

# 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：古文献の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	<a href="http://hdl.handle.net/20.500.12000/2768">http://hdl.handle.net/20.500.12000/2768</a>

# 연세대학교 중앙도서관에서의 디지털 컬렉션 구축 사례

## - 고문헌 DB 구축을 중심으로

金永元\*, 許永錫\*\*

### 1. 서론

#### 1.1. 연세대학교의 고문헌 컬렉션

연세대학교는 개신교 선교사들이 창립한 학교이지만 초기부터 한국의 역사, 언어, 문화를 연구하는 국내의 대표적인 기관으로 꼽혀왔다. 이들 연구를 뒷받침하기 위하여 동 중앙도서관에서는 고문헌자료들을 적극적으로 수집하여 지금은 국내 대학 도서관 중 가장 큰 규모의 컬렉션을 이루고 있다. 연세대학교 내의 모든 고문헌들은 모두 중앙도서관 국학자료실에 집중되어 있다.

중앙도서관 국학자료실 소장 고문헌 자료는 대략 1910년을 기준으로 하여 그 이전에 인쇄되거나 필사된 [古書] 컬렉션과, 1910년 이후부터 1953년 사이에 간행된 양장본 [新書貴重本] 컬렉션으로 크게 나누어진다. 1910년이라는 기준은 한일합방이라는 큰 정치, 사회적 변동과 함께 한국의 출판사에서 인쇄와 장정 방법이 서구화되고, 출판된 서적의 내용이 근대적 학문으로 바뀐 것을 반영한 것이다. 간행/필사 시기 뿐만 아니라 수록된 내용이 전통사상에 관한 것이고 인쇄 방법, 장정, 제책 등의 형태가 전통적인 것을 [고서]로 인정하고 있다. 본 도서관에서는 현재 25,000여 종의 약 10만 책의 [고서]를 소장하고 있다. 이 중에는 국가문화재로 지정된 보물 5종을 포함하여 국내 유일본, 희귀본 자료들이 다수 포함되어 있다.

\* 연세대학교 중앙도서관 국학자료실 부장

\*\* 연세대학교 중앙도서관 열람부 주임

[신서귀중본]은 개화기에서부터 한국 전쟁기(1950~1953) 사이에 출판된 자료로 한국에서 간행되거나 한국과 관련된 자료의 컬렉션이다. 한국에서는 19세기 말 개화기 이후 근대적 출판물이 생산되기 시작하였으며, 일제 식민시기, 한국전쟁기까지 간행된 자료도 지금은 희귀자료에 속하므로 본 도서관에서는 이 시기의 자료를 [신서귀중본]으로 지정하여 별도의 컬렉션으로 관리하고 있다. 이 [신서귀중본] 컬렉션에는 단행본이 23,600여 책, 잡지가 650종 14,000책이 소장되어 있으며 신문 50여종이 포함되어 있다.

[표1]연세대학교 국학자료실 소장 고문헌 현황

구분	범위	장서수	비고
고서	- 1910년 이전 간행/필사된 자료 - 내용/형태상으로 전통적인 자료	10,670책  (귀중고서 951종 2,881책 포함)	국가문화재  (보물) 5종
신서 귀중 본	- 개화기부터 한국전쟁기 사이에 간행된 자료 - 한국에서 간행되거나 한국 관련 외국 자료	24,334책  -단행본 10,284책 -잡지 650종 14,000책 - 신문 50종	기독교 관련 고문헌 포함

## 1.2. 디지털 컬렉션 구축 배경

앞에서 간략히 소개한 바와 같이 국학자료실에는 국학 분야를 연구하는 연구자들에게 꼭 필요한 좋은 자료들이 많이 소장되어 있다.

한편 이용의 측면에서 볼 때에, 본 대학 내에서는 국학연구의 특성화를 대학의 주요 목표 중의 하나로 책정하여 연구를 촉진하면서 국학자료실 소장 소장자료에 대한 이용 요구가 크게 늘어났고, 또 대학 내의 연구자 뿐만 아니라 외부 이용자도 자주 방문하여 자료 수집 요구를 하였다. 반면 자료 관리 측면에서는 자료가 물리적으로 많이 손상되어 있기 때문에 열람이나 복사에 있어서 이용자에게 충분한 서비스를 제공하기 어려웠다. 귀중한 원본 자료를 잘 보존하는 것도 고문헌 관리자에게 부과된 큰 임무이기 때문이다. 고문헌 자료가 시간적으로 오래되었다는 측면도 있지만, 우리 도서관에서 수집한 고서들은 중간 소장자들의 손을 거쳐서 입장된 된 것들이 다수이어서 상대적으로 더 많이 훼손되고 위약화되어 있다는 점도 이용의 제한점으로 작용하고 있다.

이러한 자료의 이용 요구와 보존 욕구의 상반된 입장에서 생기는 문제를 해결하고자 이전에 마이크로필름화 작업이나 훼손자료 보존처리 작업이 진행되어 오기는 하였으나 소극적이고 소규모 범위 내에서 이루어지는데 그쳤다. 1990년대 후반 이후 도서관 업무에 관련된 우수한 전산 시스템, 장비, 기술의 보급으로 디지털 환경이 조성되면서 마침내 고문헌 디지털콘텐츠 구축 작업을 본격적으로 추진할 수 있게 되었다.

연세대학교 중앙도서관에서의 도서관업무 디지털화 과정을 고서관련 사항을 중심으로 살펴보면 아래와 같다.

- 1990 년에 DOBIS/E 시스템을 도입하여 서지정보의 전산화작업에 착수하였다.

- 2001년에는 통합형 디지털도서관시스템(YDL, Yonsei Digital Library)을 가동하기 시작하였다. 동시에 디지털콘텐츠 구축사업을 진행하였는데, 학위논문, 연속간행물기사, VOD 등과 함께 고서 원문 DB 구축 사업을 진행하였다.
- 2004년부터는 교육부에서 추진한 ‘역사자료정보화사업’에 참가하여 서지정보와 원문정보 DB 구축사업을 수행하였다.

## 2. 통합형 디지털도서관 시스템의 개발과

### 한국학 기본 고서의 원문 DB 구축 (2000~2001년)

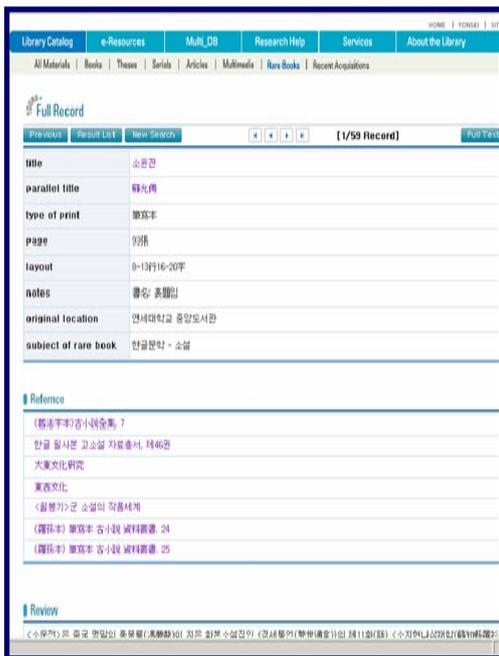
#### 2.1. 통합형 디지털도서관(YDL) 시스템 개발

연세대 도서관에서는 1999년부터 디지털도서관 구축사업을 시작하여 2001년 5월에 서지정보시스템, 원문관리시스템, VOD 관리시스템, 강의/연구지원정보시스템, 웹에이전트시스템, 통합정보검색시스템 등으로 구성된 통합형 디지털도서관시스템(YDL)을 가동하기 시작하였다. 전체 시스템을 새로 개발한 것은 아니고 패키지를 도입하여 연세대 중앙도서관에 적합하도록 커스터마이징한 것이다.

YDL의 하위 시스템 중 고문헌 관련 부분은, 서지정보시스템과 원문관리 시스템이었는데, 서지정보시스템에서는 당시 새로 제정된 고서용 KORMARC을 시스템적으로 지원할 수 있게 구현하였고, 원문관리시스템은 고서, 학위논문, 학술지 기사 등의 원문을 제공할 수 있는 시스템으로, 도서관 자체에서 구축한 DB, 외부에서 반입한 DB, 대학원에서 제공받는 학위논문 DB 등을 관리할 수 있도록 설계되었다. 이 원문관리시스템에서는 다양한 이미지 / 디지털 포맷을 PDF 포맷으로 변환하여

서비스할 수 있도록 하였으며 XML/SGML 등 다양한 구조화 문서형태를 지원할 수 있도록 하였다.

고서 자료의 경우, 자료에 대한 다양한 접근점을 제공하기 위해 원문 DB 만을 구축하는 것이 아니라 부차적인 내용을 함께 제공할 수 있게 하였다. 즉, [해제]에서는 해당 고서 자료에 대한 전문적인 설명을 텍스트로 첨부하여 연구자는 물론, 일반인들도 자료에 대한 정보를 쉽게 얻을 수 있도록 하였고, [관련문헌]에서는 그 책에 관련된 단행본, 학술지 기사들을 링크하여 이용자에게 다양한 접근점을 제시하고자 하였다. 또한 [관련사이트]에서는 그 책과 관련된 정보를 제공하고 있는 URL 정보를 제공하기도 하였다. 말하자면 한 고서에 대해서 가능한 한 모든 정보를 제공하고자 한 것이다. (그림 1)



[그림 1] 검색결과 화면



[그림 2] 원문보기 화면의 TOC 정보

원문 이미지를 이용할 때에도 이용자가 방대한 분량의 내용 중에 필요로 하는 부분을 바로 찾아갈 수 있도록 TOC 를 작성하여 이용자가 원하는 내용을 찾기 위해 일일이 파일을 넘기는 수고를 덜도록 하였다.(그림 2)

이와 같은 내용들은 고문헌 디지털콘텐츠 구축에 적합하도록 시스템을 개발할 때에 유의 하였던 것들이었다.

## 2.2. 고서 원문 DB 구축

### 2.2.1. 대상자료

DB 구축 대상자료를 선정하면서 우리가 세웠던 2 가지 기본적인 원칙은 첫째, 우리 대학의 국학연구 전통과 밀접하게 관계되고 본 도서관 소장 고서의 성격을 잘 나타내는 자료를 선정하며, 둘째, 각 주제 분야에서도 이용빈도가 높은 기본적인 고서를 선정한다는 것이었다. 이러한 원칙에 따라 실학, 국어학, 한글문학 세가지 주제분야의 큰 범주를 설정하고 각각에 해당하는 고서에 대해 기초조사를 거치고 해당 분야 교수의 자문을 얻어 디지털화 대상자료로 최종 선정하였다.

#### 1. 실학

실학이란, 조선시대 후기 학술사에 나타난 개혁적이고 근대지향적인 학문 경향을 일컫는 것이다. 해방(1945년) 이전부터 실학 연구를 하는 유명한 학자들이 본 대학에 재직, 연구활동을 하게 되면서 본 도서관에서는 실학자들의 저술 및 관련 고서를 집중적으로 수집하여 비교적 풍부한 장서를 갖추게 되었다. 그리고 이용 측면에서도 실학분야, 혹은 인접 분야의 주제에 대한 연구자 수가 많고 자료 이용 요청도 많은 편이다. 우리는 본 도서관에 소장되어 있는 실학 관계 저술을 다시 歷史書, 地理書, 經世書, 科學技術書, 文集/全書, 類書/雜纂 등 6가지 소주제로 나누어 원문 DB를 구축하기로 하였다. 여기에는 丁若鏞의 『與猶堂全書』, 『我邦疆域考』, 朴趾源의 『熱河日記』, 『燕巖集』, 柳馨遠의 『磻溪隨錄』, 金正喜의 『阮堂集』 등 66종 653책의 실학자들의 저술이 망라되어 있다.

## 2. 국어학

연세대학교의 한국학 연구업적 중에 가장 특출한 부분이 국어학 분야라고 할 수 있는데 이것은 물론 이전에 재직하던 崔鉉培, 金允經 교수 등 뛰어난 학자들의 연구업적에 따른 부분도 있지만 도서관에 소장되어 있는 한글 관련 고문헌들이 풍부하게 소장되어 있는 것과는 어느 정도 관련이 있을 것이다. 처음 우리 대학만의 특색 있는 디지털콘텐츠를 구축한다고 하였을 때 가장 먼저 손꼽힌 대상 주제 분야가 바로 국어학이었다. 우리는 소장하고 있는 국어학 관련 고서를 ‘한글硏究書’, ‘譯語’, ‘經典’, ‘實用書’, ‘倫理書’, ‘文學/歷史書’ 등 7 가지 소주제로 나누어 대상자료로 선정하였다. 한글연구서는 한글 창제의 배경과 제자(制字) 원리 등에 대한 이론적 접근에 대한 것이고, 역어는 주로 조선시대 譯官(지금의 통역사)들이 외국어를 익히기 위해 만든 교재이며, 국어학의 나머지 분야는 한문으로 된 자료를 한글로 번역한 것이거나 한글로 저술된 자료들이다. 申景濬의 『訓民正音韻解』, 柳僖의 『諺文志』, 『四書三經諺解』, 佛教 經典 諺解本 등을 포함하여 모두 78 종 160 책이다.

## 3. 한글문학

한글 문학은 한글로 지어진 소설, 노래이다. 한문으로 쓰인 자료들이 주로 양반 사대부의 생활을 다루고 있다면, 한글 문학은 일반 서민들이 향유했던 문학 작품을 통해 당시 사회상을 엿볼 수 있는 좋은 자료이다. 우리가 DB 구축대상으로 삼은 자료들은 중고등학교 교과서에 실리거나 소개된 자료를 중심으로 하여 비전공자 누구에게나 익숙한 것들이다. 한글문학은 詩歌와 小說의 2 가지 소주제로 구성되어 있다. 여기에는 『歌曲源流』, 『東國歌辭』 등의 時調와 歌辭, 『春香傳』, 『洪吉童傳』,

『九雲夢』 등의 소설을 포함하여 여러 異本들을 망라하였으며 모두 68 종 127 책이다.

### 2.2.2. 기술적인 고려

고서원문 DB 구축 실제업무에 들어가서는 아래와 같은 기술적인 문제들을 고려하였다.

1. 아날로그 형식의 보존용 마이크로필름과 디지털 형식의 이미지 파일을 함께 제작한다.
2. 비용과 원본재현성을 고려하여 이미지의 해상도 기준은 8 bit GRAY scale 로 하되, 이미지 상으로 종이의 지질을 구별할 수 있고, 인쇄된 부분과 후에 필사로 가필된 부분을 구별할 수 있도록 한다.
3. 파일 사이즈와 접근성을 고려하여 보존용 파일은 JPEG 파일로 하고 서비스용은 원본파일에서 압축된 PDF 파일로 변환한 것을 사용한다.
4. 매 페이지의 이미지 마다 축척(scale)과 학교로고를 하단에 삽입하여 스캔하여 디지털화된 이미지만 보고도 책의 원래 크기를 짐작할 수 있도록 하고 저작권이 본 도서관에 있음을 나타내었다.

## 3. 한국 기독교 고문헌 원문 DB 구축(2004년)

### 3.1. 추진 배경

2002년 이후 우리 도서관에서는 고서 서지 DB 구축, 고잡지 서지 DB 및 기사색인 DB 구축, 귀중고서 마이크로필름화 작업을 수행하여 왔다.

이 사이에 한국에서는 지식정보자원관리법(법률 6232호, 2000. 1. 23)의 발효에 따라 고문헌 자료에 대한 DB구축사업이 국가사업으로 대규모로 진행되고 그 결과물이 한국역사정보시스템(<http://koreanhistory.or.kr/>)을 통하여 일반 시민에게 서비스되기 시작하였다. 이에 맞추어 교육부에서는 대학 소장 역사자료 정보화사업 5개년 계획(2004~2008년)을 세우고 각 대학 도서관 소장의 역사자료에 대한 서지 DB구축, 원문이미지 DB구축사업을 진행하게 되었다. 우리 도서관에서는 고서 서지 DB와 한국기독교고문헌DB 구축이라는 내용을 가지고 2004년 교육부 공모사업에 참가하여 정부예산을 지원 받아 사업을 수행하였다.

### 3.2. 대상자료

연세대학교는 1885년 한국에 처음 들어 온 미국 개신교 선교사 언더우드에 의해 창립되었을 뿐만 아니라 학교의 창립이념도 기독교 정신의 구현에 두고 있다. 우리 도서관 소장 [신서귀중본] 컬렉션에는 기독교 관련 고문헌들이 많이 소장되어 있어서 국내에서 기독교 문헌으로는 가장 규모가 큰 컬렉션이다.

여기에는 카톨릭을 포함한 기독교 각 종파를 망라한 번역 성서, 교단별 선교보고서, 교단별 회의록, 성서주석, 찬송가, 잡지 등이 포함되어 있다. 우리 도서관 소장 기독교 고문헌 중 다른 기관에서 이미 DB구축한 자료를 제외한 단행본 1,048책, 잡지 79종 2,020책을 디지털콘텐츠 구축대상으로 하였다.

### 3.3. 기술적인 고려

2004년 사업에서는 시스템을 기존에 만들어 놓은 상태에서 시작하였으므로 콘텐츠 구축에 주력할 수 있었다. 다만 아래 몇가지는 우리가 특별히 유의하였던 부분이다.

기독교 자료 중 잡지의 경우에는 잡지에 수록된 기사 단위들이 검색이 가능하도록 기사 DB 가 구축되어 있었기 때문에 원하는 기사가 검색되면 바로 해당 이미지를 볼 수 있도록 하는 동시에 각 권호마다 원문을 링크시켜 잡지 전체를 볼 수 있도록 하였다. 단행본 자료와 영문 잡지의 경우에는 고서와 마찬가지로 TOC 파일을 통해 원하는 내용에 접근할 수 있도록 하였다.

2004년 사업부터 서비스용 이미지 포맷을 PDF에서 Djvu로 변경하였다. PDF 포맷이 파일용량이 너무 커서 접근하는데 속도가 너무 느리고, 한 서지당 탑재할 수 있는 이미지 파일 수에 제한이 있었다. 보존용 파일은 그대로 JPEG파일로 하되 서비스용 파일은 Djvu 포맷으로 바꾸고, 2001년 구축하였던 PDF파일도 모두 Djvu로 변환하였다. 칼라이미지를 기준으로 5~10배 정도 서비스 속도의 향상을 가져왔다.

또한 원문의 불법유통/불법복제 방지를 위하여 DRM 기술을 적용하였다. DRM 을 통하여 이용가능시간, 이용가능횟수, 저장기능 사용여부, 출력가능 여부 등을 도서관 정책에 따라 통제 가능하게 되었다. 또한, 이용자의 접근권한 제어를 위해 저작권 시스템(신분별, IP 대역별 접근권한관리)을 적용하였다. 그리고 원문보기 및 출력시 연세대학교 소유의 저작물임을 명시하기 위하여, 워터마크를 이미지에 삽입하였다.

#### 4. 일반 고서 원문 DB 구축

2006 년에 본 도서관에서는 교육부의 역사자료정보화 사업에 고서 서지 DB 구축 및 원문 이미지 DB 구축 사업을 신청하여 선정되어 역시 정부의 지원을 받았다. 대상자료는 우리 도서관의 필사본 시가와 소설류가 대부분이고 일부 역사자료를 포함하였다. 필사본 시가와 소설류는 연구자들이 이본대조를 위해 많이 찾는 자료이면

서 상대적으로 낮은 자료가 많아서 시급히 디지털화할 필요가 있었기 때문이다. 모두 357 종 769 책이다. 대체적으로 이전과 같은 방식으로 DB 를 구축하였다.

여기서 우리 도서관에서 이미지 DB 를 구축하면서 유의하였던 사항을 2006년 사업을 중심으로 정리하여 보면 다음과 같다.

#### ○ 스캔 준비작업

- 기 구축된 서지레코드와 원본 대조
- 원본자료의 페이지별/책별 성격 분류(칼라, 사진 및 그림, 본문, 지질, 제본 두께)
- 해상도판 및 반사판, 표준 칼라 바(bar) 스캔(촬영)
- 칼라 이미지 구현할 수 있는 오버헤더 스캐너 및 스캐너별 북 홀더(book holder) 준비

#### ○ 스캔 작업

- 작업 중 원본자료 훼손방지 대책 마련
- 대상자료의 지질이나 인쇄상태, 내용(텍스트, 사진, 그림 등)에 따라 밝기와 농도를 조정하며 작업 진행
- 굴곡에 의한 이미지 왜곡 현상은 북홀더의 높이를 조절하여 보정함

#### ○ 이미지 보정 및 파일포맷 변환 작업

- 이미지 테두리 부분의 노이즈 제거. 본문 중의 변색 부분이나 잡티 등은 수정 없음
- 기울기 조정 및 중심 보정
- 양면 이미지를 단면 이미지로 분할
- JPEG 원본 파일을 Djvu 포맷으로 변환시 이미지 손실이 가지 않도록 유의

## ○ TOC 작성

- 원본 자료의 언어 그대로 TOC 작성
- 목차페이지가 따로 없는 경우 3 단계까지 목차 생성
- TOC 와 이미지 파일과의 링크방식 등 세부사항은 주관기관과 협의하여 결정

## ○ 시스템 업로드

- 서지 단위별로 본 도서관의 시스템에 서비스용 파일을 업로드함.
- 웹서비스용 이미지 파일의 무단 자료유출 방지를 위한 기술적인 조치(워터마킹 등 DRM 관련 기술) 시행

## 5. 결 론

지금까지 연세대학교 중앙도서관에서 진행하여 온 그간의 고문헌 디지털컬렉션 구축 사업을 사업 단위의 사례 보고 형식으로 정리하여 보았다.

고문헌을 관리하는 도서관인들은 누구나 느끼는 것이겠지만 자료의 열람/복사에 따른 원본 훼손 우려와 이용자에 대한 만족할 만한 서비스 제공이라는 두가지 목표 사이에서 겪어오던 갈등들이 원문 디지털화로 해결하게 된 것은 크게 고무적인 일이라고 하지 않을 수 없겠다. 이와 함께 고문헌 원문을 한번 디지털화시켜 두면 다양한 방면에서 활용할 수 있게 되는 편리함도 있다. 연구 대상으로 활용하는 경우는 물론이고, 강의 교안 작성에 활용하는 경우도 자주 보게 된다. 본교 교육개발지원센터에서 인물 200 선, 필독도서 200 선 등과 함께 고서 200 선을 선정하여 홈페이지에서 제공할 때에도 본 도서관에서 구축한 고문헌 디지털컨텐츠를 그대로 가져다가 활용한 경우도 있다.

2000 년 우리가 처음 사업을 시작할 때보다 지금의 디지털 환경은 훨씬 더 발전되어 있어서 지금 와서 돌아보면 금석지감이 없지 않다. 당시만 하더라도 국내에 컬러이미지용 오버헤드 스캐너가 널리 보급되지 않고 비용이 비싸서 좀더 저렴한 gray 형식으로 스캔을 하였는데, 지금은 고문헌을 거의 컬러로 스캔하여 DB 를 구축하지 gray 형식으로 DB 구축하는 경우가 거의 없는 것으로 알고 있다. 앞에서 설명하였던 파일 포맷 변경(PDF -> Djvu) 같은 경우도 당시에는 PDF 포맷이 압축을 등에서 문제가 있었으나 지금은 많이 업그레이드 된 제품을 출시하고 있다. 디지털 기술의 발전에 발맞추어 꾸준히 시스템을 업그레이드해 나가야 하는 것은 우리에게 새롭게 주어진 과제라고 할 것이다.

이 밖에 디지털 콘텐츠의 보존성, 디지털화와 마이크로필름화와의 병행 여부, 모든 고문헌 소장기관들이 네트워크로 연결되는 상황에서 동일자료의 중복 디지털화 우려 등도 쉽게 해결할 수 없는 과제로 우리 앞에 놓여 있는 것으로 보인다.

