

원문이미지정보시스템 비교분석

백 성 규

(강남대학교 중앙도서관)

최 재 화

(단국대학교 경영학과)

요 약

인터넷의 활용과 컴퓨터 기술환경의 새로운 패러다임 변화는 디지털도서관의 등장을 가져오게 하였으며, 동시에 이미 온라인 서비스를 실시해오던 여러 분야의 정보센터들로 하여금 어떠한 유형의 데이터도 저장, 관리, 제공하며 정보라는 상품을 생산해 내는 정보시장(Information Marketplace)의 역할 자로서 요구하고 있다. 본 논문에서는 1980년대부터 진행되어온 도서관 서지자료의 Database화 작업과 병행하여 도서관 이용자에게 보다 광범위하고 효용성 있는 정보를 제공하기 위한 디지털도서관의 이론적 배경에 대하여 기술하고, 디지털도서관의 주요 소장정보인 원문이미지정보의 처리기술에 대한 이론을 정리하며, 현재 국내에서 추진중인 원문이미지정보시스템의 분석을 통하여 향후 디지털도서관 구축시 비교 분석 모델을 제시해보고자 한다.

I. 서 론

최근의 정보 및 통신기술의 혁신은 정보 수집과 유통에서 일대 변혁을 일으키고 있다. 특히, 21세기 정보화 사회의 기반 구조로서 초고속통신망의 구축, 개인용 컴퓨터의 급속한 보급과 멀티미디어 기술의 발전은 과거에 생각할 수 없었던 정보 서비스의 새로운 장을 열고 있다. 정보의 저장 및 전달매체가 종이에서 디지털 형태로 전환되고 네트워크와 통신기반 구조가 진보함에 따라서 종래의 도서관서비스 기능이나 제공정보의 형태들도 전자화, 디지털 화되고 있는 이른바 디지털 도서관의 등장이 이루어지고 있다. 디지털도서관의 특징은 원격지 소장 정보에 대한 용이한 접근과 텍스트, 이미지, 영화, 비디오, 그래픽, 오디오 등 다양한 형태의 멀티미디어 정보검색이 가능하다는 점에서 점차 우리의 관심을 끌고 있다.

즉, 인터넷의 활용과 컴퓨터 기술환경의 새로운 패러다임 변화는 디지털도서관의 등장을 가져오게 하였으며, 동시에 이미 온라인 서비스를 실시해오던 여러 분야의 정보센터들로 하여금 어떠한 유형의 데이터도 저장, 관리, 제공하며 정보라는 상품을 생산해내는 정보시장(Information Marketplace)의 역할 자로서 요구하고 있다.¹⁾ 여기서 정보는 도처에 편재하여 정보의 소비자가 쉽게 접근이 가능하고 어떠한 유형의 데이터도 얻을 수 있는 반면, 소비자 스스로의 정보 발굴 및 개발 활동 역시 점차 중요하게 되었다. 따라서 대부분의 도서관이나 데이터뱅크에서 종래에 제공해 오던 정보형태(서지정보 위주)의 데이터베이스로는 이용자들의 원문정보에 대한 요구를 만족시킬 수 없게 되었다. 기하급수적으로 증가되고 있는 자료의 관리에 소요되는 인적, 공간적 비용을 최소화 할 뿐만 아니라 검색의 신속성 및 효율성을 높일 수 있는 최적의 방안을 찾고자 한 것이다. 종전의 Microfilm 시스템은 인쇄매체 자료의 영구보존 및 활용을 위한 필요성에서 등장하였으나, 자료의 신속한 검색이나 여러 이용자를 대상으로 온라인 공동 활용이 거의 불가능함은 물론 영구 보존상의 한계를 안고 있다. 이러한 문제점의 해결방안으로 원문을 이미지 방식으로 디지털화 하여 DataBase를 구축하는 경향이 증가되었다.

고해상도의 스캐너와 팩시밀리 등의 이미지 입력장비(Device)는 도큐먼트나 그림데이터의 하드 카피된 디지털 또는 래스터 표현을 가능케 하였다. 또한 비트맵 이미지등 대용량의 저장매체를 효율적으로 지원해주는 고밀도 광저장 매체의 개발, 그림 데이터의 출력 등을 위한 컬러모니터 및 프린터, 비디오 또는 이미지 데이터와 같은 멀티미디어 정보의 보급을 위한 광대역 통신, 멀티미디어 어플리케이션을 위한 CPU의 처리 능력과 메모리의 증가 등에 관한 기술 개발이 이러한 문제를 해결하는데 있어서 매우 중요한 요소인 것이다.²⁾ 이미 국내의 몇몇 도서관에서는 서지데이터 뿐만 아니라 全文데이터의 디지털화를 시도하고 있으며, 이를 컴퓨터통신이나 광학적 매체를 이용하여 제공하고 있다. 정보의 저장매체가 다양해짐에 따라 최근에는 정보의 온라인 전송과 임의 가공까지 고려하여 처음부터 디지털화 한 출판물이 나타나고 있으며 SGML(Standard Generalized Markup Language)문서는 이의 대표적인 사례이다.

컴퓨터공학과 통신기술의 발달에 따라 새로운 형태의 정보와 정보수록매체가 계속

1. David M. Choy , "The Alamaden Distributed Digital Library System", NewYork : Springer-verlag, 1995, pp. 155-172.

2. 홍기채, "멀티미디어 기술에 기반한 도서관 최신정보서비스시스템의 설계 및 구현", 충남대학교 대학원 석사학위논문, 1993, pp.12-21.

등장하고 있지만, 도서관에 소장되어 있는 기존 자료들의 정보이용가치를 결코 무시할 수는 없다. 그러므로 디지털도서관의 구축은 디지털 화하다 생산된 자료, 새로운 형태의 정보자료에 대한 디지털 화하다 Paper 자료에 대한 디지털화가 조화롭게 진행되어야 한다.³⁾

본 논문에서는 1980년대부터 진행되어온 도서관 서지자료의 Database화 작업과 병행하여 도서관 이용자에게 보다 광범위하고 효용성 있는 정보를 제공하기 위한 디지털도서관의 이론적 배경에 대하여 기술하고, 디지털도서관의 주요 소장정보인 원문이미지정보의 처리기술에 대한 이론을 연구하며, 현재 국내에서 추진중인 원문이미지정보시스템의 분석을 통하여 향후 디지털도서관 구축시 비교분석 모델을 제시해보고자 한다.

II. 원문이미지정보시스템

다양한 형태의 원문이미지정보를 효율적으로 저장하고 이용자가 원하는 정보를 정확히 전달하기 위한 요소기술로서 원문의 디지털화(이미지 스캐닝 또는 OCR변환)를 들 수 있다. 스캔되어진 이미지데이터는 원 문헌의 구성요소에 따른 특징을 그대로 유지하여 이 이미지정보를 조작하고 사용하는 이용자가 원문이미지데이터에 익숙해 질 수 있도록 해야한다. 이때 원문이미지의 저장방법은 문헌의 논리적 구조만을 표현하는 텍스트 문헌의 표준데이터 포맷인 SGML(Standard Generalized Markup Language), 문헌의 구조와 내용의 구분이 명확한 문헌기술의 모형인 ODA(Open Document Architecture), 멀티미디어 문헌의 구조화를 강조하는 SGML의 확장된 의미라 할 수 있는 Hytime(Hypermedia Time-based Structuring Language)등과 같은 원문 저장방법의 국제적 표준화 도구를 수용해야 한다. 그러나 이미지데이터는 텍스트 데이터에 비해 데이터의 크기가 방대하여 대규모 수준의 정보를 저장하고 네트워크를 통하여 고속으로 검색, 전달하기 위해서는 저장용량, 검색속도 등 전문검색의 경제성을 높일 수 있는 적합한 압축 및 복원기술이 필수적이다. 실제로 내용 기반 검색을 위한 색인의 경우에는 기존 데이터베이스시스템에서 키 값과 주소만으로 구성되는 B-tree 등과는 달리, 역화일 색인 등 그 크기가 실제 데이터의 크기에 접근하는 경우가 많다.

또한 원문이미지정보시스템은 이용자 우호적인 인터페이스가 지원되어야 한다. 진정

3. 우학명, "Paper 자료의 디지털화 형태에 따른 구동시스템 및 비용·효과 분석", 국회도서관보 34(1), 1997, pp.32-33.

한 의미의 이용자 우호적인 인터페이스란 서로 이질적인 시스템들 사이에서의 정보 검색이나 정보전달, 또는 정보를 표현할 때 이용자들이 시스템의 사용에 어려움을 느끼지 않도록 함을 의미하기도 한다. 즉, 실시간으로 고속의 통신경로와 다양한 네트워크 방식을 통하여 원격 검색이나 로컬검색이 이루어지도록 하고, 이용자에게 친근하고 사용하기 쉬운 그래픽 인터페이스(GUI : Graphic User Interface)를 지원하며, 텍스트 형식의 정보뿐만 아니라 이미지의 형식을 자연어로 표현하여 이용자가 손쉽게 정보에 접근할 수 있도록 하는 것이다.

원문이미지정보시스템은 위와 같이 여러 가지 기법이나 비율로 압축되어 저장된 광범위한 정보에서 사용자가 필요로 하는 최적의 정보를 찾을 수 있는 진보된 검색방법을 지원하는 기능을 갖추어야 한다. 우선 인터넷을 통하여 시간과 장소에 제한 없이 관내는 물론 외부 어디에서나 손쉽게 접근하여 각종 정보를 검색하고 활용할 수 있어야 한다. 그리고 인쇄매체인 문헌을 대상으로 독자적인 분류 및 색인체계에 익숙해진 이용자들에게 효과적으로 정보를 탐색해 줄 수 있는 에이전트 기반의 인터페이스가 필요하다.⁴⁾ 이러한 기능은 무한한 정보공간에서 이용자가 자신이 원하는 정보를 효율적으로 검색할 수 있도록 지원하는 가장 강력한 수단으로 작용할 것이다.

이밖에 원문이미지정보시스템이 갖추어야 할 조건으로 첫째, 인증(authentication)과 권한부여(authorization) 프로세서를 제공해야 한다. 인증은 어떤 사람의 신분을 확인하여 시스템에 접근할 수 있도록 허용하는 처리과정이며, 권한부여는 어떤 오퍼레이션을 허용할 것인지 허용하지 않을 것인지를 결정하는 과정이다. 이 두 가지 처리과정은 모든 컴퓨터 응용시스템에서 요구되는 공통적인 사항이라고 할 수 있다. 원문이미지데이터의 추가나 갱신, 수정작업을 위하여 이러한 기능이 반드시 제공되어야 한다.

둘째, 스캔된 원문이미지정보를 식별할 수 있는 식별기호 체계가 마련되어야 한다. 이미지정보의 식별기호는 정보 공간에서 해당 이미지정보로의 일관성 있는 참조를 가능하게 한다. 식별기호 체계는 식별기호를 부여하는 이미지정보의 단위나 정보공간에서 이미지정보의 위치 등을 고려하여 이미지정보의 식별기호에서 해당 이미지에 대한 정보를 추측할 수 있는 체계를 유지해야 한다.

셋째, 이미지정보로 구축된 문헌에 대한 메타데이터가 유지되어야 한다. 메타데이터

4. 에이전트란 일종의 비서와 같은 역할을 수행 해주는 프로그램으로 사용자의 기호나 습관 등을 학습하여 사용자가 어떤 행위를 하고자 할 경우 매뉴얼한 작업을 미리 대행하여 추리해 주는 것이다.

란 문헌의 일부분인 데이터가 아닌, 문헌에 관한 데이터를 뜻하며 네트워크로 연결된 정보 자원들의 기술(description)과 조직, 검색, 접근에 대한 핵심적인 역할을 담당한다. 메타데이터를 구성할 때에는 문헌의 특성을 표시하는 표준적인 속성을 사용함으로써 서로 이질적인 여러 시스템을 대상으로 정보를 검색할 때, 오류가 발생하지 않도록 해야 한다.

넷째, 원문이미지정보에 대한 refreshing이 지속적으로 이루어져야 한다. 이는 컴퓨터 기술의 진보에 따른 것으로서, 만일 refreshing 작업이 수행되지 않는다면, 새로운 기술의 도입시 기존의 시스템에 대한 기술적인 차이에 의하여 이전 시스템을 새로운 시스템에서 사용할 수 없게 될 수 있다. refreshing 작업은 기존 시스템에 수록되어 있던 데이터의 포맷을 현재의 시스템의 데이터 포맷으로 전환시켜 앞으로 등장하게 될 새로운 기술에 대비한다. 5) 따라서 지금까지 각 도서관에서 진행해온 도서관전산화 작업의 주요 핵심인 서지데이터 표준 포맷 즉, MARC 데이터베이스는 물론 CD-ROM이나 외부 데이터베이스의 서지정보와도 연계하여 활용할 수 있도록 하는 시스템이어야 할 것이다.

이와 관련하여 원문이미지정보 시스템의 외형적인 부분에 대한 시스템 모듈, 시스템 구조, 정보검색, 그리고 관련기술에 대한 기준을 설정하면 다음과 같다.

모듈화란 소프트웨어가 얼마나 잘 정의된 모듈로 구성되어 있는지를 나타내는 정도이다. 모듈의 질은 모듈분해력(modular decomposibility), 모듈결합력(modular composibility), 모듈이해력(modular understandability), 모듈연속성(modular continuity), 모듈보호(modular protection)로 평가될 수 있다. 질 좋은 모듈을 생성하기 위해서는 언어의 단위를 모듈의 단위로 사용하고, 모듈간에 작고 좁은 인터페이스를 사용하며, 또한 명시적인 인터페이스를 사용하고 모듈내의 정보를 은폐하도록 하는 모듈 생성원칙을 적용해야 한다.6)

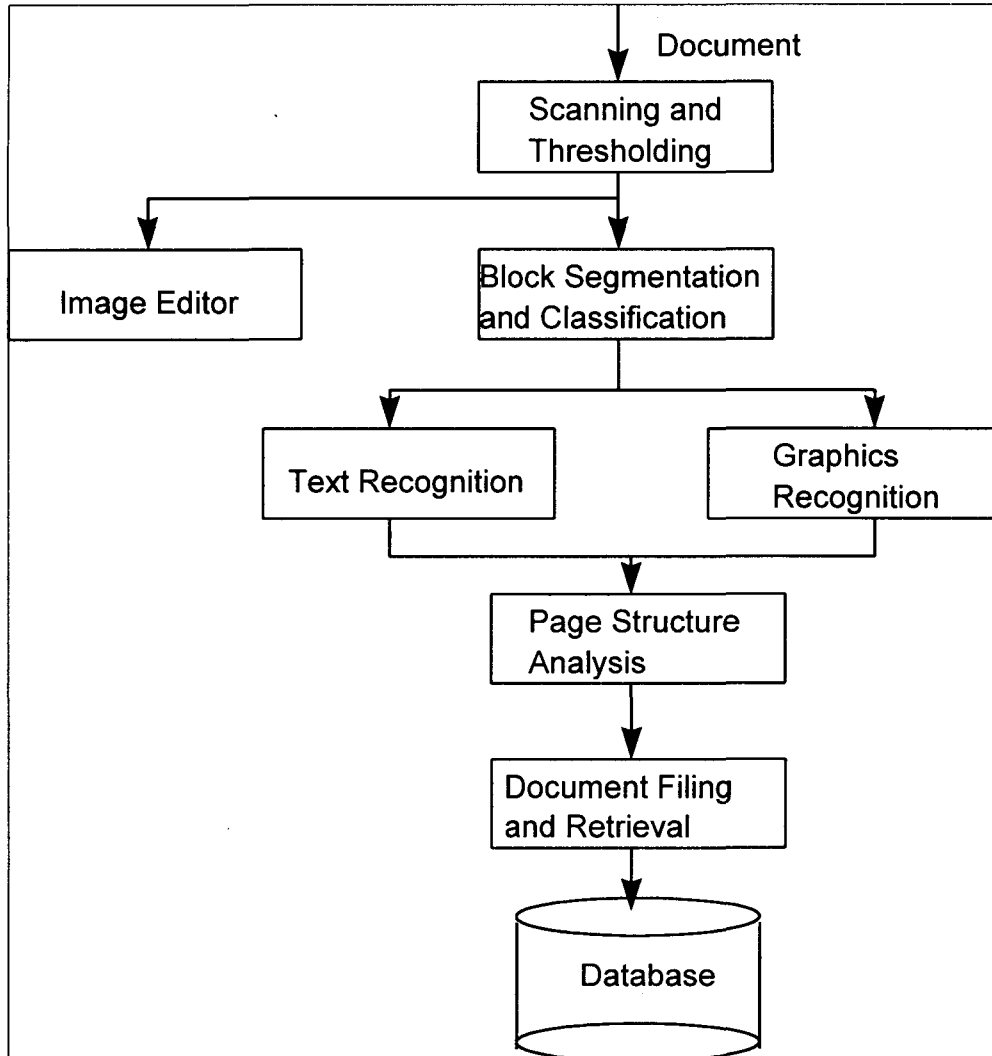
원문이미지 정보시스템은 [그림 1]과 같이 몇 개의 주요 모듈로 구성되며, 각 모듈에 대하여 설명하면 다음과 같다.

○ Scanning and Thresholding

문서를 스캔하고, Bitmap 형태로 표현하는 모듈

5. 김성혁, “디지털도서관의 문헌 특성 및 관리,” 한국문헌정보학회지 31(1), 1997

6. “KAERI 문서관리시스템 구축을 위한 문서이미지시스템 기술현황 분석,” 한국원자력연구소, 1991



[그림 1] 시스템 모듈 구조

(출처 : 한국원자력연구소 기술보고서, 1991)

- Block Segmentation and Classification
자동으로 이미지를 블록(block)으로 세그먼트화하여, 각 블록을 텍스트, 그래픽스
혹은 이미지로 분류하는 모듈
- Image Editor
스캔된 문서를 편집하는 모듈

○ Text Recognition

텍스트 블록을 프로세싱하고, 개개의 문자를 인식하는 모듈

○ Graphics Recognition

Bitmap 형태로부터 그래픽스의 블록 세그먼트를 벡터나 고급 그래픽스 기술 (Higher-level graphics description)과 같은 그래픽스 기호로 변환하는 모듈

○ Page Structure Analysis

Heading, summary, paragraph와 같은 복합문서 구조를 인식하는 모듈

○ Document Filing and Retrieval

문서 인덱스를 생성하고, 검색을 위한 질의어 인터페이스를 생성하는 모듈

III. 원문이미지정보시스템 비교분석 기준

원문이미지정보시스템은 도서관에 소장된 자료를 스캔(scan)하고 그 이미지를 저장(store), 검색(retrieve)하는 정보시스템으로써 1960년대까지의 마이크로필름 기술과 1970년대의 마이크로필름화된 문서의 인덱싱(indexing)과 검색을 지원하는 컴퓨터 보조시스템(Computer Aided Retrieval System)이 그 시효이다. 1980년대 초 DMS(Database Management System)는 멀티미디어 데이터베이스와 함께 연구 되었으며, 1980년대 중반부터 시장이 활성화되어 현재 그 응용이 급격히 확산되는 추세이고, 이미 연구소나 전문. 특수도서관을 중심으로 국내에도 여러 가지의 원문이미지정보 제공 시스템들이 개별적으로 구축되어 운영되고 있다.

본 절에서는 원문이미지정보시스템을 비교 분석하기 위한 기본적인 틀을 소프트웨어의 품질평가 방식을 참고하여 시스템의 기능 및 활용성을 평가하는 기준을 마련함으로써 서지데이터베이스 구축을 완료한 도서관에서 준비하고자 하는 원문이미지정보 제공 서비스의 질적인 부분에 있어서 선택의 방향성을 제시하고자 한다.

참고적으로 소프트웨어의 품질평가요소에는 사용자에게 뚜렷이 드러나는 요소로서 처리속도나 사용편이성(usability)과 같은 외부평가요소와 재사용성(reusability) 모듈화(modularity)과 같은 내부평가요소가 있다. 소프트웨어의 내부평가요소는 소프트웨어의 정확성을 높이기 위해 매우 중요하며, 내부평가요소들이 높은 소프트웨어를 개발하는 것이 소프트웨어공학의 목표이듯, 원문이미지정보시스템 또한 디지털도서관을 구축하기 위한 기본적인 도구로서 그리고 사용자위주의 범용성을 갖는 시스템을 선정하기 위한

평가요소에는 재사용성, 모듈화, 유지보수 용이성, 확장 용이성, 신뢰성 등을 고려한 평가방식이 구현되어 야 할 것이다. 무엇보다도 중요한 것은 시스템을 평가하는데 있어 냉정하고 객관적인 자세로 일관해야 하는 것이다.

원문이미지정보 시스템은 대용량의 데이터베이스를 구축하여 처리하는 것이 주된 기능이다. 따라서 원문이미지정보 시스템은 입력, 가공, 검색, 출력 및 열람, 시스템 유지 관리 또는 데이터베이스관리와 같은 기능 등으로 구성되어 있어야 한다.

각 기능별 세부내용으로는 첫째, 입력 및 가공기능은 종이형태의 정보를 디지털 정보로 변환하는 기능으로서 이미지의 스캔 및 압축 그리고 색인을 생성하여 데이터베이스에 전달하는 기능을 수행하게 된다. 따라서 시스템의 입력기능에는 다음과 같은 요소기술이 적용되어야 할 것이다.

1) 다양한 형태의 자료입력

스캐너를 통해 생성된 문헌 이미지파일 뿐만 아니라 추가적으로 다양한 파일형태자료 즉, 텍스트자료, Word자료(MS Word, HWP, PDF) 등을 수용하여 데이터베이스화 할 수 있어야 한다.

2) 다양한 이미지파일 형식을 지원

인쇄된 문헌자료를 이미지 형식으로 디지털화할 경우 이미지 파일 형식은 범용성을 갖는 다양한 형식을 지원해야 한다.

3) 서지정보와의 연계성

표제면, 저자소개, 목차, 색인, 삽화, 도표, 북자켓 등의 이미지는 이미 구축된 KORMARC 및 USMARC 서지정보의 모든 세부사항을 수용하며, 관련 태그(tag)와의 관계를 표시할 수 있어야 한다.

4) 문자인식(OCR) 및 색인(Indexing)기능

OCR(Optical Character Recognition) 방법은 시중 소프트웨어를 이용하여 이미지로 읽은 정보를 문자로 변환하므로 서지색인정보를 입력하기 위한 시간을 줄일 수 있도록 한다. 또한 인덱싱 방법에 있어서 다음과 같은 기능들을 수용 할 수 있어야 한다.

- 키 엔트리(Key entry) : 키워드(keyword)가 스캐닝할 때 이미지 필드(image field)와 연관되게 하는 방식
- 데이터베이스 필드를 이용하는 방식(database field) : 서명, 저자명, 출판사, 청구 기호, 주제어 등
- 자동 키 생성방식(Automatic key creation) : 자동적으로 생성된 특정키가 연관되

계하는 방식

- 키 인식방식(Key recognition) : 키워드, 숫자가 지능적 문자인식에 의해 이미지로 부터 추출하는 방식
- Markup Language : 이미지가 특정 키워드, section, 원 문헌의 페이지를 Markup 언어에 의해 연관되게 하는 방식
- 하이퍼미디어 링크에 의한 방식(Hypermedia Link) : 이미지가 아이콘(Icon), 버튼 (Botton), 키워드 등을 이용하는 하이퍼미디어 도구(Hypermedia Tool)를 통해 연 관되게 하는 방식

5) 데이터파일의 크기를 축소하는 기능

이미지데이터는 크기가 방대하여 저장용량을 최소화하여 대량의 데이터를 최적화하여 전송속도를 증가시키기 위한 고도의 파일 압축기술을 포함한 시스템이어야 한다.

6) 시소러스(Thesaurus) 지원

이미지와 오브젝트를 연결할 수 있는 통제 어휘를 사용하여 이미지정보 검색의 효율성을 제고하여야 한다.

7) 외부자료 반입

자관에 소장되어 있으나 이미지파일로 변환하지 못한 정보에 대하여 도서관 분담목록 기능에 의거한 교환 이미지데이터를 자관 데이터베이스로 변환할 수 있어야 한다.

8) 이밖에 스캐너 입력형식, 입력용지 크기, 입력밀도, A4사이즈 입력시 소요시간 등에 대한 세밀한 배려가 있는 시스템이어야 한다.

둘째, 검색기능은 스캔되어 저장된 이미지정보를 빠르고 쉽게 Access하기 위하여 다음과 같은 기능을 포함하여야 한다. 현재까지 시판되어지는 이미지검색시스템은 원문이미지의 내용을 기반으로 하는 검색보다는 이미지 자체에 대한 서지정보를 별도로 보관하고 있어 이를 통한 검색이 대부분이다. 따라서 다음에서 구술되는 것은 텍스트 색인 검색이 주가 되는 환경이다.

1) 인터넷(Internet) 환경 지원

인터넷의 웹(World wide web)을 이용하여 시간, 장소에 제한 없이 이용자가 원하는 자료를 검색할 수 있어야 한다.

2) 메타데이터와 연계하여 검색

원문이미지정보의 이미지 전체를 메타 데이터와 연계하여 관리 및 검색할 수 있도록

한다. 원문이미지정보는 문헌의 특성에 따라 임의로 구조화할 수 있도록 한다.

3) 네트워크 표준프로토콜 수용

원격데이터베이스 레코드 접속 및 서로 다른 컴퓨터에 접속하여 검색의 호환을 이룰 수 있는 국제표준규격의 Z39.50을 수용하여야 한다.

4) 서지데이터베이스와 연동

현재까지 구축되어 온 서지데이터베이스 검색과 연동하여 관련 문헌의 원문이미지정보에 접근 할 수 있도록 한다.

5) 다양한 검색 연산자 사용

서지데이터베이스를 검색할 때와 마찬가지로 검색어를 Boolean 논리연산자 및 위치 연산자, 관계연산자 등을 사용하여 원하는 정보를 신속하게 검색할 수 있도록 한다.

6) 시소러스 검색

원문헌에 이용되고 있는 자연어와, 색인자 및 이용자가 사용하는 자연어에서 보다 통제된 시스템언어로 변환하는 경우에 이용되는 시소러스를 구축하여 이용자 검색과정에서 효율적인 통제어휘집을 제공하여야 한다.

셋째, 입력 및 검색기능과 같이 필수적인 시스템 기능 이외에 데이터베이스관리 와 시스템의 보안에 대한 기능들이 부가적으로 고려되어야 할 사항인데, 그 내용은 다음과 같이 구분할 수 있다.

1) 이용상태 파악

현재 검색중인 데이터베이스, 로그인 시간, 로그인 상태 등에 대한 사용자정보 관리를 통하여 데이터베이스의 효율적 관리를 수행해야 한다.

2) 통계 정보 제공

데이터베이스별 접속회수, 활용도 등을 나타내주는 관리정보를 그래픽이나, 수치별로 제공하여 프로그램 관리자에게 항상 신속한 시스템 관리기능을 부여한다.

3) 데이터베이스 관리 기능

시스템 에러나 네트워크상의 제반 문제 등에 의한 자동 백업(Backup), 데이터베이스 원상복구 기능 등을 포함해야 한다.

4) 보안기능

무자격자에 의한 데이터베이스의 수정, 유출을 방지하기 위해 사용자별 접근 범위를 지정하여 각 기능별로 통제할 수 있도록 한다.

이와 같이 원문이미지정보시스템이 갖추어야 할 필수적인 요소 외에 시스템의 구성 기기와 구성호환성에 관한 사항과 시스템의 확장성도 충분히 검토되어야 할 것이다. < 표 1> 은 원문이미지정보 시스템 비교분석시 검토되어야 할 항목에 대하여 정리하였다.

<표 1> 원문이미지정보 시스템 분석기준

구 분	세 부 기 능	비 고
입력 및 가공	<ul style="list-style-type: none"> ① 이미지 파일 뿐만 아니라 텍스트, Word 형 태자료의 입력기능 ② 이미지 파일 형식의 범용성 ③ MARC 형식으로 구축된 서지 데이터와 의 연계성 ④ OCR기능 및 인택싱 방법의 다양화 ⑤ 이미지 데이터의 압축/복원 기술 ⑥ 시소러스 지원 기능 ⑦ 분담목록, 즉 이미지데이터의 호환성 수용 ⑧ 스캐너 및 용지등에 구애받지 않는 범용성 을 지닐 것 	
검색기능	<ul style="list-style-type: none"> ① 인터넷 환경 지원 ② 이미지 전체에 대한 메타데이터와 연계하여 검색 가능 ③ 네트워크 표준 프로토콜 수용 ④ MARC 데이터와 연계 검색 가능 ⑤ 검색연산자의 다양화로 이용자 검색 수준정 도에 따라 신속검색 지원 ⑥ 시소러스 검색 Tool 제공 	
데이터 관리 및 보안 기능	<ul style="list-style-type: none"> ① 데이터베이스 이용 상황 파악 ② 데이터관리자를 위한 각종 통계 정보 생성 ③ 자동 백업 및 복구 기능 ④ 시스템 보안 기능 ⑤ 이미지 데이터의 압축/복원 기술 	

IV. 상용 원문이미지정보시스템의 분석

4.1 TG-Imagebase

삼보정보시스템과 오름컴퓨터가 공동으로 개발한 시스템으로 현재 국내 대학 및 전문, 특수도서관에서 사용중인 도서관자동화 프로그램인 TG-VintageLAS의 편목, 검색 모듈과 연동하여 원문이미지자료를 입력, 편집, 저장, 검색할 수 있도록 설계되어 있다. 따라서 원문이미지정보의 독자적인 구축을 위한 모듈화가 진행되어야 할 것이다.

TG-Imagebase는 도서관이 새로운 기능을 도입하여 그 활동 영역을 확장하고, 광범위한 정보봉사(Comprehensive Information Service)가 가능하도록 할 뿐 아니라 업무의 단절이나 두 개의 별도 시스템을 운영하는 부담을 줄이고, 기존의 자료 및 정보를 지속적으로 활용하고, 비용 대 효과면에서 경제적인 디지털도서관을 단계적으로 구축하기 위한 첫 번째 단계의 시스템이 될 수 있다.

이 시스템은 기존의 도서관 자동화 시스템이 지원하는 모든 기능과 디지털 기술이 갖는 정보 추적, 검색, 통신 등의 다양한 장점들이 종합되어 이루어진 하나의 시스템으로 사용자 인터페이스를 고려하여 메뉴 접근 방식이나, 검색식의 조합 등은 윈도우 환경을 최대한 활용하는 기능들을 수용하고 있다.

또한 KORMARC 및 USMARC의 모든 세부사항을 수용하고, 각 자료별 특성에 따라 자관용 표시기호(9XX)를 추가로 지정하여 사용할 수 있다. 작업대상 자료는 도서관에서 소장하는 단행본, 연속간행물, 각종기사 등 모든 자료를 망라적으로 포함하여 처리할 수 있다.

4.2 Image-Plus

(주)한국파일링에서 1994년 초반기부터 멀티미디어 대용량의 저장장치인 CD-ROM Jukebox, CD-Recorder, CD-Duplicator 등의 광파일링 시스템을 LG전선, 문화체육부, 현대중공업등에 공급하면서 축적한 기술을 바탕으로 개발된 이 시스템은 인터넷을 통하여 시간 및 공간의 제약 없이 손쉽게 정보에 접근하여 다양한 정보요구(인쇄매체, 멀티미디어 형태의 동영상 등)를 자연어 인터페이스로 하여 하이퍼텍스트(Hyper-text)나 하이퍼미디어(Hyper-media Link)검색이 가능하도록 구현된 시스템이다.

Image Plus는 표준 검색프로토콜인 Z39.50을 채택한 전문검색엔진을 지원함으로써 향후 디지털도서관간의 상호운영을 위한 자료공유시 이용자는 인터넷을 통해 한번의 질의어로 이 프로토콜을 채택한 모든 도서관이 구축한 DB의 목록과 원문이미지정보까지 검색할 수 있으며, 스캐너로 읽어들인 이미지에 대하여 문자인식(OCR)을 수행함으로써, 자료를 문자화하는데 드는 시간과 비용을 최소화 할 수 있다는 특징을 갖고 있다.

Image Plus는 이미지 정보 이외에도 ASCII 텍스트 파일, HWP, Exel 파일 등 다양한 형태의 파일들을 한 시스템 안에서 통합적으로 관리해주며, UNIX계열 운영체제 및 각종 관계형 데이터베이스 관리시스템을 기반으로 모두 작동될 수 있는 호환성 및 확장성을 지니고 있다. 그리고 이 시스템은 서지데이터베이스 Server와는 별도의 이미지 Server를 독립적으로 운영하므로 원문이미지정보관리 전용 시스템이라 할 수 있다. 따라서 기존의 MARC을 이용하여 구축되어 온 서지데이터베이스와의 연계를 위한 별도의 모듈이 설치되어야 한다는 단점을 지니고 있으나, 현재 각 단위 도서관별로 시행중이거나 계획중인 디지털 도서관 프로젝트를 실현하는 데는 별 문제가 없는 것으로 분석된다.

4.3 비교분석 및 평가

지금까지 본 고에서는 과학적 정보관리 및 네트워크를 기반으로 하는 정보공유 체계를 유지하고, 시간과 공간을 초월한 원격 서비스 체계를 구축하여 서비스 중심의 도서관으로써 정보의 영구적 보존은 물론 정보 유통의 신속성, 망라성, 편의성, 접근성을 효율화하기 위한 디지털 도서관 구축의 방법으로 원문이미지정보 시스템 구축에 필요한 이론적 고찰과 국내에서 개발되어 도서관에 설치되고 있는 시스템을 살펴보았다.

컴퓨터 기술과 매체의 발달로 디지털화의 방법이 가장 자연스럽게 떠오르는 Solution으로 대두되었고, 이미 몇몇 도서관에서는 대용량의 하드 디스크나 광 자기 디스크를 이용하여 학위논문이나 연구보고서를 원문이미지화 하여 보관하고 검색하고자 하는 단순한 저장논리의 시스템들이 선보이고 있다. 그러나 이러한 방식들은 단순히 저장 공간의 절약과 보존의 영구화라는 측면에서 고려된 것이므로 이용자측면에서 손쉽게 상세한 정보검색을 할 수 없는 CD-ROM화에 그치고 있다.

원문이미지정보관리 시스템은 방대한 양의 이미지 데이터를 저장하고 처리하는 광디

스크, Juke Box 및 주 프로세서, 입출력 장치인 고성능 디스플레이어 및 스캐너 등의 하드웨어와 DBMS, 이미지 관리 소프트웨어 등으로 요약된다. 본 고에서 비교 분석되어진 두 시스템은 하드웨어의 특성에 따라 그 기능의 활용성을 표현할 수 있었다. 그리고 현재 하드웨어의 대부분이 미국이나 일본의 제품을 사용할 수밖에 없는 실정이므로 한글처리와 관련된 소프트웨어 기술의 다양성을 모색해야 하며, 단순히 메타정보만을 위주로 검색되는 단순 이미지데이터베이스 형태의 시스템에서 원문의 글자를 인식할 수 있는 전문검색(Full Text)과의 연계성도 고려되어야 할 과제인 것이다.

1990년대 들어서면서부터 본격적으로 시장이 형성된 광과일링시스템에 비하면, 초기 단계로 접어드는 원문이미지정보관리 시스템은 장기적 자체 계획에 의한 개발과 꾸준한 시스템 업그레이드 및 기존 광과일링 시스템으로 축적된 데이터와의 호환성 등을 추구하는 지속적인 개발 노력이 요구된다.

본 절에서는 디지털 도서관이라는 대형 사업을 원활히 수행할 수 있도록 대학도서관 또는 전문, 특수도서관에서 원문이미지 정보시스템 구축시 최선의 선택을 지원하기 위하여 현재 시점을 기준으로 국내에서 개발되어 도서관 또는 자료실 등에 보급되고 있는 Package S/W인 두 시스템 즉, TG-Imagebase 와 Image Plus로 제한하여 디지털 도서관이 갖는 기능 및 이용자 인터페이스를 기조로 하여 분석한 것이다. 본 고에서 서술하였던 원문이미지정보 시스템이 갖추어야 할 기능을 중심으로 분석한 결과를 <표 2> 와 <표 3> 으로 압축하여 설명하였다.

본 시스템 비교표는 각 회사에서 제작한 기술제안서와 현재 시스템을 활용하여 정보 서비스를 제공하고 있는 인터넷상의 시스템에 직접 접근하여 본 고에서 제안하는 비교 기준에 근거하여 작성되었다.

V. 결 론

초고속망을 기반한 정보통신 분야의 디지털 도서관은 우선적으로 정보 전달에 대한 병목현상을 해결할 것이며, 정보 전달에 대한 구문(Syntactic)과 정보 표현에 대한 의미론의 표준화로 정보의 투명성 및 공유가 보장되어 정보축적 형태별로 보는 도구가 존재해야 하는 문제와 검색엔진에 의존한 정보검색의 비효율성이 해결 될 것이며, 시공간을 초월한 정보검색이 가능하게 될 것이다. 그러나 디지털 도서관을 생각할 때 모든 자료, 기술, 업무 즉, 영구적인 자료와 일시적인 자료, 불변의 자료와 가변의 자료, 디지

<표 2> 입력기능 비교

구분 \ 시스템명	TG-Imagebase	Image Plus
자료 입력 (다양한 형태의 자료 입력)	지원 가능	지원 가능
다양한 이미지 파일 형식 지원 (TIF, PCX, JPG, BMP 등)	지원 가능	지원 가능
서지정보 연계	서지정보관리 시스템인 TG-VintageLAS와 연계	별도의 이미지 서버로 원문이미지정보만을 관리 하는 방식임
문자인식(OCR) / 색인(Indexing)	아직까지는 연계되지 못 하고 있음	시중 S/W를 사용하여 원문정보의 초록을 OCR 처리하여 검색 키워드로 제공
데이터 압축기능	흑백문서 : TIFF 컬러문서 : GIF	- CCITT G3/G4지원 - JPEG 지원
시소러스 기능	서지정보시스템 내장	지원함
외부자료 반입	지원함	지원함
스캐닝 방식	고속(A4, A3)	고속(A4, A3)

털기술과 인쇄매체기술, 개인에 의한 정보활동과 협동에 의한 정보활동 등 모두를 고려하는 도서관을 생각하게 된다. 하지만 이와 같은 종합적인 디지털 도서관의 구축이란 거의 불가능하며 의미가 없는 경우가 많다.

본 연구에서는 디지털 도서관을 구축하기 위한 기본적인 정보구축 및 서비스를 제공할 수 있는 원문정보의 이미지데이터베이스화를 통하여 단순히 OPAC에 의한 소장정보와 사서를 통한 원문제공 서비스에 만족할 수밖에 없었던 정보 이용자에게 새로운 정보 제공기술을 구현 하고자 하였다. 원문이미지정보 시스템은 디지털 도서관의 기반이 될 수 있으나, 지금까지 도서관에서 기술하여 온 MARC형식을 수용하지 못하는 범

<표 3> 검색기능 비교

구 분 \ 시스템명	TG-Imagebase	Image Plus
인터넷 지원	제안서 상에는 LAN용, WEB용으로 구분되어 있으나, 현재 활용 사례가 없음.	CGI를 지원하며, 인터넷 상에서 검색서비스 제공중 (LG상남도서관, 홍익대 도서관)
메타데이터 연계	지원함	지원함
네트워크 표준프로토콜 수용 여부	개발중	Z39.50지원
서지데이터베이스와 연동 검색	서지 데이터베이스인 TG-VintageLAS와 연동하여 검색 가능	순수 이미지 검색지원
검색 방법	Boolean검색, 일치검색, 절단검색, 히스토리검색, 등록번호 검색	Boolean검색, 절단검색, 히스토리검색, 인접어검색, 제한검색, Pattern Matching 검색, 퍼지(Fuzzy) 검색
시소러스 검색	지원함	지원함

위에서는 단순히 이미지데이터베이스 시스템으로 불릴 수밖에 없다.

따라서 정보 네트워크나 도서관자동화 등의 전자정보 서비스 환경은 갖추어져 있지만, 그 내실을 다질 수 있는 데이터베이스의 구축 및 이용자 인터페이스 개발, 도서관 시스템 환경의 개선, 원문이미지 정보에 대한 마인드 개선 등이 도서관에서 갖추어야 할 부분이며, 기술적인 면에서는 지금까지 이어온 MARC 데이터베이스와 연동하여 데이터를 구축하고, 검색할 수 있는 표준안이 마련되어야 할 것이다.

본 연구에서 인용된 TG-Imagebase와 Image Plus 시스템들은 이제 막 시작하는 우리나라의 디지털 도서관 구축에 동참하여 외국의 신기술과 우리나라만의 전통적인 기

술의 접목이 필요하며, 사진이나 그림자료에 대한 검색을 위하여 내용을 기반으로 하는 검색시스템이 추가되어야 할 것이다. 또한 현재의 환경에서 데이터의 축적은 많은 시간과 막대한 비용이 드는 일이므로 분야별, 매체별, 서비스별, 지적재산권의 특성별로 많은 도서관들이 데이터의 디지털화를 분담하여 원문이미지정보를 빠르고, 표준형태로 구축하는 것이 우리 도서관의 과제로 남을 것이다.

참 고 문 헌

- 김성혁, 디지털도서관의 문헌특성 및 관리, 한국문헌정보학회지 제31권 제1호, 1997.
- 김순자, 원자력연구소의 전자도서관 구축전략, '96과학기술정보관리협의회 연례학술세미나 발표자료집(2), 1996. ,
- 김태운 외, 데이터의 유사성을 이용한 실제 색상화상의 압축기법, 한국정보처리응용학회 논문지, 제1권 제3호, 1994
- 박재영, 전자도서관 모형 및 구축현황에 관한 연구, 연세대학교 대학원 석사학위논문, 1995.
- 박태진, 브라우징과 전문검색을 지원하는 이미지 검색시스템의 설계 및 구현, 연세대학교 대학원 석사학위논문, 1994.
- 박한중 역, 화상정보처리기술, 성안당, 1996.
- 박현제, 인터넷 입문, KRNET '96특강자료집, 1996.
- 석창호, 멀티미디어 데이터베이스에서의 검색 방안에 관한 연구, 광운대학교 대학원 석사학위논문, 1993.
- 신병석 외, 특징 기반의 3차원 물체정보 검색시스템 모델, 정보과학회지 3(1), 1995.
- 오현주, 이미지정보 처리기술, 한국기업전산원, 1993.
- 우학명, Paper 자료의 디지털화 형태에 따른 구동시스템 및 비용.효과 분석, 국회도서관보: 34(1), 1997.
- 유양근, 도서관 환경의 변화와 미래도서관, '96 한국디지털도서관 포럼 자료집, 1996.

- 윤석민, 멀티미디어 데이터베이스의 보안 특성, 동양공업전문대학 논문집 제19집, 1996.
- 이상의 외, 광 기술을 이용한 연구보고서 관리시스템 구축, 정보관리학회지 제9권 제1호, 1992.
- 임미경, 화상정보시스템 구축에 관한 연구, 연세대학교 대학원 석사학위논문, 1996.
- 정현수, 정보통신분야의 디지털라이브러리 구축과 전망, '96과학기술정보관리협의회 연례학술세미나 발표자료집(2), 1996.
- 최기호 외, 내용을 기반으로 한 이미지 검색 데이터베이스 시스템, 정보과학회지 3(1), 1995.
- 최석두, 전자도서관의 개념과 발전추세, 제24회 전국도서관대회 발표자료집, 1996.
- 최학성, JPEG 압축 표준안을 이용한 멀티미디어 화상저장 시스템에 관한 연구, 광운대학교 대학원 석사학위논문, 1993.
- 한국과일링, 디지털도서관 구축기술, 한국과일링, 1996.
- 한국원자력연구소, KAERI 문서관리시스템 구축을 위한 문서이미지시스템 기술현황 분석, 한국원자력연구소, 1991.
- 홍기채, 멀티미디어 기술에 기반한 도서관 최신정보서비스시스템의 설계 및 구현, 충남대학교 대학원 석사학위논문, 1993.
- A. C. Hung, Image Compression : The Emerging Standard for Color Image, IEEE Computer 22(12), 1989.
- David M. Choy et al.(1995), The Almaden Distributed Digital Library System, IN : Advance in Digital Libraries, Preliminary version, New-York, Springer-verlag, 1995.
- "Digital Libraries : Issues and Architectures, "http://www.csd1.tamu.edu/DL95/papers/nuernberg.html.
- David R. Guenette, Document Imaging, CD-ROM and CD-R, CD-ROM Professional, April, 1996.

David Volk Beard, Computer Human Interaction for Image Information Systems,
JASIS 42(8), 1991.

T.Sheter, Birth of the BLOB, Byte Feb, 1990.

W.B.Green, Introduction to Electronic Document Management System, Academic
Press, 1993.