

대학도서관의 학위논문 전문DB 구축방안

Design for Creating Full-Text Database of Korean Dissertation

방 준 필(Jun-Pil Bang)*

목 차

- | | |
|---------------------|-----------------|
| 1. 서 론 | 3. 고려대학 도서관의 상황 |
| 2. 전문DB 구축을 위한 고려사항 | 4. 전문DB 구축원칙 |
| 2.1 파일형식 | 5. 전문DB 구축방향 |
| 2.2 뷰어 | 6. 결 론 |
| 2.3 검색 | |
| 2.4 저작권 | |

초 록

대학도서관에서 전자도서관 서비스를 위하여 전문(Full-Text) DB, 특히 학위논문 DB를 구축하는 방법을 제시하였다. 알파벳 문서와 달리 한글 문서는 전문DB로 구축하는 데 범용성을 가진 표현화된 방법이 많지 않다. 따라서 관련기술의 발전을 고려하여 적절한 DB 구축방법을 설계해야 한다. 이를 위하여 텍스트 형태와 이미지 형태의 파일 포맷, 각종 솔루션, 뷰어, 검색, 저작권을 고려하여 전문DB 구축원칙을 정하고, 고려대학교 도서관을 모델로 적용하여 전문DB 구축 원칙과 방향을 설정하였다.

ABSTRACT

The purpose of this study is to design the solution for creating full-text database of Korean dissertation. After considering file formats for text based and image based database, Viewer, Search, Copy Right, Abstracts and Indexes, situation of Korea University Library, decided the principles of creating database. And suggested the design to produce the database for Korea University Library, that is easy to get file format conversion in case of the introducing new technology for the future.

* 성균관대학교 강사
접수일자 1998년 11월 24일

1. 서론

현대의 도서관은 거의 모두 자동화시스템을 도입하여 자료 처리와 서비스의 효율을 높이고 있다. 도서관의 자동화는 기존의 도서관 업무가 전산 처리되어 목록카드를 생산하지 않는 시스템으로 나타나고 있지만, 정보를 담고 있는 자료 자체 즉, 실물에 대한 메타데이터를 제공하는데 그치고 있다. 이러한 약점을 보완하기 위하여 근래 인터넷 시대에 구축되기 시작한 전자도서관은 시간과 장소에 관계없이 실물의 내용 즉, 전문을 제공하는 기능을 가지고 있어 각광을 받고 있다. 자동화시스템이 가동되는 도서관은 인터넷, 홈페이지, 웹서버, DB 서버 등의 환경이 갖추어져 있어 관련 장비와 프로그램, DB만 갖추면 곧 전자도서관 서비스를 할 수 있다. 전자도서관의 구축은 서비스 확대차원이 아니라 이제 도서관을 평가하는 기준이 되었으며 도서관 간의 경쟁에 있어서도 고려되는 사항이다.

대학도서관은 서비스 수준을 높이기 위하여 전자도서관 사업을 진행중이거나 착수할 준비를 하고 있다. 그런데 전자도서관은 물리적인 시스템 환경보다도 DB가 더 중요하다. 아무리 시스템의 설계가 잘 되었다 하더라도 이용할 수 있는 전문DB의 양이 적고, 이용하는 데 불편하다면 전자도서관을 구축한 효과가 감소할 것이다. 전자도서관에서 구축하는 DB 가운데는 전문DB 외에도 음향DB, 화상DB도 있지만 특수한 목적을 가진 도서관 외에는 대부분 전문DB를 구축하는 것이 우선일 것이다. 전문DB를 구축하는 데에는 저작권 문제가 있어 해당 대학에서 생산되는 자료, 특히 학위논문을 대상으로 하고

있는 실정이다. 그런데 관련 기술의 발전 및 변화로 인한 시행착오, 전문DB 구축상의 과도한 비용 문제 때문에 시스템 설계가 어려워질 수 있다.

본 연구는 이러한 상황에서 대학도서관의 전자도서관 서비스를 위하여 학위논문 전문DB를 효과적으로 구축하기 위한 방안을 제시하는 데 목적이 있다. 연구에서 고려하는 솔루션은 현재까지 개발되어 실제로 이용되고 있는 것을 대상으로 하며, 논의의 배경이 되는 환경은 고려대학교 도서관을 모델로 한다.

2. 전문DB 구축을 위한 고려사항

2.1 파일 형식

전자도서관에서 전문DB를 구축하기 위해서는 어떤 파일 형식으로 제작할 것인지를 결정해야 한다. 전문DB는 이미지 형태나 텍스트 형태로 구축할 수 있다. 텍스트 파일을 입수할 수 있는 경우는 전문DB를 텍스트 형태로 혹은 텍스트 형태와 이미지 형태를 동시에 구축하고, 텍스트 파일을 입수할 수 없는 경우 이미지 형태로 구축하는 것이 일반적이다.

2.1.1 텍스트 형태

텍스트 형태의 DB는 먼저, 내용이 수정, 조작될 수 있으므로 전문DB의 경우 내용의 완전성이 보장되어야 한다. 두 번째, 인쇄본의 모습을 완벽하게 재현하지 못할 수 있다. 이는 화면상의 디스플레이 문제와 텍스트의 편집 과정이 포함되기 때문이다. 세 번째, 문헌의 인용시 인

쇄본의 경우처럼 출처의 정확성을 나타내지는 못한다. 그럼에도 불구하고 텍스트 형태의 전문 DB가 많이 구축되는 것은 단점을 보완하는 장점이 크기 때문이다. 텍스트 형태의 데이터는 화면에 디스플레이되거나 프린터에 출력될 경우 미려하다. DB구축 비용이 상대적으로 저렴하고, 파일 저장용량이 작고 이미지 파일에 비해 전송속도가 빠르다. 그리고 내용 전체를 대상으로 검색이 가능하다는 점이 가장 큰 장점이다. 텍스트 형태로 전문DB를 구축하는 데 있어서는 융통성이 많다. 이는 어떤 파일 형식이든 전문이 확보되어 있는 한 필요시 파일을 변환하여 서비스 방법을 개선할 수 있기 때문이다.

파일 형식에 대해서는, 전자도서관이 인터넷 상에서 웹 브라우저를 통해 서비스되고 있기 때문에 이와 관련한 것을 먼저 고려한다. 웹 브라우저 자체가 지원하는 파일 형식은 HTML과 txt인데, 워드프로세서 파일을 HTML이나 txt 파일 형식으로 변환하면 원문의 모습을 완벽하게 재현하지 못하기 때문에 인터넷 상에서 원문의 모습을 잘 나타내도록 지원하는 파일 형태로 변환하는 것이 일반적이다. 여기에 해당하는 파일 형태가 sgm이다. sgm 파일은 텍스트 파일을 SGML 규정에 따라 편집하여 만든 것이다.

SGML(Standard Generalized Markup Language)은 ISO에서 정한 문서 표현 기준으로 제정된 것으로, DTD에 따라 문서의 구조를 지정하는 마크업 언어이다. 웹 브라우저가 사용

하는 HTML(HyperText Markup Language)언어는 SGML을 기반으로 한 언어의 일종이다. 마크업이란 편집자나 인쇄디자이너가 색인, 텍스트 및 그 외의 부분에 대해서 서체나 사이즈 등 체제에 관한 지정을 원고 상에 손으로 기입하는 방식이다.¹⁾

SGML은 문서 데이터베이스의 작성이나 유지, 보수에 유효할 뿐만 아니라, 교환 유통에도 적당하다. 개개의 시스템에 독립되어 있을 뿐만 아니라 구조가 엄격히 정의되어 있어, 개개의 시스템을 위하여 변환할 수 있기 때문이다. 나아가 SGML에서 만들어진 문서 데이터베이스는 전자도서관 시스템의 자원으로 사용된다.²⁾

DTD(Document Type Definition)는 SGML(Standard Generalized Markup Language)의 규정에 따라 문헌의 문단을 분리하고 표목을 확인하는 등의 방법과 절차에 관한 코드(혹은 markup)를 정한 명세서이다. 문헌을 DTD로 전송하면 DTD 리더(혹은 SGML 콤파일러)를 통하여 그대로 화면에 나타내거나 출력할 수 있다. 이는 표준적인 SGML 콤파일러 하나가 여러 가지 마크업 코드를 사용한 다양한 문헌을 처리할 수 있음을 의미한다. 콤파일러는 DTD에 따라서 화면이나 프린터에 출력한다.³⁾

SGML로 구축된 전문DB는 국립중앙도서관, 국회도서관, 한국과학기술원 과학도서관, 한국과학기술원 연구개발정보센터 및 한국학술진흥재단 등의 5개 기관에서 공동으로 97년도에

1) 유사라, 정보화사회와 도서관 정보네트워크, p. 450

2) bid. p. 451.

3) <http://www.whatis.com/>

시행한 전자도서관 시범사업⁴⁾에서 볼 수 있다. 구축된 DB의 내용은 국회도서관의 국회사(7,933면), 국정감사결과시정 및 처리사항(10,116면), 연구개발정보센터의 정보과학회지(355편), 한국과학기술원 과학도서관의 박사논문(80책) 및 교수연구논문(115편)이다.⁵⁾

조왕근은 SGML 파일로 변환하는 과정을 제시하였다.⁶⁾ 그러나 변환결과를 검증하는 시간과 노력이 많이 필요하다. 따라서 원문을 완벽하게 재현하는 SGML 자동변환기가 개발되기 전에는 많은 양의 학위논문을 전문DB로 구축하기는 현실적으로 어렵다. SGML의 대안으로 SGML과 HTML의 중간 성격을 가진 XML 방식을 고려할 만하다.

XML(Extensible Markup Language)은 WWW, 인트라넷 등에서 정보 포맷을 생성하고 포맷과 데이터를 공유하는 도구로서, 현재 WWW 컨소시움에서 웹의 유용성을 높이기 위하여 공식적으로 권하고 있다. XML은 문헌이나 파일을 기술하는 마크업 기호를 포함하기 때문에 웹 페이지, HTML 등의 언어와 유사하다. 그런데 HTML이 웹 페이지의 내용(문장과 그림의 배열, 관계)을 기술하는 데 비해 XML은 데이터가 기술되는 용어로 내용을 기술한다. 예를 들면 <PHONENUM>은 전화번호라는 데이터를 나타낸다. 이는 XML 파일이 프로그램에 의해서 단순히 데이터로서 처리될 수 있거나 다른 컴퓨터에 유사한 데이터로 저장될 수

있거나 HTML 파일처럼 출력된다는 것을 의미한다.⁷⁾ XML이 SGML과 HTML의 단점을 극복하기 위하여 개발되었으므로 이에 대한 기대감은 높지만 아직은 국내 한글 환경에서 널리 사용되고 있지 않다.

근래 급속히 보급되고 있는 전문DB의 파일 형식으로 PDF가 있다. PDF(Portable Document Format)는 윈도우나 맥킨토시, 유닉스 등 어떤 타입의 컴퓨터 시스템 환경에서도 전송과 읽기가 가능하도록 지원되는 포맷으로서 이러한 포맷은 그래픽 소프트웨어 회사인 Adobe사에서 PostScript 언어에 기반을 두고 만들어졌으며, PDF에는 자체의 압축기능을 포함하고 있어 인터넷에서 파일을 작은 사이즈로 만들어 전송할 수 있다. 또한 On-line 환경이나 Off-line 환경에서도 여러 전송 수단을 통하여 문서정보의 공유 및 전송 등의 여러 장점을 가지고 있는 File Format이다.⁸⁾

PDF는 원본과 동일한 모습을 보이고 서체 및 페이지 포맷정보를 내장하고 있으며 이미지를 포함하기 때문에 텍스트 파일과 이미지 파일 모두를 PDF 파일로 변환할 수 있는 것이다. 북마크, 하이퍼링크, 전문의 내용 검색 등이 가능하고 뷰어도 인터넷 상에서 쉽게 받을 수 있다. 아직 일부 워드프로세서 파일밖에는 PDF 파일로 변환할 수 없기 때문에 국내에서 서비스되는 대부분의 PDF 형식의 전문DB는 인쇄본을 스캐닝한 것이다.

4) <http://ultra.nl.or.kr/>

5) 이치주, 전자도서관: 자료의 디지털화 및 이용환경의 정비, 제2차 한일 국립도서관간 업무교류 발표자료, 1998. 8. p. 7.

6) 조왕근, 학위논문 SGML DTD 및 데이터베이스 구축에 관한 연구, p. 72-82.

7) <http://www.whatis.com/>

8) 김휘출, 인터넷을 이용한 참고봉사 한국디지털도서관포럼, 1988. p. 237.

PDF 방식과 유사한 XLX 방식이 국내에서 개발되었다. 파일의 포맷을 PCL6 (PCL XL)로 변환하여 XLX 파일 형식으로 통합하여 관리하는 것이다. 즉, XLX는 윈도 어플리케이션 상의 내용을 프린터에서 파일로 출력한 결과이다. XLX 문서는 데이터를 코드값으로 가지고 있으므로 페이지 단위까지 검색이 가능하며 전송 속도가 빠르다. 줄긋기, 도형 그리기(원, 네모), 형광펜 등의 마크업 기능이 있다. 워드프로세서 파일을 XLX 형식으로 변환하면 파일 크기가 워드프로세서의 종류에 따라 다소 차이가 있지만 대체로 줄어든다. 개발된 지 얼마 안 되기 때문에 일부 대학도서관에서 이러한 파일 형식의 전문DB를 구축하고 있을 뿐이지만 모든 워드프로세서 파일을 XLX 형식으로 자동 변환시키는 프로그램의 개발로 보급률이 높아질 것으로 전망된다.

워드프로세서 파일을 다른 형태로 변환하지 않고 그대로 전문 DB로 구축하는 방안도 있다. 이러한 방법은 DB 구축비용이 저렴하지만, 여러 가지 파일 형식의 데이터를 하나의 뷰어로 볼 수 없다는 이용상 불편하다는 문제가 있다. 따라서 도서관의 홈페이지에서 파일 형식을 지원하는 전용 뷰어를 각각 제공하거나, 이용자가 파일을 전송받아 개인 PC에서 해당 워드프로세서로 보는 수밖에 없다. 후자의 방법은 전자도서관의 서비스로서 적합하지는 않지만, 임시방편으로 일시 사용할 수는 있을 것이다.

2.2.2 이미지 형태

전문DB를 구축하기 위한 텍스트 파일이 확보되지 않았거나, 적합한 텍스트 형식의 전문DB 구축방법이 확립되지 않았을 때는 학위논문

인쇄본을 스캐닝하여 이미지 형태의 전문DB를 구축한다. 이미지 형태의 전문DB는 텍스트 형태보다 파일의 크기가 크고 구축비용이 높고, 내용을 검색할 수 없는 단점이 있지만 원문의 모습을 전하기 때문에 인용면에서 장점이 있으며, 범용성이 뛰어난 솔루션이 이미 개발되어 있다.

이미지 형태의 전문DB를 구축하기 위해서는 파일 형식을 고려해야 한다. 원문을 스캐닝하여 생산된 이미지는 저장하는 압축방식에 따라 파일 형식이 정해지는데, 문헌의 이미지로는 JPG와 TIFF 형식이 많이 사용된다.

JPG는 JPEG(Joint Photographic Experts Group)라고도 하는데, 문헌을 스캐닝하면 페이지 수만큼의 파일을 생성한다. 이 경우 학위논문 한 타이틀을 구성하는 많은 파일을 관리해야 하는 단점이 있지만 내용을 웹 브라우저 상에서 직접 볼 수 있는 장점이 있다.

TIFF(Tagged Image File Format) 형식에는 JPG처럼 한 페이지를 각각 파일로 저장하는 Single TIFF와 전체 페이지를 하나의 파일에 저장하는 Multi TIFF가 있다. Multi TIFF는 이미지 정보에 목차, Page Matching 정보 등의 표시가 가능하며, 이미지 원문에서 처리하기가 불가능한 목차정보를 텍스트로 처리할 수 있다. 따라서 학위논문을 이미지 형태로 전문DB를 구축하는 데 있어서 Multi TIFF 방식이 선호되고 있다. TIFF 방식의 데이터는 웹 브라우저 상에서 직접 볼 수 없기 때문에 전용 뷰어를 별도로 제공해야 한다.

이미지 파일의 범용성을 높이기 위하여 TIFF 파일을 PDF 형식으로 변환하여 전문DB를 구축하는 방법도 있다. PDF가 텍스트 파일

뿐만 아니라 이미지 파일을 수용하고, 마크업 기능 등의 편집이 가능하기 때문에 근래 널리 사용되고 있다.

2.2 뷰어(Viewer)

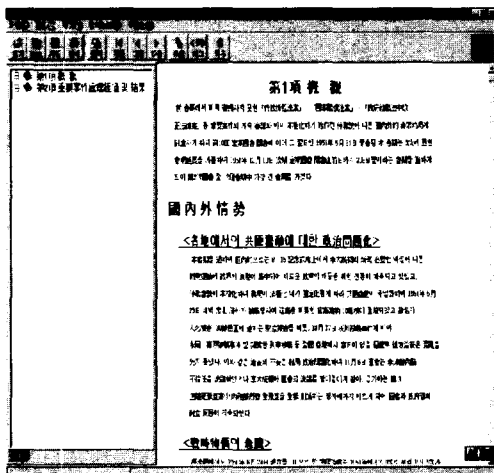
SGML 형식의 파일을 보여주는 뷰어는 <그림 1>과 같다. SGML 뷰어는 하이퍼링크의 기능을 갖는 목차 프레임과 내용을 보이는 프레임으로 구성된다. 파일을 쪽별로 보이지 않고 전체를 한꺼번에 보이기 때문에 로딩하는 시간이 길고, 페이지 이동 키가 기능을 못한다. SGML 태그가 있는 소스를 볼 수 있으며, Front, Body, Back 등의 구조별로 볼 수도 있다.

워드프로세서 파일을 XLX 형식으로 변환하여 구축한 전문DB는 Jet Document 뷰어로 볼 수 있다. Jet Document 뷰어는 파일이 전송되는 동안 먼저 도착한 부분부터 보인다. 전문 내용을 검색할 수 있으며, 페이지 이동, 화

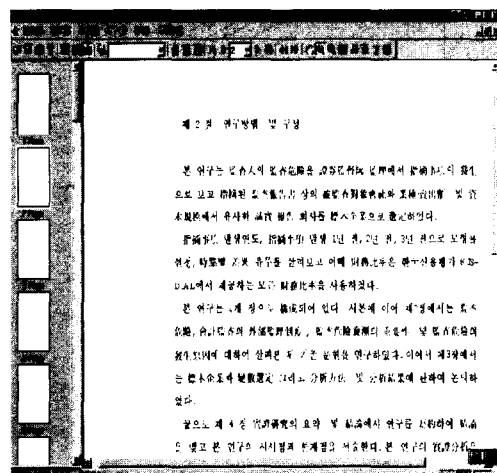
면 확대 및 축소, 하이퍼링크를 포함한 편집 기능이 있다. XLX 형식 외에 TIF, PCL, PRN 파일 형식도 볼 수 있다.

워드프로세서 파일을 전문DB로 구축할 경우 이를 볼 수 있는 뷰어 가운데, HWP 파일을 볼 수 있는 것은 나모 HWP 뷰어 이다. Windows 시스템에서 한글 프로그램 없이 웹 브라우저로 한글 문서를 직접 읽을 수 있는 나모 HWP 뷰어는 원본의 모습을 거의 그대로 재생한다. 페이지 이동, 점프, 확대, 쪽 내용 보기, 쪽 윤곽 보기, 여백 및 안내선 보기 기능이 있다. 편집할 필요가 있을 때 한글 프로그램을 실행시킬 수 있다.

이미지 형태의 전문DB도 파일 형식에 따라 뷰어가 다르다. JPG 파일의 전문DB는 웹 브라우저 상에서 직접 볼 수 있어 별개의 뷰어가 필요없는 장점이 있다. JPG 파일의 데이터를 보이는 <그림 4>의 경우, 원문 한 장의 내용이 하나의 파일 단위로 저장되고 디스플레이 되기



<그림 1> SGML 뷰어



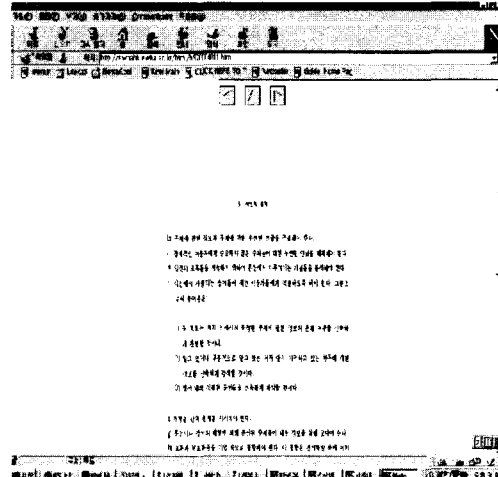
<그림 2> JetDocument 뷰어

(별첨)

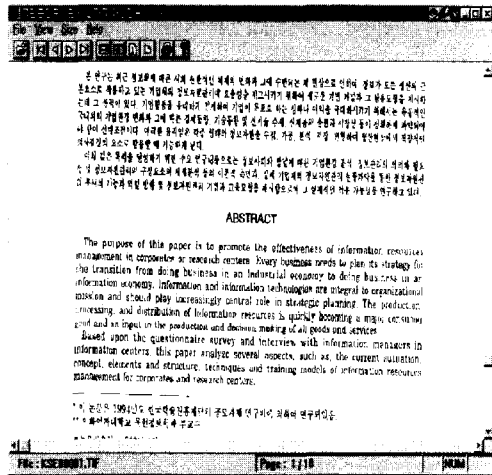
진료비 감면기준표

번호	구분	의료보험		건강보험		보험·고정		임 환	
		본인	직계	본인	직계	본인	직계	본인	직계
1	의료원·본교교직원·병원에서 일 직원	30	50	50	40	30	20	30	30
2	의료원·본교교직원·병원에서 일 직원의 배우자·부모	30	-	40	-	20	-	30	-
3	의료원·본교교직원·병원에서 일 직원의 형제자매	10	-	30	-	10	-	20	-
4	연년(영)복지 고지원	30	30	50	30	30	20	50	30
5	본교 조교·간사·보급기관 직원	10	-	50	30	30	10	30	20
6	대학 및 대학원 재학생	10	-	30	30	10	-	25	-

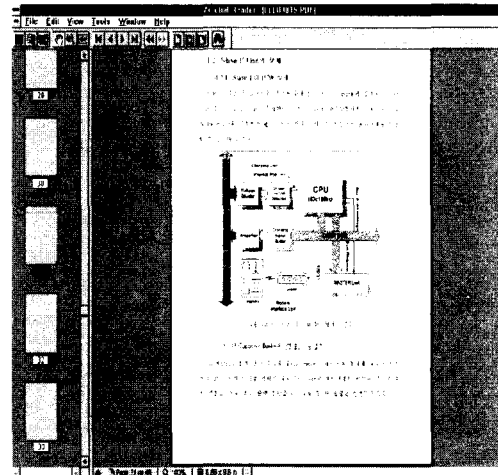
<그림 3> 나모 뷰어



<그림 4> JPG 형식의 전문DB 뷰어



<그림 5> TIFF 형식의 뷰어



<그림 6> PDF 뷰어

때문에 논문 전체 출력 혹은 복사, 한 화면 상에서의 페이지 점프, 확대 등의 기능이 없다. 대신 목차 페이지로 이동하여 원하는 부분을 볼 수 있다.

TIFF 파일 전용 뷰어의 경우 <그림 5>와 같

이 페이지 점프, 확대 및 축소 기능이 지원된다.

PDF 파일의 전용 뷰어인 Acrobat Reader기는 Internet 상에서 쉽게 구하여 설치할 수 있다. <그림 6>과 같이 페이지 이동, 확대 및 축소 기능이 있으며 페이지 단위의 프

런터 출력이 가능하다. 긴 파일의 첫 페이지를 먼저 받고, 나머지 Background cache로 내려 받기 때문에 기다리는 시간이 짧다.

위에서 제시되었듯이 뷰어는 전문DB의 파일 형식과 직접 관계가 있어 파일 형식의 특성에 따라 뷰어의 기능이 결정된다. 그리고 뷰어의 디자인은 프로그래밍에 의해 개선될 수 있는 것이다.

2.3 검색

전문DB를 검색하기 위해서는 색인이 구축되어야 한다. 색인은 전문의 내용 중에 원하는 항목이 어디에 위치하는지를 알려주는 기능과 검색어로 사용되는 기능이 있다. 색인의 대상은 저자, 인명, 주제어, 초록 단어 등으로 구분한다. 색인은 도서관 혹은 전문업체에서 직접 구축하거나 논문 작성자가 논문 파일과 색인 파일을 함께 제출하도록 하는 방법이 있다.

초록은 원문의 내용을 파악하거나, 원문을 볼 것인지 판단하는 도구로서 검색에 필요한 요소이다. 전문 DB를 구축하는 데 있어서 초록이 우선적으로 구축되어야 한다. 즉, 전문은 없어도 초록은 반드시 있어야 하는 것이다. 초록에는 연구의 목적, 연구방법, 결과 등의 내용을 알리는 통보적 초록, 연구주제 혹은 방향 등을 알리는 지시적 초록 그리고 이 두가지를 혼합한 형식이 있다. 그리고 초록의 길이는 학위논문의 경우 한 페이지 이내의 분량인 600자 정도가 적절하지만, 전자문서의 경우 보통 300자 내외의 길이를 선호하고 있다. 대개의 논문 제출자는

초록작성법에 대하여 교육을 받지 않고 각자 나름대로 작성하기 때문에 형식과 길이가 일정하지 않다. 따라서 도서관에서는 논문파일을 제출 받기 전에 초록의 형식과 길이를 정하여 이에 따르도록 유도하고, 제출된 초록을 확인하여 필요시 수정해야 한다.

검색은 전문DB 가운데 원하는 자료를 찾기 위한 것과 전문 가운데 특정 단어나 어구를 찾는 것으로 나누어 고려할 수 있다. 전자의 경우, 일반 서지DB 시스템과 전문DB 시스템을 통합과 분리의 여부에 따라 다르다. 통합된 경우 서지DB 시스템과 같은 서버에 색인DB를 구축하고 같은 검색엔진이 사용되지만, 분리된 경우 별도의 서버와 전용 검색 엔진이 사용된다. 전문의 내용을 대상으로 검색하는 경우, 텍스트 파일을 기반으로 구축된 DB에는 뷰어에서 제공하는 검색기능이 지원되지만 이미지 파일로 구축된 전문은 검색이 되지 않는다.

OCR을 이용하여 전문의 내용을 검색할 수도 있다. 원문 상의 글자 이미지를 텍스트로 인식시키는 OCR 기술은 문자 인식률을 얼마나 높이느냐가 관건이다. PDF의 경우 스캐닝할 때 OCR 방식의 옵션으로 설정하면 내용 검색이 가능하지만 한글이 지원되지 않아 국내 학위논문을 대상으로 이용할 수 없다.

엑스칼리버사(Excalibur Technologies Corporation)⁹⁾에서 개발한 Excalibur RetrievalWare FileRoom은 스캐닝한 자료의 색인어를 추출하는 데 있어서 검색기인 Excalibur RetrievalWare 기술을 기반으로 한다. Adaptive Pattern Recognition

9) <http://www.excalib.com>

Processing에서 제공되는 퍼지기능으로 스캐닝 과정에서 색인어가 불명확하게 인식되는 부분을 코드의 패턴을 인식하여 저장한다. 검색어를 기입할 때 철자법이 틀린 단어도 코드의 패턴을 인식하여 인접한 색인어를 부르게 된다. 이러한 검색엔진에 한글처리가 가능하도록 하였으며, 한글 형태소 분석기를 적용하여 색인을 추출하여 전문 내용을 검색할 수 있다. 또한 OCR의 가장 큰 약점인 한글 인식을 문제는 패턴인식 처리과정을 통해 정확률이 높은 결과부터 먼저 보여줌으로써 해결한다.

이용자가 검색을 하도록 제공되는 화면의 디자인과 기능 즉, 유저 인터페이스는 별도의 교육 없이도 바로 사용할 수 있도록 우호적이어야 한다.

진정한 의미의 이용자 우호적인 인터페이스란 서로 이질적인 시스템들 사이에서의 정보 검색이나 정보전달, 또는 정보를 표현할 때, 이용자들이 시스템의 사용에 어려움을 느끼지 않도록 함을 의미하기도 한다. 이는 서로 이질적인 시스템에 수록된 서로 다른 데이터 포맷을 갖는 정보들도 이용자가 정보의 결절을 느끼지 않고 자유롭게 검색하여 이용자의 편의에 따른 포맷에 맞추어 이용자에게 디스플레이 되거나 제공

되어야 함을 뜻한다.¹⁰⁾

2.4 저작권

전문 DB를 구축하면서 고려해야 하는 법적 인 문제가 저작권에 관한 사항이다.

저작권법은 저작자의 권리와 이에 인접하는 권리를 보호하고 저작물의 공정한 이용을 도모함으로써 문화의 향상발전에 이바지함을 목적으로 한다.¹⁰⁾ 여기서 저작자란 저작물을 창작한 자를 말하는데, 저작물이란 문학·학술 또는 예술의 범위에 속하는 창작물을 말한다.¹²⁾ 저작자는 저작인격권¹³⁾과 저작재산권¹⁴⁾을 가지며, 저작권은 저작한 때부터 발생하며 어떠한 절차나 형식의 이행을 필요로 하지 아니한다.¹⁵⁾ 저작재산권은 저작자의 생존하는 동안과 사망후 50년간 존속한다. 다만, 저작자가 사망후 40년이 경과하고 50년이 되기전에 공표된 저작물의 저작재산권은 공표된 때부터 10년간 존속한다. 공동저작물의 저작재산권은 맨 마지막으로 사망한 저작자의 사망후 50년간 존속한다.¹⁶⁾

공표란 저작권을 공연·방송 또는 전시 그밖의 방법으로 일반공중에게 공개하는 경우와 저작물을 발행하는 경우를 말한다.¹⁷⁾ 복제는 인

10) 김성혁, 디지털도서관의 문헌 특성 및 관리, p. 60.

11) 저작권법 제1조

12) 저작권법 제2조 제1항 및 제2항

13) 저작권법 제11조부터 제13조에 의하면, 저작자는 저작물을 공표하거나 공표하지 아니할 것을 결정할 권리를 가진다. 제3자에게 이용을 허락한 경우에는 저작물의 공표를 동의한 것으로 추정한다. 저작자는 저작물의 원작품이나 그 복제물에 또는 저작물의 공표에 있어서 그의 이름을 표시할 권리가 있다. 저작자는 저작물의 내용·형식 및 제호의 동일성을 유지할 권리를 가진다.

14) 저작권법 제16조부터 제21조까지 규정한 바에 의하면, 저작자는 그 저작물을 복제할 권리를 가진다. 저작자는 저작물의 원작품이나 그 복제물을 배포할 권리를 가진다. 저작자는 그 저작물을 원저작물로 하는 2차적저작물 또는 그 저작물을 구성부분으로 하는 편집저작물을 작성하여 이용할 권리를 가진다.

15) 저작권법 제10조

16) 저작권법 제36조

쇄·사진·복사·녹음·녹화 그밖의 방법에 의하여 유형물로 다시 제작하는 것을 말한다.¹⁸⁾ 배포는 저작물의 원작품 또는 그 복제물을 일반 공중에게 대가를 받거나 받지 아니하고 양도 또는 대여하는 것을 말한다. 발행은 저작물을 일반공중의 수요를 위하여 복제·배포하는 것을 말한다.¹⁹⁾

공개된 저작물을 영리를 목적으로 하지 아니하고 개인적으로 이용하거나 가정 및 이에 준하는 한정된 범위 안에서 이용하는 경우에는 그 이용자는 이를 복제할 수 있다.²⁰⁾ 도서관의 경우는 다음과 같이 저작권법 제26조에서 규정한 범위 내에서는 복제할 수 있다.

- ① 조사·연구를 목적으로 하는 이용자의 요구에 따라 공표된 저작물의 일부분의 복제품을 1인 1부에 한하여 제공하는 경우
- ② 도서관등이 자료의 자체보존을 위하여 필요한 경우
- ③ 다른 도서관등의 요구에 따라 절판 그 밖에 이에 준하는 사유로 구하기 어려운 저작물의 복제물을 보존용으로 제공하는 경우

저작권의 기본적인 목적인 저작권자의 노동에 대하여 보호하려는 것이 아니라 과학과 실용 예술의 발전을 촉진하려는 것이다. 이러한 목적

에서 저작권은 저작권자에게 그들의 독창적인 표현에 대한 권리를 보장하는 대신 다른 사람들이 저작물에 대한 아이디어와 정보를 자유로이 이용할 수 있도록 유도한다.²¹⁾

현행 법규상 도서관에서 Text 형태나 이미지 형태로 전문 DB를 구축하는 경우 자료를 보존하기 위해서라면 면책되지만 이용을 위해서라면 허용되지 않는다.²²⁾ 초록의 경우 저작권의 취지에 따라 전자도서관에서 이를 구축하고 이용해도 무방하다. 따라서 학위논문 제출자에게 전문 DB를 구축하고 이를 이용시켜도 좋다는 동의서를 받은 경우에 한하여 작업 대상으로 삼아야 한다. 이에 대한 법적 근거는 다음과 같다.

저작권법 제54조 제1항은 “저작물을 복제·배포할 권리를 가진 자(이하 “복제권자”라 한다)는 그 저작물을 인쇄 그밖의 이와 유사한 방법으로 문서 또는 도서로 발행하고자 하는 자에 대하여 이를 출판할 권리(이하 “출판권”이라 한다)를 설정할 수 있다”고 하였으며 제2항에서는 “출판권을 설정받은 자(이하 “출판권자”라 한다)는 그 설정행위에서 정하는 바에 따라 그 출판권의 목적인 저작물을 원작 그대로 출판할 권리를 가진다”고 하였다. 여기서 발행이란 저작물을 일반공중의 수요를 위하여 복제·배포하는 것을 말한다.²³⁾ 또한 동 제42조에 의하면, 저작재산권자는 다른 사람에게 그 저작물의 이용을 허락할 수 있으며 허락을 받은 자는 허락받은 이용방법 및 조건의 범위 안에서 그 저작물을 이

17) 저작권법 제2조 제17항

18) 저작권법 제2조 14항

19) 저작권법 제2조 제15항

20) 저작권법 제27조

21) 황찬현, 전자도서관과 저작권, <http://www.nanet.go.kr/nal/3/3-1-4/nal96123.htm>

22) 저작권법 제28조, 동법 시행령 제3조

23) 저작권법 제2조 제16항

용할 수 있다고 하였다. 따라서 도서관에서 저작권자로부터 출판권을 인정한 동의서를 받으면 전문DB를 구축하여 이를 이용시키는 데 있어서 문제가 없다.

3. 고려대학 도서관의 환경

고려대학 도서관은 중앙관과 세 곳의 분관으로 구성되며 총 160만 권의 장서를 소장하고 있으며, 6개 부서에 직원이 64명 수준인 대규모 도서관이다. 1994년부터 자동화시스템이 가동되었으며, 인터넷으로 목록검색과 CD-ROM 자료 서비스를 지원하기 위하여 1996년부터 홈페이지가 구축되었다.

'98학년도 전기 졸업 일반대학원 석·박사 학위논문 489건의 전문DB를 시범적으로 구축하여 1998년 3월부터 전자도서관 서비스를 시작하면서 인터페이스와 문제점을 개선하였다. 전문DB는 Winframe 서버에 설치하고, 색인은 전용 NT서버에 설치하여 MS Access로 검색 프로세스를 수행한다. Winframe은 윈도우 어플리케이션을 동시에 다수 이용자가 이용할 수 있도록 지원하는 서버로서, 데이터 자체를 화면값으로 전송하기 때문에 전문을 화면 상에서 보는 데는 신속한 장점이 있다.

현재 '98학년도 후기 졸업생의 학위논문 전체(427건)와 전기 졸업생 학위논문 가운데 미구축분(318건)에 대한 전문 DB를 구축하려고 하고 있다. 이를 위하여 전문DB가 구축된 이후 현재까지 솔루션을 포함한 관련기술, 지원업체의 상황이 달라져 전반적인 검토를 다시 해야하는 시점을 맞이 하였다.

원문 파일의 경우 논문 제출자로부터 디스크으로 받았는데, 파일 제출에 대한 강제성이 없기 때문에 '98학년도 일반대학원 전기 학위취득자에 대한 파일 수집율이 70%밖에 되지 않는다. '98년 후기 학위취득자에 대한 파일 수집율은 82.2%이다. 제출된 파일을 분석한 결과 하나의 파일로 편집해야 하거나 원본의 모습을 보이지 않는 불완전한 파일이 60%를 초과한다. 앞으로 교육과 홍보를 지속하면 논문 파일의 완전성을 높일 수 있을 것이다. 고려대학교 도서관의 홈페이지의 URL은 "http://kulib.korea.ac.kr"이다.

4. 전문DB 구축원칙

전문DB를 구축하는 데 있어서 위에서 제시한 여러 사항과 도서관의 환경을 함께 고려한 결과, 다음과 같은 전문DB 구축의 원칙을 정하고 이에 가장 적합한 설계를 하였다.

가. 표준과 범용성을 따른다.

전자도서관의 서비스는 인터넷 상에서 이루어지므로 전세계에서 이용할 수 있도록 표준방식과 범용성이 뛰어난 방식으로 DB를 구축, 서비스한다. 더구나 각 대학도서관에서 구축한 전문DB가 공동으로 이용될 것을 고려한다면 통합에 유리한 표준을 따른다.

나. 이용자의 편의를 높인다.

이용자가 웹으로 인터넷을 이용할 수 있는 PC 환경이 486급 이상이며 웹 브라우저가 설치된 것을 기준으로 전문DB 시스템을 설계한다. 그리고 특정 워드프로세서가 없어도 이용하는

데 지장이 없도록 한다.

다. 전문 DB를 구축함에 있어서 초록을 먼저 제공해야 한다. 이용자는 초록을 참조함으로써 전문을 볼 것인지 결정한다. 모든 DB는 색인 → 초록 → 전문 제공의 순으로 설계되는 것이 원칙이다. 초록의 형식과 분량을 지정하여 논문 파일 제출시 이에 따르도록 하고, 제출된 초록을 검토하여 필요시 수정, 편집한다.

라. 전문DB 시스템은 서지DB 시스템과 인터페이스가 통합되어야 한다. 이용자 입장에서는 검색시스템에서 일반 자료를 이용할 때와 동일한 인터페이스를 갖는 것이다. 예를 들면, 자료의 소장 여부를 확인하는 검색 프로그램 상에서 학위논문 검색을 한 후 디스플레이되는 간략 검색 사항에 전문이 구축되어 있는지를 표시한다. 또는 검색 옵션에서 전문DB를 지정하여 검색한다. 검색 결과 원하는 자료에 대한 초록을 보고 전문을 볼 것인지 결정한다.

마. 학위취득자가 논문 파일을 제출하는 제도를 만들어 파일을 전량 수집한다. 디스켓에는 논문 내용 전체가 한 개의 파일로 작성되도록 하고, 색인과 초록도 별개의 파일로 작성하도록 사전 교육과 홍보를 한다. 파일이 확보되면 텍스트 형태의 전문 DB를 구축한다.

바. 텍스트 파일이 없거나 소급분에 대해서는 이미지 형태로 전문 DB를 구축한다. 이미지 포맷은 현재 범용적으로 사용되는 Multi TIFF나 PDF 형식으로 구축한다. 예산이 지원된다면 텍스트 형태의 전문DB가 구축된 것도 원문의 모습을 보여주기 위해서 이미지 형태의 전문DB를 구축한다.

사. 저작권에 대한 보호장치를 마련한다. 이를 위해서 논문 제출자가 본인의 논문을 도서관

에서 전문DB로 구축하고 이를 이용시키는 것에 대하여 허락하는 내용의 동의서를 받는다.

5. 전문DB 구축방향

워드프로세서 파일을 SGML 형식으로 변환하기 위하여, 파일을 일일이 태깅 작업하는 것은 시간과 경비 면에서 비효율적이다. SGML 형식으로 전문DB를 급히 구축해야 할 이유가 없다면 국립중앙도서관에서 전자도서관 사업의 하나로 추진하는 SGML 한글 변환기 개발을 기다리는 편이 유리하다.

세계적으로 PDF 방식으로 전문DB가 구축되고 있는 경향에 따라 PDF 방식의 전문DB를 구축할 수 있다. 모든 워드프로세서 파일을 PDF 형식으로 변경할 수 있는 한글 변환기가 개발되기를 기다려, 텍스트 파일과 이미지 파일을 같은 PDF 형식으로 전문DB를 구축하면 전문DB의 파일형식을 통일되어 하나의 뷰어로 서비스할 수 있다. 이러한 설계는 세계적인 조류에 따르게 되지만, 마이크로소프트사의 경우처럼 PDF 프로그램 제작사인 Adobe사의 독점에 의한 종속현상을 발생시킬 수 있으므로 이에 대한 우려가 된다.

워드프로세서 파일을 전부 XLX 방식으로 변환하여 전문DB를 구축하고 이를 볼 수 있는 뷰어를 제공하는 시스템으로 나아갈 수도 있다. 각종 워드 파일을 한꺼번에 자동으로 변환시킬 수 있으며, 파일의 크기가 작아 저장과 전송에 있어서 유리하다. 전용 뷰어가 TIFF 파일을 지원하기 때문에 텍스트 파일이 없는 소급 DB의 경우 TIFF 형식으로 전문DB를 구축한다. 텍

스트와 이미지 파일의 형식은 다르지만 하나의 뷰어로 이용할 수 있기 때문에 PDF 형식의 전문DB 시스템과 개념이 유사하다. PDF 형식과 달리 범용성이 낮아 다른 기관의 전문DB를 같은 뷰어로 보지 못하는 경우가 많겠지만 자관의 DB를 당장 구축하기는 유리하다.

워드프로세서 파일을 다른 형식의 파일로 변환하지 않고 그대로 전문DB로 구축할 수 있는데, 가장 저렴한 비용으로 DB를 구축할 수 있는 장점이 있다. 파일 형식에 따른 전용뷰어를 지원할 수도 있으며, 뷰어를 제공하지 않고 파일만 제공하는 서비스에 그칠 수도 있다. 그러나 이용 편의를 극대화하는 전자도서관의 기능을 고려한다면, 저작자가 제출한 파일을 그대로 이용자에게 제공하지 않고 쉽게 볼 수 있는 파일 형식으로 변경하는 편이 좋을 것이다.

이미지 데이터 전문의 내용 검색을 위하여 OCR 방식의 솔루션에 대하여 계속 검토한다. 마지막으로, 전문DB 시스템의 운영 방식을 서지DB 시스템과 독립적으로 운영할 것인지에 대한 결정은 서둘러야 하는 편이 좋다. 서지DB 시스템 자체의 안정성이 먼저 확보되어야 하며, 통합 시스템 프로그램과 검색엔진에 대한 검증도 해야 한다. 또한 전문DB의 양이 아직 많지 않고 아직 표준화된 솔루션이 결정되지 않았기 때문에 서지DB 시스템과 독립적으로 운영하다가 여러 가지 여건이 완비될 때 통합하는 편이 유리할 것이다.

6. 결 론

전자도서관을 둘러싼 기술의 발전속도가 신

속하기 때문에 현재 개발된 기술을 근거로 장기 계획을 수립하기가 어렵다. 국가나 전문단체에서 전문DB 구축을 위한 표준을 만든다고 해도 기술개발의 속도를 고려하면 이에 대한 수명을 보장할 수 없다. 신기술이 개발되었다고 이를 먼저 적용하여 다량의 전문DB를 구축한다면 시행착오를 범할 가능성도 높다. 전자도서관마다 다른 방법으로 전문DB를 구축하면 전자도서관 간의 DB 공동이용에 있어서 장애가 될 것이다. 그러나 이러한 환경 때문에 전자도서관의 기본인 전문DB를 구축하지 않고 표준이나 좋은 솔루션이 개발될 때까지 기다릴 수만은 없다. 따라서 개발되는 기술을 신속히 적용할 수 있도록 파일 형식을 쉽게 변경할 수 있으면서 안정된 시스템으로 설계하여 DB를 구축하고 이를 서비스해야 한다.

SGML과 PDF 한글 변환기가 개발되지 않은 시점에서, XLX 파일 형식의 안정성에 대한 점검이 완료되면 워드프로세서 파일을 XLX 방식으로 변환한다. 이미지 DB는 같은 뷰어에서 지원되도록 Multi TIFF 형태로 구축한다. SGML, PDF 한글 변환기가 개발되면 이를 이용하여 새로 설계할 수 있는 시스템과 기존의 서비스를 평가하여 시스템 변환 여부를 결정한다. 전문DB 시스템은 서지DB 시스템과 독립적으로 운영하다가 서지DB 시스템 자체가 전문DB 시스템과 통합될 정도로 안정성이 있다고 검증되면 그때 통합한다.

마지막으로, 전문DB를 제공하는 서비스에는 초록이 먼저 제공되어야 한다. 초록작성에 대하여 훈련된 사서가 일정한 초록 형식이 유지되도록 제출된 초록 파일의 내용을 검토, 수정한다. 기 구축된 전문DB 가운데 초록이 제공되지 않

는 데이터, 아직 DB로 구축되지 않은 논문에 대한 초록 파일을 작성하고 이를 시스템에 반입하여 초록 내의 단어 검색이 이루어지도록 한다. 위와 같은 방안은 현시점에서 학위논문 전문

DB를 구축할 수 있는 하나의 해결책이다. 이는 제한된 정보, 기술을 고려한 결과이므로 새로운 정보, 기술이 발견되거나 개발되면 이것들을 검토하여 새로운 해결책이 제시될 수 있다.

참 고 문 헌

- 권영하, 1997, PDF를 이용한 디지털 도서관 구축 방안, 디지털도서관, 봄호, pp. 86-95.
- 김성혁, 1997, 디지털도서관의 문헌 특성 및 관리, 한국문헌정보학회지 31:1 (1997. 3) : 53-70.
- 김용, 남궁황, 1998, 전자문헌 개발도구에 관한 고찰 : SGML, HTML과 PDF를 중심으로, 정보관리연구, 29:1(1998. 3) : 1-19.
- 김용훈 외, 1988, SGML 문서 관리 시스템의 설계 및 구현, 한국문헌정보학회지, 32:3 (1998. 9) : 157-176.
- 김휘출, 1998, 인터넷을 이용한 참고봉사, 한국 디지털도서관포럼.
- 박홍서, 1997, 전자문서의 표준 SGML, 디지털도서관, 가을호, pp. 72-82.
- 유사라, 1966, 정보화 사회와 도서관 정보네트워크, 나남출판.
- 이치주, 1998, 전자도서관: 자료의 디지털화 및 이용환경의 정비, 제2차 한일 국립도서관 업무교류 발표자료.
- 조왕근, 1997, 학위논문 SGML DTD 및 데이터베이스 구축에 관한 연구, 한양대학교 교육대학원 석사학위논문.
- 홍재현, 이두영, 1998, 디지털저작물에 대한 저작권 보호의 국제 동향: 데이터베이스를 중심으로, 중부대학교 인문사회과학논문집, pp.101-115.
- 홍재현, 1997, 네트워크 환경에서의 디지털 복제와 공정사용 법리 적용의 문제점, 한국문헌정보학회지, 31:4 (1997. 12) : 139-164.
- 홍재현, 1996, 디지털 시대의 전자복제와 관련된 저작권 문제, 도서관, 51:4, 341 (1996 겨울) : 56-80.
- 황찬현, 전자도서관과 저작권,
<http://www.nanet.go.kr/nal/3/3-1-4/nal96123.htm>
<http://www.lg.or.kr/newhome/digilib/lecture/view0004.htm>
<http://www.lg.or.kr/newhome/digilib/lecture/view0008.htm>
<http://www.excalib.com/products/rwfile/rwfile.html>
<http://pdf.cadinfo.co.kr/pdfindex.htm>