資料，及西部地區文學作品。除文字資料外，Perseus 也納入地圖、像片的圖型資料。



【圖 15：Perseus 數位館藏清單】

Perseus 所面臨的議題

一、如何將文獻以結構化的方式組織呈現

Perseus 計畫所設定的目標不單是將文獻予以影像化，同時也希望能夠提供結構化的全文檢索。因此一個很重要的議題是將資料數位化之後，要以什麼樣的方式，資料架構組織數位資訊。除了能夠顯示文獻的原貌，也要能夠透過文獻標示，表現資料的語意關係。可以從中找出文獻內容的結構與意義。

二、如何使研究古典文獻的學者可以更容易的整合不同的資源

Perseus 計畫的目標是以建立一個可以支援古典文獻研究的數位圖書館，除了文獻原件外，參考工具、解釋資料都是在研究過程中可以提供研究者相當具有價值的資訊。因此除了文獻外，Perseus 小組成員要羅列相關的輔助工具，同時也要建立彼此間的關係架構，將資料做一整合處理。

研究參考工具是另一個 Perseus 計畫目標典藏之一，提供文獻中所涉及名詞、概念的說明資訊，使利用者可以更容易掌握研究資料中各個相關資訊。例如文獻中討論與機械有關的資訊，透過工具的連結可以進一步的解釋定義、操作原理，

利用圖示或動畫顯示機械的外觀及操作狀況。經由這樣的方式，研究者或學習者可以對所討論的概念有更進一步的認識。

三、如何整合處理不同的語文資料

由於主題因素，Perseus 系統中除了英文資料外，還需要處理古希臘文及拉丁文。如何在資料呈現與檢索上，同時處理不同語文，也是 Perseus 計畫希望達到的目標之一。

四、資料檢索問題

從語文的角度來看，希臘文以及拉丁文的檢索需求較不同於英文資料，一般的搜尋引擎無法處理較為複雜的語詞變換。搜尋引擎在語文詞型的處理上多是以處理字尾變化為主，而古典語文處理，動詞可以有不同的變化型式，並非只侷限於字尾的變化。因此 Perseus 計畫另一個需要處理的問題是建立一個可以處理希臘文及拉丁文在檢索上的需求。

五、個人化的資訊服務

如同實體圖書館一樣，Perseus 數位圖書館預期服務的讀者具有不同層次的學科背景，研究興趣及資訊需求。經由系統設計，Perseus 計畫希望能夠達到依使用者的背景程度，提供不同層次的資訊，包括資訊的內容與組織方式(圖)。

六、從既有資料中找出新資訊

Perseus 計畫的另一目標是希望能夠透過資料的數位化以及程式的運算功能，以不同的條件與角度分析現有的資料，從中找出新的資訊。例如以時間及地理因素為分析軸，分析、比對文獻中各個部份的關係，透過運算找出文獻中可能隱含的資訊，提供不同的思考方向。例如從文獻中找出同一時間，同一地點的訊，或是同一地區在不同時間軸上的情況，把這些從文獻中抽離出來。同時與透過與參考資源的結合，可以經由連結提供文獻中各個部份，如地名、時間，更進一步的資訊。將資料從二度擴增為三度，一份文獻所能提供的不單僅限於文獻本身。



【圖 16：系統參數設定】

未來計畫

一、擴展語文版圖

有處理希臘文及拉丁文的經驗，Perseus 工作小組，希望可以將處理語文的範圍擴大到其他不同語文。具計畫主持人 Crane 教授表示，正與韓國方面有所接觸，希望可進一步獲取處理東方語文的經驗。目前 Perseus 計畫以整合不同語文的資料，為了協助研究者對資料的處理，同時對研究時代有更進一步的了解，Perseus 將往文獻翻譯發展，初期以拉丁古文獻開始。館藏中的法律及公共服務相關文獻，由於文獻結構明確，語意較易掌握，將以此為先導計畫的實驗物件。

二、相關文獻的連結

除了以時間、地區方式整合、羅列目前典藏資料，Perseus 計畫希望可將整合的功能擴及到概念文獻的連結、整合。Perseus 希望經由詞彙出現頻率的計算，判斷文獻內容的相關度，以主題關連方式提供資料的檢索。

Perseus 是一個由個人研究興趣，進而希望建立一個可支援學術研究數位館藏資訊所發展的計畫。由於館藏語文以及館藏建立的需求，其中所涉及的問題從非英語資料的處理、圖像資料的建立以及資料的整合連結均在計畫過程中逐一發展出來。而建立的數位圖書館也由原有個人研究支援，擴及到各個不同層面學習者的資訊資源，研究機構的角色也走入不同學習社群，展開不同的知識交流。

心得與建議

從所考察的幾個不同計畫的推動來看，數位圖書館的規畫以及資訊服務架構的建立，尤其是大學及學術環境中均趨向朝跨學科合作、與研究教學結合及本土資訊建立等方向發展。同時在使用者導向的影響下，視覺化的設計及使用者研究的推動，亦在各計畫中扮演相當重要的角色以及提供機制設計的基本資訊。

跨學科的合作發展

隨社會的發展、對資訊需求的了解以及資訊環境的改變，資訊檢索機制以及對資訊精確度的要求，資訊檢索系統的建立需要融入各種不同的專業知識。以往各學科領域的研究多是以學科本位為範圍，較少跨越學科界限進行不同的研究。但在對資訊需求要求的改變以及資訊技術給予個人及自行發展計畫的工具，使得學科距離縮短。跨學科的研究不單有其環境，同時也有其必要性。所考察的數個計畫其共通性均是在計畫以及架構資訊服務模式的過程均整合不同的專業性，包括對主題分析的學科專家、掌握資訊技術的科技人員、對資訊組織以及使用行為有所研究的圖書資訊工作者以及可視覺化使用介面呈現資訊的影像專家。

建立與研究相關的數位圖書館計畫

與所屬單位或是個人研究相關的數位圖書館計畫的建立，在各方面的資源支援的取得的困難度較低，更重要的是數位圖書館的建立所帶來效益對相關研究的人員能夠有立即的影響。有鑑於此，所考察的幾個單位在架構其數位圖書館及資訊服務架構均是由與其所屬單位或社群所關心的議題開始。

本土資訊的建立與典藏

本土、社群相關資料的保存與充份利用同樣也是被關心重視的課題。需要處理的資訊相當豐富，在資源及使用限制的考量，與對所屬社群關懷因素的影響下，與社群相關的記錄是優先被考慮以不同方式典藏及提供使用的資訊之一。

圖像視覺的資訊檢索

文字型式的資訊檢索方式在使用上的限制，使得使用者與資訊系統及資料間一直有所距離。圖像視覺技術的運用，提供了一個不同的展現方式。除了藉由圖

像視覺使操作介面更具親和力，各個計畫均運用圖像視覺的方式表現文獻資訊的關連性以及暗喻系統操作方式，甚至經由這樣的方式詮釋所要展現的資訊，改進文字描述語文的限制。

使用者導向發展及個人化資訊服務

圖書資訊的服務不單只是系統設計者的產品，系統的設計或是圖書館的資訊服務應以讀者為導向，考慮讀者的資訊需求以及個人需求的差異，是資訊服務設計應有的發展方向。

不論是由國內目前環境，或是國外相關計畫的經驗，圖書館在數位圖書館建置及架構資訊服務機制能夠朝幾個方向發展，規畫與學校研究相關的數位圖書館計畫、建立學科合作機制、使用者研究的推動、技術開發與應用。同時相當重要的一點是在政策、經費以及資源上必須要能夠提供一個可以永續經營的環境。