

## 디지털도서관의 구성요소에 관한 연구

### Requirements for the Digital Library Design

최석두, 오영화  
(이화여자대학교 문헌정보학과)

Suk Doo Choi, Young Hwa Oh  
(Dept. of Lib. & Inf. Sci., Ewha Womans Univ.)

디지털도서관은 데이터를 디지털화하여 수집·축적하고, 네트워크를 통해 이용가능한 도서관이라 할 수 있으며, 계산, 데이터축적, 통신관련 기술을 이용하여 완전한 정보서비스를 제공하는 것이 목적이다. 다양한 정보를 디지털화하기 위해서는 멀티미디어기술, 대용량의 데이터축적기술, 통신을 위한 컴퓨터네트워크기술 등이 기반이 되므로 여러 분야와 관련이 되는 것은 틀림없으나 결국 도서관의 역할을 하는 시스템을 만드는 것이 목적이다. 본 연구에서는 디지털도서관 설계시 개념화하고 범주화하여야 할 디지털도서관의 기능, 구조, 범위, 요소기술, 고려사항 등에 대하여 논한다.

#### I 서론

도서관은 점차 컴퓨터, 전자출판, 네트워크, 멀티미디어 등의 발전과 함께 "소유"에서 "접근"의 개념으로, 시설로서의 도서관에서 지적체계로서의 도서관 개념으로 변화하고 있다.

시대는 세계의 도서관이 소장하고 있는 다양한 형태, 다양한 내용, 그리고 대량의 데이터를 불특정다수의 다양한 이용자에게 고도의 정보검색기능으로 즉시 제공하도록 요구하고 있으며, 도서관의 기능을 단순한 정보의 수집과 이용 뿐만 아니라 교육적 기능과 학술활동의 장으로의 확대를 요구하고 있다. 최근의 전자출판, 전자문헌전송, 이용자탐색, 계산과 텔레커뮤니케이션기술의 발달로 이와 같은 정보요구를 충족시킬 수 있게 되었으며, 이 도서관혁명은 디지털도서관이라는 개념하에 활발히 진행되고 있다.

본 연구에서는 디지털도서관 설계시 개념화하고 범주화하여야 할 디지털도서관의 구성요소를 도서관 측면, 요소기술 측면, 처리과정 측면, 사회적 측면, 표준

화 측면을 중심으로 기능, 구조, 범위, 요소기술, 고려사항 등에 대하여 논하고자 한다.

#### II 디지털도서관의 정의

##### 1 용어

미래의 도서관에 대하여, "벽이 없는 도서관"(library without walls) 혹은 "논리도서관(logical library)"이라는 용어를 많이 사용하여 왔으나, 최근에는 "전자도서관(electronic library)", "버추얼라이브러리(virtual library) 혹은 "디지털도서관" (digital library)이라는 용어를 사용하고 있으며, "버추얼리얼리티라이브러리(virtual reality library)를 사용하기도 한다. 이들은 모두 동의어로 취급할 수 있다. 최근의 용어사용을 보면 "virtual library"와 "digital library"라는 두 용어가 집중적으로 사용되고 있으나 "digital library"라는 용어가 더 빈번하게 사용되고 있다. "virtual library"는 "가상도서관"(假想圖書館)이라고 번역하는 경우도 있으나 우리말의 "가상"은 "실제로는

존재하지 않는 것을 임시로 존재하는 것으로 상정하는 것"에 사용되는 용어이므로 의미를 정확하게 전달하고 있지 못하다. 따라서 본 고에서는 "digital library"의 역어 "디지털도서관"을 사용하기로 한다.

## 2 정의

1980년 A.J. Harley(1980)는 가상기계(假想機械, virtual machine)의 개념을 빌려, "비추얼라이브러리란 다수의 동시 사용자 개개인이 전체의 장서를 혼자만 소유하고 있는 것처럼 느낄 뿐만 아니라, 다른 도서관과 연결함으로써 눈에 보이는 정보자원보다 더 많은 정보자원을 자신의 책상앞에서 즉시 액세스할 수 있는 도서관"이라고 정의하고 있다. 즉, 누구나 어디에서나 세계의 모든 정보자원을 액세스할 수 있는 도서관이다. 이에 대해 전자도서관은 정보자원을 위한 공간, 사람, 시설은 있으나 물체로서의 책이 눈에 보이지 않는다는 것을 강조한 것이며, 비추얼라이브러리의 진행과정이나 구성요소로 보기도 한다(Coller et. al., 1993).

디지털도서관은 정보자원을 이용하기 위해서 데이터의 디지털화가 필수적으로 선행되어야 하므로 이를 강조하기 위한 용어이다. Gladney 등(1994)은 "전통적인 도서관이 종이나 기타 매체를 이용하여 제공하여 온 수집, 목록작성, 정보의 발견과 유통이라는 서비스를 계현, 모의, 확장하는 데 필요한 내용이나 소프트웨어를 계산, 축적, 통신의 기계장치와 적절히 조합한 것이다. 완전한 서비스를 제공하는 디지털도서관은 전통적인 도서관이 제공하여 온 필수적인 서비스를 실현하지 않으면 안되며, 디지털화된 정보축적, 검색 및 통신의 잘 알려진 장점을 살리지 않으면 안된다"라고 정의하고 있다.

## III 디지털도서관의 구성요소

### 1 도서관 측면

멀티미디어전자도서의 출현으로 현재의 형태인 "책"이 살아남을지 의문이 간다. 디지털도서관에서 자료가 디지털화 된다는 것은 전자화도서를 중심으로 서비스한다기보다는 전자정보, 디지털정보 자체를 서비스하게 된다. 현재에는 CD-ROM을 중심으로 정보제공이 증가하고 있으나 곧 온라인화 자료가 급속도로 증가할 것이라 예측하고 있다. 즉, 개개 도서관의 장서가 눈에 보이지 않거나 없어진다는 것이다. 당연히 사용자가 도서관에 갈 일이 없어지며, 거대한 건물도 필요 없어진다. 그러면 도서관은 왜 필요한가, 사서는 무엇을 할 것인가, 어떤 훈련과 어떤 자질이 필요한가?

우선, 도서관의 조직, 사서의 훈련 및 정보의 제공 방법을 각 디지털도서관에 맞는 체제로 바꾸어야 할 것이다. 이를 위하여 의사결정, 교육, 오락, 연구, 새로운 정보의 생산 등 도서관과 이용자의 정보요구에 따라 디지털도서관의 봉사모형을 결정하여야 한다. 둘째, 이용자 인터페이스를 GUI(Graphical User Interface)로 바꾸어야 한다. 예를 들면, 검색의 결과를 자료가 서가에 꽂혀있는 상태로 보여주며 서명, 저자와 함께 크기, 면수에 따라 다르게 보여 줄 수도 있다(Sugimoto et. al., 1995).

셋째, 이용자가 디지털도서관에서 세계의 다양한 정보를, 최선의 방법으로, 저렴하게 액세스할 수 있는 환경을 만들어야 한다. 현재 인터넷에는 무료자원이 계속 공개되고 있으나 이것들이 언제까지나 무료라는 보장은 없다. 대단위 정보소유자들이 사용료를 받게 될 것이며, 네트워크, 하드웨어, 소프트웨어 등이 계속 발전하게 되면 이에 적합한 환경을 만드는 것은 쉬운 일이 아니다.

네째, 전술한 바와 같이 자료가 대량, 다양화, 분산, 국제화, 멀티미디어화됨에 따라 이용방법이 복잡해져 전문가(사서)의 도움이 필요하게 된다. 물론 이 도움은 네트워크를 이용하여 컴퓨터의 지원하에서 공동으로 작업하는, 소위 CSCW(Computer Supported Cooperative Work)가 될 것이다(Fox, 1993: 197-199). 이와 같은 시스템으로 최근에는 "네트워크참고봉사 시스템"도 등장하고 있다. 전통적인 도서관의 知的서비스는 언제나 필요하게 되며, 도서관은 "네트워크사서"(networking librarian, network specialist)라 불리는 전문가를 양성해야 할 것이다.

### 2 요소기술 측면

디지털도서관과 관련되는 연구영역은 여러 분야에 광범위하게 관련되고 있으며 일별하면 다음과 같다. 임의형식의 데이터의 입력에 관한 연구로 OCR, 음성인식, 도형이해 등이 있다. 다양한 형식의 전자정보를 조직화하는 연구로 색인, 내용이해, 분류, 목록, 다언어색인, 하이퍼미디어 등이 있다. 다양한 정보에 대한 검색, 필터링, 이미지들 포함하는 대규모 데이터의 초록 등에 관한 연구로 검색이론, 검색모델, 문서의 형식구조, 지적문서처리, 자연언어처리, 이해 등이 있다. 다양한 정보의 가시화와 열람에 관한 연구로 화상인식, 이미지분류, 멀티스케일디스플레이, 데이터가시화, 하이퍼미디어상의 네비게이션 등이 있다. 네트워크와 관련된 연구로 네트워크의 보안, 프로토콜디자인, 데이터압축 등이 있으며, 분산된 네트워크상의 데이터베이스이용을 단순화하기 위한 연구로 에이전트, 지적계

이터키퍼, 오브젝트지향 데이터베이스, 이용방법에 관한 모델 등이 있다(杉本重雄, 1994).

### 3 처리과정 측면

사서들에 의해서가 아니라 정보과학자들에 의해 붙여진 디지털도서관은 과연 "도서관"이라 불러야 하는가에 대해 의문을 던지는 사람도 있다(Miksa and Doty, 1994). 그러나 디지털도서관은 디지털화된 정보의 집적이며 도서관의 역할을 하는 시스템이다. 다만 전통적인 도서관이 갖는 성질을 반드시 만족시키지 않는 부분도 있으며, 전통적인 도서관이 갖는 제약에 속박되지 않는 면도 많다. 디지털도서관의 주요 구성요소를 처리과정 측면에서 보면 <그림 1>과 같다.

디지털화된 데이터는 바로 색인되고 축적되어 이용

이와 같이 축적된 대단위 데이터는 정보검색시스템과 이용자인터페이스를 통하여 이용자와 연결되게 된다. 다양한 검색기능과 자연언어처리기법을 통합한 지능형 정보검색시스템이 활발히 연구되고 있으며, 부정확하고 불충분한 질문에 대하여 정확한 정보를 검색할 수 있는 시스템을 목표로 하고 있다. 또한 전자저작권관리 시스템(ECMS: Electronic Copyright Management System)을 두어 지적재산권을 보호해야 한다.

모든 정보검색시스템은 네트워크와 여러 가지 브라우저들을 통하여 액세스하게 되며, "정보의 소장장소", "이용자의 위치", "정보의 형태"라는 제약을 뛰어 넘어 세계적으로 이용할 수 있다. 현재는 주로 인터넷을 이용하고 있으며, 인터넷으로 연결된 다양한 시스템을 네비게이트할 수 있는 시스템들이 속속 개발되고 있

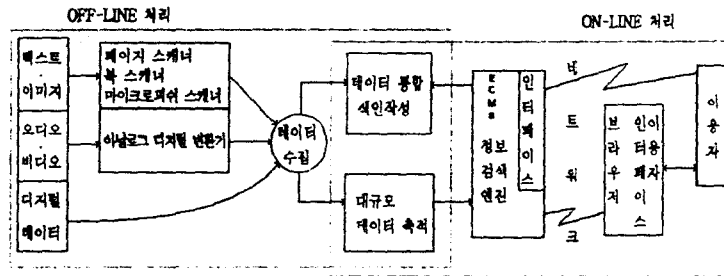


그림 1 디지털도서관의 주요 구성요소

할 수 있다. 텍스트, 이미지, 오디오, 비디오 등과 같은 데이터가 책자형태, 마이크로필름형태, 테이프형태 등으로 디지털화되지 않은 상태로 있을 때는 여러 종류의 스캐너로 디지털화하고 텍스트 부분은 OCR을 이용하여 ASCII 텍스트로 바꾸어야 한다. 물론 이 때 다양한 문헌, 다양한 폰트, 페이지의 레이아웃, 표, 그래프, 차트 등을 인식, 이해하고 분석할 수 있어야 한다(Srihari et. al., 1994).

직접 입력되거나 OCR의 결과로 수집된 데이터는 색인과정을 거치게 되며, 관련이 있는 데이터끼리는 그룹화되어 링크를 갖게 된다. 여기에는 이미지와 그 색인, 해설이 있는 사진에서 해설과 사진 등의 링크도 포함된다. 완벽한 링크를 가짐으로써 이용자는 모든 정보자원을 자유롭게 네비게이트할 수 있게 된다. 자동색인을 위하여 통계, 확률, 자연언어처리, 시소러스 등 여러 가지 기법들이 동원되고 있으며, 동시에 이와 같은 색인과 원정보를 포함하는 대단위 데이터를 효율적으로 검색하기 위하여 새로운 데이터 하부저장기법도 강구되고 있다.

다. 복수의 서버와 클라이언트로 구성된 WAIS(Wide Area Information Servers)와 분산형 하이퍼시스템인 WWW(World Wide Web)를 중심으로 anonymous FTP, archie, finger, gopher, telnet, whois 등이 있다.

이 중 가장 각광을 받고 있는 WWW는 WAIS와 마찬가지로 서버와 클라이언트로 구성된다. 서버에는 링크된 텍스트, 화상, 음성이 축적된다. 텍스트중에는 링크 이외에 문자의 크기나 字體를 지정할 수 있으며, 이것을 지정하기 위한 태그의 형식은 SGML(Standard Generalized Markup Language)에 준거한 HTML(Hypertext Markup Language)에 따르고 있다. HTML로 기술된 텍스트를 서버와 클라이언트간에 전송하기 위하여 HTTP(Hypertext Transfer Protocol)을 사용하며, 링크를 나타내는 태그에는 서버의 호스트명과 그 호스트상의 해당 텍스트를 식별하기 위한 이름이 URL(Uniform Resource Locators)이라 불리는 표준형으로 들어가 있다. 클라이언트의 브라우저로는 NCSA Mosaic과 Netscape가 많이 사용되고 있다.

#### 4 사회적 측면

지적재산권, 저작권, 개인정보의 보안, 디지털도서관의 영향, 이와 같은 환경에서의 출판 등이 관련이 된다.

특히 저작권은 저작권소유자, 저자, 출판자가 관련된다. 문헌이 한번 스캐닝되면 디지털이미지가 만들어지며 시스템이용자는 누구나 복사할 수 있으므로 다음과 같은 문제가 생긴다(Chen, 1995). 첫째, 디지털도서관은 이 디지털이미지를 소유하는 데 필요한 허가와 보상을 어떻게 할 것인가? 둘째, 디지털도서관을 통해 이 이미지를 전송하는 데 관련된 문제는 무엇인가? 셋째, 이용자가 이 디지털이미지를 사용하는 데 필요한 허가와 보상을 어떻게 할 것인가, 그리고 2차 이용자에 대한 제한은 무엇인가? 넷째, 전통적인 도서관의 권리는 무엇인가, 그리고 디지털형식 문헌의 정당한 이용이란 무엇인가?

이와 같은 문제를 해결하기 위하여 디지털도서관 환경하에서 필요한 저작권법 제정과 기관단위 연회비, 복사단위 징수 등 여러 가지 경우에 대한 징수방법을 강구해야 할 것이다. 예를 들면 CNRI에서는 ARPA와 LC를 연계시켜 디지털도서관 환경에서의 전자저작권 관리시스템(ECMS)을 개발하고 있는 것을 볼 수 있다 (Garrett and Lyons, 1993).

#### 5 표준화 측면

다행스럽게도 정보처리를 위한 표준화는 서지데이터형식 관련 표준 이외에도 상당히 안정되어 있다. 중요한 것만 예를 들면 다음과 같다(Fox, 1993: 7). 가장 중요한 데이터코딩 표준으로 문자코딩용으로는 ASCII, 문헌표현용으로는 전술한 SGML이 있다. 문서구조의 정의는 문서의 형을 정의하는 DTD(Document Type Definition)를 SGML의 구문으로 기술하게 된다. 메시지처리표준으로는 전자메일의 X.400, 디렉터리의 X.500, 정형화된 검색요구의 Z39.50이 있다. 정보검색 시스템의 일관성 있는 통신언어로 범용명령언어인 Z39.58가 있으며, 전자데이터교환의 X.12가 있다. 멀티미디어와 관련해서는 두 가지의 표준이 중요한 역할을 하고 있으며, 컬러이미지의 JPEG표준, 비디오와 오디오의 MPEG표준이 그것이다. 상기 표준들은 모두 미국 국가단체 ANSI, NISO, 국제단체 ISO, IEC, CCITT 등과 같은 표준관련기관에서 인정되고 있다.

#### IV 결론

정보는 "찾기 쉽고, 읽기 쉽고, 싸야" 한다. 분명한 일이지만 이를 위하여 다양한 기술을 집약시켜 디지

털도서관을 만들고 있다. 그러나 어느 기술까지를 포함시킬지 아직 미지수이다. 따라서 디지털도서관이란 과연 어떤 도서관인가에 대한 정확한 정의는 현재로서는 무리이며, "바로 이것이 디지털도서관이다"라고 할만한 디지털도서관의 출현까지는 상당한 시간이 걸릴 것이다. 다만 우리는 정보를 찾는 데 도움을 주고, 비공식적인 정보도 찾을 수 있으며, 꼭 필요한 정보만 골라서 찾고, 만남의 장이 되는 그런 사회적인 기능을 갖는 디지털도서관을 만들어야 할 것이다.

#### 참고문헌

- 杉本重雄(1994). "Digital Librariesへのアプローチ - 米國で開催された Workshop, Conferenceに出席して -". 「デジタル図書館」, no.1, 5-22.
- Chen, Su-Shing(1995). "Technologies for Digital Library," 1995 Digital Libraries Conference: Moving Forward into the Information Era, 27-28 March 1995. 39-45.
- Collier, Mel W., Anne Ramsden and Zimin Wu(1993). "The Electronic Library: Virtually a Reality?," Opportunity 2000: Understanding and Serving Users in an Electronic Library. Essen University Library. 135-146.
- Fox, Edward A.(1993). Source Book on Digital Libraries. Version 1.0. National Science Foundation.
- Garrett, John R. and Patrice A. Lyons(1993). "Toward an Electronic Copyright Management System," JASIS. 44(8):468-473.
- Gladney, Henry M. et. al.(1994). "Digital Library: Gross Structure and Requirements: Report from a March 1994 Workshop," Proceedings of DL '94. 101-107.
- Harley, A.J.(1980). "Towards the Virtual Library," The Nationwide Provision and Use of Information: Aslib/MS/LA Joint Conference Proceedings. London : Library Association. 163-166.
- Millsa, Francis L. and Philip Doty(1994). "Intellectual Realities and the Digital Library," Proceedings of DL '94. 1-5.
- Srinari, Sagur N., Stephen W. Lam and Jonathan J. Hull(1994). "Intelligent Data Retrieval from Raster Images of Documents," Proceedings of DL '94. 34-40.
- Sugimoto, Shigeo et. al.(1995). "Enhancing Usability of Network-based Library Information System Experimental Studies of a User Interface for OPAC and of a Collaboration Tool for Library Service," Proceedings of DL '95. 115-122.