

行政院國家科學委員會專題研究計畫 成果報告

設計以 XML 技術為基礎之文獻傳遞雛型系統架構

計畫類別：個別型計畫

計畫編號：NSC92-2416-H-009-011-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學資訊管理研究所

計畫主持人：陳瑞順

計畫參與人員：張展志，蔡永順，余顯強

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 10 月 4 日

行政院國家科學委員會補助專題研究計畫成果報告

※※※

※ 設計以 XML 技術為基礎之文獻傳遞雛型系統架構 ※

※ ※

※※※

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC92-2416-H009-011

執行期間：92 年 8 月 1 日至 93 年 7 月 31 日

計畫主持人：陳瑞順 國立交通大學資訊管理研究所

計畫參與人員：張展志, 蔡永順, 余顯強

執行單位：國立交通大學資訊管理研究所

中 華 民 國 93 年 10 月 5 日

行政院國家科學委員會專題研究計畫成果報告

設計以 XML 技術為基礎之文獻傳遞雛型系統架構

An XML-Based Prototype System for an Electronic Document Delivery System

計畫編號：NSC 92-2416-H009-011

執行期間：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：陳瑞順 國立交通大學資訊管理研究所

中文摘要

電子出版的普及與網路的發達，使得圖書館的館藏政策從「擁有」(ownership)轉為「取得」(access)為導向。這種說法的實質意義乃在提醒圖書館：使用者能否得到資料原件，比圖書館是否擁有該資料更重要。另外，在數位資訊的時代，文獻提供者往往採用電子媒體形式以租代賣。是故善用網路科技，可節省資訊購買及空間儲存的成本，尤其在電子文獻傳遞數量逐漸增加的今日，如何有效縮減資訊檢索與文件傳遞的程序，提供讀者快速取得原件的管道，變成了一個重要的課題。

另一方面，延展式標示語言((Extend Markup Language, XML)的彈性、單純和同時具備了機器與人類可讀的特性，能夠提供不僅只有資料處理的應用範圍，XML 簡潔的文法與明確的結構，使其非常適合在大型的專案中應用。因為，XML 比現有的資料格式更容易地傳遞、調整、處理、分解和重製。

因此，本研究要目的是探討使用 XML 技術整合後端各文獻提供者的電子文獻傳遞系統架構，結合現有圖書館自動化系統，

並提昇 OPAC 的功能，其研究重點在於(1)規劃電子文獻系統作業流程(2)建構一個電子文獻傳遞之資訊處理系統(3)分析、規劃系統細部架構。透過圖書館中介的服務，提供使用者能夠上線查詢到所需的書目並需要傳遞所需的資料時，能夠將需求直接轉送至文件擁有者或出版商的系統內。無論是否收費，或是由圖書館、還是使用者付費，使用者均可透過此系統立即獲得電子文件全文，達成了查詢求現率。而且，圖書館不一定需要具備館藏，也能達成相同的線上服務，間接的往無人圖書館服務邁向一大步。

關鍵詞：電子文獻傳遞 延展式標示語言
電子資料交換 資料加密

Abstract

This research explores the architecture, which new XML technical to design the new electronic document delivery system combine the existing library system, assist the user to query by library OPAC and retrieve information entity rapidly in Web. In order that, not only accelerate the development for electronic publishing, but

also position the library role of information mediation between user and provider, archive information circulation and provide secure copyright-fee through the efficient electronic document delivery and payment mechanism. The proposed electronic document delivery system have three contribution: (1) resolve the electronic document payment, (2) shorten the time between enquiry, retrieve electronic document entity, and promote the recall-rate of enquiry, (3) anticipate the changing libraries to develop into the role of Web, with reduce the load of collection

Keywords: Electronic document delivery, XML, Library mediation, EDI.

一、前言

當使用者要查詢所需的文獻時，通常會利用圖書館的目錄，或書目中心的聯合目錄來檢索是否有收藏所需的資料。近幾年來，Z39.50 資訊檢索協定⁽¹⁾在圖書館自動化系統及書目資料庫檢索方面的應用，提供使用者以一種檢索方式、指令，即可檢索不同的書目資料庫，並且以相同的畫面顯示檢索結果，大大地提昇了資料檢索的範圍。但是，隨著電腦及網路的發展，讀者對於資料的獲取更著重方便性及時效性。因此只提供書目性服務而無全文資料的傳統圖書館線上目錄(Online Public Access Cataloging, OPAC)，已無法滿足讀者即查即得的需求。

電子出版的普及與網路的發達，使得圖書館的館藏政策從「擁有」(ownership)轉為「取得」(access)為導向。這種說法的實質意義乃在提醒圖書館：使用者能否得到資料原件，比圖書館是否擁有該資料更重要。另外，在數位資訊的時代，文獻提供者往往採用電子媒體形式以租代賣。是故善用網路科技，可節省資訊購買及空間

儲存的成本，尤其在電子文獻傳遞數量逐漸增加的今日，如何有效縮減資訊檢索與文件傳遞的程序，提供讀者快速取得原件的管道，變成了一個重要的課題。

電子文獻傳遞服務,有下列發展特性:

1. 整合資源檢索,館藏記錄註記暨全文傳遞的服務
2. 無仲介的文獻傳遞服務(non-mediated document delivery):如 CARL 公司的 UnCover 系統, OCLC 的 Content/Article First 服務, Silver Platter 的 ERL 系統, 我國國家圖書館資訊網路的遠距圖書館服務, 飛資得公司代理的 OVID 系統等等,除提供期刊文獻索引摘要的檢索和部份全文外,並有線上訂購全文服務,允許使用者自行線上檢索,自行線上訂購全文,於 24-48 小時內經傳真或網路取得全文資料,成為無仲介的文獻傳遞服務.
3. 運用 Internet 作為文獻傳遞的工具:Internet 提供使用者和資訊之間的連接管道,為文獻傳遞的良好工具,無論以 Ariel 傳遞館際複印表單和全文文獻,以 E-mail 傳遞館際複印申請單,或線上訂購全文文獻,皆可運用 Internet 作為文獻傳遞的工具,提供及時性的文獻傳遞服務.
4. 及時性服務(just in time)之提供
5. 兼顧個別化及整體性的需求滿足

電子文獻傳遞系統的概念,可以回溯到最早電腦應用的時期。Cawkell 在此領域的專案計畫報告中,敘述在 1960 年代中期開始使用傳真複製的技術。現今,主要的電子文獻傳遞系統大多是由商業性線上資料庫、期刊目次、文獻摘要等供應商、以及圖書館所提供。基本上依據其提供服務的架構可以區分為三代:線上訂購系統、非整合性全文影像系統、以及整合性

獨立的全文影像系統。但由於現有電子文獻傳遞系統大多著重於「如何處理電子期刊」和「如何由紙本轉換成光碟版本(CD-ROM)」,而不是「如何從查詢到遞送文件的完整解決方案」。

電子文件供應商可以透過圖書館的線上目錄,達到文件線上電子交易的行為。不僅圖書館可以減少館藏徵集負荷,不須購買過多的複本;甚至還可由轉介的服務,向電子文件供應商收取適當的轉介費用,循環使用來擴大圖書館其他的服務。而且,電子文件供應商也可藉由使用者下載文件而獲得實質收益,使得文件發行電子版本的銷售管道更為快速,連帶的影響各種電子媒體與電子書的整合,如此必能加速電子化出版品的發展。

雖然電子文獻傳遞所需的基礎技術已存在有一段時間,但是實際應用的比率仍舊不高,出版商發行的資料仍是以紙本為多。

因此在文獻傳遞系統上必須考慮版權、付款方式和如何在使用者端顯示各種不同媒體形式的文件內容。在此種概念之下,如何快速將文獻送達至讀者手中,便比電子資料的格式與其硬體配備還要來的重要。當使用者能透過整合性的一次檢索,便能獲得所需的書目描述及所有具備全文的單位,且當檢索結果的資訊呈現在使用者端的瀏覽器上時,使用者便能夠直接當場決定要下載的文件來源。藉由這種線上即時查詢訂購下載全文文件的交易,並結合電子商務安全的遞送與付款機制,可帶動相關的電子出版發展計畫與趨勢。

本研究以電子文件傳遞系統模型的架和系統應用 XML 框架文件檢索上,透過三個主要關鍵點解決現有傳遞上的缺失:

1. 將檢索的結果(result set)轉換成為 XML 文件格式。
2. 建立檢索結果的 metadata。
3. 透過 XML 文件能夠超連結進入文件

供應商系統。

且在電子文獻傳遞系統模型中必須能夠整合使用者、圖書館自動化系統、和電子文件供應商現有的系統,透過系統分析方式,提出四個必要項目的解決方案:

1. 瀏覽器、圖書館系統和電子文獻提供者之間的資料交換:藉由 XML 來封裝所需的資訊與文件內涵。
2. 版權保護:包括加密與浮水印處理。
3. 付款方式與機制:包含免費、圖書館付費、使用者自行付費等處理機制的流程
4. 瀏覽器之 plug-in 程式:解決不同的媒體型態;閱讀限制;付款方式等應用需求。

本研究使用 XML 標示技術所提出的電子文件傳遞系統模型,同時結合圖書館與文件供應商既有的系統,本系統模型能夠將電子文獻傳遞系統完全整合至圖書館自動化系統,透過有效的電子文獻傳遞與付款機制,不僅加速電子出版品的發展,還可提昇資訊流通與提供安全的版權保護,促進提供圖書館扮演良好的 Web 資訊中介者的角色。

1. 在讀者服務方面:本模型提供了前端使用者一個整合文件查詢和檢索的服務,並且解決了電子文件在使用上付費的方式,以及縮短查詢至取得電子文件實體所需的時間,並提昇了資料查詢的求現能力。
2. 在圖書館方面:本模型能夠將圖書館定位在 Web 資訊中介人的角色,並且降低圖書館館藏徵集的負荷。

二、文獻探討

現今,主要的電子文獻傳遞系統大多是由商業性線上資料庫、期刊目次、文獻摘要等供應商、以及圖書館所提供。基本上依據其提供服務的架構可以區分為三代:線上訂購系統、非整合性全文影像系

統、以及整合性獨立的全文影像系統。

第一代系統：

線上訂購系統，DIALOG 與 EAS/IRS 提供這一類的服務系統，其運作方式是經由專屬的應用程式連結到參考的資料庫，產生需要文獻傳遞的工作清單。此種系統最主要的缺點是缺乏有效率的處理方式，不管文獻是否被重複申請，每次均需依賴人工至架上取出資料影印，相當耗費時間與人力。

第二代系統：

非整合性全文影像系統，第二代文獻傳遞系統引進新的供應驅動 (supply driven) 方法，並利用影像來儲存文獻內容的方式，以解決效率的問題，ADONIS 系統即屬於此種類型。這一代系統主要的缺點是遞送的過程仍然需要透過人工處理。

第三代系統：

整合性獨立的全文影像系統，第三代系統大多仍然是將文件以影像方式處理為主，但採用了需求驅動方法來降低運作上的負荷與花費。當使用者按下電腦上的下載按鈕時，即可將該全文內容傳輸至使用者端。符合第三代系統特性的單位包括 CARL 的 Uncover、OCLC 的 ContentFirst、ArticleFirst、ContentsAlert, Faxon Finder、SWETSCAN、EBSCOdoc、ISI 的 Current Contents 等。影像為主的系統有一個主要的缺點是不夠開放 (open)，以致於無法適應未來的發展。而另一個缺點是無法整合館際互借及既有的文件傳遞程序，使用者必須逐一至各資料庫中查詢所需的資料，再逐一決定是否執行遞送或下載的程序。

三、研究方法

針對本研究內容，其各項主研究方法如下：

1. 規劃電子文獻傳遞系統作業流程。

分析電子文獻傳遞系統在使用者、圖書館和電子文件供應商現有的系統之間的整合需求，落實電子文獻傳遞的流程及需求，使得架構能夠符合電子文獻在網路環境傳遞與稽核控管的需求，也能同時符合建立數位典藏與虛擬圖書館的組織發展目標。本研究透過實際流程的分析及文獻的參考，來進行規劃。

2. 分析並建構電子文獻傳遞系統架構

架構將以物件導向的方式來分析，在 OOD (Object-Oriented Design) 階段時將參考現有的 Design Pattern。運用良好的 Design Pattern，可以使得系統架構更優良也可以更快完成，對於後續的 OOP、測試、維護，都會有很大的幫助。並此時考量電子文獻在實際運作中的特性，將所有可能需要的功能及系統規格做初部的規劃，並建構一個系統雛型，確定軟體功能、資料庫以及使用者介面需求。

3. 分析、規劃系統細部架構

在這個階段中，配合 Rational 公司所提出 4+1 View，以 UML 來描述本系統的步驟流程及架構雛型，並將上個階段所分析軟體功能、資料庫及使用者介面做細化的工作，也就是將各部份的功能都分割到相當於一個程式模組的程度為止。在過程中可能會因系統開發過程中所產生的回饋，而必需不斷

的修正系統細部功能。

4. 實作並測試系統

在這個階段，根據前面之系統模式分析及規劃的結果，以物件導向語言（Object Oriented Programming, OOP）來開發虛擬團隊系統，每個子功能根據所選定之程式語言的特性來做撰寫的工作，以驗證之前所提出之架構。如果在設計中發現一些局部問題，即回去修改之前細部分析時的一些錯誤；而測試則是從每個子功能完成之後就開始進行，先從「單元測試」（Unit Testing），再依據之後完成模組的程度，進行「整合測試」（Integration Testing）。然後，逐步將測試層次提高到「系統測試」（System Testing）階段，包括績效測試、安全測試、與反覆測試等工作。

本計劃的研究步驟如下：

1. 收集及研讀相關論文報告。

搜集電子文獻傳遞、知識管理、協同合作工具及代理人程式等相關的文獻；從研讀文獻中，逐步釐清研究主題，並從資訊技術的觀點，研究企業運作模式中所需之資訊技術。

2. 雛形系統之初步分析。

首先定義電子文獻傳遞的建構流程，再根據其流程來定義系統所應具備的功能，及各部分組成的模組。當這些前置工作完成時，就可以先建立一個概念性基本系統架構。

3. 從實務中做資訊交換分析。

將基本系統架構運作在實際的流程機制上，將回饋所得到的資訊，修改部份系統架構。

4. 實作架構確定。

經由不斷的測試及改良，獲得最後完成的系統架構。

5. 系統設計、測試。

實作系統，並作完整的測試。

6. 成果分析和比較。

分析成果，與預期的結果作個比較，找出相異的地方，探討其生成原因。

7. 後續研究探討及報告撰寫。

找出可供將來後續研究的方向，及目前系統不足的地方，完成報告。

四、結果與討論

本研究主要是使用 XML 技術整合後端各文獻提供者的電子文獻傳遞系統架構，本研究之主要成果包括：

1. 規劃資料流程格式

(1)將檢索的結果轉換成為 XML 文件格式。

(2)建立檢索結果的 metadata。

(3)透過 XML 文件能夠超連結進入文件供應商系統。

2. 實作電子文獻傳遞系統架構

本系統模型採用瀏覽器作為前端使用者與 HTTP 伺服器溝通的介面。在伺服器端，既有的圖書館自動化系統結合一個 Web 版本的圖書館線上目錄(OPAC)，和提供電子文獻傳遞系統的前台(front-end)查詢功能，由現有圖書館系統增加運算和網路處理功能，並將圖書館與文件供應商之間的資訊交換透過 XML/EDI 代理人程式處理。透過圖書館中介的服務，提供使用者能夠上線查詢到所需的書目並將需求直接需要傳遞至文件擁有者或出版商的系統內，使用者可透過此系統立即獲得電子文件全文，達成了查詢求現率。電子文件供應商可以透過圖書館的線上目錄，達到文

件線上電子交易的行為。圖書館可以減少館藏徵集負荷，還可由轉介的服務，向電子文件供應商收取適當的轉介費用。電子文件供應商也可藉由使用者下載文件而獲得實質收益，使得文件發行電子版本的銷售管道更為快速。

參考文獻

- [1] Juha Hakala, "Z39.50-1995 Information Retrieval Protocol: An Introduction to the Standard and It's Usage", available over the Internet at URL: <http://renki.helsinki.fi/z3950/z3950pr.html>
- [2] Judy Luther, "Electronic Book '98- Turning a new page in knowledge management: NIST Conference", Library Collections, Acquisitions, & Technical Services, 1999, Vol. 23, No. 2, pp.179-181
- [3] "Extensible Markup Language (XML) 1.0", W3C Recommendation, World Wide Web Consortium, 1998, URL: <http://www.w3.org/TR/1998/REC-xml-19980210>
- [4] Cawkell, A.E., "Electronic document supply systems", Journal of Documentation, 1991 Mar, vol.47, no.1, pp.34-37
- [5] Hans Roes, Joost Dijkstra, "Ariadne: the next generation electronic document delivery systems", The Electronic Library, Jan 1994, available over the Internet at URL: <http://www.kub.nl/~dbi/users/roes/articles/ariadne.htm>
- [6] Hal Berghel, "Value-Added Publishing", Communications of the ACM, 1999 Jan, Vol.42, No.1, pp.19
- [7] Pamela Mccauley-Bell, "Intelligent agent characterization and uncertainty management with fuzzy set theory: a tool to support early supplier integration", Journal of Intelligent Manufacturing, 1990, Vol.10, pp.135-147
- [8] Webber, David R, "Introducing XML/EDI Frameworks", EM - Electronic Transactions. EM - Electronic Markets, Mar 1998, Vol. 8, No. 1, available over the Internet at URL: http://www.electronicmarkets.org/netacademy/publications.nsf/all_pk/804
- [9] "Introducing XML/EDI: the e-Business framework", available over the Internet at URL: <http://www.geocities.com/WallStreet/Floor/5815/start.htm>
- [10] ISO/IEC 10181-1 "Information technology- Open Systems Interconnection- Security frameworks for open systems: Overview", Aug 1996, pp.11
- [11] "Access & Technology Program/NAI LDD Project", Association of Research Libraries, available over the Internet at URL: <http://www.arl.org/access/naildd/naildd.shtml>
- [12] "The Ohio Library and Information Network", available over the Internet at URL: <http://www.ohiolink.edu/>