

## 國家圖書館古籍影像檢索系統—古籍文獻典藏數位化計畫系統建置概介

宋慧芹 國家圖書館特藏組

### 一、前言

「國家圖書館古籍文獻典藏數位化子計畫」中，預計將館藏重要善本古籍，以及金石拓片等珍貴館藏總計約 1 萬 2 千餘種數位化，並且於民國 90 年初步開發建置完成「國家圖書館善本古籍 Metadata 書目資料庫」、「古籍影像檢索系統」，將館藏古籍相關書目訊息以詮釋資料 (Metadata) 格式提供查詢。

為使該系統能夠更有效的提供服務，提升資料整理時的深度，以及日後資訊檢索服務的廣度，本館於民國 91 年初，提出系統擴充功能計畫，至 91 年底完成該案，交出了一張亮麗的成績單。以下就去91完成之系統擴充功能中，有關缺字處理、簡繁對譯、影像瀏覽、序跋全文檢索、人名權威、機讀格式 (marc) 結構轉換為詮釋資料 (Metadata) 結構等項逐一介紹。

### 二、系統介紹

#### (一) 缺字處理

數位化工作的苦惱之一為缺字問題：一方面建檔時需先做好造字管理，才能控管建檔品質；一方面使用者查詢時又必須先安裝好配合該系統使用之造字檔，才能順利閱讀內容。如此造成使用上的不便，且各系統造字檔各不相同，成為資料交換和管理上的難處，加上造字空間有限，超越之後，勢必造成碼位的重疊或衝突，在處理古籍文獻資料上更是有此擔憂。

本系統因為資料來源為館藏書目資料，其內碼為 CCCII，因此轉為 Big5 碼時，造成不少的缺字。為了避免造字的問題，以及讓一般使用者不需安裝任何造字檔或額外程式，決定不自行造字。本系統採用如下的做法：

1. 資料存檔：從書目資料庫轉至本系統時，若有Big5 碼所沒有的缺字，則以CCCII 內碼方式儲存。
2. 缺字記錄：用程式將每筆書目資料中含有CCCII 內碼的作記錄，如此可以知道哪個缺字出現幾次，出現在哪筆書目中，亦可知道各筆資料含有多少缺字。如此在作管理時可知道需優先處理哪些缺字，以達到最大時間效益。
3. 缺字管理：提供管理者可以彈性選擇每一缺字在顯示時要用相似的 Big5 字代替，或者用中

央研究院的漢字構字式處理。（見圖一）

4. 查詢顯示：系統在顯示資料時，會判斷內容是否含有缺字，然後選擇用 Big5 代替字或漢字構字式來作顯示，若是用漢字構字式，則會用圖片的方式顯示該缺字（在網頁中內嵌中研院漢字構字式的 Java Applet）。（見圖二）
5. 保留彈性：由於目前尚無中文缺字處理的正式標準之制定，且電腦軟硬體又與日俱新，因此本系統採用的做法，可以配合目前作業環境，日後轉置到新的標準時又可把轉置成本減到最低。
  - (1) 日後 win2000,winXP 普及，而 win98 已經淘汰時，系統內碼要轉為 unicode，此時書目資料可由 Big5 批次轉為 unicode；另外再從缺字記錄中找出已經出現在 unicode 的缺字，系統批次把書目資料中該缺字轉為 unicode。至於 unicode 仍無的缺字，因為漢字構字式尚未支援 unicode，可用程式預先轉為漢字構字式圖片的網址直接顯示，不再用 Java Applet。
  - (2) 若中研院漢字構字式已經支援 unicode，則仍和前述的做法一樣，只是 unicode 仍無的缺字還是用 Java Applet 方式來呈現缺字的圖片。
  - (3) 前述以上兩種轉置，主要花人工在比對 big5 缺字有無出現在 unicode，因為已經統計出約有七百多缺字，所以估計人力可以完成之，批次轉換內碼及缺字的程式亦不難完成，因此本系統採取目前之缺字處理方式。

## （二）簡繁對譯功能

本系統之內碼為 Big5，但為方便海內外眾多華人使用，特加入簡繁對譯功能，使系統能照使用者所選擇的內碼顯示網頁內容，擴大本系統的服務範圍。簡述其特點如下：

1. 使用者可以在使用介面上選擇繁體版或簡體版，系統會自動將網頁內容以 Big5 或 GB 碼顯示，方便使用者閱讀，不用擔心看不懂簡體字或繁體字。（見圖三、圖四）
2. 使用者選擇簡體版時必需輸入 GB 碼的檢索值來查詢，系統會自動把查詢結果以 GB 碼顯示。因此使用簡體版的使用者其操作方式和繁體版的使用者是一樣的，查詢結果也相同，只是內碼不同而已。
3. 本系統加入此簡繁對譯功能後，只需維護 Big5 碼資料，使用者在簡體版查詢時，系統會即時自動作轉換，如此就不會有繁體、簡體兩份資料維護不同步時所產生的問題，亦可節省硬碟空間，且程式只需一套，不必簡繁體各一套查詢程式，如此也可節省系統的維護成本。



圖一 缺字「」(CC0II視為 22B59)之轉字式如上圖



圖二 查詢端之轉字式「」已被轉換為圖片，將游標移到該圖片上方，會顯示其轉字式



圖三 書號 10231-0002 的書目之簡覽版顯示



圖四 書號1029-0002的書目之詳細版顯示



圖六 輸入帳號、密碼後可以點選任一卷名



圖五 瀏覽影像前需先輸入帳號、密碼



圖七 卷之五的各篇目，在此可點選任一篇目或者其它卷名

### (三) 影像瀏覽機制

本系統除了讓使用者查詢著錄的古籍書目資訊之外，另有掃描影像提供使用者瀏覽，可避免使用微捲片的不便，更方便在網路上開放研究使用。

下列幾種方式可以瀏覽影像

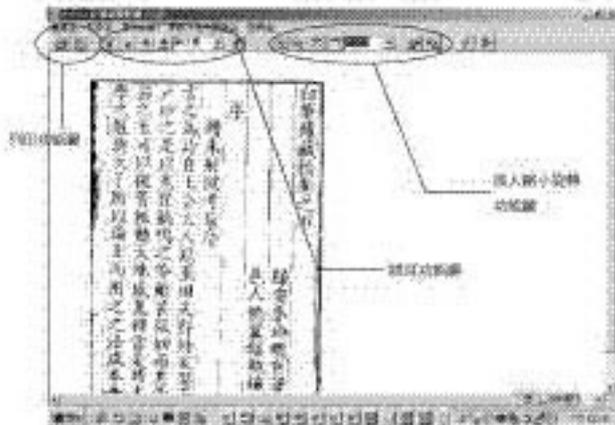
#### 1. 查詢方式：(查詢子功能)

- (1) 輸入查詢值後，系統以條列方式顯示查詢結果，此時有影像的書目資料前方會顯示「此篇目有影像資料，如要看影像請登入」。(見圖五)
- (2) 進入書目之詳目式，除顯示該書目的相關欄位，還會顯示各卷的卷名，點選某卷則會顯示該卷的篇目，再點選篇目則會顯示其影像。(見圖六、圖七、圖八)
- (3) 可在詳細查詢或更多查詢畫面的只查有影像這個檢索欄位選擇是，如此查詢結果就會是只有影像的書目。(見圖九)

#### 2. 瀏覽方式：(影像瀏覽子功能)

- (1) 在此子功能輸入查詢值即可查詢到有影像的書目。

- (2) 利用標題層次、標題分類、版本、篇目等方式來瀏覽有影像之書目。因為系統已先將資料作分類，使用者瀏覽時可以做觸類旁通的研究。（見圖十）
3. 在閱覽影像時皆可以放大、縮小、旋轉、列印、跳頁，以方便使用。（見圖八）
4. 本系統除了可以看單篇的影像之外，另提供可瀏覽整卷影像的「古籍電子書」功能，只要點選篇目名稱，系統會自動跳至其對應之頁數顯示影像，此操作方式如同翻閱紙本時一樣，更方便使用。（見圖十一）



圖八 「虞未射放考最序」的第一頁影像，使用者可以自行選擇頁數瀏覽



圖十 點選「影像瀏覽」可依系統分類做觸類旁通的研究



圖九 可以查詢「只有影像」的書目，以節省檢索時間



圖十一 點選篇目名稱可自動翻頁的「古藉電子書」

#### (四) 序跋全文檢索

本系統亦提供序跋全文可讓使用者作進一步的研究，下列幾個步驟來閱讀序跋全文：

1. 查詢：在詳細查詢、更多查詢的畫面有「序跋全文」檢索欄位，輸入檢索值後系統會對所有書目的已建檔之序跋全文作查詢，然後條列其結果。（見圖十二）
2. 結果顯示：在詳目顯示時，系統會顯示該筆書目資料的序跋全文，並且若有其對應的影像，亦可提供閱讀，方便使用者同時比對全文和影像。尤其有些題跋為藏書者草書

手寫，若非對草書有研究者無法明白，且古籍之內容因無標點符號難以閱讀，本系統加入此功能後，方便使用者的閱讀及研究。（見圖十三）

3. 新增功能：92 年度計劃將序跋全文和其對應之影像一起呈現，使用者不用個別去點選，系統會自動作判斷。

#### （五）人名權威機制

由於書目資料是依據原古籍資料來著錄的，因此其創作者、序跋者等人名，有可能同一人在不同書目資料用不同名字著錄，因此本系統在人名查詢時作權威處理，方便使用者查詢同一著者的資料。作法如下：

1. 建置權威資料：由本館書目資料庫轉置過來，並建置索引。
2. 查詢時對人名欄位的檢索值先作處理，和權威資料比對，若出現在人名權威資料庫，則將該人物的所有名字如：字、號、別號等一起當作檢索值來查詢。例：趙孟頫有數個字號：子昂、文敏、水晶宮、水晶宮道人、松雪齋、松雪道人、歐波、魏國公、鷗波亭、趙子昂、趙文敏公、趙松雪。使用者用趙子昂在人名欄位查詢，系統會先對應出此名是屬於趙孟頫的字號，再把趙孟頫的本名連同其所有字號當作查詢值一起查詢。（圖十四、圖十五）



圖十二 有序列表全文和其對應影像時會一起呈列，供使用者閱覽



圖十四 在人名欄位輸入：趙子昂



圖十三 可以將序列表全文和影像作對照研究



圖十五 系統作人名權威處理可查出趙孟頫的書目資料

## (六) 詮釋資料 (Metadata)

本系統資料來源為館藏書目資料，其資料結構為機讀格式，現已依據國家圖書館 Metadata 研究小組編撰之「中文詮釋 (Metadata) 格式彙編」初版一書中「古籍善本詮釋資料及著錄範例」為基礎，訂定 Metadata 所需與相對應的 Tag 欄位，另根據古籍原件或「國家圖書館善本書志」補充機讀格式書目檔中所沒有的 Metadata 欄位，訂出古籍資料的 DTD，以作為資料著錄及驗證和匯出、匯入的依據。該 DTD 乃以 Dublin Core 15 個基本欄位為主要的架構，但是為了描述資料豐富的屬性、更精確表達描述藏品的語意，在相關欄位下，搭配欄位修飾語 (element qualifier) 的使用，不僅可以擴大其應用的範圍，同時也具有國際性。

本系統參考其它數位典藏計畫的做法，除了將資料結構由機讀轉換至 Metadata 之外，亦提供了 XML 的匯入與匯出之功能 (圖十六、圖十七)，方便與其他單位作資料交換。又由於 DTD 在宣告上並不具備資料格式、特殊處理轉換、建索引等欄位定義，所以提供讓管理者可以彈性設定每個欄位的資料屬性，如：日期、日期區間、檔案、URL 等，亦可設定每個欄位是否要重複、必備、系統自動產生、建索引、條列顯示、詳目顯示等，這些設定可讓管理者依據使用者的反應作調整，隨時提供使用者更好的服務，卻可不必修改程式。(圖十八、圖十九)



圖十六 匯入功能：可匯入符合目前古藝 DTD 之 xml 檔案，系統會將資料批次載入



圖十八 系統列出古藝資料的欄位一覽表，可再選擇任一欄位作詳細設定



圖十七 資料匯出：選擇 DTD 名稱，並輸入系統識別號碼或異動日期區間，以作資料輸出



圖十九 對單一欄位作更詳細的設定

### 三、結 語

本計畫預計待館內重要典藏影像數位化完成後，批次將其全數轉入上述系統之資料庫中。屆時，不但可提供學林資料檢索，對於一般社會大眾而言，亦可藉由古籍文獻的欣賞、金石拓片文獻的流傳，無遠弗屆地弘傳我國悠久的藝術與文化。各種充實的網路內容，亦可提供海內外華人利用，為後代子孫留下珍貴的數位化史料。