

電子圖書館文獻與博物館藏品數位化技術與應用研究-總計畫

A study on the technology and applications of the digital library and museum

子計畫一-電子圖書館文獻與博物館資料管理及儲存系統設計之研究

Study of contents management and storage system design in the digital
library and museum

計畫編號：NSC88-2213-E-002-030

執行期限：87/08/01~88/07/31

主持人：歐陽彥正

處理方式：

可立即對外提供參考

, 一年後可對外提供參考

, 兩年後可對外提供參考

(必要實, 本會得延展發表時限)

執行單位：國立台灣大學 資訊工程研究所

中華民國八十八年十一月

電子圖書館文獻與博物館資料管理及儲存系統設計之研究

計畫編號：NSC88-2213-E-002-030

計畫期限：87/08/01~88/07/31

主持人：歐陽彥正 台灣大學資訊工程學系

一、中文摘要

提出本計畫的動機源自我們自 86 年 1 月以來即著手建立“台大電子圖書館與博物館”的經驗與心得。第一版的“台大電子圖書館與博物館”已於 86 年 10 月完成，並建立網站 (<http://ntudlm.ntu.edu.tw>)，一個電子圖書館/博物館的系統大致可分為兩大部分，一是 search engine 的部份；一是 repository system。Repository system 負責電子圖書館/博物館 contents 的儲存與管理，同時必須根據 search engine 的 requests，迅速有效率的提供所需的資料。本計畫的目標是研究 repository system 設計上的三個關鍵議題。

1. Repository system 的內容管理 (contents management)
2. Repository system 的檢索效率
3. Repository system 的儲存系統設計

英文摘要

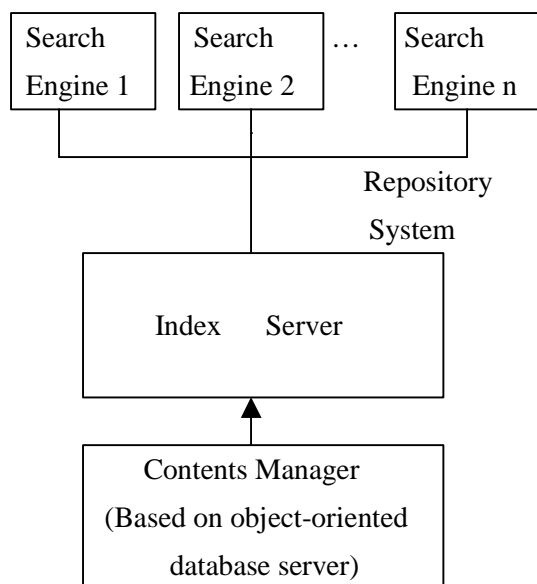
The motivation behind proposing this project originates from the experiences accumulated during the design of the “National Taiwan University Digital Library and Museum” (NTUDLM). The first version of NTUDLM was completed in October 1997 and has been on the Internet since then. A digital library and museum (DLM) system comprises two major parts: the search engines and the repository system. The repository system is in charge of contents management and storage in the

DLM and must be able to retrieve the information requested by the search engines in an effective and efficiency manner. The goal of this project is to study the key issues in the design of the repository system for a DLM. Based on our experiences accumulated during the design of the NTUDLM, we identify the following key issues to be studied in this project.

1. Contents management in the repository system
2. Information retrieval in the repository system
3. The storage system

二、計畫緣由與目的

一個電子圖書館/博物館的系統可分為兩大部分（如圖一所示）。上層是 search engines，而下層則是 repository system。Search engines 的功能主要是提供使用者檢索服務的工具。Repository system 則是負責電子圖書館/博物館的內容管理，並針對上層 search engine 的 request，迅速有效率地進行內容檢索，以提供所需的資料。在本群體計畫中，有關 search engine 方面是由子計畫二“台灣史料中知識擷取與閱讀環境之研究”負責。本計畫則專注在 repository system 的設計議題上。



圖一、電子圖書館/物館系統示意圖

提出本計畫的動機源自我們自 86 年 1 月以來即著手建立“台大電子圖書館與博物館”的經驗與心得。我們體認到下列三個有關 repository system 設計的重要議題。

1. Repository system 的內容管理(contents management):

電子圖書館/博物館的 contents 有各種不同形態的資料 (如 text, image, graphics, video, audio,...等等)。而各種不同形態資料物件 (data object) 均有其屬性, 同時資料物件之間又存在種種關連性 (association)。這些屬性及關連性代表了電子圖書館與博物館所內含的知識, 因此如何以一個適合的軟體結構來管理這些 contents, 以利未來新增加的 contents 能有效地與舊 contents 整合, 並建立其間新的關連性, 以維持電子圖書館與博物館的 integrity, 便成為一個重要課題。

2. Repository system 的檢索效率:

Repository system 必須針對上層

search engine 的 requests, 迅速有效率地進行內容檢索, 以提供所需的資料。因此 repository system 的設計亦必須針對檢索效率的需求作特殊考量。

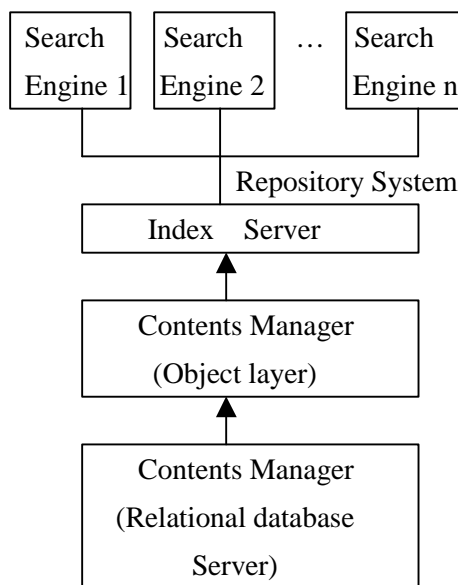
3. Repository system 的儲存系統設計:

電子圖書館/博物館的資料量相當大, 又有各種不同的形態。同時由於各種不同形態資料之間又有各種不同的關連性, 使得資料的 access locality 會跨越不同形態的資料。因此如何針對這種特性設計多階層的儲存系統, 以達到 cost/performance 的最佳化便成為一個相當重要的課題。

本計畫的目的乃針對這三個議題作深入的研究。

三、研究方法與成果

綜合前節的討論及本計畫第一年所完成之雛形系統於網路上實際運作所獲得的經驗, 我們在本年度的計畫執行中完成圖二所示的電子圖書館/博物館系統架構:



圖二、電子圖書館/博物館系統架構

茲就整個系統各模組的功能及運作情形敘述如下：

1. Contents Manager-

Contents Manager 為 Repository System 的底層軟體子系統。Contents manager 負責管理整個 digital library/museum 的資料，並負責處理負責的 query。Contents Manager 又分為兩個 layers。上層為 object layer，這一層是 Representation layer，所有在電子圖書館/博物館中各種不同形態的資料均以 object 的型態表現及管理。同時亦透過 object 的屬性 (attributes) 及彼此間的關連 (association) 記錄儲存於電子圖書館與博物館中的知識。Contents manager 的底層則為一 relational database server。整個 object layer 是建構在這個 relational database server 上。Relational database 則是整個電子圖書館/博物館軟體層面的資料儲存系統。

2. Index Server-

Index Server 為 Repository System 的上層軟體系統。由於大部分使用者的 query 均極為簡單，若這些 simply query 均由 database server 處理，顯然並不合乎成本效益。因為 database server 之設計均強調功能強大可處理 complicated query，因此若 simple query 亦交由 database server 處理，則顯然對系統處理資源的使用形成浪費。基於這項原因，本計畫在系統架構的設計上，便增加 Index Server 一個層次的軟體模組。Index Server 運作於由 Contents Manager 所 dump 出來的 indexing structure。這些 indexing structure 的設計著眼於能迅速有效處理 simply query，以有效使用系統處理資源。至於 Index Server

所不能處理的 complicated query，則交由 Contents Manager 的 database server 處理。

3. Search Engine-

在 digital library 上為配合不同背景使用者的使用行為模式，必須設計不同的 search engine。譬如說一個一般的使用者與一個專家，對 search engine 所提供的 precision 及 recall rate 的要求便顯然不同。一般使用者會較注重 precision 而對 recall 則較不重視。但一位專家則可能較注重 recall rate，而對 precision 的程度較不重視。再者不同領域的專家所重視的角度亦會不同。以本計畫所建構的 digital library 而言，歷史學者及人類學者均是未來的使用者，但顯然歷史學者與人類學者對資料搜尋的需求與重視角度是不一樣的。因此對不同領域的專家，必須設計不同的 search engine，以發揮 digital library 的最大效益。

四、結論

經過本計畫兩年的努力與經驗累積，相信我們將可完全掌握電子圖書館/博物館中的 repository system 及多階層儲存系統設計上的關鍵技術。由於電子圖書館與博物館的建構是資訊科技領域新興的研究主題，因此本計畫的執行，對建立國內資訊學界這方面的研究實力將有相當大的注意，亦將產生許多國際上相當領先的研究成果。這些成果不僅有學術上的貢獻，同時也具有產業的價值。我們可以預期在未來資訊化的社會當中，透過網際網路做資料搜尋無疑是提升生活品質與工作效率不可或缺的一環。而電子圖書館與博物館正式這種社會活動中最重要資訊來源。因此建構電子圖書館與博物館

所涉及的資訊科技將是未來資訊產業重要的關鍵技術。此外，基於我國在全球華人社會中的地位與科技實力，實應積極朝向成為華人文化科技中心的目標邁進。本計畫的執行，將有助於建立我國在中文電子圖書館與博物館的研究地位。使得這項成為華人文化科技中心的目標更能及早實現。

Yuan Shao, Yi-Chin Tu, and Yen-Jen Oyang, "System Design of the National Taiwan University Digital Library and Museum", Proceedings of the First Asia Digital Library Workshop, Hong Kong, August 1998.

五、參考文獻

- } Jian-Hua Yeh and Yen-Jen Oyang, "Content and Knowledge Management in a Digital Libarary" Accepted by the Journal of American Society for Information Science, Special Issue on Digital Library, (Publisher: American Society for Information).
- } Jian-Huan Yen, Jia-Yang Chang, and Yen-Jen Oyang, "Exploiting of Information Processing Technologies to Extend Digital Library Applications", Journal of Library and Information Science, Vol.25, No.1, pp.1-11, April 1999.
- } Chung-Chen Lai, Juan-Hua Yeh, and Yen-Jen Oyang, " Toward Automatic Generation of Geographic Thesaurus through Data Mining in a Geographic Information System",
- } Chien-Kang Huang, Nai-Wen Lin, and Yen-Jen Oyang, " A Search Engine for Multiple-Version Chinese Buddhist Scriptures", Proceedings of the Second Asian Digital Library Conference, Taipei, Taiwan, November 1999.
- } Jian-Hua Yeh, Hsiu-Ping Lin, Chun-