

琉球大学学術リポジトリ

延世大学校中央図書館でのデジタルコレクション構築事例：古文獻のDB構築を中心に

メタデータ	言語: 出版者: 琉球大学学術リポジトリ事務局 公開日: 2007-12-25 キーワード (Ja): キーワード (En): 作成者: 金, 永元, 許, 永錫, Kim, Yeong Weon, Heo, Young Seuk メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/2768

延世大学校中央図書館でのデジタルコレクション構築事例 —古文献のDB構築を中心に—

金永元^{*}・許永錫^{**}

[日本語訳：朴花玉^{***}、監修：許永錫]

1. 序論

1.1. 延世大学校の古文献のコレクション

延世大学校は改新教の宣教者達が創立した学校であるが、初期から韓国の歴史、言語、文化を研究する国内の代表的な機関として挙げられてきた。これらの研究を支援する為、中央図書館では古文献の資料を積極的に収集し、現在は国内大学の図書館の中では最大の規模のコレクションを遂げている。延世大学校内の全古文献は全て中央図書館の国学資料室に集中している。

中央図書館の国学資料室に所蔵されている古文献の資料は大体 1910 年を基準とし、それ以前に印刷されたものや筆写された「古書」コレクションと 1910 年以降から 1953 年の間に刊行された洋装本「新書貴重本」コレクションに大きく分けられる。1910 年という基準は韓日合邦という大きな政治、社会的変動とともに韓国の出版史での印刷と装丁方法が西洋化され、出版された書籍の内容が近代的学問に変わったことを反映したことである。刊行/筆写の時期だけではなく収録された内容が伝統思想に関わっているもので、印刷方法、装丁、製本などの形態が伝統的なものを「古書」と認めている。我が図書館では現在 25,000 種の約 10 万冊の「古書」を所蔵している。この中には国家文化財に指定された宝物 5 種を含め、国内での唯一の本、珍しい本の資料など多数含まれている。

「新書貴重本」は開化期から韓国戦争期（1950～1953）の間に出版された資料として韓国で刊行されたものや、韓国と関連のある資料のコレクションである。韓国では 19 世紀末の開化期以降に近代的出版物が生産し始め、日本植民時期、韓国戦争期まで刊行された資料も今は珍しい資料に含まれる為、我が図書館ではこの時期の資料を「新書貴重本」と指定し、別途のコレクションとして管理している。この「新書貴重本」コレクションでは単行本 23,600 冊、雑誌 650 種 14,000 冊が所蔵され、新聞 50 種が含まれている。

* 延世大学校中央図書館国学資料室部長

** 延世大学校中央図書館閲覧部主任

*** 琉球大学教育学部 3 年次

「表 1」 延世大学校の国学資料室所蔵の古文献の現況

区別	範囲	蔵書数	備考
古書	- 1910 年以前刊行/筆写された資料 - 内容/形態上での伝統的資料	10,670 冊 (貴重古書 951 種 2,881 冊含む)	国家文化財 (宝物) 5 種
新書貴重本	- 開化期から韓国戦争期の間で刊行された資料 - 韓国で刊行されたものや、韓国関連の外国資料	24,334 冊 - 単行本 10,284 冊 - 雑誌 650 種 14,000 冊 - 新聞 50 種	キリスト教関連の古文献含む。

1.2. デジタルコレクションの構築の背景

上で簡単に紹介した通り、国学資料室では国学分野を研究する研究者に必ず必要な良い資料が数多く所蔵されている。一方、利用の側面からみると、本大学内では国学研究の特性化を大学の主要目標の一つとして策定し研究を促進することにより、国学資料室所蔵の資料に関する利用要求が増加し、また、大学内の研究者だけではなく外部利用者も頻繁に訪問し資料収集を要求した。反面、資料管理の側面からは資料が物理的に多く損傷しているため、閲覧や複写においては利用者に十分なサービスの提供が出来なかった。貴重な原本の資料を大切に保存する事も古文献の管理者に負荷された大きな任務であるためだ。古文献の資料が時間的に古いという側面もあるが、我が図書館で収集した古書は中間所蔵者の手を経て入蔵された物が多数であるため相対的により多く毀損され、虚弱化している事も利用の制限点として作用している。

このような資料の利用要求と保存欲求の相反する立場から生じる問題を解決するため、以前からマイクロ・フィルム化作業や毀損資料の保存処理の作業が進行されているが、消極的で小規模の範囲内に留まっていた。1990 年代後半の以降、図書館の業務に関する優秀な電算システム、装備、技術の普及でデジタル環境が造成され、遂に古文献のデジタルコンテンツ構築作業を本格的に推進できるようになった。

延世大学校中央図書館での図書館業務のデジタル化過程を、古書関連の事項を中心としてみると以下の通りである。

- 1990 年に DOBIS/E システムを導入し書誌情報の電算化作業に着手した。
- 2001 年に統合形デジタル図書館システム (YDL, Yonsei Digital Library) を稼働し始めた。同時にデジタルコンテンツの構築事業を進行したが、学位論文、逐次刊行物記事、VOD などと共に古書原文の DB 構築事業を進行した。
- 2004 年からは教育部が推進した「歴史資料の情報化事業」に参加し、書誌情報と原文情報の DB 構築事業を遂行した。

2. 統合形デジタル図書館システムの開発と 韓国学基本古書の原文の DB 構築 (2000～2001 年)

2.1 統合形デジタル図書館 (YDL) システムの開発

延世大の図書館では 1999 年からデジタル図書館の構築事業を行い、2001 年 5 月に書誌情報システム、原文管理システム、VOD 管理システム、講義/研究支援情報システム、ウェブエージェントシステム、統合情報検索システムなどで構成された統合形デジタル図書館システム (YDL) を稼働し始めた。全体のシステムを新たに開発したのではなく、パッケージを導入し、延世大学中央図書館に適合するようにカスタマイズをしたものである。

YDL の下位システムのうち、古文献の関連部分は書誌情報システムと原文管理システムであったが、書誌情報システムでは当時新しく制定された古書用 KORMARC を系統的に支援できるように具現し、原文管理システムは古書、学位論文、学術誌の記事などの原文が提供できるシステムに、図書館自体から構築した DB、外部から搬入した DB、大学院から提供された学位論文の DB などが管理できるように設計された。この原文管理システムでは多様なイメージ/デジタルフォーマットを PDF フォーマットに変換してサービスするようにし、XML/SGML など多様な構造化文書形態の支援が出来るようにした。

古書資料の場合、資料に対する多様なアクセスポイントを提供するため、原文 DB だけを構築するのではなく、副次的な内容を共に提供できるようにした。つまり、「解題」では、該当の古書資料に関する専門的な説明をテキストに添付し研究者はもちろん、一般人も資料に関する情報が容易に得られるようにし、「関連文献」では、その本に関連する単行本、学術誌記事をリンクさせ、利用者に多様なアクセスポイントを提示しようとした。また、「関連サイト」では、その本に関連する情報を提供している URL 情報を提供した。言い換えれば、一つの古書に関して出来るだけすべての情報を提供しようとした事である。(図 1)



「図1」 検索結果の画面



「図2」 「原文見る」画面のTOC情報

原文のイメージを利用する際にも利用者が膨大な分量の中で必要な部分を直ぐに探せるように TOC を作成し、利用者が必要な内容を探すためにいちいちファイルをめくる手間を省くようにした。(図2)

このような内容は、古文献のデジタルコンテンツの構築に適合できるようにシステムを開発する際に留意した点である。

2.2. 韓国学基本古書原文のDB構築

2.2.1. 対象資料

DB構築の対象資料を選定しながら我々が立てた2つの基本的な原則は、第一、我が大学の国学研究の伝統と密接に関係し、本図書館の所蔵古書の性格を十分に表している資料を選定し、第二、各主題の分野でも利用頻度の高い基本的な古書を選定することであった。これらの原則に従って、実学、国語学、ハンゲル文学の3つの主題分野といった大きな範疇を設定し、それぞれに該当する古書について基礎調査を行い、該当分野の教授の諮問を得、デジタル化して対象資料に最終選定をした。

1. 実学

実学は、朝鮮時代の後期学術史に現れた改革的で近代指向的な学問の傾向を言う。解放(1945年)以前から実学を研究する有名な学者達が本大学に在職、研究活動をするようになり、本図書館では実学者の著述及び関連古書を集中的に収集したため、比較的豊富な蔵書を備えることが出来た。そして、利用の側面からも実学分野、若しくは隣接分野の主

題に関する研究者の数が多く、資料の利用要請度も高い方である。我々は本図書館に所蔵された実学関係の著述をさらに歴史書、地理書、経世書、科学技術書、文集/全書、類書/雑纂など6つの小主題に分け、原文のDBを構築することにした。ここには丁若鏞の『與猶堂全書』、『我邦疆域考』、朴趾源の『熱河日記』、『燕巖集』、柳馨遠の『磻溪隨錄』、金正喜の『阮堂集』など66種653冊の実学者の著述が網羅されている。

2. 国語学

延世大学の韓国学の研究業績の中で一番特出な部分がこの国語学分野と言えるが、これは、以前に在職した崔鉉培、金允經教授など優れた学者の研究業績による部分もあるが、図書館にハングル関連の古文獻が豊富に所蔵されている事にもある程度関連があると思える。我々の大学だけの特色のあるデジタルコンテンツを構築するとした時に、最初に挙げられた対象主題分野がこの国語学であった。我々は所蔵している国語学の関連古書を‘ハングル研究書’、‘訳語’、‘経典’、‘実用書’、‘倫理書’、‘文学/歴史書’など7つの小主題に分け、対象資料として選定した。ハングル研究書はハングル創製の背景と製字の原理などについての理論的接近に関するもので、訳語は主に朝鮮時代の訳官（今の通訳士）が外国語を習得するために作った教材であり、国語学の他の分野は漢字で書かれた資料をハングルで翻訳したものやハングルで著述された資料である。甲景濬の『訓民正音韻解』、柳僖の『諺文志』、『四書三經諺解』、仏教経典諺解本などを含め、全78種160冊である。

3. ハングル文学

ハングル文学はハングルで作られた小説、歌である。漢字で書かれた資料が主にヤンバン士大夫の生活を取り扱っているのであれば、ハングル文学は、一般庶民が享有した文学作品で、これを通じて当時の社会世相をみることが出来る良い資料である。我々がDB構築対象として選定した資料は、中・高等学校の教科書に載っているものや、紹介された資料を中心としたもので、非専門者の誰にも馴染みのあるものだ。ハングル文学は詩歌と小説の2つの小主題に構成されている。ここでは、『歌曲源流』、『東国歌辭』などの時調と歌辭、『春香傳』、『洪吉童傳』、『九雲夢』などの小説を含め、色々な異本などを網羅して、全68種127冊である。

2.2.2. 技術的な考慮

古書原文のDB構築の実際業務に入ってからには以下のような技術的な問題を考慮した。

1. アナログ形式の保存用のマイクロ・フィルムとデジタル形式のイメージファイルを共に製作する。
2. 費用と原本再現性を考慮し、イメージ解像度の基準は8 bit GRAY scaleにし、イメージ像で紙の紙質を区別し、印刷された部分と後ほど筆写で加筆された部分が区別できるようにする。

3. ファイルのサイズはアクセス性を考慮し、保存用のファイルは JPEG ファイルにしてサービス用は原本ファイルから圧縮された PDF ファイルに変換させたものを使用する。
4. 各ページのイメージごと縮尺 (scale) と大学のロゴを下段に挿入して、スキャンしてデジタル化されたイメージだけを見ても本の本来のサイズが押し量れるようにし、著作権は本図書館にあることを示した。

3. 韓国の基督教の古文献原文の DB 構築 (2004 年)

3.1. 推進背景

2002 年以降、我が図書館では古書書誌の DB 構築、古雑誌書誌の DB 及び記事索引の DB 構築、貴重古書のマイクロ・フィルム化作業を遂行してきた。

この間に韓国では、知識情報資源管理法 (法律 6232 号、2000. 1. 23) の発効により古文献の資料に関する DB 構築事業が国家事業として大規模に進行され、その成果物が韓国歴史情報システム (<http://koreanhistory.or.kr/>) を通して一般市民にサービスされるようになった。これに合わせて教育部では、大学の所蔵歴史資料の情報化事業 5 年計画 (2004～2008 年) を立て、各大学図書館の所蔵の歴史資料に関する書誌 DB 構築、原文イメージの DB 構築事業を進行するようになった。我が図書館では古書書誌 DB と韓国基督教の古文献の DB 構築という内容として 2004 年教育部の公募事業に参加し、政府予算を支援され事業を行った。

3.2. 対象資料

延世大学校は 1885 年、韓国に初めて来た米国の改新教宣教者の H. G. Underwood により創立され、学校の創立理念も基督教精神の具現である。我が図書館の所蔵「新書貴重本」コレクションには基督教と関連する古文献が数多く所蔵され、国内では基督教の文献としては一番規模が大きいコレクションである。

ここではカトリックを含め、基督教の各宗派を網羅した翻訳聖書、教団別の宣教報告書、教団別の会議録、聖書注釈、賛美歌、雑誌などが含まれている。我が図書館で所蔵する基督教の古文献の中、他の機関で既に DB 構築した資料を除外した単行本 1,048 冊、雑誌 79 種 2,020 冊をデジタルコンテンツの構築対象とした。

3.3. 技術的な考慮

2004 年の事業では、システムを既存に作っておいた状態から始めたので、コンテンツの構築に力を注ぐことができた。但し、以下の幾つかは我々が特に留意した部分である。

基督教の資料のうち、雑誌の場合は雑誌に収録された記事単位での検索ができるよ

うに記事 DB が構築されていたため、必要な記事が検索されると直ぐに該当のイメージが見られるようにした事と、同時に各巻・号ごとに原文をリンクさせ、雑誌全体が見られるようにした。単行本の資料と英文の雑誌の場合は古書と同じく TOC ファイルを通して必要な内容にアクセスできるようにした。

2004年の事業からサービス用のイメージフォーマットをPDFからDjVuに変更させた。PDFフォーマットはファイル容量が非常に大きいため、アクセス速度が遅く、一つの書誌に搭載できるイメージファイルの数の制限があった。保存用ファイルはそのままJPEGファイルにして、サービス用ファイルはDjVuフォーマットに変え、2001年に構築したPDFファイルも全てDjVuに変換させた。カラーイメージを基準とし5～10倍程度のサービス速度の向上をもたらした。また、原文の不法流通/不法複製の防止のためDRM技術を適用させた。DRMを通して利用可能時間、利用可能回数、保存機能の使用可否、出力機能の可否などを図書館の政策によって統制可能になった。また利用者のアクセス権限制御のために著作権システム(身分別、IP帯域別のアクセス権限管理)を適用させた。そして、「原文を見る」及び出力の際に延世大学校所有の著作物であることを明示するため、電子透かしをイメージに挿入した。

4. 一般古書原文のDB構築(2006年)

2006年に、本図書館では、教育部の歴史資料情報化事業に古書書誌のDB構築及び原文イメージのDB構築事業を申請し選定され、再び政府の支援を受けることになった。対象資料は我が図書館の筆写本の詩歌と小説類がほとんどで、歴史資料が一部含まれている。筆写本の詩歌と小説類は研究者が異本対照のため頻繁に探す資料であり、相対的に古い資料が多いため至急にデジタル化する必要性が求められたからである。全357種769冊である。ほぼ以前と同じ方式でDBを構築した。

ここで我が図書館がイメージのDBを構築する際に留意した事項を2006年の事業を中心として整理してみると以下の通りである。

○ スキャンの準備作業

- －既構築された書誌レコードと原本の対照
- －原本資料のページ別/本別性格の分類(カラー、写真及び図、本文、紙質、製本、太さ)
- －解像度板及び反射板、標準カラーバー(bar)スキャン
- －カラーイメージの具現ができるオーバーヘッドスキャナー及びスキャナー別ブックホルダー(book holder)準備

- スキャン作業
 - －作業中での原本資料の毀損防止対策の準備
 - －対象資料の紙質と印刷状態、内容（テキスト、写真、図など）による明るさと濃度を調整しながらの作業進行
 - －屈曲によるイメージの歪曲現象はブックホルダーの高さを調整し補正する。

- イメージ補正及びファイルフォーマットの変換作業
 - －イメージの枠部分のノイズの除去、本文中の変色部分と細かな染みなどは修正しない。
 - －傾きの調整及び重心の補正
 - －両面のイメージを断面のイメージに分割
 - －JPEG の原本ファイルを DjVu フォーマットに変換させる際に、イメージの損失がないように留意

- TOC 作成
 - －原本資料の言語そのまま TOC 作成
 - －目次ページがない場合 3 段階まで目次生成

- システムのアップロード
 - －書誌単位別に本図書館のシステムにサービス用ファイルをアップロードする。
 - －ウェブサービス用のイメージファイルの無断資料流出の防止のため技術的な措置（電子透かしなど DRM 関連技術）試行

5. 結論

今まで延世大学中央図書館が行ってきたこれまでの古文献のデジタルコレクションの構築事業を事業単位の事例報告形式で整理してみた。

古文献を管理する図書館の人なら誰でも感じることであるが、資料の閲覧/複写による原本の毀損の恐れと利用者に対する満足できるサービスの提供という 2 つの目標の間で経験してきた葛藤が原文のデジタル化で解決できることは大変鼓舞的な事だと言えよう。これらと共に古文献の原文を 1 回デジタル化しておくことで多様な方面で活用できるという便利さもある。研究対象として活用する場合はもちろん、講義の教案作成に活用することもよく目にする。本校の教育開発支援センターで、人物 200 選、必読図書 200 選などと共に古書 200 選を選定し、ホームページで提供する時も本図書館で構築した古文献のデジタルコンテンツをそのまま持ち出し活用する場合もある。

2000 年、我々が最初に事業を始めた頃より今のデジタル環境はより一層発展していて、

振り返ってみると今昔の感に堪えない。当時は国内でカラーイメージ用のオーバーヘッドスキャナーがあまり普及していなく、費用が高かったため、より安い gray 形式でスキャンを行ったが、今は gray 形式で DB を構築するケースはあまりなく、古文献をほとんどカラーでスキャンして DB を構築する。前で説明したファイルフォーマットの変更 (PDF→DjVu) のような場合も、当時は PDF フォーマットが圧縮率などといった問題があったが、今は多くアップグレードされた製品を出している。デジタル技術の発展に合わせて、諦めずシステムをアップグレードしていくことは我々に新しく課された課題であると言えよう。

これ以外にもデジタルコンテンツの保存性、デジタル化とマイクロ・フィルム化との並行可否、全ての古文献の所蔵機関がネットワークで繋がっている状況下での同一資料の重複デジタル化の恐れなど簡単に解決できない課題が我々には残っている。

