

초보적인 구축에서 빨리 벗어나야

최근 국가 정보화 프로젝트의 일부분으로 전자도서관시스템이 개발되고 있으며 각 대학과 기업은 나름대로 전자도서관 구축에 힘쓰고 있다. 이 글에서는 최근 전자도서관 구축에 이용되는 요소기술들을 살펴본다. 특히 다양한 전자도서관 기반 기술들 가운데 현재 도서관 분야에서 적용이 가능한 기술들을 중심으로 언급할 것이다. 그리고 현재 국내에서 개발되어 운영되거나 판매되고 있는 제품들을 소개하고자 한다. 다만 여기에서는 제한된 여건상 모든 제품들을 소개하지 못하고 대표적인 제품들만을 소개하기로 한다.

■ 안현수/한국통신 연구개발본부

전자도서관 의미와 동향

1990년대초 미국의 초고속정보통신망 프로젝트의 주요 응용 서비스들 중의 하나로 시작된 전자도서관(Electronic Library)은 디지털도서관(Digital Library), 가상도서관(Virtual Library), 벽 없는 도서관(Library without Wall) 등으로 불리우며 정보제공자 및 사용자들의 서비스 요구를 만족시키는 새로운 응용분야로서 전세계적으로 활발한 연구와 개발이 진행되고 있다.

그러나 이러한 전자도서관은 물리적인 의미로는 분산환경하에서 대량의 멀티미디어 정보에 대한 저장과 검색이 가능한 전자적으로 색인된 정보저장소를 의미하지만 사용자와 정보서비스에 따라 여러가지 다양한 모습을 갖는다. 도서관 분야에서는 전통적인 도서관 자동화시스템에 이미지 데이터가 포함된 원문관리와 Z39.50, ILL 프로토콜 기능 등이 추가된 확장된 도서관 자동화 시스템을 전자도서관으로 여기고 있다.

반면 전산학 분야에서는 원거리에 분산되어 있는 대량의 멀티미디어 정보를 다양한 목적을 가진 사용자들이 편리하게 사용할 수 있도록 해주는 종합적인 기술의 구현을 전자도서관으로 생각한다. 그리고 교육학 분야에서는 다양한 교육 매체와 교수활동을 지원하는 기술을 바탕으로 하는 가상교육시스템을 전자도서관으로 생각하는데 비해 방송 및 연예오락분야에서는 다양한 종류의 멀티미디어 정보를 손쉽게 제작하고 이들을 실시간으로 처리할 수 있는 기술을 전자도서관으로 여긴다.

이와 같이 전자도서관에 대한 견해는 각 분야마다 그 모습이 조금씩 다르기 때문에 특정 분야에서의 전자도서관에 대한 범위와 정의를 보다 명확히 해야만 혼돈을 피할 수 있을 것이다.

따라서 이 글에서는 전자도서관을 도서관 분야로 그 범위를 한정하고 현재의 도서관 분야와 관련하여 국내에서 추진되고 있는 전자도서관시스템 개발 및 운영을 크게 두가지 범주로 구분하여 접근하고자 한다.

첫째는 국가 정보화 프로젝트의 일부분으로 전자도서관시스템이 개발되어 운영되는 경우이다. 대표적인 사례로 국립중앙도서관을 중심으로 7개 기관이 참여해 개발, 운영되고 있는 국가전자도서관시스템이 있다.

둘째는 대학 및 전문도서관을 중심으로 각 기관 단위로 도서관 자동화 전문 개발업체 혹은 대형 시스템통합(SI)업체와 공동으로 도서관 자료의 디지털화 및 축적, 제공을 목표로 기존 도서관 자동화시스템을 확장하는 개념에서 전자도서관 시스템이 개발되어 운영되는 경우이다. 외국의 사례에서와 마찬가지로 국내에서도 위의 두 가지 모든 경우에 대부분의 시스템 개발 전문업체가 참여하여 개발을 주도한다.

전자도서관 구축을 위한 기술

전자도서관시스템을 구현하기 위해 필요한 요소기술에는 크게 자료생성, 저장 및 관리, 검색 및 탐색, 분산처리, 저작권관리 등

이 있다.

자료생성

전자도서관의 주요 기능 가운데 하나는 정보자료의 생성기능이다. 이를 위해 워드프로세서나 스캐너로부터 시작하여 HTML, SGML, XML 등과 같은 문서 저작도구나 멀티미디어 저작도구까지 매우 다양한 기능을 제공하는 이용자 인터페이스가 요구된다. 특히 전자도서관이 갖는 자료의 보존 기능을 고려할 때 고품질의 디지털 이미지를 인쇄매체로부터 고속으로 생성할 수 있는 기술은 매우 중요하다.

원문(full text) 제공을 특징으로 하는 전자도서관에서 문헌을 디지털화 하는 방법에는 이미지 방식, 텍스트 방식, 마크업 텍스트 방식 등이 있다. 1980년대 CD-롬으로 대표되는 광디스크의 출현으로 활성화되기 시작한 이미지 방식에서는 스캐너를 통해 자료를 tiff 등과 같은 이미지 형태로 변환시킨 후 하드디스크나 광디스크 등에 저장한다.

이 방식은 비교적 구축이 용이하고 비용이 적게드는 장점을 가지고 있으나 자료 저장공간을 많이 차지하고 원문 자체에 대한 검색을 할 수 없는 단점을 가지고 있다. 원문검색 문제를 해결하기 위해 최근에는 문자인식 소프트웨어(OCR)를 이용하여 이미지를 텍스트 형태로 변환시킨 후 이들 텍스트를 대상으로 원문검색을 하고 검색 결과물은 이미지 형태로 디스플레이하는 방법을 제공하기도 한다. 다만 대상 자료가 원문검색이 불필요하거나 혹은 순수하게 보존용일 경우 굳이 OCR을 이용하여 텍스트로 변환시키지 않고 전체를 이미지 방식으로 구축하게 된다.

문헌내에 태깅(tagging) 정보를 포함하고 있지 않는 텍스트 문헌은 일반적으로 아스키 코드로 표현되어 있다. 이러한 자료는 비교적 소규모의 저장공간을 필요로 하며 다양한 용도로 변환하여 이용할 수 있을 뿐만 아니라 원문검색도 가능하다. 그러나 자료내에 구조화된 정보를 가지고 있지 않기 때문에 효율적인 정보 검색이 불가능하며 태깅 정보를 이용한 다양한 출력도 어렵다.

마지막으로 마크업 텍스트방식의 문헌에서는 원문내에 해당 문헌의 구조화 정보를 나타내는 마크업을 포함하고 있다. 이러한 방식은 일반 텍스트 문헌에 비해 추가로 저장공간을 필요로 하며 특히 마크업 정보를 입력하는데 많은 시간과 노력을 들여야하는 단점을 가지고 있지만 일단 마크업 작업을 하게되면 구조화 정보를 활용한 강력한 검색이 가능하다. 특히 최근 들어 SGML과 XML 등과 같은 마크업 언어용 편집기가 개발되어 널리 보급되

면서 이러한 마크업 작업이 매우 쉬워졌다.

저장 및 관리

일반적으로 전자도서관의 저장 구조는 색인 서버, 콘텐츠 서버, 그리고 클라이언트의 3각 구조를 유지한다. 색인 서버는 목록 데이터를 저장하고 있으면서 해당 콘텐츠에 대한 위치정보를 유지하고 있다. 콘텐츠 서버는 다양한 포맷으로 된 정보 객체 자체를 저장하고 있으며 클라이언트는 이들 정보를 이용하는 주체가 된다.

단순한 파일시스템으로 전자도서관을 구축하는 경우도 있지만 전자도서관이 갖는 자료보존의 기능을 지원하고 대용량의 자료를 효율적으로 관리하기 위해 강력한 기능을 갖는 상용 DBMS를 사용하는 것이 일반적이다. 그러나 DBMS는 신속하고 다양한 기능을 요구하는 전자도서관 이용자의 검색에는 적합하지 않으므로 자료의 원본을 저장하는 단순 하부구조로서의 역할로 한정되고 이용자의 질의처리는 전문적인 정보검색시스템을 이용하는 경우가 일반적이다. 이밖에도 대량의 정보 데이터를 안정적으로 유지 관리할 수 있는 백업기능이 제공되어야 한다.

검색 및 탐색

최근 들어 전자도서관에서의 텍스트 정보검색에서는 SGML이나 XML 등과 같은 구조화된 문서를 대상으로 문서의 특정 부분을 검색하는 기능에 대한 요구가 높아져 이에 대한 연구개발이 활발히 이루어지고 있다.

멀티미디어 자료를 내용에 의해 검색하는 기법도 전자도서관의 구축에 있어서 매우 중요한 요소이다. 이미지나 오디오 혹은 비디오 자료를 검색하는 방법은 난이도가 높은 기술로서 아직은 검색 신뢰도 측면에서 많은 문제점을 가지고 있으나 전자도서관 구축에 실제 사용되고 있다. 대표적인 예로 IBM의 QBIC(Query By Image Content)가 있다.

전자도서관의 분산환경하에서 서로 다른 검색기법을 제공하는 정보시스템들에 대하여 동일한 이용자 인터페이스를 제공할 수 있도록 해주는 Z39.50 프로토콜의 사용이 증가하고 있다. 뿐만 아니라 Z39.50 프로토콜을 확장하여 가상종합목록(Virtual Union Catalog)을 구현하려는 노력도 활발히 시도되고 있다.

분산처리

전자도서관의 특징들 중의 하나인 분산 환경하에서 투명성을

제공하기 위해서는 분산 구조 및 각 서버의 특성을 이해할 수 있는 소프트웨어 에이전트가 사용자 질의를 적절히 분배하고 그 결과를 각 서버로부터 수집하여 이용자에게 배달하는 서비스를 수행할 수 있어야 한다. 또한 분산되어 있는 이질 컬렉션의 탐색을 용이하게 하기 위한 메타정보의 표준으로 미국 OCLC를 중심으로 만들어진 더블린 코어(Dublin Core)가 널리 활용되고 있다.

멀티미디어 정보의 분산처리를 원활히 하기 위해서는 충분한 통신기반이 제공되어야 한다. 현재의 인터넷은 거의 포화상태에 이르러 대량의 멀티미디어 정보를 효과적으로 배포할 수 없는 상황이다. 이에 따라 Internet 2와 같은 차세대 인터넷 프로젝트들이 추진되고 있으며 ISDN, ADSL, 케이블 모뎀 등과 같은 초고속 네트워크 솔루션들이 상품화되어 활용되고 있다.

저작권 관리

일단 정보가 디지털화 되면 복사와 배포가 아주 쉬워지기 때문에 정보의 불법적 이용이 실질적인 문제로 대두되고 있다. 즉, 저작권이 보호되지 않는 한 가치가 높은 자료를 확보하여 전자도서관을 통해 유통시키는 것은 네트워크로 연결된 환경에서는 거의 불가능하다.

이런 문제를 해결하기 위한 수단으로 암호화(encryption)와 특수무늬표시(watermarking) 등을 사용한다. 또한 다양한 디지털 저작물들의 체계적인 관리를 위해 전자저작권관리시스템(ECMS)이 개발되어 운영되고 있다.

국내 주요 제품 동향

현재 국내에서 개발되어 운영중인 전자도서관시스템은 크게 국가 차원의 프로젝트를 기반으로 하는 시스템과 각 도서관 단위로 개발되어 운영중인 시스템으로 구분할 수 있다.

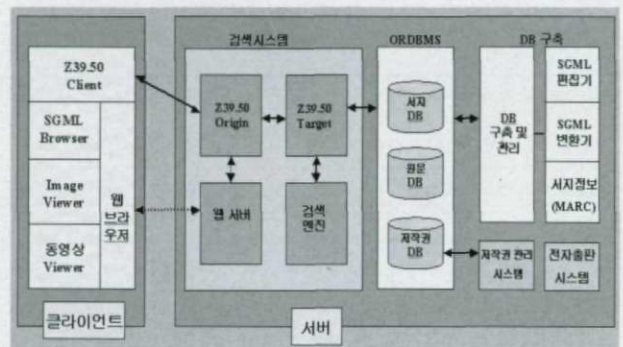
국가 프로젝트 기반 전자도서관(<http://www.dlibrary.go.kr/korean/info/content.html>)

국가 차원의 프로젝트를 기반으로 하는 전자도서관으로는 국립중앙도서관을 중심으로 7개 기관이 참여하여 수행한 국가전자도서관사업이 있다. 이 사업의 목표는 국내 주요 도서관을 연계하여 국가 정보능력을 향상시키고 지역간 균형있는 발전을 꾀하며 국가 정보자원의 공유체제를 확대 발전시켜 연구자들은 물론 일반 국민들에게까지 온라인으로 필요한 정보를 제공함으로써 21세기 지식정보사회에서 국가경쟁력을 강화하는데 있다. 이 프

젝트를 통해 구현된 주요 내용은 다음과 같다.

- ▲ Z39.50 프로토콜 기반 통합전자도서관시스템 체제 구축
- ▲ 7개 도서관으로의 사업 확대 및 DB 구축범위 확장
- ▲ 서비스중인 기존 전자도서관 시범시스템 기능 보완(DTD 추가 개발, 분산검색, 저작권관리시스템, 전자출판시스템 등 추가)
- ▲ 멀티미디어(동영상, 음성) 및 네트워크 기술을 응용하여 편리한 사용자 환경제공으로 서비스 질을 높임
- ▲ 객체관계 데이터베이스관리시스템(ORDBMS)을 적용한 SGML 문서 데이터베이스 구축

본 사업은 1997년 1월부터 1998년 11월까지 1차와 2차로 나누어 추진되었으며 참여기관은 국립중앙도서관, 국회도서관, 법원도서관, 산업기술정보원, 연구개발정보센터, 한국교육학술정보원, KAIST과학도서관 등이다.



〈그림 1〉 국가전자도서관 시스템 구성도

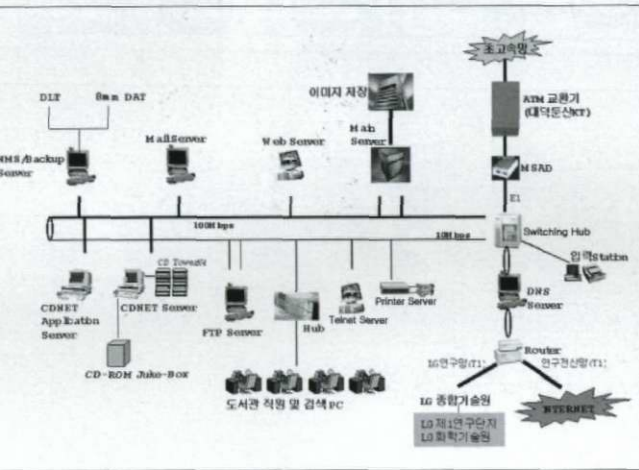
개별 도서관 기반 전자도서관

대학 및 전문도서관을 중심으로 각 기관 단위로 도서관자동화 전문 개발업체 혹은 대형 시스템통합(SI)업체와 공동으로 기존 도서관자동화시스템을 확장하는 개념에서 전자도서관 시스템이 개발되어 운영되는 경우로 현재 국내에서 이용가능한 시스템으로는 AIMS-II(코아정보), IBM DB2 Digital Library, LINNET 2000(포항공대), LG 상남도서관, 오름 Vintage LAS, SA99(SA&K), SOLARS(서울대) 등이 있다. 이 글에서는 대표적인 사례로 LG상남도서관, IBM DB2 Digital Library, 오름 Vintage LAS 등을 소개한다.

▲ LG 상남도서관(<http://www.lg.or.kr/digilib/lecture/case1.htm>).

1994년 8월에 시스템 개발을 시작하여 1996년 4월 국내 최초

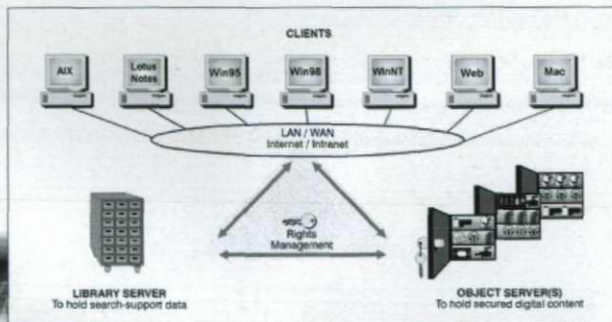
의 전자도서관으로 문을 연 LG 상남도서관은 국내에서 구하기 어려운 해외 과학기술 관련 정보를 집중적으로 수집해서 이를 디지털화하여 컴퓨터 네트워크를 통해 서비스하고 있다. 회원제로 운영되고 있으며 주 이용자는 이공계 대학 교수 및 대학원생, 그리고 연구소의 연구원 등이며 이들은 언제 어디서든 전자도서관 시스템에 접근하여 정보를 검색하고 필요한 자료의 원문까지 온라인으로 이용할 수 있다.



〈그림 2〉 LG 상남 전자도서관 시스템 구성도

▲IBM DB2 Digital Library(<http://www-4.ibm.com/software/is/dig-lib/brochure24/>)

IBM DB2 digital library는 현재 국내외적으로 널리 보급되어 운영되고 있는 대표적인 상용 전자도서관시스템으로 도서관 환경 뿐만 아니라 박물관, 영화, 방송, 오락 등 다양한 분야에서 활용되고 있다. 이 시스템에서는 비디오, 오디오, 이미지, 텍스트 등과 같은 멀티미디어 정보를 디지털 파일로 변환하여 인터넷이나 인트라넷을 통해 제공한다.

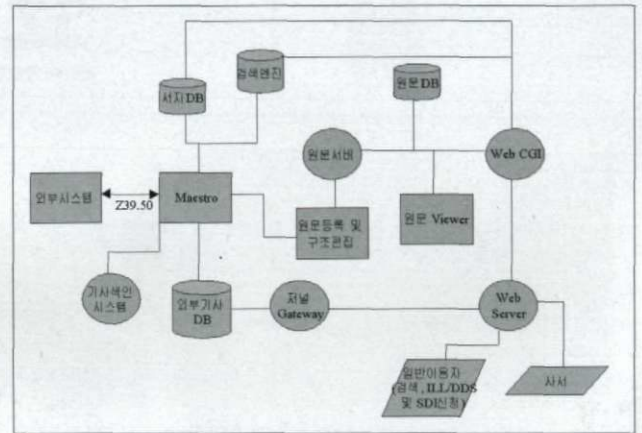


〈그림 3〉 IBM DB2 Digital Library 시스템 구성도

▲오름 Vintage LAS

오름의 전자도서관 개발목표는 인쇄매체와 전자매체, 원문데

이터와 메타데이터, 기존 서비스와 신규 서비스들을 포괄적으로 수용할 수 있는 시스템을 개발하는데 있다. 이렇게 함으로써 이용자에게 One-Stop 서비스를 제공할 수 있고 업무의 단절이나 기존 시스템과 전자도서관시스템을 별도로 운영하는데 오는 부담을 줄이면서 기술의 발전에 따라 점진적으로 진화할 수 있는 능력을 확보할 수 있게 된다.



〈그림 4〉 오름 디지털 라이브러리 시스템 구성도

맺음말

국내 전자도서관은 아직 개발 초기 단계라고 할 수 있다. 국내 대학들에서 서비스하고 있는 전자도서관은 일부를 제외하고 대부분이 아직 서지정보의 검색서비스만을 제공하는 수준이다. 이는 검색한 자료의 원문 자체를 제공하는 진정한 의미의 전자도서관과는 거리가 있다.

특히 해외 유수의 전자도서관시스템들에서 기본적으로 구현되어 활용되고 있는 Z39.50 프로토콜이나 ILL 프로토콜 등의 경우에도 국내에서는 제대로 구현이 안되어 실제 활용하는데 어려움이 있다.

현재 국내의 전자도서관 개발 및 운용사례를 살펴보면 기존 도서관의 기능을 컴퓨터화시키는 시도가 주를 이루고 있다. 반면 해외의 경우에는 미국 및 유럽을 중심으로 전자도서관 분야에 국가 차원의 체계적인 연구개발이 지속적으로 진행될 뿐만 아니라 기존 도서관 자동화 업체들은 이러한 연구결과들을 활용하여 구체적이면서도 실용적인 전자도서관 시스템들을 개발하여 운영하고 있다. 따라서 국내에서도 이러한 연구결과물들을 충분히 활용하여 보다 발전된 형태의 전자도서관시스템들을 개발하여 활용해야 할 것이다.