

# 談古籍之電子版本

謝清俊 中央研究院資訊科學研究所研究員  
臺灣大學圖書館學研究所兼任教授

陳昭珍 輔仁大學圖書資訊學系講師  
臺灣大學圖書館學研究所博士班學生

## 目 錄

壹、概論	參、工作程序
一、古籍電子化的意義	肆、文獻之一般性內容與結構
二、研究背景與範疇	伍、古籍之內容、結構與處理原則
三、古籍之界說	陸、文心雕龍之內容、結構與標誌
(一)由圖書形制之演變而言	一、實驗對象的選擇
(二)由目錄學的角度而言	二、文心雕龍版本的選擇
(三)由研究使用的角度而言	三、文心雕龍之內容與結構
(四)以電子化之目的而言	(一)紙本文心雕龍之內容與結構
四、電子版本之定義	(二)文心雕龍之標誌
貳、古籍機讀化之目標	柒、結論
一、有關古籍之版本資訊	參 考 資 料
二、有關古籍之校勘資訊	誌 謝
三、有關古籍之注釋資訊	
四、古籍之全文檢索	
五、相關參考書目與參考文件的提供	
六、研究者之眉批、筆記	

## 壹、概論

### 一、古籍電子化的意義

古籍電子化的工作，概括地說，是試圖將歷來以寫本、雕版、活字版、套版和鉛字印刷等方式所呈現的古代文獻，轉為電子媒體的形式，使古籍能活出現代風貌；其意義則在維護我中華文化的傳承。

百餘年來，西風東漸，我國社會歷經多次巨大變化，使得生活與傳統文化日益疏遠，已是不爭的事實。為了使文化不致斷絕，並能存精去蕪地重新回到生活之中，則利用現代的資訊科技（Information Technology）將古籍變為電子版本以挽狂瀾，是我文化能適應現代社會必須經歷的一個過程。

古籍的電子化，表面上看來，是一種古籍存在形式的改變，可是這種存在形式的改變也導致了對古籍的使用環境的改變。首先：它解除了古籍對傳統印刷媒體物質上的依存關係，在電子網路上，人人都可以無弗遠近地共享，再也沒有「得不到」的困難。其次，雖然古籍浩繁，然而以電子速度來檢索，卻能大大舒解了查訪不易的困難，比起人工來幾乎可以不算時間，於是，不再有「找不到」的困難。再者，可借由計算機處理的技術，把相關的資料（如注疏、版本、校評，以及字書、類書、引得、百科全書、方誌等等），甚至於白話語譯，層次分明的引出，則可解決引徵資料反復查稽困難以及語文和理解上的障礙。復次，以自然語文處理或計算語言學中的種種定量分析方法，甚至可以提供新的研究工具，新的方法，新的概念及模式來分析和了解古籍記載的資料。這些古籍應用環境的改變，事實上，已造成漢學革命性的變化。

## 二、研究背景與本文之範疇

中央研究院（台北）有鑑於此，十年前即開始從事於古籍自動化的計畫，迄今以來未曾稍歇，（請參閱：中央研究院歷史語文研究所機讀古籍總目）。在系統發展方面，為建構廿五史全文資料庫所發展的中文文章處理系統（CTP, Chinese Text Processor）已沿用近八年

【1】實在有更新之必要，本文所論者，實即 CTP 之更新構想。

目前試圖解決之問題，主要包括有：古籍之版本問題，校勘問題，注疏問題，檢索問題等等，是故有必要對古籍做一番通盤之考量，並擷取其抽象結構，以期所發展之系統能適用於一般之古籍，而發展此系統所採用之文獻，則為各版本之文心雕龍，其中又以詹《文心雕龍義證》作為發展之主軸。

本研究依照標準通用標誌語言（SGML, Standard Generalized Markup Language）設計古籍之標誌語言，並將標誌好之古籍檔案轉換為全文資料庫及關連式資料庫，以構成超文件（Hypertext）式的系統結構和檢索功能，（本文將Hypertext譯為超文件，乃取其「性能超越目前印刷品的文件」之含意）。不過，本文之目的只在說明古籍轉換之處理程序、原則、目標、標誌前及標誌後之文獻訊息，同時因系統程式部分尚未完成，故有關SGML、超文件之技術，請參閱周亞民之文章【2】，系統軟體部分在本文中暫不介紹。

## 三、古籍之界說

何謂古籍？這是一個眾說紛紜的問題。由不同領域的學說及資料整理的角度而言，都各有不同的看法，甚至，在同一領域裡對古籍的定義也顯得籠統模糊。本文試由不同角度入手，對電子化古籍之提出綜合性的界說：

### （一）由圖書形制之演變來看

根據嚴文郁教授在中國書籍簡史一書【3】，將我國書籍之發展與形制，依資料記載之材料與印刷技術，分為五個時期：

- 萌芽時期：上古之文字流傳工具有甲骨、金石、縑帛，都是就地取材，材料既不一致，故不定型。
- 第二期：從戰國到東漢末，此時主要的書籍書寫材料為竹木簡，將簡編連成策，是我國書籍發展階段。
- 第三期：由戰國到唐末，書籍以抄寫的紙本為主。紙書的行式為卷軸，在帛與紙卷中貫上軸以便舒展，故亦稱為卷軸時期。
- 第四期：從五代到道光間中英鴉片戰爭，書籍由寫本進入印本時期、雕版印刷是主要的生產方法，書籍形式也由長卷發展為旋風裝、蝴蝶裝、包背裝、線裝之冊葉方式。
- 機器印刷時期：由道光二十年到現在，印刷術從手工業進到機械工業，這是受西方技術傳入的影響，冊葉制度雖仍保持，但書籍的形式已由線裝變為平裝和精裝，印刷技術也漸轉為西洋的凸版、平版、及凹版印刷了。

故若由圖書形制而言，古籍之定義乃指機器印刷以前之書籍，也就是以中國傳統技術所製作的書籍。它在時間的劃分上並不是道光年間，因為清朝可說是中國雕版印刷的興盛時期，私家雕版刻書者相當多【4】，甚至在清末民初，仍有不少線裝書出現。

### （二）由目錄學的角度而言

將中國典籍作有系統之整理分類者，始於西漢劉向，向卒後，子歆繼承父業，總括群篇，撮其旨要，著為七略。故有輯略、六藝略、諸子略、詩賦略、兵書略、術數略、方技略，大凡收書一萬三千卷。這是中國漢朝、先秦以前典籍之內涵與數量。自漢以後典籍之分類有四分法、五分法、六分法、九分法及十二分法。基本沿襲劉向父子之傳統，只不過略有增減罷了。

【5】在各種分法當中，以四分法為歷代目錄最主要的方法，經、史、子、集及其下所含蓋的細目，歷朝學者無不耳熟能詳。

在西學未傳入之前，中國典籍皆能納入四庫之範圍（佛、道之典籍另有專門分類），以清朝最大的一套目錄，四庫全書總目而言，收錄的書有三千四百七十種，七萬九千一十八卷，四庫未收之存目有六千七百六十六部，九萬三千五百五十六卷。【6】自鴉片戰爭以後，海禁大開，西學輸入，洋書原本源源而來，非復四、七諸法所能概括，以今日仿杜威之中國圖書分類法而言，其內容範圍與四庫已相去甚遠，故一般目錄學家，大都認為，依傳統之四部等分類法所收之書為中國古籍。

所以，今日各藏有善本、線裝等古籍之圖書館，大都採用適合古籍之分類或編目等方式，將古籍另行處理。故北京圖書館學系和武漢大學圖書館學系即合編圖書館古籍編目一書，其中將古籍定義為：「1911年以前歷朝之刻本、寫本、稿本、拓本等，從圖書館古籍編目工作的實際情況看，1911年以後的影印、排印的線裝古籍，如四部叢刊、四部備要等書也都屬古籍。要從時間上截然劃分是困難的，以1911年為下限，也只能說大致符合圖書內容及形制的實際情況。」【7】

### （三）由研究使用的角度而言

有關古籍的實物處理，為工作方便，必須由形制或年代上加以規定，以便遵循。但若由研究使用之立場而言，凡有關一主題之資料，無論其形制、媒體等有何不同，都希望能放置一處，以便合併參考。是以，就研究使用者而言，所謂的古籍研究，一般即指對中國先人智慧之探究，他們除了會參考古版刻書外，也需參考今人之解釋，今日出版之資料。

### （四）以電子化之目的而言

綜合以上觀點可知，不同的領域有不同的立場，所以對古籍之定義不易一致。將古籍電子化之目的主要在提供研究者使用，所以在此古籍應指前清以前先人智慧研究有關之古今資料。由於一書經過多年之流傳，後人對於前人之著作每有不解或獨到之領悟處，必定加於原書上，並於再版重新開離時，加印於離版上，是故今日我們所見之古籍，或今日印刷之新版，其內容均已不只是作者原著而已，而是加上了後人之注、疏、校勘、評、句讀、白話翻譯等之言論，換言之，我們要處理的古籍已非作者一人之著作，而是歷代多人集體創造之著作，我們不只要轉換古籍的某一版本，而是轉換和此古籍相關的多種版本及參考資料，所以其內涵不只和研究使用有關，也和圖書形制、目錄學、版本學的觀點有關，除了能將各種不同版本之資料整合利用、連結檢索外，還需保持原書之形式，這是本研究 and 二十五史全文資料庫差別最大之處。由於這種古籍具備了Hyperbook之性質，故以超文件之理念設計之。

## 四、電子版本之定義

電子版本一詞，望文生義，乃指能以電腦處理之數位化文獻(document)，或稱為機讀版本(machine readable version)。就技術意涵而論，Peels等人曾參考ISO之文獻處理標準，將文件(text)一詞定義如下：【9】

Text is all information for human comprehension that is intended for presentation in a two-dimensional form. Text consists of symbols, phrases, or sentences in natural or artificial languages, figures, formulars and tables.

由此可知，文件乃指以二維形式來表現、闡釋的所有資訊，它包括自然語言及人工語言中的符號、片語、句子、圖、表格、公式等。換言之，凡用文字符號所表達的知識即為文件。整合來說，文件是由文件元素(document element)構成，而文件除了這些元素之外，還包含了作者所要表達的文件或概念，文獻是文件的集合名詞，而電子版本所要呈現的就是有完整意義的文獻。在本研究中，我們將重點放在文獻而非單一之文件。

雖然目前有關文件編輯、桌上排版、超媒體等技術均有長足的發展，但和數據處理(Data Processing)技術相較之下，文獻處理(document processing)在計算機領域裡的發展仍屬遲緩。值得注意的是，目前人類社會中仍有99%的資訊是紙本文獻【10】，而根據Tsichritzis的說法，在辦公室環境中，有70%的資訊是非數據性資訊。【11】換言之，人類大部份的文化遺產、智慧結晶，即文獻，都尚未電子化。

早期的文獻處理系統，由於未能對文獻之邏輯結構、版式及形制作合適的區分，故傳統上皆將文獻以循序檔來儲存(Sequential file)。這種檔案結構，隨應用環境而變(application-dependent)。當系統將這種檔案轉換到另一應用環境時，往往會遺失原書部份資訊，同時也無法和其他電子文獻整合為共用資訊。

今日，已有一些文件編輯、桌上排版及超文件系統，可直接檢索文獻中的某一片段，且其機讀文件的形式也不因應用及設備而異(application and device independent)，使得儲存的資訊得以分享、交換，又能滿足不同的應用需求。【8】之所以能達到這種境界，就是將文獻做了標準化的標誌(Markup)的工作。

文獻標誌並非始於電子化作業，早在人工編輯出版年代，編輯人員即一枝色筆在手，將文獻作印刷格式、字體安排和版面設計之注解，這就是標誌的一種。在電子文件中，標誌通常是對文件元素及特定屬性之內容做標示記號(tagging)或編碼(coding)，可惜的是，目前幾乎所有的電子出版系統，全文資料庫、超文件超媒體系統等，都有其專屬的標誌符號，故一位使用某程式處理文件的作者或出版者，通常無法將該文件送到另一系統上作編輯，或和他人交換文稿。【12】這就是國際標準組織為何要制定SGML的原因。

綜合上述觀點，在此，我們以電子文件應具備之特性，對電子版本作進一步之定義：

1. 它是以目前發展的標誌語言能夠做標示的古籍文獻（詳請參閱下文），所以它是一經過文獻結構及內容標誌之數位化文獻。
2. 它的形式不因應用環境及設備而異。
3. 它必需可以互相交換，以供大家共享。

## 貳、古籍電子化之目標

設計古籍系統時可以利用的知識有：中國語言文字之知識，文件結構之知識，統計知識，古籍背景知識及相關之電腦知識等五大類【1】。僅利用這些知識的目的，是希望設計出來的系統，能滿足一般古籍電子化之需所以，本計畫之系統程式，非專為某一部古籍、某一版本而設計，乃是一通用於古籍之程式。茲將涉及之知識綜合如下：

### 一、有關古籍之版本資訊

版本資訊是研究古籍不可或缺的資訊，刻工的良窳，校勘的審慎與否，在在影響內容的正確性。是以研究國學者，莫不以目錄、版本學為治學的基礎，前者在辨章學術，考鏡源流，後者則在探尋善本，以窮究原作之真貌。過去由於版本四散，有關版本之研究難以有系統的進行，而今善本古籍大都已集中於各地圖書館，故若能藉電腦之力，設計一易於使用之版本系統，相信對於目錄、版本與國學之研究，都極有助益。

利用版本資訊，我們希望本研究的發展系統能夠支援下列的工作：

#### (一) 建立古籍各版書誌資料庫

書誌乃在說明各版之刊刻者、刻工、刊刻地點、刊刻年代、版式行款、圖記、收藏沿革、版本優劣等問題。其性質猶如西方之 Annotated bibliography。我們將以 SGML 來標誌書誌，並定義出書誌之文獻結構，使書誌能作完整的呈現及靈活的檢索。

#### (二) 反向建立古籍之版刻沿革表

將某書歷代刻版之傳承沿革作系統化的呈現，是版本學者之目標。後代所刻之書必以前代所傳為底本，有時底本不只一部，且參考狀況有多寡輕重之別，故在上述之資料庫中，有關底本之欄位資料是可重複的，同時也必需定義出參考量之多寡。在此，我們須將兩千年中西曆對照表建為貫串版本之縱軸，對映到上述資料庫中，如此可成一清晰有系統的沿革表。

### (三)雕版原貌之呈現

歷代各朝或各家所刊刻之版，皆各有特色。版本不僅記載先人智慧，其本身也是中國藝術與印刷技術的表現。雖然有書誌的描述，但百聞不如一見，特別是有關套色、印記、牌記、字體、書法、版式行款等等說明，若有實物可印證，則更相得益彰。所以我們計劃利用影像技術處理，這將使系統具有多媒體 (Hypermedia) 的特色。

## 二、有關古籍之校勘資訊

古籍輾轉傳刻愈久，訛誤之處愈多，歷代學者往往花甚多心力在校勘考證上。故古籍電子化時，校勘資訊的提供非常重要。用電腦來處理這部份資訊時，希望能夠支援下列之工作：

### (一)可呈現不同版本間之差異

各版間之文字差異，一般而言約有用字不同、某版有多字、某版有漏字、文字顛倒等狀況。在此我們可能用電腦比對，或自動偵錯與改正之技術來處理。

### (二)校勘意見之呈現

凡文字有差異處，即有學者參酌他書或逕書己學而作的校證，此意見與正文連結參考，是研究者之習慣。故注釋中有關版本之考證者我們將予標出，賦予校勘之屬性，以便應用。

## 三、有關古籍之注釋資訊

注釋資訊的提供是本系統之核心部份，完整的注釋(即匯集各家注解之會注)是讀者理解原書不可或缺者。所以系統須能將本系統所用的各版書之注釋，去其重複，針對某一句正文提供給檢索者。除了會注功能外，檢索者也可由注釋者或書名檢索，以瞭解各家之特色。此外注釋的性質也有所不同，有解釋詞義、注明出典、分析名物、說明制度、闡發義理、參證史實、確定地理、賦予音切、通釋大意、及貫串講解者，目的各有不同。若能以注釋性質來檢索，將可便利不同領域之學者的研究。是故有關注釋資訊之提供希望能達到下列文件檢索功能：

- (一)有關某句正文之瀏覽(browsing)、導閱(navigation)、檢索(retrieval)，及其所有注釋的呈現
- (二)有關各家注釋的瀏覽、導閱、檢索，同時也可了解注釋之傳承
- (三)按注釋性質檢索各類注釋

## 四、古籍之全文檢索

全文檢索是全文資料庫必具備的基本功能，而在古籍方面我們希望能達到下列主要功能：

- (一)可用各種屬性檢索，例如：書目資料中之書名、作者、地名等
- (二)可用自然語言關鍵字詞檢索。在此我們將應用 Vector Space Model [14]、Cover Coefficient [15]。相關係數，和統計等數學模式來處理。
- (三)可做古籍輯佚之工作

有甚多的古籍至今我們只見其名，不見其書，不知何時即已亡佚，但由前人著作中卻可窺其吉光片羽，所以若能有系統的將古籍機讀化，借助電腦整合檢索的功夫，相信能使不少古籍重現部份原貌。

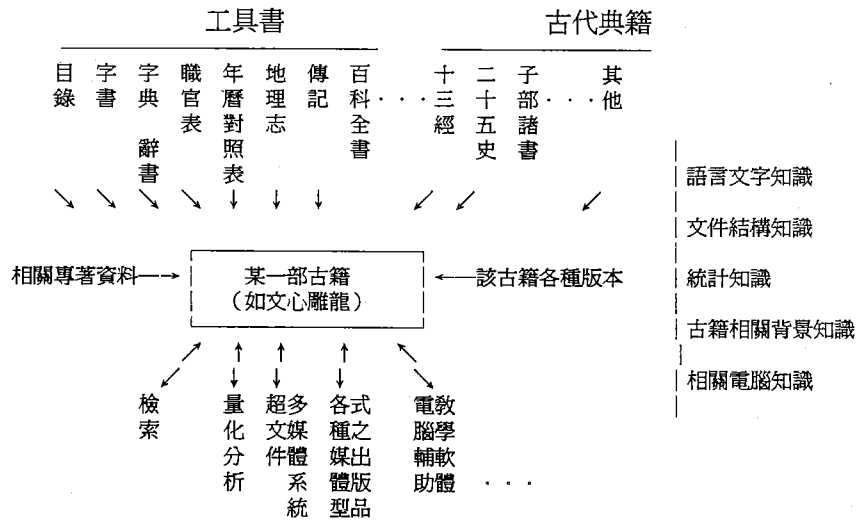
## 五、相關參考書目與參考文件的提供

若能將古籍全部轉換為電子資料，則當某書之研究必須參考另一資料時，隨即可尋獲連結起來，而不必再動手動腳找資料，上窮碧落下黃泉，浪費時間了。例如，目前台灣已有二十五史全文資料庫，將來文心雕龍中有提到史書處，即可以網路及檢索介面找到二十五史中之相關資料。

此外完整的參考書目，也是研究時不可或缺者。所以我們將蒐集參考書目建入此系統中，改變資料庫向來將書目資料庫與全文資料庫分開建檔的作法，提供使用者完整的資訊檢索服務。

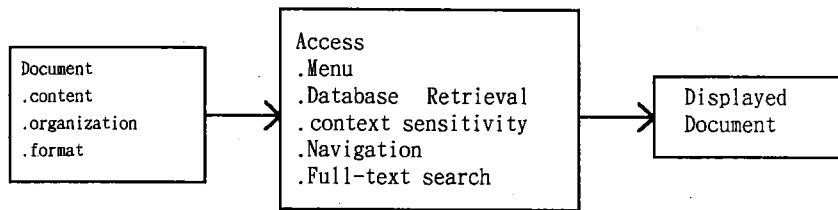
## 六、研究者之眉批、筆記

學者研讀文獻時，每有批評或心得處，都習慣於書旁寫下眉批，或於私人簿本上紀錄心得。設計電子文獻系統時，若能提供電子筆記本或卡系 (note card system) 供私人使用，則研究者對於心得眉批的整理應相當得心應手。以上需求乃就使用者之角度而言，若再加上其他領域之相關應用，則古籍電子化之目標與環境可以下圖表示之：

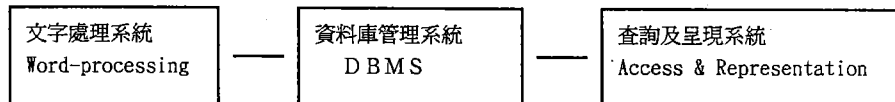


## 參、工作程序

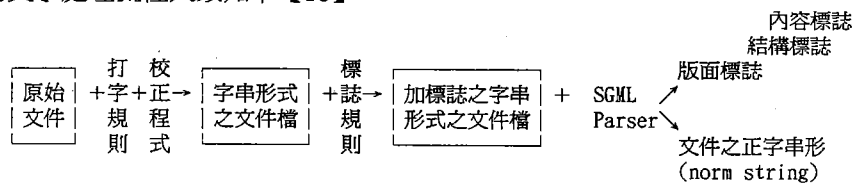
電子文獻系統一般包含二個必要的部份，一為儲存之數位化資訊，二為該儲存資訊的查尋方法，茲以下圖說明之：【14】



若以工作單元而論，電子文獻系統大概可分為下列三個部份：



上圖中的每一系統，都包含非常複雜的動作。例如 Salton G. 認為有關文字處理工作通常包括文件產生、文件修改、文件校正、文件排列、文件格式化等。【14】本研究中有關資料的文字處理流程大致如下【13】：



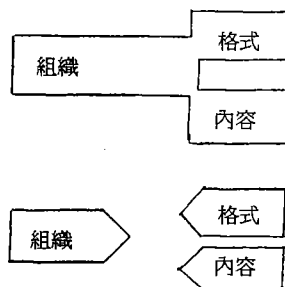
## 肆、文獻之一般性結構與內容

任何文獻，無論其儲存媒體為何，都可視為一完整的樹狀結構實體。【16】若此實體為傳統之某一本書籍，則此書可能分解為：章→節→段→句→字，之樹狀結構。任何文獻所包含之資訊內容，一般而言，大致不外三種類型：

1. 內容(content)、2. 格式(format)、3. 組織(organization)。

內容是文獻之主體，它可能是文字、圖片、或聲音等。格式則決定內容如何被顯示，它包括字體、版式、色彩等等。組織則說明文獻內之論題(topics)的順序及其交互關係。

在一紙本文獻中，上述三種資訊內容是不可分離的，如右圖所示：【14】



但在一線上電子版本中，則將三者完全分開，系統設計者必須分別管理這三種元素，如右圖所示，但又需使之能組合在一起應用。

若由使用者的角度而言，文獻通常包含兩種結構：一為明顯結構 (explicit structure)，一為隱晦結構 (implicit structure)。【14】明顯結構是一文獻的正式組織 (formal structure)，例如不同層次之標題 (headings)，明晰之參照 (cross reference)，及目次等均屬之；隱晦結構則指文獻內容之屬性及其相互關係，例如作者由兩個不同的觀點來探討同一主題，正文之注釋具某類別屬性等等。隱晦的部份通常是作者真正希望讀者了解的部份，同時讀者也需要具備相當之背景知識，才能完全心領神會，不過在轉換紙本為電子版本時，這種隱晦的資訊內容往往容易被忽略而漏失掉。所以在一紙本文獻的轉換過程中，樹狀結構的分析與儲存；如何將內容、格式及組織作合理的區分與組合；如何將文獻之隱晦結構轉變為明顯結構，則是系統設計之成敗關鍵。

## 伍、古籍之內容、結構與處理原則

### 一、古籍之內容與結構

古籍之內容如前所定義，已非一單人著述，而是集體創作；已非一本書，而是一套書；若一古籍資料庫要滿足各類學者之需求，則轉換連結的對象亦非一個版本、一種古籍所能滿足。

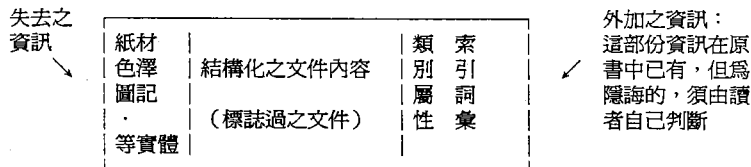
古書結構和現代書籍相同，也是一樹狀結構，但在內容的表現上，卻有其不同之特色，例如：在標題方面，多為小題在前，大題在後；文章不句讀；行文簡潔，語體為文言文，少有贅字，若以關鍵字作全文檢索，則幾乎字字皆為關鍵；古籍之序多和本文合併，如文心雕龍之序志篇即為全書之總序，而作者將之置於第五十篇；此外，古書本文多無明顯之篇章段落結構；注疏資訊是古籍最大之特色，它除了具有現今書籍之註解 (footnote) 性質外，尚有解釋詞義，分析名物，說明制度，闡發義理，參證史實，確定地理，賦予音切，通釋大義，貫串講解者。最有趣的是，這些注解都不是著者寫書時加上的，而是讀該書的人賦予的，注釋的呈現形式大約有：隨文附釋、單疏或單注本、注疏合刊本、注文寫在書籍之天頭地腳、注文與正文分開但合刻(印)在同一本書中等等，不一而足。

此外，古籍的研讀以選擇優良的版本為首務，古書由於時代的區隔，抄書者與刻工的良窳等因素，造成版本品質上的差異，是以不同版本之文字會有不同。【19】這和凡有新版出現，舊版即失去價值之現今圖書不同，因為今日出版之書，無論新版、舊板，均為同一人之作品，但不同版本之古籍其智識內涵差異頗多，多有不同之注疏者、評者、為序者、校勘者等等，後人的智慧不斷的溶入同一本書中。

## 二、古籍之處理原則

在古籍之電子化的過程中，最好是不遺漏原書任何訊息，同時也不要參雜任何的訊息，這是古書電子化時相當重要的原則。然而，這原則卻無法忠實地完全做到。是故如何在電子化環境限制下，以求其盡善，就成了設計上優劣的關鍵。

古籍電子化時，會失去原書的部份訊息，如原版之紙張、色澤、書法、刻書者之牌記等實體資訊，雖然這些資訊也可利用影像科技等彌補，但仍究是一虛擬實體(virtual reality)；此外，也會因檢索及解說之便，將隱晦的信息轉變為明顯的信息，而無法避免地將加入些其他訊息，如類別屬性、索引詞彙等，茲以下圖說明之



此外，建立資料庫時若不賦予資料主題性質，將資料作系統化的分類，就好比一座汗牛充棟、森然有序的圖書館，卻無圖書目錄一般，不知如何應用，所以無論是書目性資料庫、數據式資料庫、全文資料庫，都會賦予資料適當的主題性質。此外，在注釋部分，我們也將對各條注釋賦予類別性質。文心雕龍雖然是一部有關文學理論的書，但其注釋卻可分別提供訓詁學家、聲韻學家、版本學家、文學家、史學家、美學家、哲學家等等對古代文字、聲韻、版本、義理、制度、思想、價值觀等之研究參考。故將注釋依性質區分是必要的作法。

## 陸、文心雕龍之內容結構與標誌

### 一、實驗對象的選擇

本研究之最終目標乃在建立一個可處理任何古籍的共通系統，研究中則以文心雕龍為實驗對象。

文心雕龍一書，據清儒劉毓崧的考證，約寫成於南北朝時代南齊之末，等於西元六世紀初年【11】，至今約有一千五百年的歷史。在這漫長的歲月裡，它歷經寫本、雕版、活字版、和鉛字版等印刷技術之演變，及各代天災人禍之衝擊而流傳至今。原來它只是劉勰一人的著作，但隨著時間的增長，為了瞭解闡釋它，各版中慢慢的加上注疏校勘等資料，形成了一本自己充足的智識群聚。從今天的義證、校注各書來看，文心雕龍已非劉勰一人獨著，其間並有注者、疏者、評者、校勘者的心血結晶。選擇文心雕龍為實驗對象，主要著眼於它完整的將南北朝前之著作做了極有系統的分析，他在書中序志篇云：

「詳觀近代之論文者多矣：至於魏文述典，陳思序書，應璩文論，陸機文賦，仲治流別，弘範翰林，各照隅隙，鮮觀衢路，或臧否當時之才，或銓品前修之文，或泛舉雅俗之旨，或撮題篇章之意。」【18】

這裏，他提出的魏文帝曹丕典論論文，今見於昭明文選卷五十二，陳思王曹植與楊德祖書，今見於同上書卷四十二，應瑒文質論，今見於藝文類聚卷二十二，陸機文賦今見於昭明文選卷十七，摯虞(仲治)文章流別論及李充(弘範)翰林論，分別亡佚於唐宋之世，其殘文今見於清嚴可均所輯全晉文卷七十七及卷五十二者皆不及原書之十分之一。此外，桓譚(君山)新論，嚴輯於全後漢文卷十三至十五，然有關論文之語不多，劉楨(公幹)論文，除劉勰於風骨及定勢二篇引述者外，已不可復見。應貞(吉甫)遺說，載籍久亡；陸雲(士龍)論文，今但見其與陸機的書信中一些偶發意見而已。如果據劉勰當時所「詳觀」的前人之論，不管是尚存或已散佚，在篇帙與內容組織上，可說沒有一個比得上文心雕龍，而歷代以來，研究文心雕龍者雖多，但卻未見相等的著作出現，所以說這部書是一空前且絕後之作，實不為過。【17】

而若以目錄學的角度來看，此書亦可說是先秦著作目錄。它讓我們了解南北朝前之重要著作有那些，可謂即具索引參考價質。所以以計算機來處理這部富多層意義的古書，在實質意義與電腦技術上，都極具代表性。



## 二、文心雕龍版本的選擇

文心雕龍版本甚多，選擇何版作為系統建立的對象，亦為一重要課題。基於我們已訂之目標，選擇下列三版為建檔基礎：

### (一)范文瀾 《文心雕龍注》【19】

辛亥革命以來，在大學講授文心雕龍始於劉師培，黃侃繼之。劉師培未發講義，當年羅常培先生曾用速記法作了記錄，整理出來發表的只有兩篇，取名左龔文論。黃侃在講述過程中寫了文心雕龍札記，雖然沒有編完，但極具功力。范文瀾從黃侃受業，以清黃叔琳文心雕龍輯注為底本，而又附錄了鈴木虎雄、趙萬里、孫蜀丞諸家校語，先編成文心雕龍講疏，後改寫為文心雕龍注，成為在注釋方面貢獻最大的一部，五十多年來，文心雕龍研究者，大都以這部書為依據，來進行探索。是故，本研究中擇其為底版之一。

### (二)詹 之 《文心雕龍義證》【20】

抗日戰爭發生後，楊照明在郭紹虞、張孟劬指導下，于燕京大學研究院以文心雕龍研究為畢業論文，一九五八年刪訂出版，取名為《文心雕龍校注》。王利器在這部書稿之基礎，於校勘方面加以擴大，寫成文心雕龍新書，一九八〇年修訂出版，改名文心雕龍校證。詹 之《文心雕龍義證》即以此書為底本，兼採各家之說，取材甚廣，對於近代各種資料，無論聽課筆記，斷簡殘篇，已刊未刊，多有採擷。本身已具會注、集解之性質，是以選擇此書為系統底版之二。

### (三)張立齋 《文心雕龍考異》【21】

此書不知為何人何年所出版，在詹瑛之序例中亦未提及。但作者自序中知其成書年代應在范文瀾、楊照明、王利器之後，詹瑛義證之前。此書難得之處在於它是一本有關文心雕龍之校勘專書，其據以參考校證之本有：1·英京所藏唐人寫本殘卷影片，自原道篇贊曰以下，至諧隱篇題止；2·涵芬樓影印之嘉靖本；3·萬曆楊升庵批點梅慶生音註本，哈佛圖書館所藏；4·凌雲本，原刻極精，五色套印，亦哈佛所藏；5·黃叔琳輯注及紀評合刻本，為道光十三年初刊於南廣節署，硃墨兩色套版，哥大圖書館所藏。作者據此五本，逐篇互校，手自甄錄，旁參御覽及范文瀾本、楊照明本、王利器新書，所引諸舊刻，皆翻檢再三，才下己說。作者為學態度嚴謹、取材有國內未見之本，是以本研究選擇此書為系統底本之三。

## 三、文心雕龍之內容與結構

### (一)紙本文心雕龍之內容與結構

文心雕龍是一本講寫作的書，序志篇開宗明義曰：「夫文心者，言為文之用心也。」可見「文心」是探索作文用心，而「雕龍」是指雕繙成體，乃就文采而說。全書的文藝理論主要是重各體文章的寫作和對各體文章代表作家之評論總結而來。

全書共分十卷五十篇。第五十篇序志篇可說是全書的總序，在此他說：「蓋文心之作也，本乎道，師乎聖，體乎經，酌乎緯，文之樞紐，亦云極矣。若乃論文敘筆，則囿別區分，原始以表末，釋名以章義，選文以定篇，敷理以舉統，上篇以上，綱領明矣。至於剖情析采，籠圈條貫，摛神性，圖風勢，苞會通，閱聲字，崇替於時序，褒貶於才略，悵於知音，耿介於程器，長懷序志，以馭群篇，下篇以下，毛目顯矣。位理定名，彰忽大易之數，其為文用，四十九篇而已。」在此，劉勰指出全書之結構與各篇之主題性質。

第一部分是文之樞紐，就是寫作的關鍵問題，也就是他的批評標準，這部分包括：原道、徵聖、宗經、正緯、辨騷五篇。

第二部分是文體論，分論文和敘筆，論文是講有韻文的各種體裁，含：明詩、樂府、詮賦、頌贊、祝盟、銘箴、誄碑、哀弔、雜文、諧讖十篇；敘筆則談無韻文之各種體裁，含：史傳、諸子、論說、詔策、檄移、封禪、章表、奏啓、議對、書記十篇。

第三部份剖情析采是創作論，結合論文敘筆，探索創作上各種問題它包括1·剖情析采的根本：神思、體性、風骨、通變、定勢五篇；2剖情析采的結合：含情采、物色、熔裁三篇；3·剖情析采的方法：含聲律、章句、麗辭、比興、夸飾、事類、練字、隱秀、指瑕、養氣、附會十一篇；創作論序言：總術一篇。

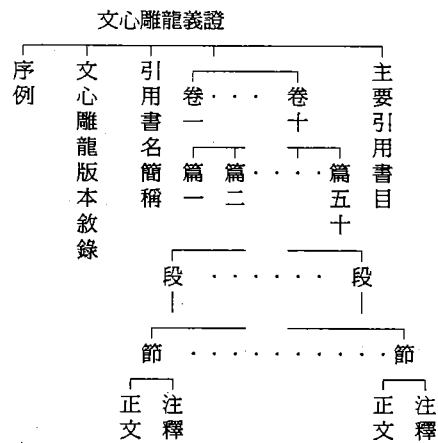
第四部分是文學史、作家論、鑑賞論。含時序、才略、知音、程器。

第五部分序志篇是全書的總序。

原書正文部分之目次形式如下：

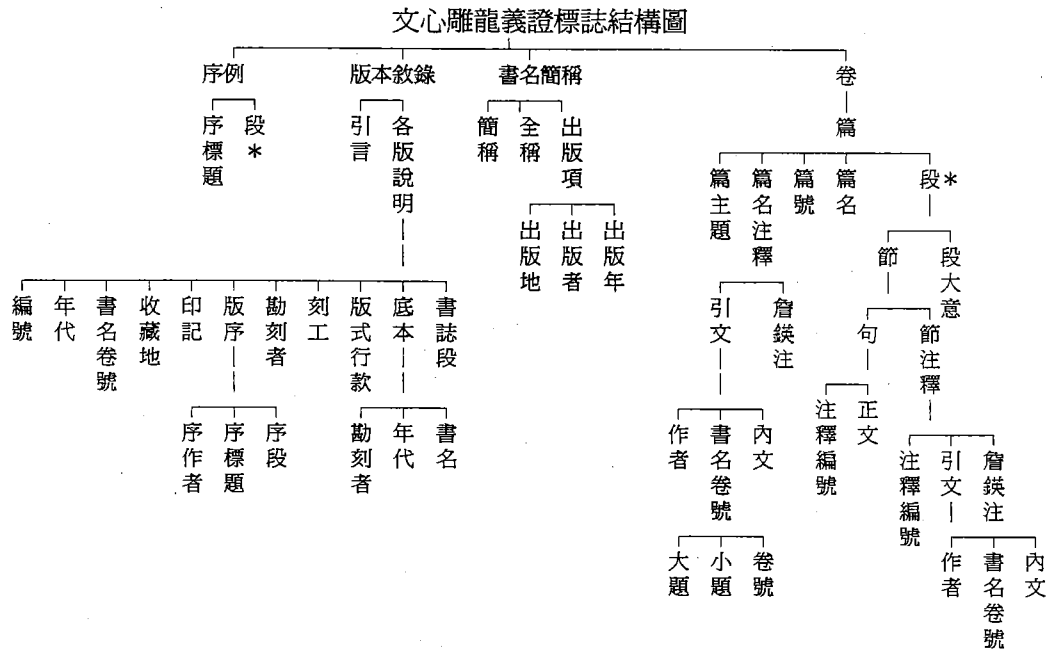
卷一	卷二	.....	卷十
原道第一	明詩第六	.....	
徵聖第二	樂府第七	.....	
宗經第三	詮賦第八	.....	
正緯第四	頌贊第九	.....	
辨騷第五	祝盟第十	.....	序志第五十

這是文心雕龍原書之內容形式，也就是劉勰獨力完成的部份，但後來之雕版中，內容都不僅如此而已，以本研究所採用之詹瑛義證而言，除在篇下又分段，段後有注釋、段大意外，正文前有序例、文心雕龍版本敘錄、引用書名簡稱、正文後附有主要引用書目等。茲以簡單之樹狀結構圖表示如下：



## (二) 文心雕龍之標誌

鍵入電腦之文心雕龍，若要滿足系統所要達成之目標，必需作更細緻的建檔處理。而建檔之前，我們選用ISO的標誌制語言SGML，對文獻作結構上的標誌【22】，茲以下圖說明文心雕龍標誌結構：



上圖之標誌結構是系統設計時希望能達成之目標，換言之，其中所列之項目是超文件連接的結點，也是讀者查詢的檢索點(access point)。不過真正在標誌文件時，正文部份只標到節，其他細節則以系統用字址(address)作比對連接，而注釋部份也會再分析各項特徵，教導系統辨識之能力，例如若能收集各書中有關校勘之固定用語，如詹義證通常用某某校傍作旁，張立齋用某某本咄作差，范文瀾則將校勘資訊放在正文下之雙行小字，這些資訊都有規則可循，也可以設計為系統處理之規則。如此一來，文字的標誌即可省事不少。同時，為了查詢之便利及不同版本之溶合，相關之對照表的建立也是必要的，如各書所用之書名簡稱與全稱對照表，注釋編號形式對照表等等。

實際之標誌文件可以由下列之文件類別定義 (DTD, document type definition) 窺得一般：【8】

```

<!Element 文心雕龍 -- ( 序例, 版本敘錄, 書名簡稱, 卷 +)>
<!Element 序例 -- ( 序標題, 序段+)>
<!Element 序標題 -- (# PCDATA)>
<!Element 序段 -- (# PCDATA)>
<!Element 版本序錄 -- (引言, 書誌)>
<!Element 引言 -- (# PCDATA)>
<!Element 書誌 -- (編號, 書名卷數, 書誌段+)+ (年代 | 收藏地* |
    印記* | 版序* | 勘刻者* | 刻工* | 版式行款* | 底本*)>
<!Element 編號 -- (# PCDATA)>
<!Element 書名卷數 -- (# PCDATA)>
<!Element 書誌段 -- (# PCDATA)>
<!Element 年代 -- (# PCDATA)>
<!Element 收藏地 -- (# PCDATA)>
<!Element 印記 -- (# PCDATA)>
<!Element 版序 -- (# PCDATA)>
<!Element 勘刻者 -- (# PCDATA)>
<!Element 刻工 -- (# PCDATA)>
<!Element 版式行款 -- (# PCDATA)>
<!Element 底本 -- (# PCDATA)>
<!Element 書名簡稱 -- (簡稱, 全稱, (出版地 | 出版者 | 出版年))>
<!Element 簡稱 -- (# PCDATA)>
<!Element 全稱 -- (# PCDATA)>
<!Element 出版地 -- (# PCDATA)>
<!Element 出版者 -- (# PCDATA)>
<!Element 出版年 -- (# PCDATA)>
<!Element 卷 -- (篇+)>
<!Element 篇 -- (篇名, 篇號, 篇主題+, 篇名注釋+, 段)>
<!Element 篇名 -- (# PCDATA)>
<!Element 篇號 -- (# PCDATA)>
<!Element 篇主題 -- (# PCDATA)>
<!Element 篇名注釋 -- (引言* | 詹注*)>
<!Element 引言 -- (# PCDATA)>
<!Element 詹注 -- (# PCDATA)>
<!Element 段 -- (節+, 注釋+1, 段大意)>
<!Element 節 -- (# PCDATA)>
<!Element 注釋 -- ( 編號, (引言* | 詹注*))>
<!Element 段大意 -- (# PCDATA)>

```

## 柒、結論

中國典籍浩如煙海，然古籍之書寫方式和白話文不同，致使豐富的先人智慧多被現代年輕一輩所忽視。而古人知識有前後呼應之特色，是故，無論是春秋戰國的諸子哲學、漢魏各家的傳經事業、韓柳歐蘇的道德文章、程朱陸王的心性義理、清儒的訓詁考據等等，在典籍上都可見其傳承軌跡，所以古籍或先人智慧與歷史等研究，特重在融會貫通，而後成一家之言，這種宏觀的境界需從最基本的博覽群籍做起，故如何將古籍電子化，使學者們能在群籍中游走閱覽，實在甚具研究價值。

雖然本文未對系統作完整說明，但對系統設計的重要思考已報告如上，我們懇切的希望獲得同道之指正，更希望能拋磚引玉，期望各界能凝聚力量，對傳統文化遺產有所貢獻。

## 參考資料

1. 謝清俊, <從廿五史全文資料庫的經驗談中文文件檢索系統設計的考量>, 第三屆中文信息處理國際會議 專題技術報告, 北京, 1992年10月26-28日
2. 周亞民, 《使用SGML與物件導向資料庫轉換古書為Hypertext》(私立元智工學 碩士論文 1993年, 改寫後發表於本會議)
3. 嚴文郁, 《中國圖書簡史》(台北市:商務, 1992)
4. 呂彼得, <我國歷代版刻的演變>, 王國良, 王秋桂合編, 《中國圖書文獻學論文集》(台北市:明文, 1986)
5. 《中國古典文獻學》(台北市:木鐸, 1983): 54.
6. 呂紹虞, 《中國目錄學史稿》(台北市:丹青, 1986): 154.
7. 北京大學圖書館學系, 武漢大學圖書館學系合編, 《圖書館古籍編目》, (北京:中華書局, 1985)
8. <ISO DRAFT International Standard DIS8879. Vocabulary Part two of Eighth Working Draft> International Standard, Information Processing System, Text Preparation and Interchange, Processing and Markup Language (ISO TC97/SC18/WG8 N4, International Organization for Standardization, Sep. 1984).
9. Peels, Arno J. H. M., Janssen, Norbert J. M. and Newijn, Wop, <Document architecture and text formatting> ACM Transactions on Office Information Systems, vol 3. no.4 (October, 1985): 349.
10. Schouten, Han, <SGML \* CASB: The storage of documents in database> SGML User's Group Bulletin. vol 4, no.1 (1989):1.
11. Tsichritzis, D. B. ed. 《Office Automation: concepts and tools..》1985.
12. Reynolds, Louis R. 原著; 吳傑中譯, <電子書> O&I (1992年6月)
13. 本研究所用之工作技術規範:
  - a. 陳瓊婷 張翠玲, 詹鏗本《文心雕龍義證》打字規則, 中央研究院資訊科學研究所文件處理研究室, 1993
  - b. 陳瓊婷 <文書校對程式>使用說明書, 中央研究院資訊科學研究所文件處理研究室, 1993
14. Salton, Gerard, 《Automatic Text Processing》(Reading, Massachusetts: Addison-Wesley, 1989): 275-312.
15. Can, Fazli & Ozkarakan, Esen A., <Computation of term document discrimination values by use of the cover coefficient concept> Journal of the American Society for Information Science, 38(3), (1987): 171-183.
16. Horton, William K. 《Designing and writing online documentation: Help files to hypertext》(New York: John Wiley, 1990): 4.
17. 王夢鷗編撰《文心雕龍》(台北:時報文化, 1987): 1.
18. 周振甫注《文心雕龍注釋》(台北:里仁, 1984): 916.
19. 范文瀾注《文心雕龍注》(香港:商務, 1960)
20. 詹鏗義證《文心雕龍義證》(上海:古籍, 1989): 4-5.
21. 張立齋《文心雕龍考異》(出版項不詳)
22. SGML 文獻請參考:
  - a. Herwijnen, Bric van, 《Practical SGML》(Dordrecht: Kluwer Academic Publishers, 1990)
  - b. Goldfarb, Charles F., 《The SGML Handbook》(Oxford: Clarendon Press, 1990)
  - c. Bryan, Martin, SGML: 《An author's guide to the Standard Generalized Markup Language》(Wokingham, England: Addison-Wesley, 1988)

## 誌謝

本文承張翠玲小姐排版, 及林晰和莊德明先生之各種計算機技術支援, 始竟全功。有關《文心雕龍義證》之打字工作則由中研院計算中心和長江打字行分別承擔。關於打字稿之校對則由陳瓊婷、李佩芝小姐使用楊允言和莊德明設計的<文書校對程式>而完成, 故有極佳之品質。此外, 美國密西根大學漢學研究計畫也幫忙選定詹本, 並提供一萬二千五百美元的補助, 這些助力都是使得本計畫能順利進行的幕後英雄。謹在此一併致衷之謝忱。