

전자도서관(Digital Library)의 개념과 발전추세

최 석 두

〈이화여자대학교 문헌정보학과 교수〉

目 次

1. 서 론
2. 디지털도서관의 연구개발 동향
3. 디지털도서관의 특성
4. 디지털도서관의 전망
5. 디지털도서관에 대한 우려
6. 결 론

1. 서 론

1990년대에 들어서면서 선진 제국은 국가경쟁력 강화라는 목표하에 새로운 사회기반구조로서 초고속정보통신망을 건설하는 데 국력을 경주하고 있다. 미국은 클린턴 행정부에 의해 추진되고 있는 국가정보기반구조(NII) 프로젝트에서 2005년까지 가정에 광케이블을 연결한다는 계획으로 정보고속도로를 구축하고 있으며, 아울러 전세계를 연결할 수 있는 세계정보기반구조(GII)를 구축하고 있다. 일본도 이와 같은 사업을 2010년까지로 완료한다는 계획을 세우고 진행하고 있다. 우리나라도 1994년 5월 정보통신부를 주축으로 2015년까지 전국의 가정과 기

업에 멀티미디어 정보를 전송하는 초고속 대용량의 정보통신망을 구축하기 위한 ‘21세기를 대비한 초고속 정보통신망 구축 종합계획’을 확정, 발표하였다. 개발에 중점을 두고 있는 응용서비스로 교육관련서비스, 의료서비스, 화상회의서비스, 문화행사서비스, 그리고 디지털도서관서비스 등이 있다.

디지털도서관서비스는 현재 도서관을 중심으로 보유하고 있는 모든 문헌에 대한 데이터베이스를 구축하고 통신망을 통해 하나로 연결함으로써 국가적인 데이터베이스를 구축하여 일반 이용자에게 다양한 정보를 제공한다는 것이다. 각국이 여기에 국력을 기울이는 것은 ‘21세기 국가 경쟁력은 정보력을 통해서만 확보할 수 있

다'는 사실에 기인한다.

이와 같은 추세는 컴퓨터를 이용한 데이터처리기술의 발전으로 거의 대부분의 문헌(텍스트, 영화상, 소리 등)을 디지털화할 수 있다는 결론에 있다. 국내외를 막론하고 많은 기관들이 디지털도서관을 구축하고 있으며 구축계획을 세우고 있다. 본고에서는 디지털도서관의 개념을 정의하고 특성을 규명하며 앞으로의 발전추세에 대하여 논하고자 한다.

2. 디지털도서관의 연구개발 동향

2.1 우리나라

국내의 디지털도서관 구축사례는 정보기술의 부족으로 초보적인 단계에 있다고 할 수 있다. 한성대학교는 전자정보관이라는 미래형 도서관을 96년 5월 10일 개관하였다. 데이터베이스 구축계획은 한성대학교가 소장하고 있는 주요 외국학술지에 대한 이미지데이터베이스 구축, 대학 논문집, 연구소논문집, 한성50년사 등의 CD-ROM제작, 한성대학보의 데이터베이스화 등이며, 추후 ASCII텍스트검색이 가능토록 할 계획이다. 또한 교육의 질적 향상을 위한 VOD 데이터를 제작할 계획이다.

LG상남도서관은 Journal류 1000종과 Proceedings자료 150-200건 등 약 20만건의 이미지데이터를 기반으로 하고 있다. 클라이언트/서버시스템이며, 네트워크 프로토콜은 TCP/IP를 이용하고 있다. 도서관건립에 50여억원을 투자하고 연간 13여억원을 투자하여 운영하고 있다.

이외에 디지털도서관과 관련하여 추진되고 있

는 계획을 요약하여 보면 다음과 같다.

- 연세대 한상완 교수팀이 국립중앙도서관의 수탁과제로 한국형 디지털도서관 구축을 위한 종합설계과제 수행 완료
- KAIST 최기선 교수팀이 학위논문과 연구보고서의 全文 및 이미지 처리기술이 추가된 전자도서관의 모델 연구 완료
- IBM Digital Library Package가 국내의 디지털도서관을 위한 솔루션으로 소개
- 숙명여대 김성혁 교수가 정보통신부의 초고속응용사업으로 학술저널 디지털도서관 구축을 위해 Illustra와 SGML을 이용한 1차년도 모델 개발 완료
- 충남대 맹성현 교수가 정보통신부의 초고속응용사업으로 자연어검색, Z39.50, 가중치에 의한 문서랭킹, 정보검색 엔진과 ORDB의 복합운영 등을 토대로 한 디지털도서관 모델 개발
- 한국방송통신대가 고려대 컴퓨터교육과와 공동으로 VOD를 이용한 원격교육 및 정보검색시스템의 모델 개발
- LG정보통신이 서울대 사회과학대학 도서관에 디지털도서관 공동개발 제안
- 한국과학재단이 전자도서관의 정책적 방향 연구 수행중
- 정보통신부에서 시범기관을 선정하여 디지털도서관의 모델을 구축하고자 계획수립. 디지털도서관 시범사업을 주요사업으로 지정하였으며, 대상기관은 국립중앙도서관, 국회도서관, 서울대도서관, KAIST과학도서관임.

- 2차년도가 진행중인 초고속 시범사업에 전자박물관, 전자미술관, 전자문화재정보, 대법원 판례전문정보 등이 참여하고 있음.

2.2 미국

미국은 많은 기관들이 디지털도서관을 계획하고 진행중에 있다. 그 중 미국국회도서관과 DLI(Digital Library Initiative)의 두 가지의 프로젝트를 간단히 소개하여 무엇을 하려고 하는지, 디지털도서관을 구축하기 위하여 어떤 기술이 필요한지를 일별하고자 한다.

미국국회도서관은 1994년 10월부터 1997년까지 국가디지털도서관 프로그램의 첫 단계를 수행하고 있다. 2000년까지 전 국민의 공통관심사이며, 국회도서관이 소장하고 있는 특수한 자료 500만건의 자료를 페이지화상으로 축적하고 초고속통신망을 통하여 전국에 서비스함을 목표로 하고 있다. 재원은 2000년까지 총 6천만달러(약480억원)를 사용할 계획이다.

National Science Foundation(NSF)/Advanced Research Projects Agency(ARPA)/National Aeronautics and Space Administration(NASA)가 공동으로 제공한 2,440만달러(약 195억원)로 1994년 가을부터 4년간의 프로젝트 Digital Library Initiative(URL 1)를 시작하였다. 이 프로젝트의 가장 기본적인 목적은 국가정보하부구조가 될 기초과학과 기술과의 연계를 구축하기 위한 것이다. Carnegie Mellon Univ.의 'Informedia(tm) Digital Video Library Project'(URL 2), Univ. of California Berkeley의 'An Electronic Environmental Li-

brary Project'(URL 3), Univ. of Michigan의 'Univ. of Michigan Digital Library Project'(URL 4), Univ. of California Santa Barbara의 'Alexandria Digital Library Project'(URL 5), Stanford Univ.의 'Stanford Digital Library Project'(URL 6), Univ. of Illinois Urbana-Champaign의 'Interspace'(URL 7)의 6개 대학이 참여하고 있다.

각 참여 대학은 연구분야에 따라 시범시스템을 구축하고 개발기술의 공유 및 통합을 목적으로 한다. 연구분야는 다음과 같다(Chen, 1995).

- 모든 가능한 유형정보 획득 : OCR서식인식, 음성인식, 오디오분할/분석, 파일변환, 방송캡춰/디지털화화, 그래픽인식(이미지, 드로잉, 그래프 등)
- 다양한 형식으로 수록된 전자정보의 범주화 및 조직화 : 색인/정보내용 해석/분류/목록, 문맥과 인용에 근거한 다언어 색인법, 하이퍼미디어구조 및 문헌 연결, 지식표현을 위한 그래픽인터페이스, 대규모 지식공간 브라우징 기술 등
- 정보탐색/필터링/요약소프트웨어 : 데이터 모델 및 검색이론, 평가방법, 문헌/텍스트/질의언어에 대한 공식적 구조, 지능형 텍스트처리와 문헌관리, 특성기반 이미지분석 및 분류/패턴인식, 다중패스검색, 수정가능한 색인시스템, 비디오의 공간-시간 특성 색인, 필터링/라우팅/선택적 배포, 클러스터링/요약/초록작성, 자연언어처리/생성시스템, 어휘/시소러스/개념공간 생성, 학습시스템, 문맥기반 패턴매칭/검색, 부정확

데이터에 적용할 강력한 매칭기법, 신속한 탐색과 질의최적화

- 이미지데이터의 신속한 브라우징을 위한 가시화 및 상호작용 기술 : 회화적 요소 인식 및 이미지 분류, 인간의 인지구조, 다양한 규모의 디스플레이 및 줌잉, 데이터가시화, 상호작용 가시화 제어, 시물레이션, 네비게이션, 하이퍼미디어, 가상현실에서 사용가능한 메타포 개발, 대규모 이미지데이터 조작시 빠른 수행성 보장기법 등
- 네트워크 프로토콜과 표준화 연구 : 네트워크 보안, 프로토콜 디자인, 데이터압축, 이용자 증가에 따른 정보처리요구 증가 대처 방안
- 분산 전세계 정보보존소 이용의 편의화 : 중개자/지능형관리자, 분산 이질 객체지향데이터와 정보시스템 통합, 사적 뉴스/잡지/저널 서비스, 장애인 정보서비스, 검증/인증, 모델링/시물레이션, 멀티미디어 정보교환 및 멀티유저환경에서의 정보처리기법
- 디지털도서관의 개인/그룹행위, 사회 경제적 이슈 : 지적소유권, 사생활 침해/보안, 학문수행상 디지털도서관의 영향, 디지털환경 하의 출판, 저작권 소유 문헌의 사용요금 지불장치 등.

2.3 일본

학술정보센터의 'NACSIS-ELS'는 학술논문을 대상으로 하는 document delivery system이다. 독자적으로 개발한 브라우저를 이용하고 있다. 잡지를 모두 페이지이미지로 축적하고 있는

며, 네트워크의 고속화가 중심 연구과제이다. 국립국회도서관의 '디지털도서관 프로젝트'는 역사적으로 귀중한 일차자료를 전자화하는 것이 주목적이다. 현재 실험시스템을 개발하고 있으며 브라우저는 Mosaic과 Netscape를 사용하고 있다. 京都대학의 'Ariadne'는 목차정보를 이용하여 계층적 검색, 사용자 지원을 위한 기계번역, 자동낭독 등을 특징으로 하고 있다. Mosaic을 확장한 브라우저를 사용하고 있다. 이외에도 100여개의 국립대학이 정부의 지원하에 디지털도서관 모형을 구축하고 있다.

2.4 유럽

ELSA(Electronic Library SGML Applications) 프로젝트(URL 8)가 있다. 이 프로젝트는 프랑스의 Jouve사, 영국의 De Montfort Univ., 홀랜드의 Elsevier Science사의 공동프로젝트이다. SGML문서와 HTML문서를 대상으로 하고 있으며, 요소기술의 개발과 사용자의 요구분석을 중심과제로 하고 있다. De Montfort Univ.는 디지털도서관의 연구개발에 적극적으로 참가하고 있으며, ELINOR라는 별도의 프로젝트를 진행하고 있다. ELINOR(Electronic Library and Information Online Retrieval) (URL 9)는 1992년부터 대영도서관과 IBM의 지원을 받아 시작된 프로젝트이다. business information course에서 사용되는 텍스트나 잡지를 대상으로 하고 있다. 소프트웨어는 주로 시판용을 사용하고 있으며 OCR출력의 全文텍스트를 대상으로 모호검색이 가능하다. 이외에도 많은 디지털도서관이 구축 중에 있다.

3. 디지털도서관의 특성

미래의 도서관을 지칭하는 용어로, logical library, electronic library, virtual library, paperless library, electronic digital library, digital library, virtual reality library, virtual digital library, networked library, multimedia digital library, library without walls; borderless library; wall-less library 등을 사용하고 있다. virtual은 시각과 공간을, digital은 디지털데이터를, network은 통신을, multimedia는 매체를 강조하고 있으나 이들은 모두 동의어로 취급할 수 있을 것이다. 최근에는 디지털도서관(digital library)이라는 용어가 주로 사용되고 있다. 디지털도서관과 대응되는 전통적인 도서관을 지칭하는 용어로, library 이외에 classical library, conventional library, current library, non-electronic library, paper library, physical library, present library, print-oriented library, real library, traditional library, visiting library 등을 사용하고 있으며, 디지털도서관의 사서를 지칭하는 용어로 digital librarian, networking librarian, network specialist, on-line reference librarian, virtual librarian 등을 사용하고 있는 것을 보면 디지털도서관의 의미를 엿볼 수 있을 뿐만 아니라 디지털도서관에 거는 기대를 느낄 수 있다.

현 도서관과 디지털도서관의 차이를, 1) 시설 중심 대 서비스 중심, 2) 인쇄매체 중심 대 전자매체 중심, 3) 정보의 소재과약 대 지식의 직접 전달, 4) 자료의 소유 대 서비스의 공유, 5)

정보생산과 소비의 분산화 대 정보생산과 소비의 일원화, 6) 도서관의 공유화 대 도서관의 개인화 등으로 비교하기도 하며, 디지털도서관의 특성을, 1) 정보를 디지털형식으로 축적한다는 점, 2) 정보의 입수시 디지털통신기술에 의존한다는 점, 3) 자료의 원본은 하나이며 하나의 원본을 이용한다는 것으로 논하고 있다. 따라서 디지털도서관이란 컴퓨터와 통신네트워크를 통하여 멀티미디어형태로 수용된 분산 지식 창고에 개인이 편리하게 이용할 수 있도록 지원하는 새로운 정보활동 지원시스템을 지칭하는 것이다.

그러나 디지털도서관의 가장 근본적인 특성은 ‘全文데이터를 처리의 대상으로 한다’는 데 있다. 상기 여러 가지 차이는 모두 이 특성에서 기인하는 것이며, 현 도서관이 컴퓨터의 도움을 받는 도서관이라 한다면 디지털도서관은 컴퓨터가 중심이 되는 도서관이라는 점을 강하게 내포하고 있다.

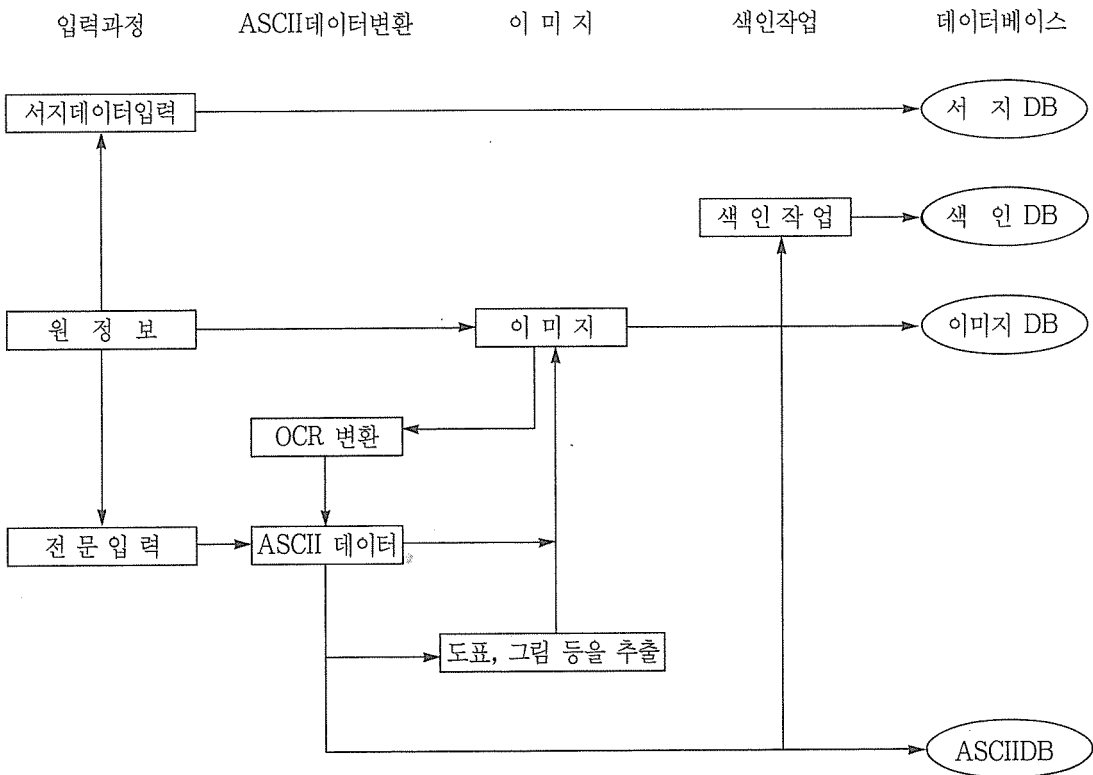
3.1 축적 대상자료 측면

데이터는 단행본, 학술논문, 기술보고서, 교과서, 고서, 繪畫, 도면, 지도, 비디오, 오디오 등으로 지금까지 도서관에서 수집, 보관, 서비스의 대상이 되어 왔던 것과 거의 일치한다. 그러나 디지털도서관에서는 서지정보 뿐만 아니라 지금까지 도서관에서 축적대상이 되지 못했던 영·화상정보를 포함하는全文데이터가 대상이 되며, 디지털도서관은 전문데이터베이스시스템이 중심이 된다. 이에 따라 자연스럽게 복합적인 멀티미디어정보가 디지털도서관의 대상정보로 들어오게 된다.

디지털도서관의 형을 생각할 때는 대상이 되는 분야와 이용자에 따라 그 구성이 크게 달라진다. 우선, 텍스트 중심 데이터와 영화상 중심 데이터로 나눌 수 있다. 전자는 일반적으로 도서, 논문 등을 대상으로 하며, 후자는 사진, 지도 등과 같은 정지화상과 소리데이터, 비디오이미지 혹은 애니메이션과 같은 동화상을 대상으로 한다. 全文데이터베이스는 ASCII텍스트모형, SGML텍스트모형, 페이지이미지모형, 통합모형 등 어떤 모형을 사용해도 좋으며, 全文데이터베이스를 구축하는 처리과정을 유형측면에서 단순화시켜 보면 그림 1과 같다.

우선, 원정보에 대하여 입력과정이 있게 된다. 서지데이터는 키보드에서 입력, 서지유틸리티에서 다운로드 그리고 기관에 따라서는 모기관 도서관에 서지데이터가 작성되어 있을 수도 있다. 어느 쪽이나 색인과정을 거쳐 색인데이터베이스를 만들게 된다. 全文이 ASCII파일로 되어 있지 않을 때는 모두를 입력하게 되며, 1) 키보드에서 입력할 수도 있고, 2) 이미지를 OCR로 변환할 수도 있다. 원정보를 이미지로 만들었을때, 1) 그대로 이미지데이터베이스만 만들 수도 있고, 2) OCR로 변환하여 ASCII데이터도 만들면서 이미지데이터베이스를 만들 수도 있다. 처

〈그림 1〉 全文데이터의 처리 과정



음부터 입력된 ASCII데이터는, 1) 이미지로 다운시켜 이미지데이터베이스를 만들 수도 있고, 2) 도표, 그림 등만 추출하여 이미지데이터베이스를 만들고 나머지 ASCII데이터는 ASCII데이터베이스를 만들 수도 있으며, 3) 이미지로 다운시켜 이미지데이터베이스를 만들고, 도표, 그림 등은 무시하고, 나머지 ASCII데이터만 ASCII데이터베이스를 만들 수도 있다. 서지데이터와 ASCII데이터는 색인과정을 거쳐 색인데이터베이스를 만든다. 원정보가 동화상, 정지화상 혹은 소리데이터인 경우에는 이를 디지털화하여 이미지데이터를 만든 후 색인과 이미지데이터베이스를 구축하게 된다.

처음부터 이미지를 생성하던 ASCII데이터를 이미지로 다운시키는 이미지데이터베이스는 가져야 할 것이며, 사용자에게 서비스할 때는 이미지데이터베이스를 주로 하여야 할 것이다. 왜냐하면 ASCII데이터를 표준화하여 서비스하는 것은 완전하지 않기 때문이다. 다만 ASCII데이터는 색인용으로 사용할 수 있으며 추후 새로운 처리에 대응할 수 있기 때문에 이미지데이터베이스와 병행하여 ASCII데이터베이스도 갖는 것이 바람직 할 것이다. 다만 처음부터 ASCII데이터만을 갖는다면 색인이 거의 불가능하므로 검색에 제한이 따를 것이다.

3.2 색인 및 검색방법 측면

全文을 대상으로 하기 때문에 색인과 검색방법이 다를 수 밖에 없다. 현재의 도서관에서는 목록을 대상이 되는 서지정보 혹은 키워드검색이 주였다. 따라서 모든 대상 데이터를 동일한

관점에 놓고 처리한다. 각국의 MARC형식을 보면 명확히 알 수 있다.

디지털도서관에서는 全文을 대상으로 하는 색인 및 검색의 연구가 주를 이루고 있다. 따라서 디지털도서관에서는 모든 대상 데이터를 색인의 관점에 보아야 한다. 색인은 검색요구를 반영하는 것이다. 그러나 실상황에서의 검색요구만을 수용하는 것은 문제가 있다. 왜냐하면 지금까지 특정한 정보요구가 없었던 것은 대부분 그 정보요구를 충족할 수 없다는 것을 이미 알고 있었기 때문이다. 全文이 대상이 되었을 때의 잠재정보요구를 파악하지 않으면 이미 처리된 데이터를 차후 다시 조정해야 하는 일이 생길 수 있다. 즉, 현 도서관은 메타데이터를 중심으로 하는 거시적인 색인 및 검색기능이 강하며, 디지털도서관은 全文데이터를 중심으로 하는 미시적인 색인 및 검색기능이 강하다고 보아야 한다.

문헌을 영화상 중심 데이터를 구별하여 색인 특성을 보고자 한다. 다만 텍스트 중심 데이터에 영화상 데이터가 일부 포함되고, 영화상 중심 데이터에도 텍스트 데이터가 일부 포함될 수 밖에 없다.

3.2.1 텍스트 중심 데이터

지금까지 도서관에서 처리하고 있는 내용을 보면 비주제색인(청구기호, 서명, 저자명 등)이 주이며 일부 키워드나 주제명표목에 의한 색인이 전부였었다. 그러나 全文이 대상이 되면 단행본, 학위논문, 학술지논문, 연구보고서, 신문기사, 특히, 표준, 컴퓨터파일, 문서 등에 따라 색인방법이 달라져야 한다.

예를 들면, 초록, 목차, 본문, 각주, 인용문헌, 권말색인 등의 색인과 함께 본문의 장, 절, 페이지, 문단, 문장 등의 구분에 의한 색인이 필요하게 된다. 뿐만 아니라 본문에 나오는 표와 그림의 처리도 문제가 된다(후술하는 SGML에 따라 기술하면 이와 같은 각 요소가 확연히 구별된다). 이와 같은 관점에서 색인이 가능하면 다음과 같은 사용자의 질문도 처리할 수 있다.

초록 혹은 결론 부분에서 ...에 관해 논한 문헌을 보고싶다

목차나 권말색인에 해당하는 페이지를 보고싶다

단어 A와 B가 같은 문단에 출현하는 문헌을 보고싶다

비무장지대에 관한 그림이나 사진을 보고싶다

동일한 단행본이어도 주제나 목표에 따라 색인의 방법은 달라질 것이다. 성서, 문학작품, 전문서, 교과서, 아동도서, 화집, 시집, 실용서, 연감, 고서, 지도, 각종 辭典, 백과사전, 학위논문, 기술보고서, 인적정보, 특허, 신문기사, 학술지 논문, 표준 및 규격 등에 따라 색인의 방법은 달라지게 된다. 성서인 경우에는 성서 자체의 문자열 뿐만 아니라 문자열이 갖는 의미상의 관계까지 요구할 것이며, 관련되는 미술작품, 예술작품, 사실, 유적 등에 대한 이미지데이터도 요구하게 될 것이다. 문학작품도 역사소설, 전기, 아동작품, 픽션, 논픽션 등에 따라 문자와 이미지 정보가 필요할 것이다.

전문서가 되면 그림, 도표, 수식, 용어 등에 관련된 정보가 중요하게 될 것이다. 화집이 되

면 그림 하나하나에 대한 작품관련 인물, 배경, 재료, 시대, 대상, 주제 등에 따라 다양한 정보가 필요할 것이다. 시집이 되면 다른 문학작품과 다른 시어, 비유 등과 같은 새로운 정보요구가 있을 것이다. 국어사전이나 전문용어사전이 되면 통합, 공유, 그림, 용어관계, 어의정보, 문맥정보, 용례정보 등의 정보가 필요하게 될 것이다.

인적정보를 보면, 개인에 관한 사항(사진, 이름, 생몰년, 시대, 특기, 전공, 학력, 경력, 저작, 상벌 등), 다른 사람과의 관계(부모를 포함하는 각종 가족관계, 친구, 스승, 단체 등), 업적 등이 포함되는 매우 복잡한 내용을 가지게 된다. 개인사진인 경우, 나이, 장소, 표정(웃는 얼굴, 근심하는 얼굴 등), 사진의 종류(흑백, 컬러, 만화 등)에 따라 선택의 범위와 수는 다양할 것이다. 이름만 하더라도 한자명, 영문명, 자, 호, 시호, 휘호, 아명, 속명, 법명, 세례명, 별명 등이 있으며, 학력, 경력, 상벌, 단체, 저작, 업적 등에는 연대가 필요하게 되고, 저작에는 시, 곡명, 소설, 서예작품 등의 구별이 포함되어야 한다. 이들은 관련 있는 것은 모두 서로 링크되어 있어야 한다. 심지어는 교수명으로 찾는 경우, 상기 정보 이외에 현재 재직하고 있는 학교, 학과, 담당과목 뿐만 아니라 담당과목과 링크되어 있는 강의계획표까지 보고싶어 할 것이다.

3.2.2 영화상 중심 데이터

지금까지의 도서관에서는 영화상이 주인 全文도 일반도서의 처리방법과 거의 유사한 방법으로 처리해왔다. 그러나 문자가 주인 全文과 달

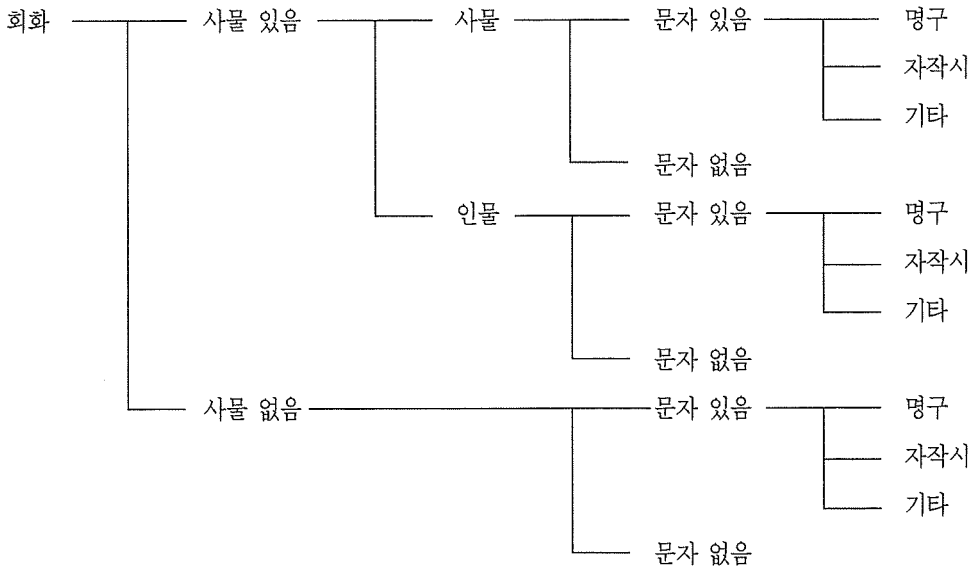
리 영화상이 추가 되면 색인의 방법은 전혀 달라진다. 동화상이 대상이 되는 경우를 보면, 일반영화, 전기, 드라마, 기록영화 등에 따라 색인이 달라진다. 일반영화인 경우에는 줄거리, 캐스트, 스태프 등의 문자데이터를 이용하여 검색하는 것은 당연하나 특정 장면을 검색하여 그 장면부터 영화를 보는 경우는 드물 것이다. 그러나 새로운 드라마를 만드는 사람이 어떤 장면에 특정 분위기의 배경영상을 넣고 싶으나 시기적, 시간적으로 그 장면을 촬영할 방법이 없을 때, 기존의 비디오에서 그런 장면을 찾아야 한다. 이를 위하여 모든 특징적인 장면이 색인되어 있어야 한다. 또한 동식물의 생태관찰영화나 강의실의 강의비디오는 특정 내용부터 보려할 것이다. 동식물의 생태에 관한 것은 화상속에 색인의 대상이 있지만 강의실의 강의비디오는 동일한 장면이 계속된다. 따라서 강의실 비디오의 색인 대

상은 강의내용, 즉 소리데이터가 되어야 할 것이다.

화상정보가 추가 되는 대표적인 예로 전자미술관(디지털미술도서관)이 있다. 지금까지 미술작품을 볼 때는 繪畵, 건축, 조각, 서예, 사진, 문화재, 설치미술 등의 관점에서 범주화하고 이를 처리하려고 한다. 그러나 전자미술관을 생각할 때는 전술한 바와 같이 색인의 관점에서 보아야 한다. 회화작품인 경우에는 동·서양화로 볼 것이 아니라 그림 속에 색인의 대상이 되는 사물이 있는가 없는가로 구분하기 시작하여야 한다. 예를 들면 그림 2와 같다.

사물이 있으면 사물명(산, 나무, 사과 등)이 색인어로 사용될 수 있으며, 인물이 있으면 개인명이 필요할 경우가 있다. 문자가 있는 경우에 한글인지 한자인지, 그리고 그 字體나 내용

〈그림 2〉 색인측면에서 본 회화



에 대한 색인여부를 결정할 수 있다. 각 경우에 대하여 사인이나 낙관이 있는지 없는지가 다음으로 문제가 된다. 심지어는 화상을 대상으로 '기분이 밝게 되는 그림이 보고싶다'와 같이 감성에 기본을 둔 감성검색도 연구대상이 되고 있다.

건축에는 현대건축과 문화재로서의 건축이 있으며 단일건물로 이루어진 건축과 복수의 건물로 이루어진 건축으로 나눌 수 있다. 어느 것이나 수 많은 색인측면이 생기게 된다. 조각은 개별적인 특성을 갖는 것과 일반적인 특성을 갖는 것으로 나눌 수 있다. 인물에 관한 조각이라면 예를 들어, 남자, 여자, 가족 등과 같이 나눌 수도 있으며, 특정 인물을 표현한 것, 추상화된 인물을 표현한 것 그리고 일반적인 특성을 갖는 것으로 나눌 수 있다. 장승, 불상 등은 일반적인 특성을 갖는 것이지만 다시 좌불, 입불, 와불 및 조각방식에 따라 나눌 수 있다. 또한 불상에는 석가모니불만 있는 것은 아니다. 불상에 대한 정보요구가

옆모습을 보고싶다

좌불중 손의 위치가 특이한 것을 보고싶다

뒷모습을 45도 각도로 위에서 보고싶다

등이 있다면 불상이라 하더라도 정지화상으로서 불가능하며 결국 동화상이어야 하고 모든 프레임에 대하여 색인이 되어야 한다는 결론이 나오게 된다.

서예작품이 되면 字體, 내용 등이 중심이 되므로 글자 한자한자에 대한 색인이 필요하게 된다. 사진인 경우에도 인물이 있는 경우, 없는 경

우로 나뉜다. 인물이 있는 경우라면 조각과 같이 특정 인물을 표현한 것, 추상화된 인물을 표현한 것 그리고 일반적인 특성을 갖는 것으로 나눌 수 있다. 개인명이 필요한 경우와 필요없는 경우가 생기기 때문이다. 문화재인 경우의 예로 사찰 중 해인사를 보면, 이상에 논한 대부분이 포함될 뿐만 아니라 건물(각종 암자 포함), 작품, 유적, 유품 등이 복합적으로 복잡한 관계를 갖는다는 것을 알 수 있다. 해인사의 전경에 대하여 계절별로 보고싶은 정보요구도 생각할 수 있다. 특히 설치미술인 경우에는 작품의 특성상 일정기간이 지나면 없어지는 경향이 있다. 이를 몇장의 사진으로 보존한다는 것은 불가능한 일이다.

3.3 열람방법 측면

지금까지 디지털도서관이라 할 때 대규모 디지털정보에 색인과 링크를 붙임으로써 부가가치를 높여 모아둔 집적장소라는 느낌이 강하다. 그러나 이외에도 디지털도서관의 기능은 다양하다.

첫째, 네트워크액세스용 터미널이나 네트워크 접속기기 등의 하드웨어와 소프트웨어, 이들을 둘 시설, 유지보수 등이 준비될 것이다. 브라우저는 사용자와 친근한 지적 인터페이스를 중심으로 독자적으로 개발하는 방법과 Mosaic, Netscape, Explorer 등 기존의 브라우저를 이용하는 방법으로 나눌 수 있다. 최근에는 Sun Microsystems의 HotJava(URL 10)라 불리는 브라우저에 대한 기대가 높아가고 있다. 이와 같은 브라우저들은 점점 편리하고 다양한 기능을 갖게 될 것이며, 새로운 시스템들이 출현하게

될 것이다.

둘째, 사용자교육, 이용방법, 네트워크상황의 안내, 훈련용 소프트웨어, 이용자가 직면하는 여러 가지 문제 해결 등의 지원체제가 갖추어질 것이다. 지금까지는 신체장애자에 대한 특별한 자료봉사 이외에는 사용자가 정해진 근무시간 내에 도서관에 가야 한다. 그러나 디지털도서관에서는 거대한 건물도 없으며 시간과 공간을 초월하여 언제든지 이용할 수 있다. 시간을 절약할 수 있으며, 교통문제가 해결되며, 딸린 어린 아이를 걱정할 필요도 없다. 외출이나 출장시 포터블 무선컴퓨터만 있으면 어디서나 이용할 수 있다. 이용자가 다양화되고 지리적으로 분산하게 되면, 특히 시차가 있는 외국에서의 액세스가 증가하면 디지털도서관은 24시간 개방해야 할 것이다. 이러한 상황에서의 이용자 지원체제는 네트워크를 통한 지원이 될 것이다.

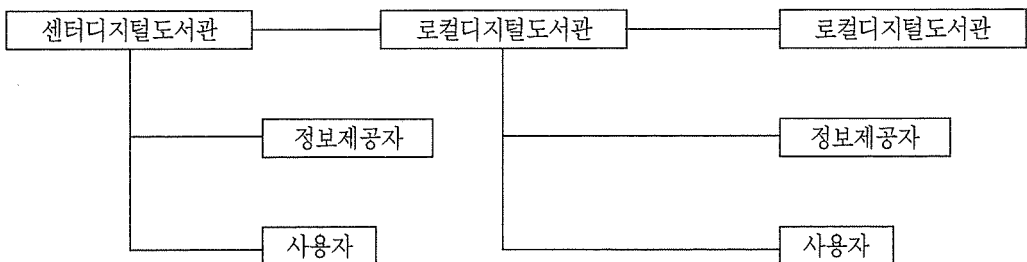
셋째, 네트워크사회에서 각 정보원에 대한 사용을 개인이 계약하고 대금을 지불한다는 것은 곤란한 일이다. 회사, 학교, 연구소, 공공단체가 운영하는 디지털도서관에서는 이를 사용자 입장에서 해결하게 될 것이다. 가장 중요한 것은 저작권의 보호와 개인정보의 보호문제이다. 지금

까지 정당하게 배포된 출판물은 최초판매이론(first sale doctrine)에 따라 저작권자의 허락없이도 2차, 3차 배포가 가능하며, 동일 저작을 몇 번 읽어도 문제가 발생하지 않는다. 그러나 디지털도서관에서는 짧은 시간내에 대량의 복제가 가능하므로 이를 철저하게 제어할 필요가 있게 된다. 물론 출력정보 중 가공이나 재이용이 허락된 것에 대하여 이를 가공하고 재이용할 수 있는 처리환경을 만들어 주어야 한다.

저작권 보호를 위하여 많은 단위도서관들이 암호화(encryption), 디지털서명(digital signature), 전자문신(stegneography), 전자태깅(electronic tagging), 복사탐지서버(copy detection server), 헤더(header), 워터마크(water mark), 디지털지문(digital fingerprint) 등의 여러 가지 기법을 고려하고 있으나 보다 근본적인 해결방안이 필요하다. 즉, 그림 3과 같이 저작권의 문제가 발생하는 관계에 대하여 원본 하나를 어느 데이터베이스엔가 축적하고 이를 통제하고 보호할 수 있어야 한다.

실선으로 표시된 부분이 저작권의 문제가 발생하는 접점이 된다. 센터디지털도서관은 하나라는 개념이 아니라 여러 가지 디지털도서관 중

(그림 3) 저작권 문제의 발생 관계



공통점을 갖는 로컬디지털도서관을 그룹화하였을 때 그 중심이 되는 디지털도서관을 말한다. 정보제공자는 상업적으로 데이터베이스를 만들어 이를 센터 혹은 로컬도서관에 제공하는 개인 혹은 조직을 말한다. 사용자에는 기관, 단체, 개인이 포함된다. 이 저작권의 보호가 완벽해지면 도서관은 종이가 없는 방향으로 급속하게 변해갈 것이다. 개인정보의 보호 문제도 저작권 보호 문제와 동일한 내용을 갖는다.

3.4 기술방법 측면

3.4.1 SGML경향

지금까지 대부분의 도서관은 MARC형식을 중심으로 서지데이터를 기술하여 왔다. 그러나 쉐문을 중심으로 하는 디지털도서관의 데이터 기술은 MARC만으로는 기술이 불가능하여 SGML형식을 고려하게 되었다. 기술적 마크업언어인 SGML은 1970년에 GML로 처음 개발되어 1986년 10월에 국제표준(ISO 8879)으로 제정되었으며, 다른 마크업언어를 정의하는 틀 혹은 규칙이므로 하나의 메타언어로 취급되고 있다. SGML에서는 문헌에서 필요한 요소와 실제로 출현하는 요소에 대한 구조화 정보를 '문헌유형 정의'(DTD)를 이용하여 처리과정과 독립적으로 기술할 수 있으며, 이 기술의 정당성은 구문 분석기로 확인할 수 있다. 처리과정과 독립적이므로 DTD만 알면 SGML로 코드화된 데이터를 상이한 하드웨어와 소프트웨어상의 여러 가지 응용시스템에서 다양한 방법으로 이용할 수 있다.

데이터의 기술언어로서 MARC 대신에 이와 같은 장점을 가지고 있는 SGML을 사용하려는

경향은 정보처리기술이 발전하고 사용자의 정보 요구가 다양화됨과 아울러 이들의 정보요구에 부응하는 정보처리를 위해서는 MARC형식이 한계가 있기 때문이다. 많은 도서관과 연구프로젝트에서 MARC의 문제점을 해결하기 위하여 신규 및 기존데이터를 SGML형식으로 변환하기 위한 연구를 수행하고 있다. TEI프로젝트, 버클리의 Finding Aids프로젝트 및 컬럼비아대학의 DIAP프로젝트를 예로 든다.

텍스트코드화 프로젝트(The Text Encoding Initiative : TEI)(URL 11)는 전자텍스트의 다양성을 줄이고 기계처리를 단순화하며 정보자원을 쉽게 공유할 수 있는 전자텍스트의 작성과 교환지침을 만들기 위해 미국, 유럽, 캐나다 등의 정부기관, 학협회, 재단 등이 참가하는 국제적인 협력사업으로 1987년 11월에 Poughkeepsie의 Vassar대학에서 시작되었다. 이 회의에서 신규문헌과 기존문헌의 전자텍스트 코드화 및 교환지침을 위한 공통코드화계획에 합의하고 'Poughkeepsie원칙'을 발표하였으며, 1988년부터 연구를 시작하여 1990년 6월에 초안을 출간하였다. 즉시 개정작업에 착수하여 1992년 4월부터 1993년 11월에 걸쳐 지침의 개정안이 장별로 발표되었으며, 1994년 5월 지침의 공식판이 결정되었다. 전자화된 텍스트는 자연언어처리, 정보검색, 하이퍼텍스트, 전자출판, 다양한 문헌 및 사적 분석, 사전편찬 등 응용분야가 점점 넓어지고 있으며, 이와 같은 응용분야 뿐만 아니라 앞으로 생길 수 있는 어떤 목적에도 사용될 수 있도록 하는 것이 TEI프로젝트의 목표이다.

'Finding Aids for Archival Collections'(URL 12)는 'The Berkeley Finding Aid Project: BFAP'라는 이름으로 1993년 가을부터 시작한 연구이며, 보존문서관, 박물관, 도서관의 코드화표준을 개발하기 위한 공동프로젝트이다. Finding Aids란 관련 자료를 기술하고, 통제하며, 액세스에 사용되는 문헌을 말한다. 집서수준의 정보엑세스나 항해를 위한 계층구조에서 Finding Aids는 서지레코드와 1차자료의 중간에 위치하게 된다. 서지레코드는 Finding Aids로 안내하고, Finding Aids는 1차자료로 안내한다. 이 프로젝트의 목표는, 첫째, SGML DTD의 형식으로 Finding Aids의 전형적인 코드화 표준을 만드는 일이며, 둘째, Finding Aids의 전형적인 데이터베이스를 구축하는 일이다. SGML과 FINDAID DTD를 사용하던 것을 1995년 후반에는 Encoded Archival Description(EAD)으로 개칭하고 EAD DTD의 개발을 완료하고 시험을 거쳐 배포하고 있다.

Columbia University Digital Image Access Project (DIAP)(URL 13)는 컬럼비아대학도서관이 개발하고 있으며 디지털이미지에 대한 서지데이터를 축적하고 액세스하기 위한 새로운 모형이다. DIAP팀은 요약서지정보 뿐만 아니라 필요하다면 상세한 계층 및 版別 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. 이 새로

운 형식의 레코드를 위한 SGML Catalog Record(SCR)가 구체적으로 제안되었다. SCR은 보다 융통성있게 데이터요소 클러스터들을 통합함으로써 계층관계요소를 분할하고 판별정보를 분리하며 개별레코드로 만들어야 하는 현재의 AACR2/USMARC모형보다 복잡한 서지정보를 기술하고 표현하는 데 적합하다고 판단하고 있다.

3.4.2 MARC/SGML 프로젝트

지금까지 도서관목록에서의 검색내용이란 저자, 서명, 주제, 주요어의 불탐색 등이 주였다. 그러나 최근에는 Web에 Alta Vista, Excite, Lycos, InforSeek, Inktomi, Aliweb, Harvest, Magellan, Open Text, Web Crawler, WWW Worm, Yahoo 등의 수많은 색인서버들이 나타났으며, 완전한 것은 아니지만 많은 시스템들이 가중치탐색, 단어의 위치나 빈도를 이용한 적합성 피드백, 자동어미절단, 자연언어 질문시스템, 개념기반탐색, 의미트리 등의 새로운 기법들이 사용되고 있다. 현재 도서관이 가지고 있는 데이터베이스를 인터넷, Web 혹은 관련시스템으로 변경하려고 할 때 전통적인 도서관시스템용으로 설계된 MARC은 여러 부분이 경직되어 있으며, 유일한 출구가 ANSI/NISO Z39.50-1995로 너무나 좁다. 따라서 MARC의 SGML 판을 만들 필요가 생기게 되었다.

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

으로 채택된 것은 아니지만 자문위원회에서 정한 설계원칙 중 중요한 것은 다음과 같다. 첫째, MARC DTD는 실제 MARC에서 SGML로, SGML에서 MARC으로 정보의 손실 없이 완전하게 양방향 변환이 가능해야 한다. 둘째, MARC/SGML레코드는 SGML문헌 내에 포함시킬 수도 있으며 메타데이터로서 독립시킬 수도 있다(Davis, 1996). 자문위원회에서는 MARC DTD를 개발하여 제안하였다. 이미 USMARC DTD(URL 14)와 변환프로그램이 Jerome McDonough 등에 의해 개발되었으며 MARC레코드와 SGML형식간에 정보의 손실을 최소화하면서 양방향으로 자동변환할 수 있도록 설계되었다.

3.4.3 MARC 형식의 보안

그렇다면 MARC형식은 더 이상 이용가치가 없으며 얼마 가지 않아 없어질 것인가? 그렇지 않다. 수십억의 MARC레코드가 시스템상에 있으며, 이들을 SGML로 바꾸는 시간과 컴퓨터 자원상의 비용은 상상도 못할 것이다. MARC은 가장 효율적으로 이용할 수 있는 방향으로 진화할 것이다. 즉, 도서관시스템에서 서지데이터를 코드화할 수 있는 형식은 이제 MARC이 전부다 아니기 때문이다(Gaynor, 1996). 미국 국회도서관은 USMARC을 개정하였으며 그 중 중요한 것은 필드 856(전자적 위치 및 액세스)을 설정했다는 점이다. 해당 정보를 포함하고 있는 전자적 위치나 그 정보를 얻을 수 있는 위치를 식별하기 위한 것이다.

제1지시기호(0 전자우편, 1 FTP, 2 원격로그

인, 3 다이얼업, 7 서브필드 \$2에서 명시한 방법, 8 기타)는 필드에 있는 데이터 이외의 데이터를 어떤 액세스방법으로 검색할 수 있는지를 정의하는 값을 갖는다. 정의된 방법은 TCP/IP 프로토콜이다. 제1지시기호의 값은 어느 서브필드를 사용할 것인지를 결정한다. 예를 들면, 제1지시기호의 값이 1 (FTP)이면, 서브필드 \$d(경로), \$f(전자적 이름), \$c(압축정보), 및 \$s(파일크기)를 사용하게 되며, 이들은 제1지시기호가 2 (원격로그인: Telnet) 일 때는 사용할 수 없다.

이 필드에는 화일의 전자적 전송, 전자잡지의 구독신청, 혹은 전자정보자원에 로그인할 수 있는 충분한 정보를 기술할 수 있다. 경우에 따라서는 원격지 호스트가 정보를 가지고 있어서 사용자는 locator table를 액세스할 수 있는 유일한 데이터식별요소만을 기술할 수도 있다. 위치 데이터요소가 다양하거나(\$a, \$b, \$d) 두 가지 이상의 액세스방법을 사용하고 있을 때는 필드 856을 반복할 수 있다. 또한 단일 자료가 온라인 축적 및 검색을 위해 여러 개의 부분으로 분리되어 있는 경우를 제외하고는 전자화일명(서브필드 \$f)이 바뀔 때마다 반복할 수 있다.

따라서 설정된 서브필드 코드 중 중요한 것을 보면, \$a 호스트명, \$b 액세스번호, \$c 압축정보, \$d 경로, \$f 전자적 이름, \$j BPS, \$k 패스워드, \$l 로그인/로그인, \$n 서브필드 \$a에 있는 호스트의 위치명, \$o 오퍼레이팅시스템, \$p 포트, \$q 파일포맷, \$s 파일크기, \$t 터미널이물레이션, \$u URL, \$w 레코드제어번호, \$2 액세스방법 등이 있다(URL 16).

4. 디지털도서관의 전망

디지털도서관은 기존 기술의 집합체에 불과하며, 연구적인 측면보다는 개발적인 측면이 강하다는 의견도 많다. 그러나 디지털도서관은 기존 기술의 집적과 함께 분산환경을 이용하기 위한 데이터베이스나 네트워크기술, 대규모데이터의 처리기술, 고해상도를 필요로 하는 데이터의 처리를 위한 고도의 멀티미디어 처리기술 등 미래의 연구테마가 수없이 남아 있다. 금후 디지털도서관의 경향을 다음과 같이 예측할 수 있다.

4.1 다양한 디지털도서관의 출현

디지털도서관을 생각할 때 모든 자료, 기술, 업무, 즉, 영구적인 자료와 일시적인 자료, 불변의 자료와 가변의 자료, 디지털기술과 인쇄매체 기술, 개인에 의한 정보활동과 협동에 의한 정보활동 등 모두를 고려하는 도서관을 생각하게 된다. 그러나 이와 같은 종합적인 디지털도서관의 구축이란 거의 불가능하며 의미가 없는 경우가 많다. 그 이유는 다음과 같다.

- 1) 매체의 통합이 불가능한 경우나 의미가 없는 경우가 많다
- 2) 지적재산권의 특성상 데이터의 집중이 불가능하다
- 3) 사용자군이 다르다
- 4) 특정 도서관에서 이 업무를 감당할 능력이 없다

따라서 집서별, 분야별, 매체별, 서비스별, 지적재산권의 특성별로 다양한 형태의 특화된 디지털도서관이 출현하게 될 것이다. 그러나 정보

의 특성을 보면 동종의 정보는 한 데 모이며, 이종의 정보는 분리된다. 따라서 다양한 형태의 특화된 정보는 이합집산의 과정을 거쳐 관련있는 정보들이 '디지털도서관'이라는 이름 아래로 집적될 것이다. 예를 들어, 각 출판사가 개별적으로 판매서가를 갖다가 이를 모을 필요를 느끼는 것과 같다. 집적의 내용, 규모, 관할권에 따라 그림 3과 같은 수많은 '센터디지털도서관'과 '로컬디지털도서관'이 존재하고 서로 협력하게 될 것이며, 포괄적인 정보서비스가 가능하게 될 것이다. 즉, 메타데이터와 원문의 연계가 통합적으로 가능하게 될 것이다.

4.2 전통적인 도서관의 중요성

디지털도서관과 전통적인 도서관을 별개로 분리하여 생각하는 경우가 많다. 전술한 바와 같이 특화된 디지털도서관과 센터디지털도서관이 출현할 것은 틀림없다. 그러나 일반적인 정보가 모두 디지털화 되기 전에는 MARC을 중심으로 하는 서지데이터가 중심이 된다. 수십억의 MARC레코드가 시스템상에 있으며, 全文으로 바뀌는 날이 언제일지 상상도 못할 것이다. 서지데이터 중심체제, 즉 全文데이터에 대한 메타데이터 처리체제으로서의 현 도서관기능은 변한 것이 아무것도 없다. 단지 특정 이용자, 특정 도서관, 특정 도서관의 정보에 의해 만들어지던 정보공간이 잘 연계된 세계의 도서관을 대상으로 이용자의 정보의 요구에 따라 시시각각으로 옮겨가고 변형될 뿐이다.

디지털도서관은 전통적인 도서관기능에 뿌리를 둔다. 따라서 현 도서관기능과 완벽하게 통

합되지 않는 디지털도서관 계획은 사상누각이 될 것이다. 미국국회도서관을 중심으로 진행되고 있는 MARC/SGML프로젝트나 MARC형식의 확장도 이를 시사하고 있다.

4.3 데이터의 디지털화 경향

디지털도서관이 사회에 영향력을 발휘함에 따라 과학기술분야를 선두로 잡지, 도서 순으로 급격히 전자화된 데이터로 변화해 갈 것이며, 과거의 자료도 전자화되어 갈 것이다. 물론 현재의 환경에서 데이터의 축적은 많은 시간이 걸리고 막대한 비용이 드는 일이므로 대단히 어려운 일이라는 하지만 장기간에 걸쳐 착실하게 변화해 갈 것이라 생각된다.

디지털도서관을 계획하는 대학을 보아도 지적재산권 처리가 무난한 해당 대학의 출판물(학위논문, 교수의 보고서, 교수의 학술지논문, 고서, 정기간행물 등)을 대단위 全文데이터베이스 구축의 우선 대상으로 하고 있다. 또한 한국문에 진흥원, 국립중앙박물관, 국립현대미술관 등에서 이미지 및 영화상데이터를 중심으로 하는 디지털도서관을 진행하고 있다. 분야별, 매체별, 서비스별, 지적재산권의 특성별로 많은 기관들이 분담하게 된다면 데이터의 디지털화는 보다 빨라질 것이다. 또한 디지털화 기술은 빠르게, 표준형태로, 저렴한 비용으로 입력하는 방향으로 연구되고 있으며, 시간이 흐름에 따라 새로운 기법이 발표되고 있다.

전술한 바와 같이 지적재산권이 전자저작권관리시스템에 의한 보호가 확실해지면 디지털도서관은 '종이없는 도서관'으로 급속도로 진행될

것이다. 독자와 저자가 가깝게 되어 읽은 것을 자신의 저작에 바로 삽입하여 인용할 수 있으며, 자신의 저작을 간단히 네트워크에 공개할 수 있을 것이다.

4.4 기술방법

초록을 포함하는 서지데이터를 ASCII데이터로 만들고 나머지 全文을 이미지로 축적하는 방법이 있을 것이며, 상당부분의 기존 정보는 이 방법을 사용하게 될 것이다. 다만 ASCII데이터로 입력된 데이터는 기술방법에 있어서 MARC과 SGML은 상당기간 공존하게 될 것이나 궁극적으로는 SGML형식으로 바뀌게 될 것이다. 미국국회도서관이 목록카드를 배포한 이래 MARC레코드와 관련 표준은 도서관 분야에 지대한 역할을 해왔지만, 전자정보환경은 도서관의 외부에서 형성되고 있을 뿐만 아니라 MARC을 중심으로 하는 도서관은 더 이상 정보혁명의 최전선이 아니라고 보아야 한다. 그러나 SGML형식의 이점이 많다고 바로 모든 데이터와 시스템을 SGML형으로 변환할 수는 없다. 이와 같은 상황에서 MARC과 SGML의 이용법은 크게 두 가지로 나눌 수 있다(최석두, 김태수, 1996).

그 하나는 MARC과 SGML의 어느 한쪽을 중점적으로 채택하는 방법이다. MARC을 채택하는 경우에는, 1) 지금까지의 형식을 그대로 사용하는 방법, 2) USMARC과 같이 필드 856 등을 설정하여 MARC형식을 개정하고, MARC형식 데이터와는 관계없이 SGML형식으로 기술된 서지 및 전문데이터는 링크를 이용하여 참조하는 방법이 있다. SGML을 사용하는 경우에

는, 1) MARC을 단순히 MARC/SGML형식으로 바꾸는 방법, 2) MARC/SGML형식으로 바꾸고 이미지를 포함하는 全文데이터를 TEI문헌으로 추가하는 방법이 있다(全文의 구조에 대해서는 ISO/DIS 12083-1994 및 ANSI/NISO Z39.59-1988 참조). 또 하나는 MARC과 SGML을 독립적으로 사용하는 방법이다. 모든 데이터는 MARC형식을 사용하고, 全文데이터로 변환되는 것만 SGML을 사용하는 방법이다. 결국 이 방법은 MARC을 채택하는 경우의 2)와 같아질 것이다. 어느 방법을 사용하더라도 이를 위하여 KORMARC/SGML을 위한 KORMARC.DTD와 양방향 변환프로그램의 개발이 시급하다고 사료된다.

4.5 색인, 검색 및 인터페이스기술의 중요성

全文데이터를 소장하는 정도에서 그치지 않고 다양한 측면의 잠재적 정보요구를 해결하기 위하여 혁신적인 색인기법 및 색인어의 추출문제가 사용자인터페이스와 함께 가장 중요한 위치를 차지하게 될 것이다. 왜냐하면 데이터베이스의 사활은 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 全文검색 및 통합검색을 기반으로 결과의 순위매김에 집중되고 있다. 또한 자연언어처리, VRML(virtual reality markup language), 최근의 JAVA 등을 이용한

사용자 인터페이스의 설계, 개발, 평가는 한 분야만의 문제가 아니라 학제적인 협력을 요하는 중요한 과제가 될 것이다.

4.6 디지털도서관에 대한 재평가

디지털도서관에 대한 우려와 함께 재평가가 있을 것이다. 디지털도서관에서 얻을 수 있는 것, 아울러 잃는 것, 디지털도서관이 아니면 안 되는 것, 디지털도서관으로 실현해야 하는 것 등에 대한 재고가 필요하다.

5. 디지털도서관에 대한 우려

현재 디지털도서관에 대한 기대와 흥미가 사회적으로 매우 커지고 있다. 디지털도서관에 대한 많은 국제회의가 개최되고 있으며, 많은 연구가 진행되고 있다. 그러나 디지털도서관에 대한 기대와 함께 우려의 목소리도 높다.

첫째, 디지털형태의 작품은 본질적으로 가소성(可塑性: plasticity)을 갖는다(Samuelson, 1994). 즉, 디지털데이터는 텍스트나 이미지를 막론하고 개작, 첨삭이 쉬우며, 편집을 용이하게 할 수 있다. 또한 하나의 매체에서 다른 매체로의 변환과 재포맷이 용이하다. 이와 같은 특성으로 인해 저작물의 원형보존이 어렵게 될 수 있다. 저작물이 아니라 인적정보나 생활정보에 대하여 첨삭이 행해진다면 사회생활에 치명적인 상황이 발생할 것이다. 심지어는 사진이나 음성의 변조까지 가능하므로 제공자와 사용자 양쪽에서 가장무도회와 같은 '전자변장'이 가능하여 무책임한 반사회적인 범죄를 유발할 수도 있다.

한편으로는 최종이용자의 정보이용 내용과 인적 사항이 공개됨으로써 불이익을 당하는 경우가 있게 된다.

둘째, 시간과 거리에 관계없이 완벽한 복제가 용이하다. 인쇄자료의 복사는 아무리 성능이 좋은 복사기라 할지라도 원본의 상태에 따라 복사 상태가 좌우되었으며, 대부분 원본보다 질이 떨어진다. 그러나 디지털데이터의 복사는 완벽한 복제가 용이하며 원본과 동일하다. 즉, 하려고만 하면 순식간에 원본을 무한정 제작 배포할 수 있다. 여러 가지 통제기법을 사용할 수 있겠지만 저작권의 보호를 위한 통제가 강력하면 학술 커뮤니케이션의, 통제가 허술하면 창작활동의 장애물이 되는 요인이 될 수도 있다.

셋째, 정보산업을 주도하는 강대국에게 악용될 가능성이 높다. 우리들이 필요로 하는 정보를 미국을 중심으로 하는 선진국들이 장악하고 있으며, 정보의 독점이라는 경향은 현실화 되고 있다. 예를 들면, 인터넷 상에서 무료로 서비스를 제공하던 기관들이 하나씩 유료화로 선회하는 추세이다. 정보선진국들은 정보후진국들에게 공짜로 정보의 맛을 보이고 맛을 들인 후 엄청난 대가를 요구하고 있다. 선진국의 정보에 주로 의존하고 있는 정보후진국들에게 보다 무서운 것은 정보선진국들이 정책적으로 어느날 모든 정보채널을 단절시키는 상황일 것이다. 이를 방지하기 위하여 우리말이 중심이 되는 정보처리 기술을 발전시키고, 처리규격을 세계표준화하며 데이터베이스산업을 육성하여 정보의 국제경쟁력을 산업적, 사회적, 문화적, 연구개발 측면에서 강화하는 수밖에 없다.

넷째, 지금까지는 사용자의 나이, 소속 등에 따라 정보의 선택적인 서비스가 가능했으나 사용자의 가명, 익명 혹은 대상을 개별적으로 식별하지 못하여 정보의 선택적 서비스가 불가능하게 된다. 기밀정보, 음란정보를 위시한 민감한 정보의 액세스에 대한 통제가 어려워 사회적인 문제가 될 것이며, 이에 대한 확실한 통제장치가 필요하게 된다.

다섯째, 디지털도서관의 데이터베이스서버는 광대역의 네트워크 인터페이스를 가질 필요가 있으나 데이터베이스시스템의 운영상 복수의 인터페이스를 갖는 것이 어려워 하나의 입출력으로 한정된다. 따라서 서버에 대한 액세스가 집중될 때에는 그 부하가 너무 커져서 쓸모없는 것이 될 수도 있다.

여섯째, 멀티미디어정보와 같은 대규모 데이터의 전송시 현재의 대용량 통신기술로 충분하다 할 수 없다. 대부분의 처리를 인터넷과 같은 전자통신에만 의존하다가 대량의 트래픽이 발생했을 때 응답시간이 늦어지거나 감당하지 못하여 국가적인 공황상태로 빠질 수도 있다.

6. 결 론

도서관의 궁극적인 목표는 이용자에게 정보원에 대한 액세스의 기회를 늘리는 일이다. 이것은 다행스럽게도 정보화사회의 진행방향과 일치한다. 많은 나라들이 국가정보기반구조로서 초고속정보통신망을 건설하는 데 주력하고 있으며 디지털도서관은 정보고속도로의 중핵이 되고 있다.

디지털도서관이 정보서비스의 중심이 되었을 때, 가장 중요한 일은 국민이나 모기관의 사용자에게 필요 정보원을 편리하게 액세스할 수 있도록 해야 하는 일이다. 이를 위하여 해당 국가, 단체 혹은 기관만이 만들 수 있는 쉐어데이터베이스, 제작에 선두주자가 되는 쉐어데이터베이스를 만드는 데 힘을 쏟아야 한다. 물론 가치가 있는 데이터베이스여야 한다. 이와 같은 경쟁력 있는 데이터베이스를 만들었을 때 데이터베이스 자체의 이용에서 뿐만 아니라 수없이 많은 데이

터베이스 소유기관과 상호제공계약이 가능하게 되어 소속 이용자가 그 혜택을 받을 수 있기 때문이다.

한편으로 주의해야 할 것은 특정 기관, 특정 국가에 대한 정보의 의존도를 줄이는 일이며, 정보화사회라는 이름으로 개인생활에 지장을 주거나 개인에게 재산상의 손해를 주어서는 안된다는 것이다. 즉, 디지털도서관계획은 국가경쟁력과 국민의 복지에 직접적인 영향을 미치는 사업으로 인식되어야 한다.

참 고 문 헌

- 최석두, 김태수(1996). MARC과 SGML의 통합에 대한 연구. 제3회 한국정보관리학회 논문대회 논문집. 67-70.
- ANSI/NISO Z39.50-1995. *Information Retrieval Application Service Definition and Protocol Specification for Open System.*
- ANSI/NISO Z39.59-1988. *Electronic Manuscript Preparation and Markup.*
- Chen, S.(1995). The NSF/ARPA/NASA Research on Digital Libraries Initiative. In *Proceedings of 1995 Digital Libraries Conference.* Singapore : National Computer Board, 48-55.
- Davis, Stephen Paul(1996). *SGML-MARC : Incorporating Library Cataloging into the TEI Environment.* (<http://www.columbia.edu/cu/libraries/inside/projects/sgml/sgmlmarc/davis9603.html> 1).
- Gaynor, Edward(May 8, 1996). *From MARC to Markup: SGML and Online Library systems.*(<http://www.lib.virginia.edu/speccol/scdc/articles/alcts-brief.html>).
- ISO/DIS 12083-1994 *Information and Documentation - Electronic Manuscript Preparation and Markup.*
- Samuelson, Pamela(1994). Copy, Digital Data, and Fair Use in Digital Networked Environments. (<http://www.droit.umontreal.ca/CRDP/Conferences/AE/Samuelson.html>).
- URL 1. <http://www.grainger.uiuc.edu/dli/national.htm>.
- URL 2. <http://fuzine.mt.cs.cmu.edu/im/informedia.html>.
- URL 3. <http://http.cs.berkeley.edu/~wilensky/proj-html/proj-html.html>.

URL 4. <http://http2.sils.umich.edu/UMDL/HomePage.html>.

URL 5. <http://alexandria.sdc.ucsb.edu/>.

URL 6. <http://diglib.stanford.e여>.

URL 7. <http://www.grainger.uiuc.edu/dit>.

URL 8. <http://www.elsa.dmu.ac.uk/>.

URL 9. <http://ford.mk.dmu.ac.uk/>.

URL 10. <http://java.sun.com/>.

URL 11. <http://www-tei.uic.edu/orgs/tei/>.

URL 12. <http://sunsite.berkeley.edu/FindingAids/>.

URL 13. <http://www.cc.columbia.edu/cu/libraries/indiv/avery/diap.html>.

URL 14. <ftp://library.berkeley.edu/pub/sgml/marcdtd/usmarc.dtd>.

URL 15. <gopher://marvel.loc.gov:70/00.listarch/usmarc/96-1.doc>.

〔 국내 발행 · 제작자료 납본 안내 〕

- 납본은 문화선진국으로 가는 첫걸음입니다.
- 납본은 법으로 규정된 발행 · 제작자의 의무입니다.
- 발행 · 제작된 자료는 국립중앙도서관에 제출합니다.

1. 누구나 자료를 발행 또는 제작한 때에는 『도서관 및 독서진흥법 제17조』와 『동법 시행령 제21조』 규정에 따라 그 자료를 국립중앙도서관에 제출하여야 합니다.

2. 제출된 자료는 국가문헌으로 영구보존하여 필요로 하는 국민에게 제공하고, 우리의 문화를 해외에 알리는 데에도 활용됩니다.

가. 제출대상자료

- 도서, 연속간행물, 음반, 카세트테이프, 비디오테이프, 비디오 CD, CD-ROM, CD-I, CD-G, 기타 CD제작물, 전산화자료, 지도, 슬라이드, 마이크로형태물 등

나. 제출자

- 국가, 지방자치단체 및 출판 · 제작자

다. 제출시기 및 부수

- 발행일, 제작일로부터 30일 이내 각 2부씩

라. 제출처

국립중앙도서관 납본과

서울특별시 서초구 반포동 산 60-1(137-702)

대표전화 02)535-4142, 직통 02)537-9648, FAX 02)533-4228

마. 기타

- 제출된 자료에 대하여는 법률이 정한 보상금을 드립니다.
- 제출하지 않을 경우 관계법률에 의해 불이익을 받습니다.