

디지털大學圖書館 設計을 爲한 理論的 考察

- 메타데이터 DB 構築과 原文檢索을 中心으로 -

A study on digital library for design in university library

장 동 춘 (Dong-Choon Chang)*

〈 목 차 〉

- | | |
|------------------------|--------------------------------|
| I. 서론 | IV. 디지털라이브러리구축을 위한 제안 |
| 1. 연구의 목적 | 1. 데이터의 디지털화 경향 |
| 2. 연구방법 및 필요성 | 2. 디지털 데이터베이스 축적 대상자료 |
| II. 선진각국의 디지털도서관의 개발동향 | V. 전자도서관 구축을 위한 환경구성 |
| 1. 일본의 디지털도서관 현황 | 1. 메타데이터 |
| 2. 미국의 디지털도서관 현황 | 2. ISO/IEC-11179 데이터요소명세 및 표준화 |
| 3. 영국의 디지털도서관 현황 | VI. 디지털라이브러리 검색과 탐색 |
| 4. 프랑스의 디지털도서관 현황 | VII. 저작권법에 의한 보호한계 |
| III. 한국의 디지털도서관 현황 | VIII. 결론 및 제안 |

초 록

종래의 전통적 대학도서관은 인터넷 기술의 발달로 인하여 새로운 형태의 디지털 라이브러리로 변모하고 있다. 인쇄자료는 국제교환표준문서 SGML 디지털 파일로, 목록은 메타데이터로, 이용자는 네트워크가 접속되는 전세계 이용자로, 자료는 무한정 복사가 가능한 형태로 바뀌고 있다. 국가 정보 하부구조를 위한 대학도서관 디지털 라이브러리 구축은 미래 정보화사회에 매우 중요한 것으로 본 논문에서는 기존의 디지털 라이브러리 기술에 대한 분석을 통하여 기하급수적으로 발생하는 전자문서를 관리하는 문제, ISO에서 문서 교환표준으로 제정한 SGML 문서의 구조 등 앞으로 추진할 디지털 라이브러리의 모습을 모델링한다.

Abstract

With the advancement of internet technology, traditional university libraries are going through a new metamorphosis into digital libraries. Digital libraries substitute digital files for printed materials, metadata for catalogues, world wide users on the network for localized patrons, while offering limitless possibilities for easy downloading of information. The aim of the study is to present a model for our digital library based on the knowledge acquired from the analysis of the state of the art technology employed in building and managing digital libraries and the management of large quantity of SGML electronic documents. This paper is our belief that the national information infrastructure for university digital libraries are mandatory for an open information society.

* 세종대학교중앙도서관 사서부장

I. 서 론

1. 연구의 목적

다가오는 21세기는 고도화된 정보화 사회가 될 것이며 이러한 정보화사회에서 필수적인 요소로는 기하급수적으로 발생하는 전자 문서를 관리하는 것이라 할 수 있다. 특히 디지털도서관 (Digital library), 인트라넷상에서의 CSCW(Computer Supported Cooperative Work), CALS (Commerce At the Light Speed) 등의 최근의 응용에서는 대량의 전자문서를 효율적으로 저장하고 관리하는 시스템이 요구되고 있다. 다양한 전자문서의 형태중 ISO에서 문서교환 표준으로 제정한 SGML은 구조정보로 사용할 수 있는 마크업(Markup)을 포함하고 있어 문서를 구조단위로 표현할 수 있다. 이는 대량의 전자문서를 저장하거나 관리하는 데 적합하다. 그러나 이용자를 위한 공중 정보통신 서비스로서 인터넷 서비스가 제공되면서 물리적인 도서관의 위치나 자료 이용의 시간적, 공간적 제한은 문제가 될 수 없으며 PC보급의 확산과 통신망의 발달로 이용자의 요구는 도서관의 서비스를 훨씬 앞질러가고 있는 등의 이용자 참고봉사의 패러다임(Paradigm)이 완전히 변화되는 상황이 되었다.

이와 같은 도서관을 흔히 전자도서관 (electronic library), 디지털 라이브러리 (digital library) 등의 용어로 부르고 있다. 전자도서관은 물리적 위치, 시간적, 공간적 제한 없는 도서관이다. 그리고 이들 도서관의 공통적인 의미는 자료의 형태에 구애됨이 없이 모든 형태의 원문데이터, 이미지데이터, 오디오, 비디오, 디지털데이터, 기계가독형 정보자료 등 멀티미디어가 데이터베이스로 구축되어, 이용자들이 데이터를 공유 공유하면서 장소에 제약을 받지 않고 컴퓨터를 통한 원격지 정보검색이 가능하여 복수의 이용자가 동시에 사용이 가능하다는 특징점이 있다.

디지털정보의 원활한 유통을 목적으로 한 전자도서관을 구축하기 위해 다수의 국내 여러 기관에서도 이에 대한 연구가 활발히 추진되고 있다. 이미 미국, 일본 등 선진국들에서는 전자도서관을 구축하기 위한 미래의 청사진을 면밀하게 마련한바 있으며, 계획에 따른 전자도서관 구축을 실행에 옮기고 있는 중이다.

특히 대학도서관은 정보를 둘러싼 내, 외부환경의 변화에 민감하게 반응해야 하는 기관이므로 더 이상 과거의 전통적인 형태의 도서관 서비스만으로는 이용자를 만족시킬 수 없게 되었기 때문에 새로운 정보기술을 받아들여 새로운 변화에 능동적으로 적응해야 한다.

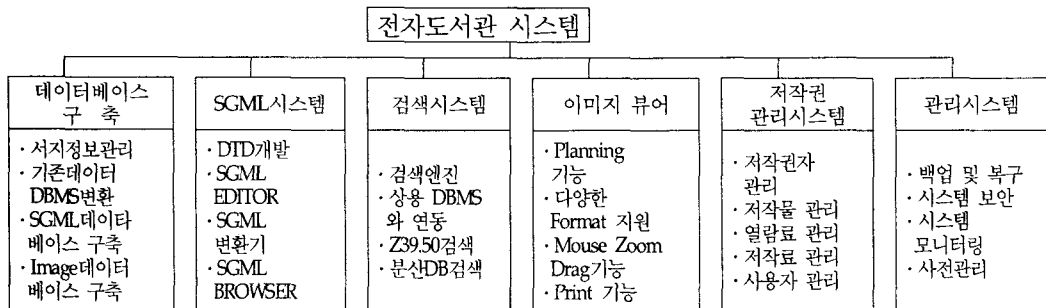
본고에서는 디지털도서관의 개념을 정의하고 그 특성을 규명하여 앞으로 디지털라이브러리의 계획과 발전을 도모함으로써 이용자에게 참고정보서비스의 질적인 개선을 하자는 데 연구의 목적이 있다.

2. 연구방법 및 필요성

오늘날 대학도서관들은 정보환경의 변화에 따른 충격과 도전에 많은 시련을 감당해야하는 입장에 있다. 다가오는 21세기는 정보화사회의 전환기를 지나 지식과 기술 집약의 지식기반 정보화사회가 될 것이며 세계의 모든 국가들은 정보화의 추진성과에 따라 21세기의 국제적 위상이 결정된다는 데 인식을 같이하고 있다. 또한 현재는 정보의 증가량이 2.5년만에 2배가 증가하지만 2010년에 이르면 70일마다 2배가 증가하리라고 예상하고 있다. 컴퓨터의 처리속도와 집적도가 18개월마다 2배씩 증가하고 있다는 '골든 무어의 법칙'은 현재도 엄청난 속도로 진행되고 있다. 정보자료도 종이매체에서 전자매체로, 정보시스템은 폐쇄형 단독시스템으로부터 개방형 웹 기반 기술로, 정보서비스는 just in case로부터 just in time으로 변화하고 있으며 정보기관의 조직도 다운사이징으로 차츰 옮겨가고 있다. 인터넷을 비롯한 통신망의 발달은 도서관의 업무에도 큰 영향을 미치고 있다. 따라서 점차적으로 증가하고 있는 전자정보자료의 구입과 이용이 도서관의 모든 기능에 영향을 주게 될 것이고 전통적인 도서관에서 전자도서관으로의 전환을 위한 개념적인 구상이 필요하게 되었다.

본 연구에서는 국내외의 전자도서관 실태와 전자도서관 설계와 구현 방안을 다룬 선행 연구문헌을 수집하여 집중적으로 연구 검토하였다. 특히 미국, 일본 등 선진국의 디지털라이브리 개발사례가 분석연구 되었으며, 국내 전자도서관 추진 현황과 실태도 아울러 비교분석 검토하였다. 그리고 이러한 변화된 상황하에서 전자도서관 개발에 필 수 요소가 되는 전자문서 기술표준인 SGML문서 구현에 의한 콘텐츠(Contents) 구축을 비롯하여, Z39.50 표준프로토콜의 정보검색 기능에 대한 원리를 다루었다. 아울러 저작권 문제와 메타데이터에 표준으로 간주되고 있는 더블린 코어의 15개 요소에 대하여서도 언급하였다. 이렇게 함으로써 오늘날 대학도서관들이 전자도서관을 계획하고 발전시키는 문제에 있어서 보다 전문성을 향상시키고 효율적으로 추진하는 데 기여하도록 기초적인 자료를 제공하려는 데 본 연구의 필요성을 두고자 하였다. 전자도서관의 주요기능은 아래 <표 1>과 같다.

<표 1> 시스템 주요기능



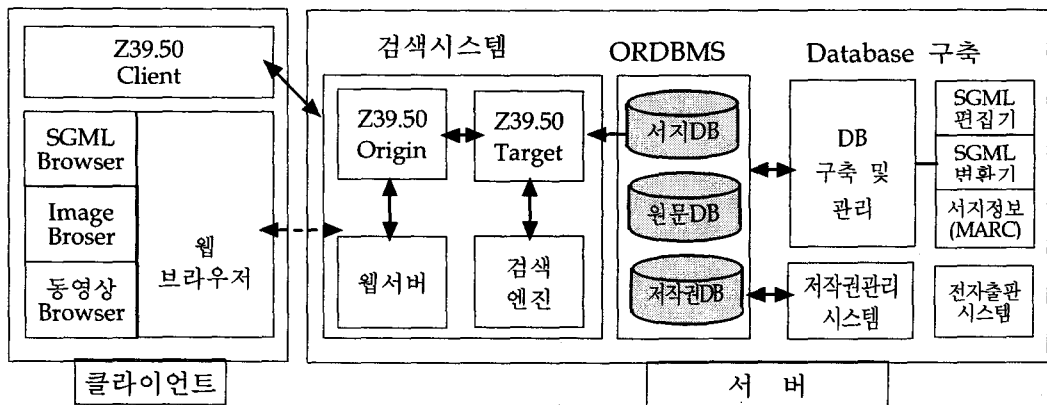
3. 디지털도서관의 특성

미래의 도서관을 지칭하는 용어로 electronic library, virtual library, paperless library, digital library, networked library, borderless library 등으로 사용하고 있다. virtual은 시간과 공간을, digital은 디지털 데이터를, network은 통신을, multimedia는 매체를 강조하고 있으나 이들은 모두 동의어로 취급할 수 있을 것이다. 최근에는 디지털 라이브러리라는 용어가 주로 사용되고 있다.

전통적인 도서관과 디지털 라이브러리와 물리적 또는 기능상의 상이한 점을 들어보면 시설중심에서 서비스중심으로, 인쇄매체중심에서 전자매체중심으로, 정보의 소재파악에서 지식의 직접전달로, 자료의 소유개념에서 서비스와 공유개념으로, 정보생산과 소비의 분산화에서 정보생산과 소비의 일원화로, 도서관의 공유화에서 도서관의 개인화 등으로 비교가 될 수 있다.

그리고 디지털 라이브러리의 특성을 좀더 분석하여 열거해 보면 우선 정보를 기존문서 종이 인쇄형태의 매체에서 SGML DTD와 같은 표준문서 형식에 의하여 구조화된 전자문서인 디지털형식으로 축적한다는 점, 정보의 입수시 디지털 통신기술에 의존한다는 점, 그리고 자료의 원본은 하나이며 하나의 원본을 이용한다는 것으로 논하고 있다. 그러나 디지털도서관의 가장 근본적인 특성은 전문데이터를 처리의 대상으로 한다는 데 있다. 전문DB는 이미지 형태나 혹은 텍스트 형태로 구축할 수 있다. 텍스트 파일을 입수할 수 있는 경우는 전문DB를 텍스트 형태로 혹은 텍스트 형태와 이미지 형태를 동시에 구축하고, 텍스트파일을 입수할 수 없는 경우 이미지 형태로 구축하는 것이 일반적이다. 또한 현 도서관이 컴퓨터의 도움을 받는 도서관이라고 한다면 디지털도서관은 컴퓨터가 중심이 되는 도서관이라는 점을 강하게 내포하고 있다. 전자도서관 시스템 구성도는 <그림 1>과 같다.

<그림 1> 전자도서관 시스템 구성도



II. 선진각국의 디지털도서관의 개발 동향

1. 일본의 디지털도서관 현황

1944년 8월 일본 정부에서는 총리를 책임자로 하는 “고도정보통신사회추진실”이라는 기구가 설립되었다. 이 기구는 1995년 2월 “고도정보통신사회 추진 가이드라인”이라는 보고서를 통해 멀티미디어 전자도서관을 위한 연구 추진과 과학정보 유통에 주요 역할을 맡고 있는 NACSIS, 대학도서관 및 기타 도서관의 지속적인 보강 및 개선을 권장하였다. 1995년8월에 발표된 교육, 과학, 체육분야 정보화 계획의 주요 내용은 다음과 같다.

- NACSIS 및 기타 기관의 멀티미디어 전자도서관에 대한 연구개발 지원

- 네트워크를 통한 디지털 정보 서비스의 향상을 위해 필요한 컴퓨터 및 기타 설비의 공급과 전자도서관 구축에 대한 지원

1996년7월에 발표된 “과학기술 기본계획”에 의하면 전자도서관 시스템의 연구와 개발 강화와 기존 대학도서관들의 전자도서관화에 대해서 언급하고 있다. 1996년7월에 발행된 대학도서관의 전자도서관화에 관한 개선 및 강화 방안이라는 보고서에는 도서관 목록과 자료, 특히 저작권과 관련이 없는 고서 및 희귀본의 디지털화와 컴퓨터 및 네트워크의 개선 등이 포함되어 있다.

1) NACSIS의 ELS

일본학술정보센터(NACSIS) 전자도서관 시스템(NACAIS-ELS)은 기존의 서지데이터베이스 검색 서비스(NACSIS-IR)와 디지털화된 학술잡지 및 회의자료를 통합한 정보서비스이다. 이 정보 서비스는 인터넷을 통하여 클라이언트/서버 아키텍처를 이용한 분산처리 환경하에서 운영되고 있으며, 이미지 데이터 처리에는 Z39.50 프로토콜의 확장 버전이 채택되고 있다.

NACSIS-ELS의 개발은 일본 정보처리학회가 모든 간행물을 디지털화하기 시작한 1993년에 시작되었다. 1997년 3월까지 디지털 작업에 참여한 기관은 29개로 늘어났으며, 총 62종의 잡지 800,000 페이지에 달했다. 이 전자도서관 시범서비스는 1995년 2월부터 무료로 시작되었으며, 1998년부터는 정식서비스를 목표로 저작권 협상을 계획하고 있다.

2) 일본정보처리개발센터의 차세대 전자도서관시스템 연구 개발 프로젝트

이 프로젝트는 일본 통산성의 예산지원을 받아 IPA와 JIPDEC이 추진하고 있으며 1996년 4월부터 2000년 3월에 끝나는 5개년 과제이다. 이 프로젝트의 목표는 “정보유통센터”로서의

21세기 전자도서관의 정보유통을 위한 인프라 구축이다.

2. 미국의 디지털도서관 현황

최근 미국도서관계에 가장 큰 관심사항은 디지털도서관의 구축이라고 해도 과언이 아닐 것이다. 이러한 추세에 대하여 도서관 현장의 사서들은 다가 올 변화에 대한 불안과 우려를 감추지 못하면서도 대단한 관심을 가지고 관련분야의 학회나 세미나가 개최될 때마다 회의장을 가득 메우며 그 열기를 더해가고 있다. 가장 대표적인 디지털도서관사업으로는 미국의회도서관의 National Digital Library Program이 있다. 이것은 의회도서관이 소장하고 있는 자료 5백만 건을 2000년까지 디지털화 하여 21세기에는 이 자료들을 개인용 컴퓨터에서 볼 수 있도록 제공한다는 것이다. 한편 디지털도서관에 대한 첨단기술은 규모가 큰 대학도서관에서 활발히 개발되고 있으며 1995년에는 국가디지털도서관연맹(National Digital Library Federation)이 설립되었다. 이 연맹의 목표는 도서관자료를 인터넷을 통하여 쉽게 액세스할 수 있도록 디지털화 하는 것이다. 그 동안 쌓아온 이론적인 연구를 배경으로 하여 앞으로는 보다 현실적인 측면에서 기관간의 협력을 통하여 디지털도서관을 구현해나가는 것이 목적이다. 연방정부의 지원 하에 수행되고 있는 대규모 디지털도서관 선도사업은 NSF/ DARPA/NASA Initiatives 이다. 이것은 미국과학재단과 국방성첨단연구사업청과 항공우주국이 공동으로 투자하고 있는 6개 대학도서관의 디지털도서관사업이다.

이것은 이미 널리 알려진 바와 같이, 멀티미디어자료는 카네기멜론대학, 컴퓨터 관련분야는 스탠포드대학, 지구우주과학분야는 미시간대학, 항공사진과 지진데이터는 캘리포니아주립대학의 산타바바라캠퍼스, 과학기술잡지는 일리노이주립대학의 어바나샴페인캠퍼스, 환경분야자료는 캘리포니아주립대학의 버클리캠퍼스 등에서 각각 연구중이다. 이 중에서 일리노이주립대학의 어바나샴페인 캠퍼스에서 수행중인 디지털도서관선도사업인 "DeLiver"는 인터넷웹상에서 과학기술정보를 효율적으로 액세스할 수 있도록 하기 위한 것이다. 현재는 과학기술 문헌정보에 대한 대규모 테스트베드를 구축하고 그 효율성을 평가하여 검색기술 향상을 위한 연구를 하고 있다. 과학기술분야의 14개 학회 및 출판사로부터 61종의 학술잡지에 수록된 논문기사를 SGML로 된 파일로 받아 축적해두고 한 개의 화면상에서 복합적으로 찾아보면서 검색할 수 있도록 통합하고 매핑하고 있다. 현재까지는 1995년도 이후 분에 수록된 기사 총 10,000건을 축적하고, 대학 캠퍼스에서만 테스트중이다. 한편 이 시스템의 인터넷서비스를 위한 소프트웨어는 대학 내에 있는 국가슈퍼컴퓨팅응용센터(National Center for Supercomputing Application : NCSA)에서 개발하고 있다. 이밖에도 미시건대학과 코넬대학이 공동으로 1800년 이후 미국역사에 대한 디지털도서관인 "Making of America"사업을 추진중이고, 콜롬비아대학

도 "Projet Janus"를 수행하고 있으며, 여타 큰 대학들의 연구개발사업들이 다수 있다.

1) "The University of Michigan Digital Libraries Research Project"

이 프로젝트에서는 University of Michigan 주관하에 대규모의 그리고 지금도 계속발전하고 있는 멀티미디어 디지털 라이브러리 텍스트 베드의 조성, 운영, 사용, 평가를 위한 종합 연구/개발을 수행한다. 이 전자도서관의 주된 대상분야는 지구과학 및 우주과학이다. 캠퍼스 사람들, 지역의 중고등학생 그리고 공공도서관 이용자 등을 포함하는 다양한 계층의 이용자들이 이 프로토타입 시스템을 테스트하고 평가한다는 점이 특기할 만한 일이다.

2) "Alexandria Project : Towards a Distributed Digital Library with Comprehensive Services for Images and Specially Referenced Information" (\$4백만 : 1994-1998)

University of California, Santa Babara에서 수행하는 Alexandria 프로젝트는 새로운 전자도서관 서비스는 물론 많은 다양한 종류의 지도, 화상자료, 회화자료 등에 편리하게 액세스할 수 있는 전자도서관을 개발하게 된다. 이 프로젝트는 지리 정보시스템을 위해 개발한 소프트웨어를 사용해서 Santa Barbara, Ventura, Los Angeles 카운티에 관한 디지털화 된 지도, 화상자료 그리고 항공사진들의 수집을 시작으로 4년에 걸쳐 여타도서관은 물론 University of California, Santa Barbara, State University of New York - Buffalo, Library of Congress, United States Geological Survey, 그리고 St. Louis Public Library 에 있는 다른 자료들도 포함시킬 계획이다.

3) 미국의회도서관의 NDLP (National Digital Library Program)

미국의 경우 의회도서관이 추진하고 있는 NDLP(National Digital Library Program) 프로그램은 앞으로 수년간 미국의 고유한 역사 자료를 디지털화 하여 학교교사, 학생, 그리고 일반 국민들에게 인터넷을 통해 서비스하는 American Memory 프로젝트이다. 아울러 이 프로그램은 다른 도서관간이나 문서관들의 참여를 촉구하기 위해 과제 공모 형식으로 American memory 프로젝트를 추진하고 있다. 이 과제는 국회도서관 예산과 개인 또는 사설 단체로부터의 기부금으로 추진되고 있다.

4) UMI사의 Digital Vault Initiative(DVI)

UMI사의 Digital Vault Initiative(DVI) 프로젝트는 지구상에서 가장 규모가 큰 디지털화 사업으로서 그들이 가지고 있는 물경 500년 분량의 마이크로폼 자료를 스캐닝을 통해 디지털

문서로 변화하는 사업이다. 수 십만권의 책, 신문, 잡지, 기타 자료 등 총 55억 페이지에 해당하는 자료의 스캐닝 작업은 지난 5월에 시작되어 앞으로 수년 동안 계속될 것이다. 이 사업을 위해 5대의 첨단 디지털 스캐너를 사용하여 24시간 3교대로 작업하고 있다. DVI의 1단계 사업으로 최초로 인쇄술이 발명된 1475년부터 1700년도까지 간행된 영문학 도서 9,600종 (22,000,000페이지)을 대상으로 하고 있다.

3. 영국

영국 국립도서관의 Initiatives of Access 프로젝트는 도서관 자료의 디지털화, 네트워크 기술 연구, 도서관 서비스 및 액세스 향상을 위한 연구, 정보의 축척과 색인, 정보검색, 데이터 전송에 관한 표준, 저작권 연구 등을 주된 목표로 하고 있다. 이 프로젝트는 130만 건의 특허 자료를 16개의 CD-ROM 주크 박스 시스템을 통해 서비스하는 Parent Express Jukebox 등 20여 개의 세부 프로젝트로 나뉘어 추진되고 있다.

4. 프랑스

프랑스 국립도서관이 수행하고 있는 전자도서관 프로젝트인 Tolbiac 프로젝트는 여러 기관에 분산 소장되어 있는 도서, 화상 및 음성자료 등 다양한 형태의 자료를 디지털화하고 이를 네트워크를 통해 서비스하는 것을 목표로 하고 있다. 1990년부터 시작된 디지털 장서 구축사업은 1998년까지 10만 권의 도서와 30만 건의 이미지를 목표로 하고 있다.

Tolbiac 프로젝트의 디지털 대상자료는 과학, 프랑스문학, 언어학, 역사분야의 참고자료와 정기간행물이며, 디지털화 형태는 비용과 기술을 고려하여 이미지 형식을 채택했다.

Ⅲ. 한국의 디지털도서관 현황

국내 전자도서관 구축 실태를 살펴 볼 때 자체적으로 디지털 장서를 구축하고 서비스하는 시스템을 구축한 기관은 많지 않으며, 대다수가 소장자료 목록이나 초록정보 서비스와 인터넷 상의 정보원내지 정보 시스템을 연결해 주는 링크 서비스 정도이다. 하지만 1990년대 중반에

들어서 정부나 민간단체의 관심이 집중되면서 몇몇 기관에서 전자도서관 사업이 본격적으로 진행되기 시작했고 그 결과물도 나타나기 시작했다.

전자도서관을 구축한 국내 대학, 연구소, 기업체 등 도서관 또는 자료실을 가지고 있는 기관들중 대다수가 도서관업무 전산화시스템과는 별개로 전자도서관을 구축 운영하고 있다. 따라서 도서관홈페이지에 전자도서관 앵커를 만들어 전자도서관 시스템에 연결해 주고 있다. 또 정보센터의 경우 자체 소장자료가 없기 때문에 정부 또는 공공부문에서 생산한 일부 문헌을 디지털화하여 서비스하고 있는 실정이다. 도서관이나 정보센터에서 운영하고 있는 전자도서관은 들어가 보면 한결같이 디지털 콘텐츠가 빈약하거나 아예 전무한 곳도 있다. 국내 전자도서관 구축은 외국의 경우와는 달리 지난 몇 년 동안 도서관 업무 전산화, 네트워크화, 자료 디지털화가 거의 동시에 추진되고 있기 때문에 디지털 콘텐츠 제작에 대한 준비와 투자가 외국에 비해 뒤떨어지고 있다.

국내의 주요 전자도서관 추진기관으로는 민간부문에서 국내 최초의 전자도서관이라할 수 있는 LG 상남도서관, 대학에서 선구자적인 홍익대학교 전자도서관, 국내 최초로 공공기관 큰소시음 형태로 추진된 국가전자도서관 시범 연계사업, 그리고 최대 투자규모를 자랑하는 한국과학기술원 주관의 첨단과학기술 전자도서관 구축 프로젝트 등이 있다. 이외에도 전자통신연구원, 이화여자대학교, 아주대학교 등이 전자도서관을 구축 운영 또는 추진중이다.

1. LG 상남도서관

1995년 12월에 국내에서는 최초로 전자도서관 정보시스템(ELIT : Electronic Library Information Tour)을 구축하여 회원제를 바탕으로 전자적 서비스를 시작했다. ELIT는 그후 1997년 웹 기반의 인트라넷 시스템으로 바뀌었으며, 학술잡지, 회의자료에 수록된 논문 원문을 이미지 형태(Tiff)로 디지털화 하여 서비스하고 있다. 전자도서관시스템은 자료입력 및 회원관리 부분은 클라이언트 서버구조로 검색부분은 웹 기반에서 이루어진다. ENTERPRISE 3000서버 위에 관계형 DBMS인 SYBASE를 탑재하였고 검색엔진은 BRS를 쓰고 있다. 상호가동성 측면에서 표준검색 프로토콜인 Z39.50을 지원하는 시스템으로 타 Z39.50 site의 목록검색이 가능하며 이를 통한 본문보기까지 구현되어 있다.

2. 홍익대학교 전자도서관

국내대학으로는 최초로 전자도서관을 구축한 대학으로서 중앙도서관에 멀티미디어센터를

설립하여 인터넷을 통해 텍스트는 물론 멀티미디어 자료를 자연어 검색을 통해 제공받을 수 있도록 했다. 이 대학 전자도서관은 소장자료검색, 기사검색, 논문원문검색, VOD검색, 학술 CD검색, 온라인DB, Z39.50, 부가정보 등의 메뉴로 구성되어 있다. 이 시스템은 표준검색 프로토콜인 Z39.50을 지원하는 시스템으로 이 프로토콜을 사용해 구축한 전자도서관 인터넷을 통한 자료와 정보를 공유할 수 있다. 홍익대학교전자도서관은 기존의 도서관시스템과는 독립적으로 운영되고 있으며 시스템 측면에서 국내 전자도서관중에서 짜임새 있고 다양한 정보서비스 제공기능을 가지고 있다.

3. 국립중앙도서관

국립중앙도서관은 1995년부터 디지털 콘텐츠를 대량으로 구축해 왔다. 국가 전자도서관 시범/연계사업 예산은 물론 자체예산, 국내학술자료화상정보 서비스사업비, 그리고 최근의 정보화 근로사업 예산 등 막대한 예산을 투입하여 방대한 양의 자료를 이미지 형태로 전자화하였다. 내년 2월말까지 구축 될 디지털 대상자료는 인문과학분야 박사학위논문, 구한말 및 조선총독부 관보, 조선총독부 발간 한국학 자료, 한국 고전 백선, 문화관광부 발간 보고서, 1950년 이전 발간 신문 및 잡지 귀중본, 한국고서, 한글판 고대소설과 딱지본 등으로 총 43,576종, 45,341책, 8,560,166쪽에 달한다. 국립중앙도서관의 전자도서관 시스템은 기존의 로컬 도서관리 시스템인 CENTLAS와는 별개로 운영되고 있다.

4. 국회도서관

국회도서관의 전자도서관 구축사업은 국가 전자도서관 시범/연계사업에 그 뿌리를 두고 있다. 타 6개 기관과는 달리 시범/연계 사업 결과 개발된 전자도서관 시스템을 가장 잘 이용하고 있다. 기존의 도서관관리 시스템과는 별개로 운영하고 있지만 대 이용자 검색 서비스는 기존의 OPAC 대신 새로 구축한 전자도서관 시스템에 의해서 제공되고 있다.

기존의 도서관리시스템을 통해서 수서, 대출/반납, 목록 등의 업무를 처리하고 생성된 목록 정보는 전자도서관 시스템에 배치방식으로 반입할 수 있도록 하였다. 시범/연계사업을 통해 구축한 디지털 장서로 국정감사자료(TIFF), 국회사(SGML, TIFF), 정부간행물(TIFF), 입법부 발간자료(SGML) 등 SGML 문서가 5종, 26,682쪽, 이미지 문서가 4종 737,000쪽으로 총 763682쪽의 디지털 문서가 이미 구축되어 있고, 정보화 근로사업이 끝나는 2월말에 총 17,292,000쪽의 이미지 형태의 디지털 콘텐츠가 구축되게 되면 약 18,000,000쪽의 원문 데이터베이스가 구축

되어 국내 최대규모의 디지털 콘텐츠를 확보하게 된다.

5. 첨단학술정보센터(KRIC)의 RISS

첨단학술정보센터(KRIC)는 RISS(Research Information Service System)라는 학술정보서비스 시스템을 구축 운영하고 있으며, 이 시스템은 대학도서관 소장도서종합목록, 해외학술정보, 학술지논문, 학술지원정보, 인터넷정보분류 등의 하부 시스템 (Infrastructure)으로 구성되어 있다. 서비스되고 있는 디지털 콘텐츠로는 학술진흥재단 추진사업을 이어 받은 이미지 형식의 외국박사학위논문 본문이 있다. 이 디지털 논문은 서지, 초록 검색을 통해 본문의 이미지를 열람할 수 있도록 되어 있다. RISS의 구조는 입력 부분은 클라이언트/서버 구조로, 검색부분은 웹으로 되어 있다. 주 서버로는 HP EPS를 쓰고 있고, Oracle DBMS와 BRS 엔진이 탑재되어 있다. 이 시스템은 Z39.50을 지원하고 있다. 또한 도서관상호대차시스템과 원문제공서비스 시스템을 개발중이며, 기타 부가적인 서비스로 선택적정보배달서비스(SDI)도 계획하고 있다.

6. 전자통신연구원 전자도서관

전자통신연구원의 전자도서관은 출연연구소중에서는 가장 먼저 전자도서관을 구축했다. 1994년에 시스템 개발을 시작하여 1995년부터 디지털화 된 본문을 웹 기반에서 서비스했다. 또한 주간기술동향을 비롯하여 정보통신 표준, 정보통신 특허, 연구동향분석 등 총 23종의 디지털 컬렉션을 구축하였는데 국내에서는 가장 많은 종류의 디지털 콘텐츠라 할 수 있다. 이러한 전자정보는 모두 일관성있게 PDF 형태로 인터넷을 통해 서비스하고 있다. 도서관리 시스템과는 별도로 운영되지만 데이터의 중복입력을 배제하기 위해 DB는 공유하고 있다. 전자도서관 시스템은 입력, 관리부분은 클라이언트/서버 구조이며 검색은 웹 기반에서 수행된다. 전자통신연구원 전자도서관 시스템의 서버는 IBM SP 이며, DBMS는 Sybase,를 쓰고 있다. 기타 부가적인 서비스로 각 신청자에게 배치방식으로 관련정보를 보내주는 반자동식 SDI 서비스를 제공하고 있다.

7. 이화여자대학교 전자도서관 사업

이화여자대학교는 1997년11월부터 전자도서관 사업을 시작하여 1998년도 12월에 완료할 예정이다. 현재까지 디지털화 된 자료는 자체 생산 박사논문으로서 약 860권이며 전량 Tiff 파일

로 구축되어 있다. 이화여자대학교의 도서관시스템인 ELIS와는 별개로 운영되지만 향후에는 하나로 통합할 계획이다. Z39.50 응용소프트웨어의 개발은 앞으로 계획중이며, Sun Enterprise 5000을 서버로 Oracle7.3, Fulcrum 검색엔진을 사용하고 있다.

IV. 전자도서관 구축을 위한 제안

21세기 세계경제는 산업사회 혹은 자원기반경제(Resource-based Economy)에서 지식기반경제(Knowledge-based Economy)로 급속히 이행하고 있는 것으로 이해되고 있으며, 이러한 지식기반경제 아래서 한 국가(대학)의 발전방향과 경쟁력은 첨단지식 및 기술의 창출과 그 활용에 달려 있다고 하겠다. 지식은 컴퓨터 및 통신망을 통해 정보로 변환되어 확산된다. 이는 궁극적으로 지식 습득 비용의 감소, 진입장벽의 붕괴 및 자연독점의 소멸을 통한 제품주기 및 기술 라이프사이클 단축으로까지 이어져 결국에는 세계화에 기여할 것으로 기대된다.

우선 전자도서관 구축은 정부차원의 적극적인 지원 없이는 DL개발은 불가능하다는 것을 외국의 사례에서 알 수 있다. 그리고 DL은 장기간 여러 관련분야의 전문가에 의한 기본연구와 응용분야연구가 선행되어야 하며 전자도서관개발을 위한 국가의 재정 및 정책적인 지원이 수반되는 국가차원의 종합계획이 필요하다는 점을 제안하지 않을 수 없다. 미국이 전자도서관 개발을 국가의 중요한 정책과제로 선정하고 있음을 의미 있게 고려하고 받아 들여야 될 것으로 생각된다. 최근 급속한 컴퓨터의 확산과 디지털 통신기술의 눈부신 발전에 힘입어 전 세계는 하나의 거대한 정보처리망, 정보교환망을 형성하게 되었고 이로 인하여 무한한 정보를 가공, 저장, 교환할 수 있게 되었다. 이러한 점에서 컴퓨터는 인간의 기억력과 사고력을 확대시키고 막대한 정보원을 제공함으로써 인간이 추구하는 발전 가능성을 확대시켜주고 있다.

통합된 체제의 디지털도서관의 이상적인 모형은 우선적으로 대부분의 기관에서 현재 구축되어 운영되고 있는 OPAC을 통한 서지정보검색 및 외부정보원에 대한 CD-ROM 네트워크 검색시스템 운영은 물론 원문 정보에 대한 이용자의 요구를 최대한 수용하기 위하여 소장자료 중 특화된 자료로서 활용도가 높고 저작권에 저촉되지 않는 자료를 대상으로 이미지형태의 원문 데이터베이스를 구축하고 나아가 PDF나 SGML 형태로 전자화, 디지털화가 완벽하게 이루어 져야한다 .

아울러 텍스트베이스로 구축된 데이터뿐만 아니라 이미지 베이스로 구축된 데이터까지도 문자인식(OCR-Optical Character Recognition)을 통해 자연어 검색이 가능한 전문 검색시스템

까지 제공되어야 한다.

그리고 음성, 영상의 디지털화 기술의 발달과 데이터의 압축 및 저장기술의 발달로 탄생된 VOD/AOD시스템을 통해 무수히 증가되고 있는 오디오, 비디오자료 및 멀티미디어 형태의 정보도 질의 확장, 재조직, 그리고 전거과정과 색인과정을 거쳐 도서관의 엔진에 의해 검색되고, 결과는 양방향 경로에 의하여 정보를 제공 받을 수 있어야 한다.

이를 위하여 서로 상이한 DBMS간에 검색의 호환을 유지하기 위한 통신규약으로 Z39.50 표준 프로토콜을 이용하며, 또한 주로 전자출판물의 데이터를 전체적으로 저장하고 전달하기 위한 구조로 편성, 제정된 표준문서형식 SGML DTD의 활용으로 전문정보가 특정의 Processing에 제한 받지 않고 각 시스템간에 교환과 저장이 가능하며, 인터넷에 확산되어 있는 HTML Format과도 호환성이 있어 Web Server를 통해 정보를 송수신함으로써 정보의 표준화를 유도할 수 있다.

디지털라이브리리는 표준문서 포맷인 SGML DTD을 채택하여야하여 데이터베이스에 구축되어 있는 문서의 포맷이 통일되어 모든 문서를 하나의 브라우저로 서비스할 수 있어야 한다. 그리고 한글도 Full Text Search가 가능하도록 완벽하게 한글을 구현할 수 있어야 한다. 데이터베이스관리 및 확장성은 물론 표준 프로토콜 Z39.50기반 전자도서관에 의하여 도서관간의 상호연계성이 뛰어나 자료검색과 상호대차 등 서비스의 극대화를 이룰 수 있어야 한다.

이상의 원리를 바탕으로 하여 보면 전자도서관은 첫째, 디지털도서관은 관련기관이나 각종 조직에서 개별적으로 설계되고 있으나 구현시에는 다양한 형식을 갖는 데이터를 저장 관리하여야 한다. 그리고 조직간에 데이터 교환 및 상호운용성(interoperability) 향상을 위해서 다수의 국내외 표준화 주체들이 메타데이터(Metadata)를 비롯한 다양한 표준화 작업을 진행해야 한다.

둘째, 타 도서관 및 관련 인접 분야와의 유기적이고 긴밀한 협조체제가 사업을 효과적으로 추진하는 데 절대적으로 필요하다. 이는 문헌정보학 뿐 아니라 컴퓨터공학, 전자공학, 정보통신학 등 기술 응용과학 쪽의 이론과 원리가 디지털라이브리리를 개발하고 추진하는 데 효과적으로 기여할 수 있다고 보기 때문이다.

셋째, 장기적인 마스터 플랜하에서 사업을 진행해야 한다. 예측 가능한 투자와 단계별 이용자 서비스를 통한 피이드백이 가능해야 사업결과를 보장할 수 있다. 물론 향후의 정보기술의 발전 추이나 정보환경 변화에 따라 마스터플랜은 갱신될 수 있어야 한다.

넷째, 자기만의 독특한 특성을 가진 전자도서관을 구축해야 한다. 콘텐츠, 적용기술 또는 아이디어 등에서 인터넷상의 고객들에게 충분히 어필할 수 있고, 다른 시스템이 갖지 않은 특징을 갖춘 전자도서관을 세워야 한다. 궁극적으로는 향후의 분산된 정보환경하에서는 이런 전문적이면서 특화된 유일한 특징을 갖는 전자도서관만이 인터넷상에서 살아 남게 되기 때문이다.

인터넷환경에서 정보를 공유하면서 확장성, 호환성, 이식성, 이동성이 자유로우면서 이용자에게 다양한 서비스가 가능해야 한다.

다섯째, 전자문서화작업으로 멀티미디어 자료를 전자문서화하는 작업에서 표준문서를 기술하는 메타언어와 문서의 유형을 정의하는 SGML DTD 그리고 분산검색을 위한 표준통신 프로토콜 Z39.50 확장서비스를 구현하는 문제 등이 연구되어야될 것이다

1. 데이터의 디지털화 경향

디지털도서관이 사회에 영향력을 발휘함에 따라 과학기술분야를 선두로 잡지, 도서, 고서, 고문서, 회귀도서 등의 어문저작물과 사진, 그림, 사운드, 동화상 순으로 급격히 전자화된 데이터로 변화해갈 것이며 과거의 자료도 소급하여 전자화되어 갈 것으로 전망되고 있다. 물론 현재의 환경에서 데이터의 축적은 많은 시간이 걸리고 막대한 비용이드는 일이므로 대단히 어려운 일이지는 하지만 장기간에 걸쳐 착실하게 점진적으로 변화해 갈 것이라 생각된다. 현재 디지털도서관을 계획하는 대학을 보아도 지적 재산권 처리가 무난한 해당 대학의 출판물(학위논문, 교수의 연구보고서, 교수의 학술지논문, 고서)을 대단위 全文 데이터베이스 구축의 우선 대상이 되고 있다. 또한 한국문예진흥원, 국립중앙박물관, 국립현대미술관 등에서 이미지 및 영화상데이터를 중심으로 하는 디지털도서관이 진행되고 있다. 분야별, 매체별, 서비스별, 지적재산권의 특성별로 많은 기관들이 분담하게 된다면 데이터의 디지털화는 보다 빨라질 것이다. 또한 디지털화 기술은 현재 관심 있는 기관을 중심으로 콘텐츠에 대하여 빠르게 디지털화하기 위한 계획에 따라 표준 형태, 편리하게 입력하는 방법이 연구되고 있으며 시간이 흐름에 따라 새로운 기법이 연구 발표되리 라고 기대되고 있다.

Columbia University Digital Image Access Project (DIAP) 는 컬럼비아대학 도서관이 개발하고 있으며 디지털이미지에 대한 서지데이터를 축적하고 액세스하기 위한 새로운 모형이다. DIAP팀은 요약서지정보 뿐만 아니라 필요하다면 상세한 계층 및 판별 관련데이터를 SGML로 코드화된 서지(메타데이터) 레코드로 축적할 수 있다는 것을 제안하였으며 이 레코드에는 실제적인 디지털문헌, 다른 관련 서지레코드, 심지어는 외부의 전자출판물, 데이터베이스, 수치파일 등과 같은 관련 디지털자료에 대한 링크를 포함시킬 수 있다고 하였다.

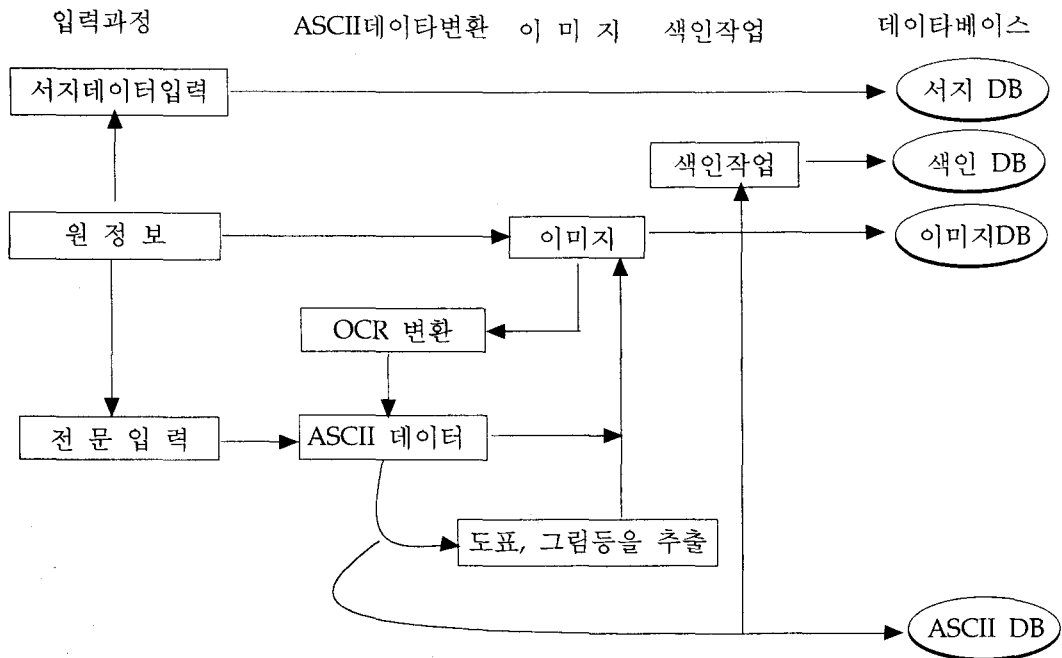
2. 디지털 데이터 축척과 대상자료

축척대상 데이터는 단행본을 비롯하여 학위논문, 기술보고서, 교과서, 고서, 繪畵, 도면, 지

도, 비디오, 오디오 학술지 기사 등으로 지금까지 도서관에서 수집, 보관, 서비스의 대상이 되어 왔던 것과 거의 일치한다. 그러나 디지털 도서관에서는 서지정보 뿐만 아니라 지금까지 도서관에서 축적대상이 되지 않았던 영상, 화상정보, 텍스트와 이미지 데이터, SGML 문서 등을 포함하는 전문데이터가 대상이 되며, 디지털도서관은 전문데이터베이스시스템이 중심이 된다. 이에 따라 복합적으로 멀티미디어 정보까지 디지털도서관의 축적대상자료로 들어오게 된 것이다. 디지털화된 자료는 원자료를 표준 전자문서인 SGML 문서나 PDF와 같이 구조화된 이미지 파일이 되기 때문에 그 자료체에 본질이나 특성이 변화되고 인덱스화 되기 때문에 대중적인 접근성이 크게 향상되고 안정성이 높아지게 된다.

디지털라이브러리의 유형을 생각할 때는 대상이 되는 분야와 이용자에 따라 그 구성이 크게 달라 진다. 우선 도서, 학위논문 등의 텍스트 중심 데이터와 사진, 지도 등과 같은 정지화상과 소리데이터, 비디오 이미지 혹은 애니메이션과 같은 동화상을 대상으로 하는 영화상 데이터로 나눌 수 있다. 전문데이터베이스는 ASCII텍스트모형, SGML텍스트모형, 페이지이미지모형, 통합모형 등 어떤 모형을 사용해도 좋으며, 全文데이터베이스를 구축하는 처리과정을 유형측면에서 단순화시키면 <그림 2>와 같다.

<그림 2> 全文데이터의 처리과정



우선 원 정보에 대하여 입력과정이 있게된다. 서지데이터는 키보드에 입력, 서지유틸리티에서 다운로드 그리고 기관에 따라서는 모 기관 도서관에 서지데이터가 작성되어 있을 수도 있다. 어느 쪽이나 색인과정을 거쳐 색인 데이터베이스를 만들게 된다. 전문이 ASCII파일로 되어 있지 않을 때는 모두를 입력하게 되는데, 키보드로 입력하는 수도 있고, 이미지를 OCR로 변환할 수도 있다.

그리고 원문정보를 이미지데이터로 만들었을 때에는 그대로 이미지데이터베이스를 만드는 방법과 OCR로 변환하여 ASCII데이터로 만들면서 이미지데이터베이스를 만들 수도 있다. 처음부터 입력된 ASCII 데이터는 이미지로 다운시켜 이미지데이터베이스를 만들 수도 있고 도표, 그림 등만을 추출하여 이미지데이터베이스를 만들고 나머지 ASCII데이터는 ASCII데이터베이스를 만들 수도 있다. 그리고 이미지로 다운로드 시켜 이미지데이터베이스를 만들고 도표, 그림 등은 무시하고 나머지 ASCII데이터만 ASCII데이터를 만들 수도 있다. 디지털 라이브러리의 특징은 서지 데이터베이스와 ASCII데이터는 색인과정을 거쳐 색인데이터베이스를 만든다는 점이다. 원 정보가 동화상, 정지화상 혹은 음성 데이터인 경우에는 이를 디지털화하여 이미지데이터를 만든 후 색인과 이미지데이터를 구축한다.

처음부터 이미지를 생성하던지 ASCII데이터를 이미지로 다운로드 시키든지 이미지데이터베이스는 가져야할 것이며, 데이터베이스로 구축된 모든 문서는 하나의 브라우저로 서비스할 수 있게 된다. 따라서 디지털 라이브러리에서는 원문 데이터베이스를 구축할 때에 모든 도서관들이 문서의 포맷을 통일하고 표준화하는 일이 중요하다. 이렇게 데이터베이스가 구축되면 추가되는 원문 데이터베이스의 관리와 확장성이 뛰어나게 된다. 이렇게 되면 도서관간에 상호연계성이 뛰어나 대 이용자 서비스를 극대화할 수 있다. 사용자에게 서비스할 때는 이미지데이터베이스를 주로 하여야할 것이다. 왜냐하면 ASCII데이터를 표준화하여 서비스하는 것은 완전하지 않기 때문이다. 다만 ASCII데이터는 색인용으로 사용할 수 있으며 추후 새로운 처리에 대응할 수 있기 때문에 이미지 데이터베이스와 병행하여 ASCII데이터베이스도 갖는 것이 바람직하다. 다만 처음부터 ASCII데이터만을 갖는다면 색인이 거의 불가능하므로 검색에 제한이 따른다.

V. 전자도서관 구축을 위한 환경구성

디지털 라이브러리는 인터넷에 분산된 자원을 사용자에게 단일의 인터페이스를 통하여 제공하며 분산된 자원을 자유롭게 수집 제공하기 위하여 메타데이터, 상호운용성, 지적소유권,

분산 정보검색 등의 기술 문제에 대한 해결을 문제로 제기하고 있다. 그리고 디지털도서관은 다양한 디지털 형태의 멀티미디어 정보를 수집, 가공하여 컴퓨터와 통신망을 기반으로 이용자에게 시간적, 공간적 제약없이 온라인으로 제공하기 위하여 메타데이터 저장기술, 압축, 색인, 정보검색, 초고속네트워크, SGML/XML/PDF 등 표준화, Z39.50 등 다양한 기술들이 도입되어야 한다.

1. 메타데이터 (metadata)

메타데이터란 데이터에 관한 데이터 즉 데이터에 대한 구조화된 데이터로서, 정보의 여러 속성을 기술하여 주는 데이터이다. 웹에서 문헌에 대한 서지정보나 도서목록 기록을 나타내고 있다. 인간이 생각할 수 있는 것의 모든 대상에 대하여 메타데이터는 존재한다고 할 수 있으며, 특히 정보의 체계적인 조직과 관련하여 그 중요성이 큰 메타데이터는 교환을 위한 표준적인 약속에 따라 해석이 가능한 형식으로 표현되는 것이다. 대표적인 것이 바로 도서목록인데 최근 인터넷상에 활발하게 이루어지고 있는 각종 메타데이터 연구도 이의 중요성을 인식하고 새로운 정보환경에 적합한 메타데이터의 체계 및 기술 방법을 찾고자 하는 데에서 비롯된 것이다. 이러한 메타데이터는 정보자원을 식별하고 그 소재를 파악하며 내용을 기술해 주는 역할을 하게 된다.

디지털라이브러리의 중요한 기능인 정보교환을 위하여 메타데이터는 상호운용성을 가져야 한다. 즉 다양한 형태의 메타데이터는 인터넷에서 일관된 형태의 검색과 교환을 지원하기 위하여 단일 의미를 표현할 수 있는 모델을 가져야 한다. W3C(www Consortium)는 워릭(Warwick)구조를 일반화한 의미 표현 구조로 RDF(Resource Description Framework)를 제안하였다. XML을 사용하여 정의하는 RDF는 다양한 형태의 메타데이터의 구조와 의미를 나타냄으로써 시스템사이에 상호운용성을 보장하여 준다. 그래서 현재 국내외의 표준화 기구 및 업계에서는 조직간의 데이터 교환 및 상호운용성 향상을 위해서 메타데이터를 비롯한 다양한 표준화 작업을 진행하고 있다

2. ISO/IEC-11179 데이터 요소의 명세 및 표준화

이 국제 표준은 데이터를 이해하고 공유할 수 있도록 만들기 위한 표준화와 등록에 대한 내용이다. 데이터 요소의 표준화와 등록은 기존의 데이터 관리 방법론들에 비하여 훨씬 적은 시간과 노력으로도 공유 데이터 환경을 생성할 수 있도록 한다. 또한 데이터 요소 기술의 관

리와 공식화에 대한 확고한 안내서를 제공하고, 데이터 요소들을 일정하고 표준화된 방법을 통하여 공식화 할 수 있도록 하고 있다. 이것은 데이터 요소 등록소(data element registry)를 위한 설명서를 제공한다.

1) SGML(Standard Generalized Markup Language) 문서관리 시스템

다가오는 21세기는 고도화된 정보사회가 될 것이며 이러한 정보사회에서 필수적인 요소는 기하급수적으로 발생하는 전자문서를 관리하는 일이라 할 수 있다. 디지털도서관에서는 대량의 전자문서를 효율적으로 저장하고 관리하는 시스템이 요구되고 있다. 텍스트, 그래픽, 오디오 및 비디오를 포함하는 멀티미디어 전자자료까지 모든 자료와 문헌의 포맷은 이 각종 시스템들간에 정보의 손실 없이 효율적으로 전송, 저장 및 자동처리할 목적으로 제정된 국제적인 표준인 SGML을 적용하여 향후 원활한 정보 및 문헌 유통체계를 구축할 수 있도록 하였다. SGML은 ISO의 물리적인 문서에서 논리구조와 내용구조를 기술하기 위한 메타언어로서 문서표현 및 교환 표준으로 제정되었으며, 구조 정보로 사용할 수 있는 마크업(Markup)을 포함하고 있어 문서를 구조 단위로 표현할 수 있다. 이는 대량의 전자문서를 저장하고 관리하는 데 적합하다. 특히 디지털 라이브러리, 인터넷상에서의 CSCW(computer supported cooperative work), CALS(commerce at the light speed) 등 대량의 전자문서를 관리하거나 구축해야하는 요구가 발생하면서 대량의 문서정보를 효율적으로 저장하고 관리하는 시스템이 요구되고 있다.

지금까지 대부분의 도서관은 MARC 형식을 중심으로 서지데이터를 기술하여 왔다. 그러나 전문을 중심으로하는 디지털 도서관의 데이터기술은 MARC 만으로는 기술이 불가능하여 SGML 형식을 고려하게 되었다. SGML에서는 문헌에서 필요한 요소와 실제로 출현하는 요소에 대한 구조화 정보를 문헌유형정의(DTD)를 이용하여 처리과정과 독립적으로 기술할 수 있으며, 이 기술의 정당성은 구문분석기로 확인할 수 있다. 처리과정과 독립적이므로 DTD만 알면 SGML로 코드화된 데이터를 상이한 하드웨어와 소프트웨어의 여러 가지 응용시스템에서 다양한 방법으로 이용할 수 있다.

SGML은 ISO에서 정한 문서표현 기준으로 제정된 것으로 DTD에 따라 문서의 구조를 지정하는 마크업 언어이다. 웹 브라우저가 사용하는 HTML언어는 SGML을 기반으로 한 언어의 일종이다. 마크업이란 편집자나 인쇄 디자이너가 색인 텍스트 및 그 외의 부분에 대해서 서체나 사이즈 등 체계에 관한 지정을 원고 상에 손으로 기입하는 방식이다.

SGML은 문서 데이터베이스의 작성이나 유지, 보수에 유효할 뿐만 아니라, 교환 유통에도 적당하다. 개개의 시스템에 독립되어 있을 뿐만 아니라 구조가 엄격히 정의 되어 있어, 개개

의 시스템을 위하여 변환할 수 있기 때문이다. 나아가 SGML에서 만들어진 문서 데이터베이스는 전자도서관 시스템의 자원으로 사용된다.

DTD는 SGML의 규정에 따라 문헌의 문단을 분리하고 표목을 확인하는 등의 방법과 절차에 관한 코드(혹은 markup)를 정한 명세서이다. 문헌을 DTD로 전송하면 DTD 리더 (혹은 SGML 컴파일러)를 통하여 그대로 화면에 나타내거나 출력할 수 있다. 이는 표준적인 SGML 컴파일러 하나가 여러 가지 마크업 코드를 사용한 다양한 문헌을 처리할 수 있음을 의미한다. 컴파일러는 DTD에 따라서 화면이나 프린터에 출력된다.

XML(Extensible Markup Language)은 WWW, 인터넷 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. XML이 SGML과 HTML의 단점을 극복하기 위하여 개발되었으므로 이에 대한 기대감은 높지만 아직은 국내 한글 환경에서 널리 사용되고 있지 않다.

SGML은 문서의 내용이나 구조를 정의하기 위한 언어로서 다양한 구조들 사이에 데이터의 위치를 표현하는 구조화된 데이터를 상호교환 하기 위한 도구로 장서 관리의 용이성, 유통 배포의 용이성, 새로운 서비스 제공, 복사의 용이성 등의 장점이 있다고 하겠다.

SGML 문서에 관련된 다양한 응용 프로그램에 대한 지원과 문서의 공유 등을 지원하기 위하여서는 데이터베이스에 기반한 SGML 문서관리 시스템이 필수적으로 요구되며, SGML 문서관리시스템을 개발하기 위해서는 다음의 네 가지 사항을 고려해야 한다.

첫째, SGML 문서의 데이터 모델로 이는 SGML 이 포함하고 있는 많은 시맨틱 (semantic) 들을 손실 없이 데이터베이스에 표현할 수 있도록 모델링하는 것이다.

둘째, 새로운 문서 구조와 내용을 정의하고 설명하는 SGML DTD를 분석하여 이에 대한 스키마를 데이터베이스에 생성하는 것으로 이는 다양한 구조의 문서들을 효과적으로 관리 할 수 있는 기반이 된다.

셋째, SGML 인스턴스를 데이터베이스에 효율적으로 저장하는 것으로 새로운 SGML 인스턴스가 들어오면, 이미 생성되어 있는 스키마에 이를 저장할 수 있어야 한다.

넷째, 데이터베이스에 저장된 SGML 인스턴스를 문서 또는 구조 단위로 꺼내는 것이다. 이는 한 문서의 공동저작이나 구조 단위의 검색을 요구할 때 필연적으로 요구된다.

넷째, 데이터베이스에 저장된 SGML인스턴스를 문서 또는 구조 단위로 꺼내는 것이다. 이는 한 문서의 공동저작이나 구조 단위의 검색을 요구할 때 필연적으로 요구된다.

전자도서관에 있어서 전문DB를 효율적으로 구축하기 위해서는 원문을 완벽하게 재현하는 SGML 자동변환기가 개발되어야 하고 이것이 DL발전을 앞당기는 요인이 될 것으로 생각된다.

2) MARC (Machine-Readable Cataloging)

MARC는 미국 의회도서관 (Library of Congress) 자동화 계획에 따라 1965년 MARC I이 완성되었고 MARC I을 수정·보완하여 1968년 MARC II가 제정되어 도서관 자동화의 중요한 개념으로 등장하였다. MARC는 초기에는 미국 의회도서관의 정부관련 고문서의 목록을 축척하기 위한 방법이었기 때문에 미국 의회도서관의 요구에 편중되어 개발된 것으로 국제적 표준형식으로 적용시키는데 한계성을 극복하기 위하여 USMARC로 개정되었고 새로운 다양한 형태의 정보자료나 통제방식을 지원하도록 확장·개발되어 왔다. 즉 고문서 전거통제의 표준적 방식을 제기했으며 기존 자료에 관한 미국내의 국가적 형식을 제공하고 기계가독 파일의 목록 규칙으로 각종 인쇄물이나 사진과 같은 그래픽 자료 등의 비텍스트 자료 처리를 위한 형식의 확장 작업으로 개발되었다.

3) MARC/SGML 프로젝트

지금까지 도서관목록에서의 검색내용은 저자, 서명, 주제, 키워드 등의 탐색이 있었다. 그러나 최근에는 인터넷 정보검색 기술이 일반화되면서 Web에서 Yahoo, Galaxy 등의 검색엔진은 주제별 디렉토리 검색을 하고, Altavista, Infoseek 등의 검색엔진은 로봇을 이용하여 인덱스 검색을 제공하며, Metacrawler, Savvysearch 등은 메타검색으로 자체 데이터베이스를 가지고 있지 않으면서 사용자가 입력한 검색 키워드를 Agent를 사용하여 여러 다른 검색엔진에 병렬적으로 질의해서 얻어진 결과를 종합하여 사용자에게 알려 주는 검색 방법 등 Web을 이용한 인터넷 검색방법이 다양하다. 이 외에도 많은 검색엔진들이 가중치탐색, 단어의 위치나 빈도수를 이용한 적합성 피드백, 자동어미절단, 자연어 질문시스템, 개념기반탐색, 의미트리 등의 새로운 기법에 의하여 검색결과를 제공하고 있다.

MARC/SGML 프로젝트는 미국국회도서관의 자문위원회가 중심이 되어 1990년부터 MARC 레코드용 SGML DTD를 개발하기 시작하였으며 개인, 기업, 기관이 참가하고 있다.

공식적으로 채택된 것은 아니지만 자문위원회에서 정한 설계원칙 중 중요한 것은 다음과 같다. 첫째, MARC DTD는 실제 MARC에서 SGML로, SGML에서 MARC로 정보의 손실 없이 완전하게 양방향 변환이 가능 해야한다. 둘째, MARC/SGML레코드는 SGML문헌내에 포함시킬 수도 있으며 메타데이터로서 독립시킬 수도 있다.

4) 더블린 코어(Dublin Core)

미국 OCLC와 NCSA가 1995년에 더블린에서 합의한 메타데이터이다. 데이터의 호환성을 유지하고 네트워크 자원의 기술에 필요한 일련의 데이터 요소를 규정하기 위함이다. 이들 자

원의 신속한 검색을 목적으로 OCLC와 NCSA가 더블린에서 개최된 워크숍에서 합의된 메타데이터를 더블린 코어라고 한다. 이 회의에서는 기존의 표준이나 프로토콜을 수정, 확장하여 네트워크 자원을 기술하고 접근하는 방안을 모색하였다. 이 회의의 주된 목적은 MARC와 같은 기존의 메타데이터로 네트워크 자원을 표현하는 데에는 구조의 경직성으로 인해 많은 비용과 시간이 소요되어 이를 대체할 수 있는 단순구조의 형식을 모색하기 위한 것이었다. 이에 따라 기본 데이터 요소의 선정과 기술의 기준을 다음과 같이 제시하였다.

- ① 고유성 - 자원의 본질적인 특성을 기술요소로 한다.
- ② 확장성 - 규정된 필수 데이터요소 이외에 부차적인 내용이나 특성을 기본 요소로 사용한다.
- ③ 구문의 독립성 - 응용분야나 표현기법을 규정하지 않는다.
- ④ 선택성 - 각 요소의 수록여부를 강제하지 않는다.
- ⑤ 반복성 - 모든 기술요소는 반복 사용할 수 있어야 한다.
- ⑥ 수정 가능성 - 한정어를 사용하여 세부사항을 조정한다.

더블린 코어 기본 데이터요소별 한정어는 Type과 Scheme으로 구분된다. 더블린 코어의 한정어는 각 응용에 필요한 도메인 특성 및 새로운 환경과 관련된 요소들을 추가 또는 생략할 수 있다. 그러나 더블린코어 기본 데이터 요소 중에서 제작자, 기타제작자, 발행처에 대한 정의가 특정 응용분야에 따라 의미가 다르게 해석되고 어떤 분야에서는 정확히 구분하기가 어렵다.

이를 위해 네트워크 환경에서 자원을 기술하고, 접근하는데 필요한 13개의 데이터요소를 기본으로 규정하였다. 그 후 1996년 영국의 워릭(Warwick)에서 열린 제2차, UKOLN/OCLC DC 워크숍에서는 더블린 코어의 입력 및 교환을 위한 구체적 구문에 대해 논의하였다. 또한 다양한 메타데이터로 구성되는 계층 구조상의 한 요소로서 더블린 코어를 위치하는 방식을 채택하고 이를 워릭 프레임워크로 제안하였다. 이 회의에서는 기존의 13개 데이터요소 이외에 두 개의 요소를 추가하여 모두 15개의 데이터요소를 확정하고 이의 실제적인 응용을 위한 이용자 지침을 제정하였다. 더블린 코어 데이터요소 15개는 아래 <표 2>와 같다.

<표 2> 더블린 코어 15 데이터요소

컨텐츠 기술요소 물리적 기술요소	지적속성요소	
Title (표제) Subject (주제) Description (설명) Source (출전) Language (언어) Relation (관련자료) Coverage (내용범위)	Creator (제작자) Publisher (발행처) Contributor (기타제작자) Rights (이용조건)	Date (발행일자) Type (자료유형) Format (형식) Identifier (식별자)

그 후 제3차, 워크숍에서는 더블린코어의 기술대상을 텍스트뿐만 아니라 이미지로 확장하여 화상자원에 대한 더블린 코어의 응용타당성에 대해 논의하였고, 1997년 4차, 워크숍에서는 각 데이터 요소의 하부구조와 로컬에서의 확장 방안을 검토하고, 데이터 요소의 의미와 기술범위를 규정하였다. 더블린 코어의 데이터의 형식과 구조를 단순화하여 원문의 저자나 발행자가 메타데이터를 직접 작성하고, 네트워크 출판을 위한 저작도구의 개발자가 이 정보에 대한 템플릿을 직접 소프트웨어에 포함할 수 있도록 하며, 작성된 데이터를 기초로 특정 분야에서 요구되는 상세한 수준으로 확장하여 사용하기 위한 것이다. 이를 위해 더블린 코어에서는 자원의 소재위치와 이러한 자원을 획득하는데 필요한 데이터만으로 요소를 구성하고, 안전이나 검증, 구독, 이용과 관련된 요소는 제외하였다. 이처럼 더블린코어에서 네트워크 자원의 식별에 필요한 데이터 요소만으로 제한한 기본의도는 데이터요소의 단순화를 통해서 유용성과 이용의 용이성을 확보하고, 필요시 다양한 목적에 따라 확장하기 위한 것이다. 그래서 특정 이용자 집단에서 중요하다고 판단되는 데이터 요소나, 특정 서지기관이 필요로 하는 어떤 데이터 요소라도 추가할 수 있다. 그러나 이러한 특징으로 인해 상세 요소의 교환이나 상호운용성을 보장하기 어려운 단점이 가장 중요한 점은 요소의 의미와 내용이 특정한 기술규칙이나 구문과는 독립적이라는 점이다.

5) RDF (Resource Description Framework)

RDF는 다양한 메타데이터들을 인코딩하고, 재사용 및 교환도구로 활용할 수 있는 자원기술 모델링 언어이다. RDF는 Web에서 기계가독정보들을 교환하려는 응용들 사이의 상호운용성을 제공하기 위하여 W3C 메타데이터 RDF워킹 그룹에서 제안한 것이다. 따라서 RDF는 메타데이터들 간의 상호운용성 및 기계 이해 가능한 의미적 정보를 제공하고 인터넷 자원발견을 위하여 일관성 있는 질의기능을 제공하며 상이한 메타데이터들의 연계 기술 언어로 사용될 수 있다. WWW의 출현은 전세계적으로 네트워크 상의 분산되어 있는 정보자원에 대한 접근을 용이하게 해주었다. 그러나, 네트워크 상의 자원의 성격은 동적이기 때문에 전통적인 목록과 같은 주제 접근이 어렵다. 따라서 네트워크 자원을 기술하기 위한 새로운 표준이나 기법이 필요하게 되었다. 메타데이터는 정보의 기술과 접근을 용이하게 하는 수단으로 이용된다. 그러나 다양한 메타데이터의 등장은 메타데이터의 활용 상의 혼란을 야기하였고, 상이한 메타데이터의 운용에 문제점을 드러내었다. 상이한 메타데이터를 효율적으로 통합하고 이용할 수 있도록 하기 위해 어의나 구문 및 구조 등의 측면을 지원하는 메타데이터 관련 요소들의 공통적 규정이 있어야 하는데, RDF는 이를 지원하는 '웹 지원을 기술한 데이터 규약'이다.

RDF의 특징은 상이한 메타데이터간의 어의, 구문 및 구조에 대한 공통적인 규칙을 지원하는 메커니즘을 제공하여, 메타데이터의 기계 가독형 정보의 교환을 지원한다. 그리고, 구문에

독립적(syntax independent)이므로 상이한 메타데이터간에 어의적인 차이를 인정하며, 그 구문에서 데이터 요소 등을 규정하지 않는다. 즉, 동의어간 어미 전달이 가능해 진다. XML에 기반하여 기술하기 때문에 XML 자체의 확장성을 보장받을 수 있다. RDF를 이용하여 자원을 기술하는 경우, 자원의 성격 및 특성에 따라 그 기술 내용을 자유롭게 확장할 수 있다.

6) Z39.50 - 정보검색 프로토콜

Z39.50 프로토콜은 이 기종들 간의 상호운용성을 보장해 주는 OSI (Open System Interconnection 개방형 시스템간 상호연결) 참조 모델의 사용자와 인터페이스를 담당하는 응용계층에 해당되는 프로토콜이다. Z39.50에 정의된 프로시저와 형식에 의하여 클라이언트는 서버에게 정보를 요청하고 이 요청에 따라 서버는 클라이언트에게 정보를 제공한다. Z39.50 표준이 지원하는 기능은 질의 기능, 정보검색 기능, 응답처리 기능, 결과처리 기능, 결과전송 기능, 관리기능 등을 지원한다.

오늘날 인터넷 정보자원 검색을 위한 표준 정보검색 프로토콜로 자리를 잡은 Z39.50은 두대의 컴퓨터가 서로 상호작동하기 위해 사용하는 포맷과 절차들을 다루는 일련의 규칙들의 집합인 네트워크 프로토콜이다. 이 프로토콜은 세션지향적 (session-oriented)이며 연속적이다. 그리고 Z39.50은 타 기관의 상이한 DBMS 간에 검색의 호환을 유지해주기 위한 통신규약이며 인터넷 통신 표준 프로토콜인 HTTP와 호환성이 있는 CGI를 이용하여 웹서버를 통한 전문 (full-text)데이터베이스의 송수신이 가능하도록 지원한다.

(1) Z39.50과 전자도서관 망 기술 현황

기존 대부분의 전자도서관은 Z39.50이라는 전자도서관 프로토콜을 기본으로 하는 클라이언트/서버형 모델을 기반으로 구축되었다. Z39.50은 전자도서관과 관련한 유일한 표준 프로토콜이라는 점에서 중요하다. Z39.50은 OSI 참조모델의 응용계층 프로토콜로서 클라이언트/서버 구조에서 질의 및 정보검색, 결과전송, 처리 및 관리 등을 담당한다.

Z39.50 은 <표 3>와 같은 11개의 서비스 기능을 제공하고 있다.

<표 3> Z39.50 프로토콜이 제공하는 서비스

기능 (Facility)	기능 내용 (서버 - S, 클라이언트 - C)
접속(Initialization)	S에 접속하기 위해 C가 요구함
질의(Search)	S에 있는 DB에 질의하여 검색된 결과를 전송 받을 수 있도록 C가 요구함
정보검색(Retrieval)	전송 : 검색된 결과셋으로부터 원하는 레코드를 전송 받기 위해 C가 요구함. 세그먼트 : 원하는 레코드를 S가 단 한번에 제공하고자 할 때 요구함
결과셋 삭제 (Result set-delete)	검색된 결과셋을 삭제하기 위해 C가 요구함

기능 (Facility)	기능 내용 (서버 - S, 클라이언트 - C)
브라우저(Browse)	DB를 편리하게 검색할수 있도록 질의할 수 있는 용어(term)를 검색할 수 있게 C가 요구함
정렬(Sort)	결과셋을 정렬하기위해 C가 요구함
접근제어 (AccessControl)	보안처리가 필요한 서비스를 사용자가 요구할 때 제어하기 위해 C가 요구함
지원관리 (Qccounting Resource Control)	자원제어 : S의 상태를 C에게 알려 주며 현재 수행 중인 서비스를 계속할지 또는 중지할 지를 요구하기 위해 S가 요구함 트리거지 자원제어 : C가 S에서 수행중인 연산을 중지할 수 있는 액션, S의 상태를 요구 할 수 있는 액션, S와 협상할 수 있는 액션 등 3가지 액션을 제공하기 위해 C가 요구함 자원리포트 : C가 S에서 수행중인 연산의 상태를 알기위해서 C가 요구함
설명(Explain)	C가 현재 연결중인 S에 대한 정보(S이름, DB이름, DB설명, 필드이름 등)를 위해 C가 요구함
확장(Extended)	C가 S에 있는 결과셋이나 질의를 저장하거나, SB를 갱신할수 있도록 C가 요구함
해제(Termination)	C 또는 S가 모든 연산을 중지하고 Z39.50 세션을 해제할 수 있도록 함

7) 상호운용성 (Interoperability)

상호운용성은 서로 다른 형태의 디지털 라이브러리 사이에 통합과 관련된 시각을 제공하는 기능이다. 즉 임의의 디지털 라이브러리가 제공하는 다양한 형태의 문헌에 대한 단일 인터페이스를 제공한다. 이러한 방식의 통합은 커다란 형태의 분산된 디지털 라이브러리를 구축할 수 있게 한다.

분산 시스템에 대한 상호운용성을 나타낸 구조이다. 디지털 라이브러리의 사용자(DL Clients)는 GUI(Graphical User Interface) 화면에서 자연어를 통하여 질의를 하면, 인터넷상에 있는 DB, Archive, Indices, 등의 시스템으로부터 원하는 정보를 찾은 후 중개자를 거쳐서 사용자에게 제공한다. 중개자는 사용자가 원하는 정보가 있는 위치, 쓰여진 언어, 문헌 형태 등에 무관하게 정보를 제공받을 수 있게 한다. 상호운용성을 제공하기 위하여 데이터모델, 중개와 통제, 질의 처리, 구현 메카니즘과 같은 기능을 제공할 수 있어야 한다.

(1) 데이터 모델

여러 서버 사이에 단일 시각을 제공하기 위하여 통합된 데이터 구조와 의미 표현이 가능하여야 한다. 데이터 모델을 기술하는 언어는 최대의 융통성과 상호운용성을 위한 존재구조적(ontological) 기반을 가져야 한다.

(2) 중개와 통제

데이터 모델 기술 언어는 새로운 데이터 형태의 추가를 허가하여야 하며, 정보 하부 구조(Infrastructure), 서비스 인터페이스, 또는 기능의 변화에 대한 중개 역할을 고려하여 설계하여

야 한다. 디지털 라이브러리의 구조 변경이나, 데이터 갱신으로 인한 불 일치성을 해결할 수 있는 시스템의 형태를 갖추어야 한다.

(3) 질의 처리

질의 처리는 질의어 형식, 다른 시스템으로 질의어 전송, 변환 그리고 결과에 대한 순위매김문제를 가지고 있다. 적합성 재전송(relevance feedback)은 질의 결과에 대한 질을 향상시켜 준다.

(4) 구현 메카니즘

디지털 라이브러리 구현은 분산 컴퓨팅 환경과 유사한 특징을 가지고 있으며, 표준화된 CORBA(Common Object Request Broker Architecture)와 같은 프로토콜을 사용하여야 한다. 상호운용성에 대하여 연구 결과로 (3)과 같은 사례가 있다.

VI. 디지털도서관 검색과 탐색

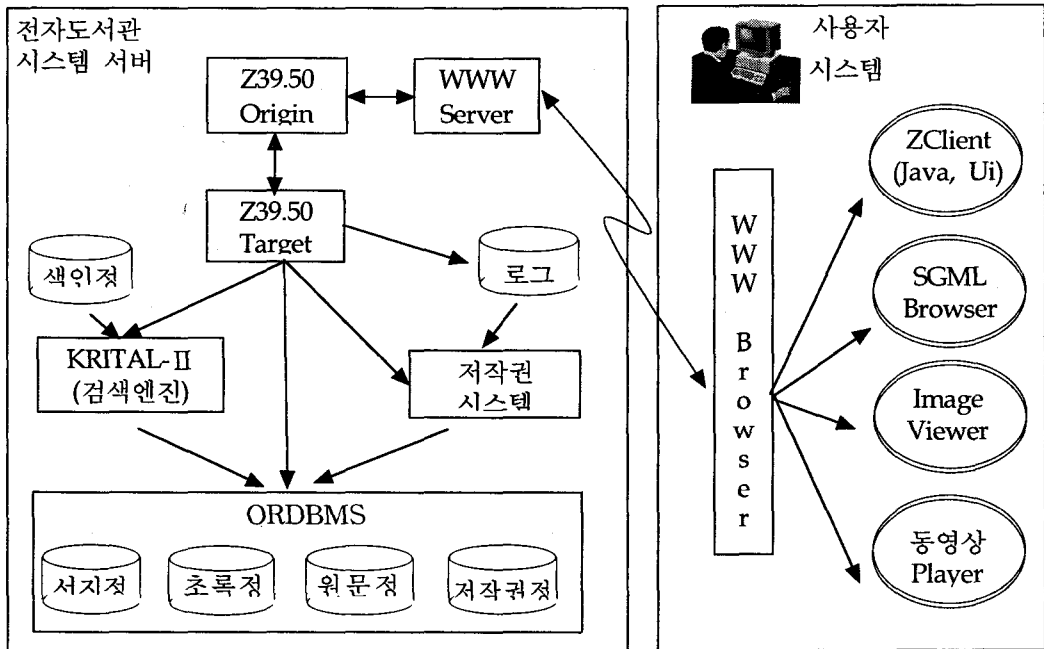
디지털도서관에서 가장 기본이 되는 검색 기능은 크게 서지정보, 디렉토리 정보, 텍스트정보를 포함하는 비 정형 자료의 검색으로 나누어 생각할 수 있다. 서지정보검색은 정형데이터용 DBMS로 구현할 수 있는 부분으로 SQL과 같은 프로토콜을 사용하고 디렉토리 정보 서비스는 X.500 이나 LDAP과 같은 프로토콜을 사용한다.

텍스트정보검색은 기존의 정보검색을 모듈로 사용하는 경우와 서비스영역에 적합한 검색엔진을 구현하기도하는 데 근래에는 SGML이나 XML과 같은 구조화된 문서를 대상으로 문서의 부분을 검색하는 기능에 대한 요구가 높아져 이에 대한 연구 개발이 이루어지고 있다. 클라이언트/서버환경인 경우 표준인 Z39.50을 사용하는 경우가 많으나 Dienst와 같은 프로토콜 혹은 HTTP를 사용한 인터넷 검색엔진을 사용하기도 한다. 멀티미디어 자료를 내용에 의해 검색하는 기법도 디지털도서관의 구축에 매우 중요한 요소이다. 이미지나 오디오 혹은 비디오를 검색하는 방법은 난이도가 높은 기술로서 아직은 검색신뢰도 면에서 많은 문제점을 가지고 있으나 디지털도서관 구축에 실제로 사용되기도 한다.

1. 색인, 검색 및 인터페이스 기술의 중요성

전문데이터를 소장하는 정도에서 그치지 않고 다양한 측면의 잠재적 정보요구를 해결하기 위하여 혁신적인 색인기법 및 색인어의 추출문제가 사용자 인터페이스와 함께 가장 중요한 위치를 차지하게 될 것이다. 왜냐하면 데이터베이스의 중요성은 색인에 있으며 해결되지 않은 문제도 대부분 색인기술과 관련이 있기 때문이다. 디지털도서관에 관한 최근의 국제회의에서 가장 중요한 관심사가 자동색인 및 검색효율이며 최근의 자연언어처리, 특히 한글 및 한국어에 대한 자연언어처리의 연구가 활발한 것은 이를 반증하고 있다. 검색기술은 전문검색 및 통합검색을 기반으로 결과의 순위매김에 집중되고 있다. 또한 자연언어처리, VRML(virtual reality markup language), 최근 JAVA 등을 이용한 사용자 인터페이스의 설계, 개발, 평가는 한 분야만의 문제가 아니라 학제적인 협력을 요하는 중요한 과제가 될 것이다. Z39.50 기반 전자도서관 정보검색 시스템 구성도는 <그림 4>와 같다.

<그림 4> 정보검색시스템 구성도



2. 분산 정보 검색/전자정보서비스

분산 정보 검색은 정보 검색을 웹을 통하여 할 수 있게 하는 기능이다. 웹 상에서 정보 검색

색이 가능하게 하기 위하여 디지털 라이브러리에서 사용하는 다양한 프로토콜에 대한 기능을 지원하는 게이트웨이 모듈을 필요로 한다. 여기서 정의하는 게이트웨이 모듈은 다양한 표준 프로토콜을 이용하여 디지털 라이브러리에서 메타데이터를 수집하는 기능을 포함한다.

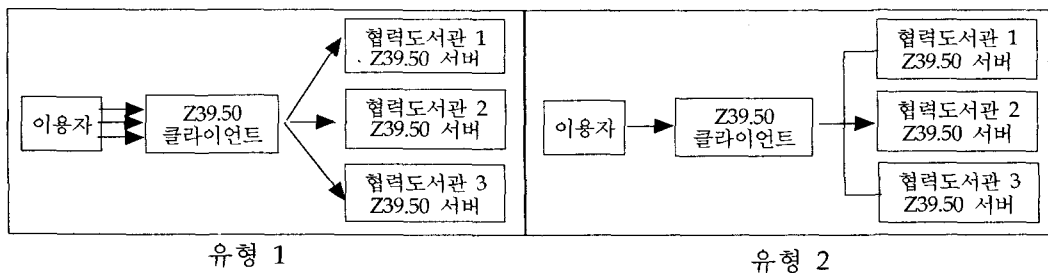
전통적 Yahoo나 AltaVista와 같은 검색기는 하이퍼링크를 통하여 연결된 여러 웹 사이트(Sites)로 부터 정보를 수집하여 기 정의된 키워드 방식에 의한 검색 서비스를 제공하고 있다. 그러나 이러한 방식은 비 공개된 가치 있는 정보를 제공할 수 없으며, 정보에 대한 구조적 검색이나 자세한 정보를 얻기가 어렵다.

메타데이터는 저작권 문제가 없는(일반적으로) 공개 자료로 누구나 추출하여 사용할 수 있으며, 표준 기술 방식인 SGML로 정의되어 있어서 정보의 정확한 추출 및 제공이 가능한 정보이다. 이와 같은 정보를 얻기 위하여 형식을 지원하여야 하며, 여러 사이트와 통신을 위한 표준 프로토콜을 이해하여야 한다.

DLI 프로젝트의 일환으로 Urbana-Champaign에 있는 일리노이 대학에서 개발한 Emerge 시스템은 하나의 디지털 라이브러리에서 다수의 인터넷 자원을 검색할 수 있는 기능을 가지며, 하부구조인 과학문헌 연합저장소(Federating Repositories of Scientific Literature)를 구축하였다.

Emerge는 Z39.50을 지원하는 Pizzazzd와 다양한 프로토콜을 지원하는 Gazebo를 통하여 웹에 분산된 정보를 수집하여 인덱스한다.

〈그림 5〉 Z39.50 탐색유형



3) 자동색인 (Auto indexing) 기능

(1) 자동색인 과정

일반적으로 자동색인 시스템은 형태소 분석에 의해 색인어 후보를 추출한 후에 통계적 기법에 의하여 색인어를 선별하고 시소러스를 이용하여 주제어를 추출함으로써 적합한 색인어 리스트를 선정하게 된다. 실제적으로 같은 주제를 설명하는 경우의 단어라고 해도 사용자의 주관에 따라 검색어가 다르게 사용되어 검색율을 떨어지게 하는 경우가 있어 이러한 문제점을

해결하기 위해 자동적으로 검색기능을 확장하여 더 많은 정보를 불러올 수 있는 완벽한 표준 ANSI시소러스를 지원한다. 시소러스는 검색하는 동안 검색요구를 시스템이 선호하는 용어로 바꾸거나 용어가 관련된 개념 계층으로 재배치한다. 만약 사용자가 선호하던 검색어를 변경해야 하는 경우에는 아주 쉬운 방법으로 사용자가 시소러스 데이터베이스의 용어를 추가, 변경, 삭제할 수 있다. 또한 사용자는 인터페이스 모듈에서 시소러스의 브라우저 기능에 직접 접근할 수 있고 관련된 용어는 검색에 자동으로 포함되어 사용자가 선택할 수 있도록 하였다. 자동색인 시스템의 과정을 보다 상세하게 기술하면 다음과 같다.

첫째, 도치파일 (Inverted File)을 만들기 위하여 문헌을 하나씩 입력받아 이들 텍스트들로부터 단어별로 나눈다.

둘째, 문헌에서 구문 분석한 단어를 색인시, 색인어로 가치 없는 불용어 리스트(Stop Words)와 비교하여 불용어 리스트 (Stop List)에서 발견되지 않는 단어는 스템밍 (Stemming) 작업을 한다.

셋째, 문헌에서 나오는 단어의 빈도수를 계산하여 검색된 문헌을 순위결정 (ranking)을 할 때 사용한다. 스템밍은 정보검색 시스템의 성능을 향상시키고 색인 파일의 크기를 줄이기 위해 스템밍은 용어를 합성시키는 방법이다. 즉 형태가 다르지만 같은 어근을 가지고 있는 용어를 같은 용어로 인식하는 것을 말한다. 스템밍을 하면 용어대신에 스템을 저장하기 때문에 색인 파일의 크기를 절반으로 줄일 수 있다.

넷째, 마지막으로 단어와 단어가 발견된 문헌번호, 단어의 빈도수 같은 관련된 정보를 데이터베이스에 저장하는 것을 자동색인이라 하고 이런 구조를 도치파일 (Inverted File)이라 한다.

VII. 저작권법에 의한 보호의 한계

1. 디지털도서관과 저작권 관계

저작권법은 저작물을 문예, 학술 또는 예술의 범위에 속하는 창작물이라고 정의하고 있다. 저작권제도의 목적은 전달하고자하는 아이디어나 메시지를 가진 사람이 다른 사람에게 정보 미디어를 공개해서 그 아이디어 메시지를 표현하고 유통시키는 것을 장려하거나 정보유통의 원활화를 기함으로써 문화, 예술, 과학, 기술의 발전을 촉진시키고 풍요로운 문화생활을 영위할 수 있도록 하는 데 있다. 그런데 저작권도 저작물의 디지털화와 이용환경의 네트워크화로 정보환경이 변화됨으로써 기존의 종이매체 환경 하에서의 저작권의 개념이 그대로 유지될 수

없게 되었다. 즉 전자출판, 하이퍼미디어, 정보관리기술을 주된 기술요소로 한 전자도서관의 출현은 정보의 생산과 배포, 이용의 과정에서 저작권과 지적재산권에 있어서 새로운 문제를 야기하였다. 그러나 도서관의 업무와 정보서비스가, 저작권법의 자유 이용의 범위 내에서 행해지던 시대는 지났다. 도서관이 취급하는 정보와 정보서비스의 형태는 변화하여, 저작권법의 자유이용의 규정밖에 해당하는 경우가 다수 출현하고 있다. 그렇게 되면 도서관은 이용자가 희망하는 서비스를 어떻게 하면 저작권자에 인정받을 수 있는가를 고려해보지 않으면 안된다. CD-ROM형태로만 출판하는 전자책 등의 정보를 이용자가 이용하고자 하는 경우는 인쇄매체와 동일하게 접근할 수 있는 방법, 또는 최신의 정보를 희망하는 이용자에게, 데이터베이스와 인터넷에서 검색해 제공할 수 있도록 하는 방법에 대하여 논의가 이루어져야 한다. 모든 콘텐츠 즉, 아날로그 정보와 디지털 정보를 막론하고 저작물을 이용하기 위해서는 저작자와 계약에 의해 합의를 얻어야만 한다. 그러기 위해, 이용자에게 대한 정보서비스는 물론이고, 도서관내에서 이루어지는 모든 업무에 대해서 즉 인쇄 매체, 전자매체를 막론하고, 저작권이 문제가 되는 경우에 있어, 저작권에 관한 지침과 규정을 정비하지 않으면 안된다.

2. 콘텐츠(Contents) 자체의 저작권문제

콘텐츠의 경우 저작권 내지는 기타 지적 재산권 법의 규정에 의한 모든 권리가 독립적으로 인정되는 것과 아울러 데이터베이스의 보호규정도 함께 적용이 되고 있다. 콘텐츠는 디지털도서관에 있어서 주로 아날로그 저장물의 디지털화를 통하여 구축될 수밖에 없다. 현재 서비스되고 있는 디지털도서관의 콘텐츠는 서지데이터정보, 키워드, 풀텍스트 등을 담고 있는 단행본, 잡지기사, 신문기사 등이 고작이다. 이 경우 어문저작물에 해당하는 콘텐츠의 경우는 어문저작물로서 보호되며, 또한 대부분의 경우에 있어서 저작권을 가지고 있지 않은 공개정보인 경우도 있다. 그러므로 콘텐츠 자체가 저작권을 가지고 있는 경우에 데이터베이스의 제작시 반드시 그 원저작자의 허락을 필요로 한다. 이에 대하여 저작권 법상 저작재산권의 제한 규정인 도서관 등에서의 복제를 규정한 저작권법 제28조에서는 도서관등의 자료를 자체 보존하기 위해 필요한 경우 일부분에 대한 이용자의 복사, 보존용의 제공 등에 대하여 복제권을 제한하고 있으므로 이에 대한 어느 정도 수긍할 수 있는 여지를 남겨두었다.

3. 아날로그 저작물의 디지털화 문제

아날로그 저작물을 디지털화하는 것이 동일성 유지권의 침해인가 하는 문제가 있을 수 있

다. 디지털 저작물은 작성 당시에 디지털로 저작하지 않은 이상 아날로그 형태를 디지털 형태로 만들 경우 그 동일성에 있어서 차이가 날 수 있다. 그 차이는 저작권법상의 저작인격권인 동일성 유지권을 침해할 여지가 있다. 왜냐하면 완벽한 변환이 일어난다 하더라도 어문저작물을 아스키코드나 한글파일로 만들 경우에 원래의 속성을 잃어버릴 수 있으므로 여기에 대한 저작자의 침해 논의가 있을 수 있다. 여기에 입법적으로 디지털화권의 저작인접권 인정과 함께 그 동일성유지권 문제를 배제하는 규정을 두는 것이 바람직하다.

VIII. 결론 및 제언

대학도서관이 이제 이용자인 교수와 학생들의 정보요구를 충족시켜주기 위해서는 각 대학도서관직원들을 디지털화하여 全文 데이터베이스를 만들고 전자도서관을 기초로한 가상도서관을 구축 운영하여 온라인 서비스를 하는 범 국가적인 종합학술 정보 시스템을 구축하여야 할 시대적인 요청을 부여받고 있다. 그러나 대학도서관 디지털라이브러리는 그 개발설계의 성격상 문헌정보학 분야만의 과제가 아니라 전자공학, 컴퓨터공학, 정보통신공학 등 인접 관련 학문 분야와의 긴밀한 연계성과 상호운용성이 요구되는 정밀하고 복잡한 연구과제라는 결론을 얻게 되었다.

따라서 전자도서관과 같은 미래 초고속 정보통신 응용서비스의 구축을 위해서는 여러 가지 기반 기술 개발이 선행되어야한다. 이러한 전자디지털기술에는 정보의 효과적 분산처리 와 그 제어를 위한 분산기술의 개발, 전자정보의 효율적인 저장, 유연한 표현을 위한 매체기술의 개발 및 정보의 다양한 특성을 보장하는 실시간 전송기술 등의 개발 과제들이 있다. 이들은 지금까지 컴퓨터 과학 및 엔지니어링이 해결하였던 어떠한 문제들 보다 개발기술 과제가 복잡하며 상당히 난해한 문제들이 포함되어 있다.

그리고 본 논문에서는 현재 메타데이터의 요소를 규정한 다수의 국내의 표준을 분석비교하여, 미래지향적인 디지털 도서관의 설계 및 구현에 지표가 될 수 있는 표준화 프레임워크를 제시하였다. 이 프레임워크는 기존의 도서목록의 디지털화에 표준인 MARC와 정보검색 프로토콜로서는 Z39.50, 그리고 네트워크 상의 데이터 호환성 향상을 위한 메타데이터 요소로는 더블린 코어를 기반으로 하고 있다. 또한 이들 표준들의 전자문서 표준형식은 SGML 구문으로 하고 그다음에 확장성이 뛰어난 XML을 채택하는 것이 바람직하다는 것을 제시하였다. 또한 위 논고에서 국내외의 주요 전자도서관 개발 현황을 집중적으로 살펴보았다. 미국의 주요 대학도서관들에서 추진하고 있는 DLI(Digital Library Initiative)나 일본정보처리개발센터(JIPDEC)에

서 추진하고 있는 차세대 전자도서관시스템 연구개발 프로젝트 등은 전산전공자들 중심으로 추진되는 프로젝트로서 전자도서관구축에 필요한 기반기술 연구가 정부 주도로 추진되면서 회원도서관 컨소시움을 통하여 온라인 분담목록 시스템, 상호대차시스템 및 슈퍼컴퓨터를 통하여 중앙집중식 원문 정보서비스 시스템을 운영하고 있다. 국내 대학도서관에 있어서 전자도서관 구축 수준은 아직은 초보적인 수준에서 이론과 기술이 머무르고 있는 상태이다. 특히 콘텐츠면에서 아주 빈약한 실정인 것을 알 수 있다. 그러나 최근 정부의 21세기 미래 정보화에 관한 집중적인 정책개발이나 예산 배정 등 적극적으로 전자도서관 개발에 투자하고 있어서 디지털 정보자원 기반구축을 크게 앞당길 수 있을 것으로 예상된다.

끝으로 기존의 대학도서관이 문헌정보학을 전공한 사서중심이었다면 전자도서관은 시스템전문가, 정보처리전문가 등 새로운 분야의 전문인력을 필요로 하고 있기 때문에 현재의 사서직에 대한 재교육은 물론, 앞으로 문헌정보학과 의 교과과정에 미래도서관 영역을 과감히 반영하고 또한 정보화사회의 급속한 발전에 부응할 수 있는 도서관 정보관리 및 서비스에 대한 법률적 뒷받침이 시급하다는 사실을 제의하고자 한다.

참 고 문 헌

- 장동춘. 1967. 대학도서관 참고봉사의 중앙집중제와 분산제에 관한 연구. 석사학위논문. 연세대학교 교육대학원. pp.5-13.
- 정영미. 오정선. 1998. "Z39.50 확장서비스를 이용한 도서관 상호대차 요청시스템의 구현." 한국문헌정보학회지 32(1) : 7-26.
- 김용훈 et al. 1998. "SGML 문서 관리시스템의 설계 및 구현." 한국문헌정보학회지 32(3) : 157-178.
- 이창열. 1998. "디지털 라이브러리 모형에 관한 연구." 한국문헌정보학회지 32(4) : 117-140.
- 김석영. 1998. "미국 도서관정보서비스의 현황과 동향." 한국문헌정보학회지 32(1) : 27-44.
- 정희경. 현득창. 이수연. 1997. SGML 가이드. 사이버출판사.
- 문헌정보처리연구회. 1998. 메타데이터의 형식과 구조. Dublin Core Metadata Elements Set, 문헌정보처리연구회 시리즈 7, pp.42-76.
- Kunze, John A. 1995. Basic Z39.50 server concepts and creation : Z39.50 mplementation experiences. NIST Special Publication 500-229.
- Francois, P. 1996. Generalized SGML repositories : requirments and modeling." Computer

Standards & Interfaces.

Maler, Eve and Andaloussi, Jeanne Ei. 1996. Developing SGML DTDs. Prentice-Hall.

Myaeng, S. H. et al. 1998. A flexible model for retrieval of SGML documents, in Proc. of the 21st ACM SIGIR Conference, Melbourne, Australia, August.

Weibel, Stuart. 1995. "Metadata : the foundation of resource description." D-Lib Magazine, July.

Won-Tae Choi. 1998. "A digital library prototype for access to diverse collections." 한국문헌정보학회지 32(2) : 295-313.