新圖書館學叢書

# 電子文件自動處理之研究

陳光華

台灣學生書局印行

# 電子文件自動處理之研究

簡 先 勝

[illegible] 年 [illegible] 月

# 電子文件自動處理之研究

陳光華　著

國家圖書館出版品預行編目資料

電子文件自動處理之研究

陳光華著.—初版.— 臺北市：臺灣學生，1999(民 88)
面；公分

ISBN 957-15—0942—6 (精裝)
ISBN 957-15—0943—4 (平裝)

1. 資訊儲存與檢索系統

2. 電子資料處理

028 88004037

# 電子文件自動處理之研究(全一冊)

著作者：陳光華
出版者：臺灣學生書局
發行人：孫善治
發行所：臺灣學生書局
臺北市和平東路一段一九八號
郵政劃撥帳號00024668號
電話：(02)23634156
傳真：(02)23636334

本書局登記證字號：行政院新聞局局版北市業字第玖捌壹號

印刷所：宏輝彩色印刷公司
中和市永和路三六三巷四二號
電話：(02)22268853

定價：精裝新臺幣二二〇元
平裝新臺幣一五〇元

西元一九九九年三月初版

02801



ISBN 957-15-0942-6 (精裝)
ISBN 957-15-0943-4 (平裝)

# 序

從小的確是立志作老師，但是沒有想過也可以寫書。寫書的經驗是痛苦的，不知道那些多產的作家是如何辦到的；完稿的時刻是美好的，大概這就是作家們追求的。據說每一本書都有個「序」，只好不可免俗地作序一番。只希望本書最後還有個墊高的功用。

# 誌謝

寫書的過程真是寒天飲冰水，點滴在心頭。雖然我還沒有當父親的經驗，卻也充分體驗老師的付出，對於父母的謝意，自不必行之於文字；但衷心感謝教育我的師長，感謝陳信希教授、王如意教授、林一鵬教授、黃宣範教授。同儕的討論也增長我的見識，感謝陳雪華教授、陳昭珍教授。我的學生江玉婷、莊雅蓁真得是很棒，協助我處理大大小小的事務，感謝她們。當然，還有背後默默支持的人，也不是誌謝就可以表達心意的。

# 目次

# 表目次

# 圖目次

# 第一章　緒　論

人類文明已經超過六千年，但從來沒有像 20 世紀這樣進展地如此快速，可以預見的是，文明會以更快的速度往前邁進，創造更令人讚嘆的世紀。在人類文明史上重要的前二個革命：印刷術革命以及工業革命，使得恭逢其時的國家或是願意改革的國家，紛紛成爲史上的強權，引領風騷數百年。如今更重要的資訊革命正方興未艾，我國的資訊工業總產值已經躍居世界第三位，如何掌握此一重要契機，提升國家的競爭力，實在是政府目前所應積極努力的一項重要課題。學術研究人員也應在資訊革命的發展過展中，貢獻一己之力量，本書將從電子文件逐漸成爲資訊傳遞形式的時空背景下，探討如何以自動化的程序，有效地進行資訊的處理，以促進資訊的生產與消費。

## 第一節　背景

網際網路作爲資訊交換的管道，由來已久，早期係屬專業人士使用的傳輸媒介，傳送具有價值與高品質的資訊。今日則由於商業體系的介入，引發「網路是否應爲國家公營的資訊管道」的論戰（註 1），促使網際網路的使用群逐漸擴展至普羅大眾。此外，加上全球資訊網（World Wide Web，簡稱 WWW）興起的推波助瀾，簡單易用的圖形化使用介面，使得不分老幼均能輕易地連上網際網路，全民上網的目標，指日可待。現在，無論政府部門、公益組織、營

利事業團體或個人，都相當積極地將資訊送入網際網路，因而使資訊累積的速度越來越快。藉著 WWW 的協助，圖書資源的傳播，經由標準的電腦使用介面，不但可以看到全文影像，亦可生動地將圖形、動畫、聲音呈現在電腦螢幕上。在可預見的未來，知識儲存的媒體將會逐漸轉換成數位化的電子媒體，而其重要性也將與日俱增。若以資訊傳播的角度觀察吾人所處的世界，它正快速的縮小。以台北到高雄爲例，實際距離爲 350km，如果傳送 500MB 的資料，表 1-1 計算以各種不同方式傳遞訊息時，所需花費的時間。若以 100Mbps 的電腦網路傳遞前述的資料，比以汽車進行郵件快遞，快了將近 580 倍。如果是從台北傳遞資料到洛杉磯，則更能讓我們體驗到天涯若比鄰的感受，所謂的「地球村」也成爲十分具體的事實，而非遙不可及的夢想。

表 1-1：資訊傳播的方式與所需時間

| 傳播方式 | 傳播時間 |
|---|---|
| 步行（5 km/h） | 70 hrs |
| 汽車（60 km/h） | 5.8 hrs |
| 飛機（270 m/s） | 0.36 hrs |
| 電腦網路（100 Mbps） | 0.01 hrs |

正由於網際網路的發展如此地驚人，各國的研究人員紛紛嘗試研發各項適用於網際網路的應用，其中最主要且最受重視的應用首推電子圖書館。1998年亞太經合會（APEC）科技部長會議，便廣泛討論電子圖書館的發展，並剖析其對國家競爭力的重大影響。

然而，目前電子圖書館的專有名詞並沒有統一，Lancaster在1978年出版的「邁向無紙資訊系統」（Toward Paperless Information Systems），認爲未來圖書館可能以電子媒體爲主要館藏，而形成所謂「電子圖書館」的概念。「虛擬圖

書館」是網路資訊聯盟（The Coalition for Networked Information）於1990年提出的，期盼藉由現有的網路環境與籌設中的網路建設，促進資訊的散佈與共享。「數位圖書館」則是美國副總統高爾在參議員任內提出「資訊基礎建設與科技法案」，首次出現的名詞。（註2）其定義也沒有普遍的共識，仍處於各自表述的階段。美國國家科學基金會在其研究報告中也並不試圖對電子圖書館下一明確的定義，而以功能面闡述電子圖書館：「電子圖書館是以經濟的方式取得掌控異質化、分散式、大量數位資料與新型態資訊的具體能力，以友善的方式進行資訊的儲存、搜尋、處理、以及檢索」。（註3）相關研究的取向其實是相當多元的，至少可由三個角度來看待：第一是圖書館學界；第二是電腦科學界；第三則是網際網路研究人員。實體圖書館的存在已有很長的一段時間，也發展出相當多的學理，以有效提供圖書資訊予讀者，因此，圖書館學界自然會以擴展本身服務的角度，思考如何運用網際網路這種新的資訊管道。電腦科學界則以資訊系統的觀點看待網際網路延伸的電子圖書館，思考如何有效地檢索資訊。另一方面，網際網路的研究人員則單純地認爲若將網際網路視爲資訊儲存的媒體，上網的使用者便能夠取得資料，很自然地形成「虛擬圖書館」。

在電子圖書館的大框架之下，有著爲數眾多的研究課題，以下是筆者認爲相當重要的項目：

- 資源的蒐集
- 資源的組織分類
- 使用者行爲研究
- 資源的分享
- 資訊與社會

資源的蒐集，指的是如何取得適當的數位館藏，無論是將現有紙本資料數位化，或是採訪電子出版品。資訊的組織與分類，指的是有效地組織並分類蒐集所得

的資訊，主要的問題是如何使用適當的詮釋資料（Metadata），或是制定合用的詮釋資料，進行資訊資源的描述。使用者行爲研究，主要指的是如何有效蒐集使用者的行爲資訊、如何分析蒐集的資料、評估使用者需求的新方法及建立人機互動的新典範。資源的分享則必須注意網路系統、通訊協定與資訊檢索等議題。資訊與社會層次的議題，指的是電子圖書館的設計、政策與實施，應更切實反映所處社會的情境脈絡。

## 第二節 研究假設與範圍

筆者身處網際網路的蓬勃發展以及電子圖書館的時空環境下，將於本書探討如何進行文件自動化的處理，而這樣的探討係基於下列幾項假設：

1. 電子文件累積的速率越來越快

   由於電子文件累積的速率越來越快，無可避免地必須採用自動化的技術，否則正如夸父追日一般，永遠無法有效處理不斷衍生的資訊。

2. 電腦硬體的效能越來越高

   如果電腦硬體的效能無法繼續提升，則本書所提出的處理技術或許將無法真正地使用於網際網路的實際應用程式中，畢竟，網路的使用者最在意的一件事就是時間。

3. 人類對於資訊服務的要求越來越高

   如果目前的使用者已經滿足於搜尋引擎與主題指引的服務，本書便無須提出文件各高層次的處理模式，因爲使用者的需要才是真正應該考慮的因素。

4. 電腦網路的頻寬越來越高

   頻寬將是未來各種網際網路研究的基礎建設，足夠的頻寬才能容納更具創意的研究，使用者也才能真正受惠。