

광기술을 이용한 연구보고서 관리시스템 구축

A Study on Construction of Technical Reports Management System Using Optical Technology

이 상 현*
김 익 철*

□ 목	차 □
1. 서론	4. 연구보고서 관리시스템
2. 이미지 처리기술	4.1 시스템 구성도
3. 국내상용시스템	4.2 서지정보 처리기능
	4.3 원문이미지정보 처리기능
	5. 결론

초 록

본 연구에서는 광기술을 이용한 보고서 서지정보와 원문정보를 관리할 수 있는 보고서관리시스템을 개발하였다. 시스템은 현재 상업적으로 보급되고 있는 문서용 광화일링시스템과의 비교 평가를 통하여 보고서 관리에 적합하고, 도서관 업무의 효율화에 직접적으로 기여할 수 있는 구조로 설계되었으며, 디지털이미지 처리기술, MARC표준, 영문OCR등의 기술이 사용되었다.

키워드 : 디지털이미지, 광기술, 전문데이터베이스, 표준화, 기술보고서, MARC, 광학문자인식

ABSTRACT

In this study, a technical report management system using optical technology is described in detail. This management system is designed for both bibliographic (character) and full-text (image) information.

Several optical filing systems already on the Korean market are scrutinized and compared with standard functions in order to build a more efficient management system for technical reports which can be easily integrated into existing KRISS library automation system.

For that purpose, up-to-date technologies (i.e., digital image processing (DIP), MARC standards, and optical character recognition (OCR), etc.) are applied to this system.

Keywords: digital image, optical technology, full-text database, technical reports, MARC, OCR

1. 서 론

보고서(technical report)는 연구 및 개발업무의 진행과정 또는 결과로 부터 얻어지는 문서로서 연구의 최신정보를 제공해 줄 뿐만 아니라 일반적으로 잡지기사, 학술회의 발표문, 혹은 일반논문보다 연구수행과정 전반에 걸쳐 보다 자세히 기술하기 때문에 정보이용자 또는 연구자의 입장에서는 특정 연구과정을 평가하고 재활용하기에 충분한 데이터를 얻을 수 있고, 진행중인 연구과제들의 추세를 확인하여 연구방향 설정, 연구과제 도출업무 등에 필수적인 정보를 얻을 수 있다.

또한 연구관리자 또는 스폰서기관의 입장에서는 연구관리 및 평가를 통한 연구활동 경제성분석, 연구정책 수립업무에 이용하므로써 중복연구, 유사연구의 방지 및 체계적 연구관리 뿐만 아니라 기관의 연구와 그 지원활동에 대한 기록문서로서 영구보존 효과를 얻을 수 있다.

특히 보고서는 연구기관의 연구결과를 전달하는 기본적인 전달매체로 이용됨에 따라 그 관리·보급시스템의 효율성에 의해 연구성과의 사회기여도가 좌우된다고 하겠다.

그러나 보고서는 그 발행목적이 주로 연구과제로부터 얻어진 정보를 소속기관, 스폰서, 또는 한정된 기술집단에 신속히 통보할 목적으로 작성되는 것이기 때문에 일반인에게 널리 보급되기에 적합한 형태나 방법을 취하지 못하는 경우가 대부분이며 그 결과 제한된 집단 밖으로의 정보전파가 불가능한 경우가 많다. [1] [2]

전세계적으로 가장 많은 보고서를 생산하고 있는 미국의 경우 보고서에 관한 정보의 주지, 유통, 보급업무는 국가적 관심사로서 NTIS같은 클리어링하우스서비스, GPO의 판매프로그램과 납본도서관프로그램등의 다양한 국가차원의 프로그램을 통하여 일반대중의 이용을 적극 지원하고 있으나, 이 흐름의 기본구조를 형성하고 있는 단위는 각 정보생산기관들이 구축한 전자사서함, 전자계시판, FTP(Anonymous File Transfer Protocol)를 통한 화일의 입수 및 다운로드가 가능한 NREN(National Research and Education Network)이용서비스 그리고 다이알업으로 이용가능한 온라인데이터베이스 서비스 등이라고 할 수 있다. [3]

국내에서는 정부출연연구기관들이 보고서의 주 생산기관이 되고 있으며 매년 10,000종 이상의 보고서를 발행하고 있고, 연 13%이상씩 발행량이 증가하고 있는 것으로 추정되고 있다. 물론 이 발행량은 미국 등의 선진각국과

1. C.P. Auger. "Reports, Patents, Standards and other Special Material," Handbook of Special Librarianship and Information Work (5th ed., London:Aslib, 1982), pp.102-110.
2. National Information Standards Organization (U.S.), Standard Technical Report Number(STRN) Format and Creation(ANSI/NISO Z39.23-1990). (New Brunswick:Transaction Publishers, 1990), p.1
3. Peter Hernon, and Charles R. McClure. "Dissemination of U.S. Government Information on CD-ROM and Other Forms," CD-ROM Professional, 5(2) (1992. 3), pp.67-70.

비교할 때 매우 적은 규모이지만 우리나라 국가과학기술투자의 약 1/3에 해당하는 연구의 결과물로서 최대한 활용하여야 할 국가적 자원이다.

그러나 국내에는 NTIS, GPO와 같은 일반 대중을 위한 국가차원의 체계적 정보서비스의 역사가 없을 뿐만 아니라, 이런 활동의 기본 단위가 되어야 할 보고서생산기관들의 자신들을 위한 체계적 정보관리시스템 개발의 노력 또한 극히 미미하였다.

더우기 보고서관리시스템을 어떤 수준에서든 운영하고 있는 일부 기관들에서조차도 보고서에 관한 서지정보처리에 대해 충분한 연구를 거쳐 시스템을 개발한 사례를 발견하기 힘들며, 더 나아가서 국가적인 보고서관리체제구축에 선행되어야 할 관련기술의 표준화연구는 요원한 실정이다. [4]

따라서 보고서정보관리시스템에 관한 한 초보적 단계인 현실점에서 적절한 표준화 기술들을 도입한 보고서관리시스템의 모델을 정립하고 이를 기본단위로 한 국가적인 체제로의 발전방향이 제시되어야만 할 것이다. 아울러 이러한 방향제시에는 다음과 같은 최근의 기술동향이 반드시 고려되어야만 한다.

이미 대부분의 보고서 또는 문서의 작성이 워드프로세서에 의해 이루어지고 있으며 이때 생산되는 화일로부터 출력한 최종 출력물을 인쇄사식과정에 사용하거나, 또는 점차적으로 화일로부터 직접 사식을 하는 전자출판 방식으로 옮겨가는 경향이 뚜렷하다. 그리고 문서의 작성, 전달, 가공, 재생산 등 관련된 전 공정을 통합환경 (Integrated System Environ-

ment)으로 구축하여 공정과 공정 사이에서 발생할 수 있는 정보의 손실 및 불필요한 중복생산을 방지할 뿐만 아니라 업무처리의 전 공정을 한자리에서 한 기기를 통해 수행하며 하나의 정보자원을 다수의 이용자가 각기 분산된 장소에서 동시에 이용할 수 있는 사무자동화가 시도되고 있다. 이에따라 문서 또는 보고서에 관한 공정의 시발점이라 할 수 있는 문서작성기술에 있어서도 과거의 문자처리에 치중했던 워드프로세서로부터 보다 고가이지만 도형, 화상 처리가 가능한 탁상출판시스템(DTP)으로 발전하고 있으며 음성 및 영상데이터까지 처리할 수 있는 멀티미디어의 실용화도 임박한 실정이다.

특히 각기 다른 문서작성시스템간의 호환성을 위하여 국제적으로 ODA(Office Document Architecture), SGML(Standard General Markup Language)등의 표준화작업이 활발히 진행됨에 따라 보고서 작성에 있어서는 조만간 현재의 기술수준에서 추구할 수 있는 최선의 목표에 근접할 수 있을 것으로 판단된다. 따라서 이러한 방식으로 작성된 문서나 보고서는 이미 작성단계에서부터 기계가독형화일로 존재함에 따라 변환처리 없이 즉시 사무자동화 또는 보고서관리시스템에서 운용될 수 있을 것이다.

그러나 과학기술정보는 일반정보와는 달리 외래어와 그리스 문자, 첨자, 특수기호뿐만 아니라 복잡한 수학공식, 화학 방정식, 도

4. 연구단지정보관리총람/연구단지정보관리협의회 기술위원회편, (서울:동협의회, 1992). p.254.

면, 그리고 사진 등을 포함하고 있다. 물론 이러한 문제를 해결하기 위하여 KS C 5601 한글카드가 제공하는 특수문자 및 기호, 아래한글과 같이 소프트웨어적으로 보다 편리하게 설계된 관련기능, 보다 고급의 문서작성 또는 그래픽 소프트웨어팩키지가 제공하는 그림 및 기호 라이브러리와 이들을 외부로 부터 도입하여 문서에 추가할 수 있는 기능, 그리고 troff과 같이 이런 정보의 처리를 위한 전용 소프트웨어팩키지가 사용되고 있다. 그러나 처리기능의 범위가 포괄적이지 못하거나, 숙달되기에 많은 시간과 관련 분야에 대한 전문지식이 필요하거나, 자신의 소프트웨어 이외에는 재현이 불가능하거나, WYSIWYG(What You See Is What You Get)이 아니기 때문에 이용자 편의성이나 친밀감에서 효율성이 떨어지는 등의 부정적인 점 때문에 제한적으로 사용되고 있을 뿐이다.[5]

따라서 앞으로도 상당기간 동안, 적어도 앞에서 언급한 기능들의 처리가 완벽한 보급형 탁상출판시스템의 대중화가 이루어질 때까지는 과학기술분야의 문서 및 보고서가 작성시부터 전문(full text)데이터베이스 구축이 가능한 완전한 디지털 데이터 형태로 생산되기 어려울 것으로 예측된다. 이는 보고서관리시스템이 보고서작성 시점부터를 기능에 포함시키기에는 시기상조이며, 결국 원문정보의 데이터베이스 구축방법도 이미지 처리방식이 되어야 함을 뜻한다.

더우기 쿠프스와 리브랜드의 연구에 따르면, 정보처리기의 광범위한 보급에도 불구하고 대부분의 기관에서 보유하고 있는 자료

중 컴퓨터에 의해 즉시 처리될 수 있는 문서는 10%에 불과한 것으로 나타났다.[6]

이는 진정한 의미에서의 사무자동화를 이루기 위해서 필수적으로 많은 양의 문서 또는 보고서를 기계가독형의 화일로 소급변환하여 입력하여야 할 뿐만 아니라 앞으로도 상당기간 동안 많은 양의 문서 또는 보고서가 비기계가독형으로 작성될 것임을 뜻한다.

결국 현실점에서 효과적인 보고서관리시스템은 대량정보의 효과적인 소급변환과 이미지정보의 효율적 처리에 두어야 할 것이다.

2. 이미지 처리기술

이미지 처리기술은 보고서관리시스템의 핵심에 해당하는 부분으로서 여기에 사용되는 기술의 종류에 따라서 시스템 전체의 설계가 좌우된다.

80년대 이후 보고서관리시스템의 주류는 원문 이미지를 마이크로필름, 마이크로피쉬, 그리고 소수이지만 COM(Computer Output Microfilm)방식으로 변환, 저장하는 것이다. 이 방식은 검색 효율성의 제고보다는 보관체제 및 공간의 효율적 활용에 주목적을 둔 것으로서 마이크로필름 또는 COM 장비에 의한 원문 이미지정보 이용과 컴퓨터에 의한 서지정보검

-
5. 유관중, 강지훈, 변승환, 데이터효율성 제고를 위한 사용문자에 관한 조사연구, (대전:충남대학교, 1991), pp.68-77.
 6. Marc R. D'Alleyrand, Image Storage and Retrieval Systems, (New York:McGraw-Hill, 1989), p.65.

색으로 이원화된 불편함에도 불구하고 아직도 생존능력을 갖고 있는 것으로 평가되고 있다.

그러나 최근 소규모 컴퓨터의 성능 대 비용의 경제성과 이미지 처리기술의 실용성이 급속히 높아짐에 따라 컴퓨터에 의한 이미지 저장 방식이 보고서관리시스템에도 채택되기 시작하였다.

이에따라 현재 보고서관리시스템에서 사용할 수 있는 이미지 저장기술에는 다음과 같은 세가지 방식이 있다고 할 수 있다.

- 1) 아나로그방식-전형적인 마이크로필름 저장방식.
- 2) 디지털방식-컴퓨터를 이용한 저장방식.
- 3) 아나로그·디지털 혼합방식-Digitized Computer-Assisted Retrieval (D-CAR)로 또는 전자식마이크로필름 제작이라고 불리는 방식으로 이미지 저장에는 마이크로필름방식을 사용하고 이미지 전송에는 디지털방식을 채택.

이들 중 3)의 방식은 매우 고가일 뿐만 아니라 시스템 구성방식에서도 앞으로의 생존력이 길지 못할 것으로 예측되고 있으나, 1)과 2)의 방식은 이미지 저장기술을 선택하고자 하는 사람들에게 아직도 비교평가의 대상이 되고 있다. [7]

이들 두 방식을 상호비교하면, 디지털방식은 1) 검색과 이미지 획득 공정이 통합되어 있고, 2) 이미지 저장, 수정 및 새로운 문서의 이미지 추가가 즉시 이루어지고, 3) 저장된 이미지의 다수 동시이용 및 전송이 즉시 이루어질 수 있는 장점이 있는 반면에, 아나로그방식은

1) 이미지저장에 소요되는 비용이 저렴하고, 이미지저장시스템을 제외한다면 이용시스템의 구축 비용도 저렴하며, 2) 이미지의 장기간 저장능력이 뛰어나고, 3) 백업이 용이하고 경제적이란 장점을 갖고 있다.

따라서 저장과 백업능력을 중시하는 경우 디지털과 아나로그방식을 혼합한 COM방식도 채택되고 있으나, 아나로그방식은 이미 더이상 시장의 확장이 불가능한 반면 디지털방식은 급격한 시장확대에 따른 성능향상과 가격하락이 이루어지고 있음을 감안할 때 디지털방식이 권고되고 있다. [8]

디지털방식의 이미지 저장기술은 이미 학술잡지 논문, 기술보고서 등의 인쇄매체의 보존 및 저장, 전송 등에 사용되고 있으며 대규모 프로토타입들이 운영되고 있다. 미국국립농학도서관과 북캐롤라이나주립대학의 공동사업인 텍스트 디지털라이징프로젝트, RLIN(Research Library Information Network)의 문서전달워크스테이션개발, 미국화학회/OCLC/Cornell CORE 프로젝트, 코넬대학교의 CUPID 프로젝트들이 개발 또는 가동 중에 있으며, 미국세청의 소득세신고서 이미지저장, 아메리칸익스프레스의 보험증권과 크레딧카드 영수증사본 이미지저장 등이 이 기술에 의존하고 있고 그 외에도 미술, 의학, 위성사진, 직물역사 등 이미지가 주요소가 되는 각 분야에서 개발사례가 늘어나고 있는 추세이다. [9]

7. Ibid., pp.63-64.

8. Ibid., p.174.

9. Clifford A. Lynch, and Lois F. Lumin, "Introduction and Overview," JASIS 42(8) (1991. 9), p.576.

이러한 디지털 방식의 이미지 저장기술은 디지털 방식의 이미지 처리를 위한 데이터의 표현, 획득, 저장, 운영, 전송, 인쇄 등의 관련 기술들과 조합되어서만 실제로 사용될 수 있으며, 다음은 이 기술에 관한 간략한 기술동향 정보이다.

2.1 이미지 표현

컴퓨터내에서의 이미지 표현방식은 라스터(비트맵) 방식과 벡터방식으로 구분된다. 라스터방식은 입력된 이미지를 점(bit)으로 분해하여 흑백을 표시하는 값(0, 1)과 위치함수로 저장하는 방식으로 사진과 도형의 동시처리가 가능한 장점이 있기 때문에 대부분의 사무용시스템에서 이 방식을 사용하고 있다. 벡터화일은 입력된 이미지를 선의 함수로 변환하여 저장하는 방식이다. CAD 등에서 사용하는 화일 즉 도면관리시스템 등에서 사용되며 도형의 확대, 축소가 자유롭고 확대시에도 이미지의 품질이 떨어지지 않으나 사진의 처리가 불가능하다. [10]

또한 흑백(단색)과 색도로도 구분되는데 단색시스템은 대부분의 사무용시스템에서 사용하는 방식으로 흑, 백 두가지만 표현하는 문자모드와 명암을 나타내는 음영모드(사진모드)로 구분되기도 한다. 문자와 도형의 처리에는 문자모드만으로 충분하나 사진모드의 경우 16-256단계의 음영기능이 가능하여야 한다. 그러나 음영을 지원시 화일의 크기가 10배 이상 커지고 문자용 압축알고리즘에서는 1:3이상의 압축이 불가능하기 때문에 사무용

시스템에서는 별로 바람직하지 않다. 색도시스템은 칼라 스캐너, 비디오카메라, 이미지 획득카드를 사용하여 칼라이미지를 입력하며 사진 수준의 화질을 지원하기 위하여는 최소한 1024×768점 이상의 고밀도 CRT가 요구되는데 현재 PC에서 사용되기 시작한 슈퍼-VGA 카드의 경우 이 조건을 만족시키며 가격도 수십만원대로 실용화가 가능하다. [11]

2.2 데이터 획득

데이터의 획득은 이미 외부에서 제작된 그림, 사진, 도표, 문서 등의 이미지를 시스템 내부로 받아들이는 것을 말하는 것으로 스캐너와 VTR 카메라, CCD(Charged Coupled Device) 카메라 등의 디지털카메라 기술이 활용되고 있다. [12] 스캐너의 경우, 해상도 300DPI가 주종을 이루고, 디지털카메라의 경우도 TV화면 보다 약간 선명한 정도인 40만 화소 정도의 품질이다. 그러나 경쟁상대인 필름을 대체하려면 스캐너의 경우 800DPI, 디지털카메라의 경우 100만 화소 이상이 되어야 하기 때문에 개선의 여지가 많다. [13]

10. 木戸出正繼, “畫像情報處におけろマルチメディアデータベース,” 情報處理 28(6) (1987. 6), pp.756-764.
11. 한글/한자 16bit super VGA board 사용설명서, (서울:가산전자, 1991)
12. Clifford A. Lynch, “The Technologies of Electronic Imaging,” JASIS 42(8) (1991. 9), pp.579-580.
13. 2005년의 정보통신기술, TIS-IS-21, (대전:한국전자통신연구소, 1991), p.13.

한편 OCR(광학문자인식)기술은 패턴인식 기술을 이용하여 입력된 이미지를 직접문자로 변환하는 방식으로서 영문의 경우 문자의 인식정확율이 92~97% 정도이고 \$1,000 이하로 실용화가 되어있다. 영문의 경우 국내에서는 Truescan, Omnipage, Textpert 등이 사용되고 있으며 Textpert의 경우 프랑스어, 러시아어 등 30여개 알파벳이 인식 가능하다. [14] 한글의 경우는 과학기술원, 삼보컴퓨터 등에서 89년부터 시제품을 발표했고 최초의 상용시스템인 Readex도 최근에 발표되었으나 인식률, 속도, 가격 등에서 영문 OCR정도의 수준이 되려면 2~3년이 걸릴 것으로 추정된다.

2.3 데이터의 저장

디지털이미지의 저장은 우리가 이미 사용중인 디지털데이터의 저장방식과 전혀 차이가 없다. 다만 이미지의 경우 그 데이터량이 매우크므로(A4크기 문서를 80자×60줄의 문자 데이터로 환산할 경우 4.8KB : 300DPI 흑백 이미지로 환산할 경우 1MB) 이를 압축하는 기술과 대용량 데이터를 저장하는 저장매체를 사용하는 것을 특징으로 들 수 있다.

2.3.1 압축기술

이 기술은 대용량의 이미지데이터를 압축하여 저장공간을 줄이고, 전송속도를 증가시키기 위해서 사용하며 이용순간에만 이를 원형으로 확장, 복원시키게 된다.

이러한 압축 및 원형복귀 과정에서 데이터의 손실이 발생할 수 있으며 이러한 손실을 어

느정도 인정해 주는 대신 압축비율을 높이는 삭제(lossy)방식과 손실을 최대한 방지하는데 초점을 두는 비삭제(lossless) 방식으로 대별될 수 있다. [15]

이러한 압축기술은 속도의 향상을 위하여 펌웨어(firmware) 형태의 하드웨어로 처리되기 때문에 표준의 확립이 중요한 과제이다. 이와 관련되어 진행중인 국제적 표준화 작업을 살펴보면 정치화면의 경우 CCITT(The International Telegraph & Telephone Consultative Committee)의 Group 3와 Group4 방식이 이미 실용화 되어 있으며 JPEG(Joint Photographic Expert Group) 표준이 ISO(International Organization for Standardization)에서 개발중에 있다. [16], [17]

움직이는 이미지의 경우 Intel이 제안한 DVI와, IBM이 제안한 MPEG (Motion Picture Expert Group) 규격등이 있으나 1초에 16~24장을 복원해야 하므로 속도의 향상이 주요 연구과제가 되고있다. 따라서 동화의 경우는 각 프레임의 차이에 관한 정보만을 수록하는 프레임간 압축(interframe compression) 기술을 사용하여 재현의 고속성을 확보하고 있지만 각 프레임에 대한 임의접근이 불가능한 문제가 있다. [18]

-
14. Textpert Catalog, (New York:CTA, 1991), p.1.
 15. Clifford A. Lynch, op.cit., p.580.
 16. Ibid., p.582.
 17. 권용목, "동화상 압축표준화동향 및 뉴미디어 분야에의 응용" 전자과학 33(통권391) (1991. 12), pp.188-197.
 18. Clifford A. Lynch, op.cit., p.580.

2.3.2 저장매체

수정이 가능한 소거형(Erasable) 디스크와 불가능한 추가기록형(WORM-Write Once Read Many) 디스크로 구분된다. 추가기록형 시스템은 일단 기록이 된 화일은 삭제, 수정이 불가능하기 때문에 보존성을 요구하는 문서, 기록 등의 시스템에서 많이 사용되는 반면에 소거형시스템은 하드디스크의 저장능력의 한계를 극복하기 위한 대체 기술로서 많이 사용된다.

소거형디스크는 기록방식에 따라 다음과 같이 광자기(Magneto-Optic)형, 상변이(Phase-Change)형, 염료중합체(Dye-Polymer)형으로 분류되고 있다. [19]

광자기형-커(Kerr)효과를 이용한 것으로서 디스크표면에 수직방향으로 자화된 기록층에 레이저빔을 조사하여 국부적으로 가열된 부분에 외부 자기장을 가함으로써 자화의 방향을 상향 또는 하향으로 조정하여 신호를 기록하는 방식.

상변이형-레이저의 강도와 폭을 조정하므로써 결정질(Crystalline)과 비결정질(Amorphous)의 상변화에 따른 빛의 반사도 차이로 신호를 기록/재생하는 방식

염료중합체형-빛에 대한 염료의 특성을 이용한 것으로 레이저를 조사한 부분과 조사하지 않은 부분의 반사도 차이로 신호를 검출하는 방식

현재 생산에 성공한 방식은 광자기형으로서 5¼인치 광자기 디스크 시스템의 경우 고용량 원체스터 디스크 드라이브에 필적하는 성능과 이용성을 보여줌으로써 시장이 확대되고 있는 추세이다.

광자기디스크는 나선형의 홈(홈간의 거리 $-1.6\mu m$; 홈의 길이 $-0.07\mu m$; 홈의 폭 $-1\mu m$)을 가지고 있는 130mm 직경의 복합탄산염 등의 내구성 플라스틱 또는 유리원판에 비결정질의 희토류-천이 금속을 기록층으로 입히고 그 위에 반사층과 보호층을 형성시킨 디스크이다.

기록층은 "0"과 "1"의 디지털신호로 기록/소거 및 재생이 이루어지는 부분이며 보호층은 기록층의 산화방지 역할과 더불어 무반사 코팅에 의해 신호의 광학적 효과를 크게 하는 역할을 한다.

또한 디스크면의 홈은 ISO의 "Continuous Composite Format" 중 서보데이터를 트랙에 삽입시키는 표준에서 규정된 것으로 각 트랙을 따라서 계속적으로 얇은 홈을 넣어줌으로써 광학헤드가 트랙에서 벗어난 것을 반사된 빔의 래디얼 시프트(radial shift)에 의해 확인하여 헤드의 위치를 정확히 수정할 수 있도록 고안된 것이다.

광자기디스크는 자기디스크나 자기테이프의 수평자화방식과는 달리 자화가 디스크막 표면에 수직으로 배열된 수직 자화막방식이기 때문에 더욱 더 고밀도의 기록이 가능하다.

19. 임만직, "광디스크 드라이브", 전자과학 32(통권368) (1990. 1), pp.144-149.

포맷되지 않은 5.25인치 디스크의 경우 단면에 325메가바이트의 정보를 수록할 수 있으며 금속제도와 쓰기 방지스위치가 달린 플라스틱 카트리지에 내장되어 사용된다. 현재까지 실험에 의하면 하루에 약 1,000번의 기록/소거/재생을 실시할 경우 40년까지의 내구력을 인정받고 있다. [20]

2.4 운영

현재의 대부분의 DBMS들이 텍스트용으로 설계된 것이기 때문에 대규모의 이미지 데이터베이스의 구축에 적합하지 않으며, 적합하다 하더라도 이미지 데이터 타입이 아니라 BLOBs(Binary Large Objects)로 저장되며 주크박스 등의 로보틱저장시스템과의 결합이 용이하지 않다. 더우기 이미지용 DBMS 와 텍스트용 DBMS를 병행 사용하는 경우 처리대상이 텍스트에서 이미지로 혹은 그 반대로 바뀔 때마다 해당 DBMS에서 빠져나와 다른 DBMS로 로그인해야 하는 단점이 있다. 따라서 보통의 DBMS에서 이미지를 처리할 수 있도록 지원해 주는 소프트웨어와 이미지획득카드를 사용하여 DB 내의 이미지, 텍스트, 수치정보를 통합 운영시킬 수 있는 전용소프트웨어가 많이 개발되고 있다. [21] 전자의 예로는 PictureWare사의 PicturePower와 Videotex System사의 T-base등을 들 수 있다. 이것은 dBase와 같이 운영될 수 있고, 값싸며 쉽게 사용할 수 있는 반면, 소장 이미지의 크기, 해상도, 처리속도등에서 대량정보 처리시스템에 적용하기 힘들고 이미지의 확대, 축소, 회전

등이 불가능한 팩키지이다. [22] 후자는 대량정보처리시스템의 일부분으로 부착되어 이미지 처리를 전담하는 펌웨어 공급방식으로서 KOFAX사의 KF-9000시리즈가 대표적이다. 이 시스템은 데이터베이스 개발자를 위해 이미지스캐닝에서 데이터베이스 구축까지의 하드웨어 및 소프트웨어의 제어기능과 사용자 인터페이스의 개발을 지원하는 개발 tool 부분, 그리고 단순검색 이용자를 위해 검색용 PC에 설치할 수 있는 검색용 소프트웨어 부분으로 구성되는 것이 일반적이며 전자에 비해 상대적으로 고가이다.

T-Base

제작사: Videotex System

사용환경: IBM PC (DOS)

화일형식: PCX

색도: 단색, 컬러(SuperVGA까지)

화일압축: 불가

사용DBMS: Dbase III, Clipper

KF-9000

제작사: KOFAX

사용환경: IBM PC (DOS)

화일형식: TIFF, PCX, IMG

색도: 단색

화일압축: 1:30

사용DBMS: KOFAX application

20. Ibid.

21. Clifford A. Lynch, op.cit., p.581.

22. T-Base 2.0 Manual, (Dallas:Videotex Systems Inc., 1990)

2.5 이미지 전송

이미지 전송을 위한 화일포맷에서는 인코딩 헤더정보(encoding header information) 즉 인코드방식, 이미지사이즈 및 기타 사용 및 가공에 필요한 정보를 수록하는 방식을 정의하며 TIFF (Tagged Image File Format)와 PCX 등 특수규격이 많이 사용되고 있지만 장래에는 현재 개발중에 있는 ODA(Office Document Architecture), SGML (Standard General Markup Language) 등이 국제표준으로 사용될 전망이다.

이러한 화일포맷을 통신망을 통하여 전송하는 방식은 다음과 같다.

- X 윈도우시스템(이미지 응용기능을 저장하고 있는 원격시스템에 로그인 하는 비트맵 표시장치(bit-mapped display device)을 이용한 표시장치로 네트워킹.
- OSI 환경내에서 TCP/IP의 FTP나 FTAM (File Transfer Access and Management)과 같은 표준화일전송프로토콜 이용. 이미지화일의 경우 화일크기가 거대하기 때문에 A4 크기의 경우 현재 공중통신망으로 사용중인 2400BPS(bit per second) 통신선을 이용하면 69분이 소요되고, 특정회선인 56KBPS라도 3분이 소요되어 장거리 지역간의 자료전송에는 실용성이 적다. 하지만 외국에서는 100MBPS의 장거리 고속전송망인 FDDI(Fiber Distributed Data Interface)가 이미 실용화되어 있고, 국가망의 경우에도 T3(45MBPS)의 건설이 추진되고 있으며 600MBPS의 속도를 지니는 SONET(Syn-

chronous Optical Network) 기술도 93년 중반쯤 실용화가 될 예정이므로 국내에서의 실용화도 조만간 이루어질 것으로 예측된다. [23]

2.6 인쇄

이미지인쇄의 경우 일반적으로 단색에서 300DPI가 상용화되어 있으며 리토그래픽(Lithographic)시스템의 경우는 1,000DPI의 초고해상도까지 가능하나 가격이 수천만원대로 아직 상용화되기에는 고가이다. 칼라의 경우는 500만원대의 장비가 나올 2-3억원 정도가 소요되기 때문에 아직 기술의 개선이 필요하다. 단색인쇄의 경우 포스트스크립 기술의 발전으로 인하여 인쇄품질이 도형의 크기에 관계없이 인쇄될 수 있도록 상당히 발전되어 있다. [24]

3. 국내상용시스템

디지털이미지저장기술이 가장 빨리 받아들여진 대표적인 분야는 사무용 이미지저장 분야이다. 사무자동화의 일부분으로 시작된 이 기술의 도입은 최근 문서보관 및 관리시스템으로 발전되어 사무자동화시스템의 중핵으로 주목받고 있을 뿐만 아니라, 일부 자료실에서 보고서관리시스템에 응용하기 시작하였고 이러한 추세가 확산될 전망이다. 따라서 본장에

23. Clifford A. Lynch, op.cit., p.581.

24. Color Copy Machine Catalog, (東京:Brother Inc., 1991)

서는 이 분야의 일반적인 현황과 국내에서 공급되고 있는 상업적 시스템들의 기능을 비교평가해 보므로써 보고서관리시스템의 표준적인 기능 설계에 참고하였다.

현재 개념의 문서보관시스템을 최초로 상품화한 것은 1982년 발표된 도시바의 TOSFile 2100이며 [25] 국내에 이 기술이 소개된 것은 1985년 경한시스템이 일본 산요사에서 12인치 추가기록형 광디스크를 사용한 광화일링 전용기를 수입하여 한국과학기술원전산센터와 공동으로 한글화작업을 하여 보급한 시점으로 보인다.

그러나 기술제공자인 일본측으로부터 핵심 기술이 입수되지 않아 1) 이용자 연계기능의 확대, 시스템의 확장등에 문제가 있었고 2) 협소한 국내시장으로 사후 서비스체제가 미비했으며 3) 비용 대 효과면에서 국내시장을 설득하지 못함에 따라 보급·확대에 실패했다. 그후 88년에 삼성전자, 서울시스템, 89년에 왕컴퓨터 등 참여업체가 늘어 현재는 금성정보통신, 대성산업, 대우정보시스템, 리스트, 뱅컴, 세양전산, 유니온시스템, 제우스컴퓨터, 정원컴퓨터, 파일코리아, 하이퍼컴퓨터, 한국전자계산등 16개 회사가 이 시장에 참여하고 있다. [26]

국내 시스템들의 기능면에서 특징은 문서의 작성에서 재활용까지 전 공정을 시스템내에 모두 포함하고자 시도한 것이다. 물론 통합환경의 관점에서 보면 문서의 작성에서부터 저장, 재활용까지의 모든 기능을 하나의 시스템에서 구현할 수 있을때 가장 이상적일 것이다. 그러나 현재 보급되고 있는 시스템

중 이러한 광범위한 기능을 모두 만족시키는 상업적 시스템을 발견하는 것은 불가능하며, 또한 기술개발 추세 자체도 문서작성을 위주로 한 탁상출판기술과 문서저장을 위주로 한 광화일링기술로 이원화되어 있는 것으로 보인다. 궁극적으로 이 두 기술은 통합된 환경하에서 운영될 수 있도록 발전되어 갈 것이지만 상당한 기간 동안은 각기의 목적에 맞는 기술로 전문화되어 갈 것으로 판단된다. 따라서 관련기술들의 도입검토에 있어서도 도입하고자 하는 목적을 문서의 작성과 저장을 양극으로 하는 스펙트럼선상에서의 위치로 정확히 설정한 후 검토를 행하여야 한다.

이는 시스템개발을 위한 기능 설계에도 동일하게 적용되는 문제로서 본연구에서도 문서관리시스템의 개념을 탁상출판과 광화일링으로 분류하여 각각의 표준기능과 현재 국내에서 보급되고 있는 대표적인 시스템들간의 비교평가를 실시하였다.

본 조사는 1990년 12월부터 91년 6월에 걸쳐 현재 광화일링시스템을 판매하고 있는 국내 16개 회사중 판매실적이 높은 5개회사를 대상으로 하였다.

조사방법은 문헌에 발표된 광화일링시스템의 표준기능들을 조합하여 작성한 기능 대비표와 5개 회사 및 그들이 추천한 시스템 도입기관을 직접 방문하여 조사한 기능을 비교하

25. 齊藤孝, "光ディスク-將來の記録媒體" 다큐멘テーション研究 34(6) (1984. 6), pp.256-265.

26. 권태열, "비상 예고하는 늑각이- 국내문서이미지처리시장", 컴퓨터월드 통권72 (1991. 10), p. 115.

였다. [27] [28] [29] [30] [31] [32] [33]

다음은 국내시스템들과 광화일링시스템의 표준기능과의 비교조사결과이며, 보고서관리 시스템의 목표를 문서저장기로 제한시킴에 따라 탁상출판기능에 대한 조사 결과는 생략하였다. (표1참조)

여기 나타난 각 시스템들의 특성은 도입하고자 하는 기관들의 환경과 도입목적에 따라 여러가지로 해석될 수 있을 것이다. 그러나 모든 시스템들이 공통적으로 갖고 있는 문제

27. 齊藤孝, 전제서, pp.256-265.
 28. 권태열, 전제서, pp.113-126.
 29. 有賀秀春, "OAのためのファイリングシステムの有効利用," 事務管理 25(3) (1985. 3), pp.17-22.
 30. 嬉山聰, "ケノファイル・フェニックス," データ通信 23(8) (1991. 8), pp.59-64. 彬
 31. 中澤俊一, "電子ファイリングシステムの動向と性能比較" 事務管. 25(3) (1985. 3), pp.23-28.
 32. 増田龍彦, "電子ファイルシステムの戰的活用" データ通信 23(8) (1991. 8), pp.30-35.
 33. 飯田恭子, "日本テレコムの手書きイメージ文書處理に威力," 事務管理 30(5) (1990. 5), pp.24-30.

표1. 사무용 광화일 시스템 표준기능

범례 : ← 좌측과 동일

해당 기능	LFS 90 경한	LQ G8000 금성통신	Marvin 대성산업	Docubase 대우정보	SF 9000 삼성전자
하드웨어 콘트롤러 하드웨어 플로피디스켓 운영체제	PC286/386 30-154MB 3.5" IRMX/UNIX	PC386 40-50MB 5¼" DOS	PC386/486 60MB ← ←	PC386/SUN 60MB ← ←	PC386/SUN 380MB ← UNIX
스캐너 속도(스캐너) 속도(화면/화일) 자동급지장치	B5 - A3 40장/분 9장/분 50매	← 24장/분 7장/분 500매/양	← 24장/분 7장/분 50매	← 19장/분 5장/분 ←	← 24장/분 7장/분 ←
레이저 프린터 인쇄밀도 인쇄속도 포스트스크립트 확대/축소 인쇄 프린터 버퍼/용량	B5 - A4 400DPI 8 PPM No Yes LEP memory	B5 - A3 ← ← ← ← ←	B5 - A4 300DPI ← ← ← ←	← 400DPI ← ← ← ←	← 300DPI ← ← ← ←
광디스크 형식 주장비 최대 접속대수 주크박스 최대카트리지/용량 " 최대접속 " 내 드라이브 수	WORM/Eras- able 2 20/12GB 3 2	Erasable ← ← ← ←	← ← 50/56GB ← ←	← ← 20/12GB ← ←	← ← 20/12GB ← ←

모니터 화면밀도(CRT) 화면 각도 소요시간 이미지 전처리 프로세서	15"/19" 2496/2496 0.5초 LSI	← 2500/2500 ← ←	← ← ← KOFAX	← 1024/768 2 초 ←	← 1680/1280 1 초 LSI
입력기 추가 검색기 추가	No Yes	No ←	Yes ←	No ←	No ←
이미지 파일 ----- CCITT 3/4 TIFF Vector Image format conversion/Numb 압축비/방식 음영처리등급	Yes No No No 1:14 MH,MR 64	← ← ← ← 1:33 MMR ←	← Yes ← Yes/5 1:14 MMR 256	← No ← No 1:14 MR 4	← ← ← ← 1:12 MR 64
검색 ----- 문서명 등록년월일 키워드/최대 설정수 항목/최대 설정수 임의접근, 직접접근 파일조작/갱신 기록 캐비넷 설정 시스템총색인 불리안 연산자 자동색인	Yes Yes Yes/32 20 No Yes No/Table Yes Yes Yes No (제한적) No	← ← 1024 byte 32 ← ← ← ← ← ← ← ← ←	← ← 70 byte 256 ← No(User) ← ← ← ← ← ← ←	← ← ← 20 ← ← ← ← ← ← ← ← ←	← ← Unlimited ← ← ← ← ← ← ← ← ← ← ←
타 시스템 접속 (H/W) ----- PC 호스트컴퓨터 팩시밀리(G3/G4) 근거리통신망 마이크로필름	Yes No Yes Yes/Ether- net No	← ← ← ← ← ←	← ← ← ← ← ←	← ← ← ← ← ←	← ← ← ← ← ←
타 시스템 접속 (S/W) ----- 탁상출판(DTP) DEMS OCR(영문/한글) CAD file LAN 도서관 자동화(MARC)	No Yes No No Yes No	Yes No ← ← ← ←	Yes ← ← ← ← ←	No ← ← ← ← ←	Yes ← ← ← ← ←

통신망 지원					
다수 동시이용	Yes	←	←	←	←
통신보안	Yes	←	←	←	←
이미지조작					
확대/축소	Unlimited	←	←	12% - 800%	Unlimited
화면 편집	Yes	←	←	←	←
블록인용/복사	Yes	←	←	←	←
회전	1 deg	←	90	90	90
경면대칭	Yes	←	No	Yes	←
반전	Yes	←	←	←	←
인덱스당 최대 이미지 수	255	Unlimited		999	Unlimited
최대 화면 window	1	6	←	4	1
페이지 검색(Top/Bott)	Yes	←	←	←	←
가 표시	No	No	Yes	Yes	No

점, 특히 도서관이 요구하는 수준의 보고서관리 업무에 이 시스템들을 사용한다고 가정했을 경우의 문제점은 제한된 소수의 정보 항목에 고정된 길이로만 입력시킬 수 있는 레코드 구조의 경직성과 다양한 검색키와 부울연산자를 이용한 복합적인 탐색기법을 구사할 수 없는 검색기능의 제한성이라 할 수 있다. 이는 사무자동화를 위한 MIS시스템이 도서관 자동화에 사용될 수 없는 것과 동일한 이유라 할 것이다.

따라서 효과적인 보고서관리시스템의 필요 조건은 서지정보처리에 적합한 정보구조의 채택이라고 할 수 있으며, 이는 곧 MARC기술의 도입을 뜻한다고 하겠다. MARC을 채택함으로써 보고서정보의 효율적 이용이 가능할 뿐만 아니라, 한 정보센터 내에서 단행본, 연속간행물 등 다른 자료들과 통합운영의 효과를 얻을 수 있고, 국가서지나 공동편목용 종합목록 데이터베이스의 구축이 원활하게 이루어

어질 수 있다.

MARC을 채택함에 있어서 고려해야 할 보고서의 특수성으로는 1) 대부분의 보고서 발행기관이 사용하고 있는 서지통제양식(bibliographic control data sheet)과 MARC사이의 관계를 정확히 해석하는 문제와 2) 국문, 영문의 두가지로 생산되는 보고서서지정보의 효과적 운영을 위한 구조설계의 문제를 들 수 있다.

첫째 서지통제양식은 ANSI Z 39.18 (Scientific and technical reports organization, preparation, and production), BS 4811 (Presentation of research and development reports), 그리고 ISO 5966 (Documentation-presentation of scientific and technical reports) 등의 국제 또는 국가규격에서 권고하고 있는 표준으로서, 보고서서지정보의 특수성과 이용행태의 특성에 입각하여 제시된 다큐멘테이션 표준이다.(그림 1 참조)

그림 1. 서지통제양식의 예 (ANSI Z 39.18의 적용예)

NST-114A (REV.3-90)		U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY		1. PUBLICATION OR REPORT NUMBER NISTSP-500200
BIBLIOGRAPHIC DATA SHEET				2. PERFORMING ORGANIZATION REPORT NUMBER
				3. PUBLICATION DATE December 1991
4. TITLE AND SUBTITLE Development of a Testing Methodology to Predict Optical Disk Life Expectancy Values				
5. AUTHOR(S) Fernando L. Podio				
6. PERFORMING ORGANIZATION(IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS) U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY GAITHERSBURG, MD 20899 Computer Systems Laboratory/Advanced Systems Division			7. CONTRACT/GRANT NUMBER	
			8. TYPE OF REPORT AND PERIOD COVERED Final	
9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP) National Archives and Records Administration Washington, D.C. 20408				
10. SUPPLEMENTARY NOTES				
11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.) <p>There are no standards for longevity of optical disks that can assist managers in the Federal government to select the right media for the storage of permanent records, and to know how long the information may be safely stored on those disks. This report focuses on research undertaken at the National Institute of Standards and Technology to develop a methodology to predict optical disk life expectancy values. In this research accelerated aging tests were run on small sets of disks and the quality parameter (the byte error rate) was periodically measured between aging cycles. These tests were used with a mathematical prediction model to develop a testing methodology.</p> <p>This report presents the results obtained. The need for standard test methods for predicting life expectancy and for measuring media characteristics is apparent. Life expectancy extrapolations derived from the experiments produced a range of values depending upon the method used for deriving the quality parameter. Recommendations are made about the implementation of a testing methodology for life expectancy predictions, and what information to include in a life expectancy specification.</p>				
12. KEY WORDS(6 TO 12 ENTRIES; ALPHABETICAL ORDER, CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMI-COLONS) accelerated aging; Arrhenius model; byte error rate; media lifetime; optical disks, stress tests; optical disks, test methods.				
13. AVAILABILITY			14. NUMBER OF PRINTED PAGES 96	
<input checked="" type="checkbox"/> UNLIMITED FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE(NTIS).			15. PRICE	
<input checked="" type="checkbox"/> ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.				
<input checked="" type="checkbox"/> ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE(NTIS), SPRINGFIELD, VA 22161.				

결국 이것은 보고서서지정보를 효과적으로 표현할 수 있는 외부형식이라 할 수 있으며, 이것을 적절히 해석하여 MARC이라는 내부형식으로 변환 또는 연계할 수 있는 기능이 있어야만 보고서라는 특수성을 보장하면서도 MARC가 제공하는 기대효과를 얻을 수 있을 것이다.

둘째, 대부분의 연구기관에서 발행되는 보고서에는 동일한 서지정보가 국·영문 두 가지로 작성, 수록되어 있다. 영문정보의 주목적은 국제적인 서지정보의 교환을 통한 국제협력강화에 있으며, 현재 이들은 완전히 별개의 서지통제양식으로 취급되고 있다.

그러나 이들을 통합하여 하나의 레코드로 처리할 경우, 1책 1서지의 관리상의 편리함과 함께 언어와 무관하게 동일한 정보가 입력되는 항목, 즉 보고서번호, ISBN, 코드화 정보 부분 등을 중복 입력하거나 저장할 필요가 없는 장점이 있다.

이런 통합운영방식내에서 국문용 서지레코드, 영문용 서지레코드를 손쉽게 분리 일치하기 위해서는 표시기호 880과 식별기호 ▼6을 이용하는 것이 최선의 해결방법이 될 수 있을 것이다.

MARC 이외의 서지정보처리상의 문제는 보고서들이 갖고 있는 평균 200자 정도의 상대적으로 긴 초록 및 요약문 정보의 효과적인 처리이다. 이 부분의 전문(full text)탐색을 고려할 경우 추출되는 색인어는 레코드(문헌)지시색인(record oriented indexing)이기보다는 용어지시색인(term oriented indexing)성격이 될 수 밖에 없기 때문에 평균적인 MARC

레코드보다 월등히 많은 색인어가 발생하게 된다. 더우기 조사의 부착이나 접사의 사용여부에 따른 변형까지 별개의 색인어로 처리할 경우 색인어화일의 크기는 거의 무한정으로 증가하는 것이 일반적이다.

따라서 이 색인어화일의 크기를 적정한 수준으로 유지하고 검색의 효율성을 높이기 위해서는 기계적으로 텍스트를 분석하여 색인대상어를 추출, 선정하는 과정에 언어학적인 방법을 적용한 자동색인기술을 채택하는 것이 바람직하다. [34]

4. 연구보고서관리시스템

4.1 시스템 구성도

본장에서는 앞에서 언급한 여러가지 기술들과, 특히 서지정보 처리면에서의 특성들을 적용하여 개발한 한국표준과학연구원의 보고서관리시스템의 구조와 기능을 소개한다.

개발된 보고서관리시스템의 기본목표는 보고서에 관한 전문(full text)정보를 데이터베이스화하여 보존문서로서 영구관리, 복사본 제작, 배포 등의 업무에 활용할 뿐만 아니라, 이 시스템의 상위에 있는 통합기술정보시스템 내에서 이용자검색서비스에 활용할 수 있도록 전달하는 것이다.

따라서 서지 및 이미지 정보의 입력, 수정,

34. 박신석 등저, 광기술을 이용한 보고서관리시스템 구축연구 - 1차년도 (KSRI-91-52-IR) (대전: 한국표준과학연구원, 1991), pp. 58-65.

갱신 등의 데이터베이스 구축, 관리기능, 보고서 원문의 복제기능, 그리고 통합기술정보시스템에서 원문의 이미지를 호출할 경우 이를 전송하여 주는 기능 등이 시스템의 주기능이 된다.

시스템은 서지정보모듈, 원문 이미지정보모듈, 그리고 관리정보모듈로 구성되어 있으며 전체적인 하드웨어와 소프트웨어의 구성도는 아래와 같다.(그림 2, 3 참조)

있는 정보를 입력, 가공하여 검색용 서지데이터베이스를 구축하고, 이를 통합기술정보시스템에 추가, 갱신하는 일련의 과정을 수행한다. 서지정보의 처리에는 한국문헌자동화목록법(KORMARC)을 채택하였으며, 국·영문의 자유로운 통합, 분리 사용을 위하여 앞에서 언급한 표시기호 880과 식별기호 ▼6을 사용하였다. 그리고 KORMARC에 규정되어 있지 않은 일부사항을 위해 USMARC이 참고되었다.

4.2 서지정보 처리기능

서지정보 모듈은 서지통제양식에 수록되어

4.2.1 표준입력기능

시스템이 제공하는 입력양식화면에 키보드

그림 2. 시스템 하드웨어 구성도

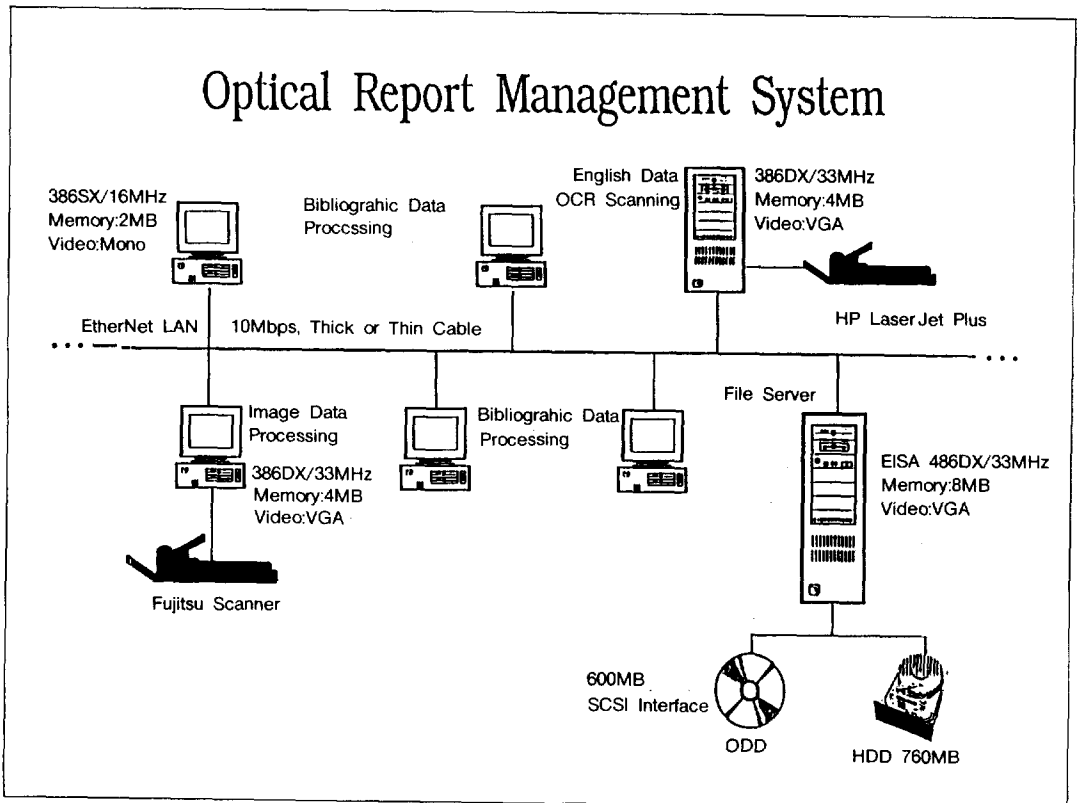
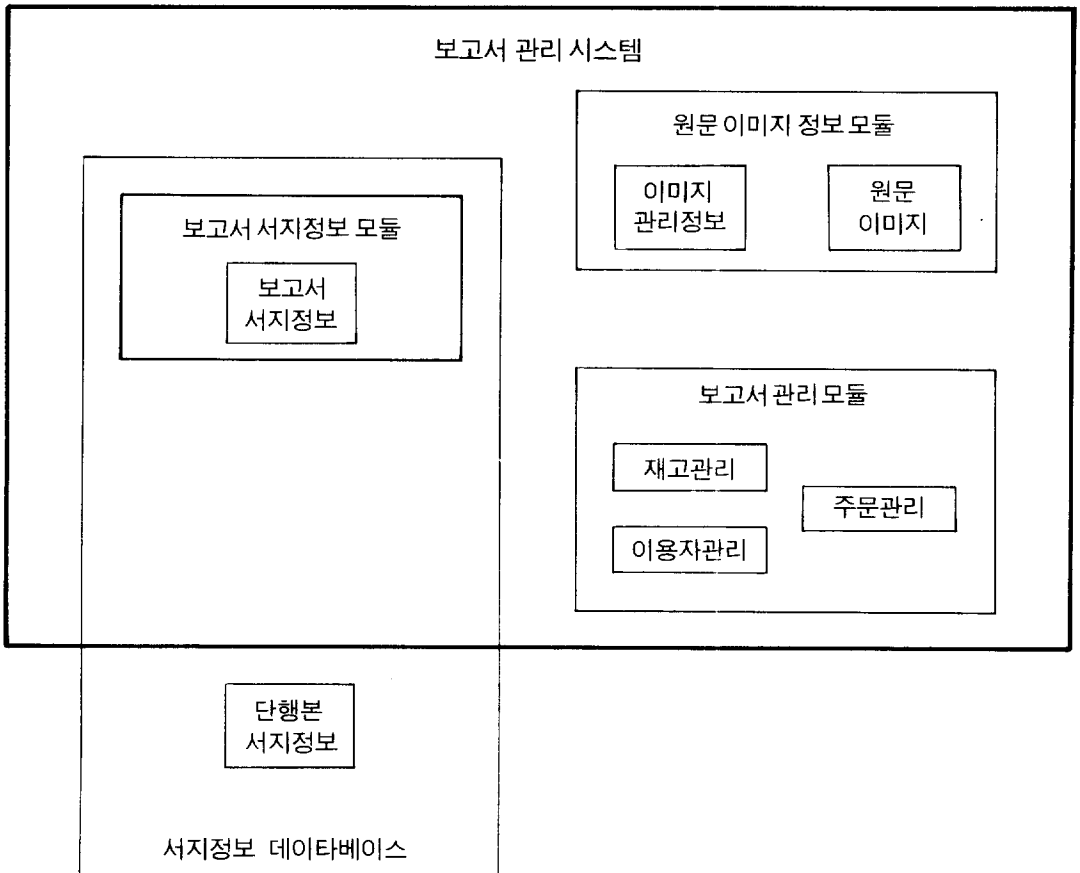


그림 3. 시스템 소프트웨어 구성도



를 통하여 서지정보를 입력, 수정, 편집하는 기능으로서 MARC방식의 화면과 보고서전용 입력화면 중 입력자가 선택하여 사용할 수 있다. 어느 경우나 리더부 및 부호화정보필드에 해당하는 데이터는 디폴트 값을 시스템에서 변수로 제공한다.

MARC화면방식은 국·영문을 한 화면에서 입력할 수 있고 필요한 태그들을 자유롭게 이용할 수 있는 탄력성이 있으나 입력자가 스스로 880과 ▼6을 부여해 주어야 하는 등 숙련된

편목사서에게 적합한 방식이다.(그림 4 참조)

보고서전용 입력화면방식은 MARC에 익숙하지 못한 입력자를 위한 것으로서 보고서 서지통계양식과 유사한 입력화면을 제공한다. 이 방식을 사용할 경우 입력자는 우선 국문용 입력화면(그림 5 참조)을 통해 국문서지통계양식에 수록된 모든 정보를 입력한 후, 영문 전용 입력화면(그림 6 참조)을 통해 국영문 공통필드를 제외한 영문 서지정보를 추가 입력한다.

그림 4. 한글/영문 요약문의 MARC 데이터 입력 화면

제어번호: KSRI00000015	상태: 신규	갱신일: 92.01.15	구분: KORMARC
유형: a	서지형식: m	입력수준: 기술형식:	입력일: 92.01.15
삽도: a	연도유형: s	발행년도: 1989,	발행국: w k 언어: kor
색인:	지식수준:	자료형태:	내용형식: t
전기:	회의간행:	기념논문:	대학간행: 문학구분:

보고서번호 > 088 ▼a KSRI-91-52-IR
 주제분야 > 072 ▼a 0502
 기본개인명 > 100 10 ▼6 880-04 ▼a 박신석
 서명저자사항 > 245 10 ▼6 880-03 ▼a 광기술을 이용한 보고서 관리시스템 구축 연구
 출판사항 > 260 ▼6 880-02 ▼c 1991.3
 대조사항 > 300 ▼a 106p.
 배포구분 > 506 ▼a 01
 보고서종류 > 513 ▼6 880-09 ▼a 1차년도 보고서
 초록, 해제 > 520 ▼6 880-10 ▼a 본 연구의 목적은 연구보고서의 출판, 저장, 이용, 보급 전공정의 완전한 자동화 시스템 구축이다. 연구보고서, 논문 그리고 기타의 연구결과로서 발생하는 출판물들은 차후 연구활동 참고용으로 뿐만 아니라 연구활동 기록으로서, 또 이를 통한 연구활동 경제성 분석과 연구소 정책수립에 의 기본자료로서 중요한 가치를 갖는다.

비록재주제어 > 653 ▼6 880-11 ▼a 광기술 ▼a 전문데이터베이스
 ▼a 하이퍼텍스트 ▼a 표준화 ▼a 멀티미디어 ▼a 기술보고서

부출개인명 > 700 10 ▼6 880-05 ▼a 이상헌.
 부출개인명 > 700 10 ▼6 880-06 ▼a 김익철.
 부출개인명 > 700 10 ▼6 880-07 ▼a 김혜진.
 부출개인명 > 700 10 ▼6 880-08 ▼a 김충렬.
 부출단체명 > 710 10 ▼6 880-01 ▼b 기술정보-SRD실.

다른문자표기 > 880 10 ▼6 710-01 ▼b Office of Tech Info & SRD
 다른문자표기 > 880 ▼6 260-02 ▼c 1991, Mar.
 다른문자표기 > 880 10 ▼6 245-03 ▼a Study on Construction of Technical Reports Management System Using Optical Technology

다른문자표기 > 880 10 ▼6 100-04 ▼a Pak, Shin Sok
 다른문자표기 > 880 10 ▼6 700-05 ▼a Lee, Sang Hun
 다른문자표기 > 880 10 ▼6 700-06 ▼a Kim, Ic Cheol
 다른문자표기 > 880 10 ▼6 700-07 ▼a Kim, Hye Jin
 다른문자표기 > 880 10 ▼6 700-08 ▼a Kim, Chung Ryul
 다른문자표기 > 880 ▼6 513-09 ▼a 1st Report
 다른문자표기 > 880 ▼6 520-10 ▼a This project is aimed at building a computerized management system which integrates publishing, storage, usage, and distribution of technical reports. Various kinds of printed materials resulting from research activities(e.g., reports, papers, memoranda, and so on) are valued not only as records and reference for future researches but also as fundamental data for policy-making and analyzing the adequacy of projects of research institution. Currently, such technical reports also serve as basic media for information flow within Korea.

다른문자표기 > 880 ▼6 653-11 ▼a Optical technology ▼a Hypertext
 ▼a Image Database ▼a Technical Report
 ▼a Automatic Indexing ▼a Multimedia

그림 5. 보고서 입력 전용화면(한글요약문)

제어번호: KSRI00000015	상태: 수정	갱신일: 92.01.25	데이터: 한글
구분: 보고서	언어: kor	입력일: 92.01.25	
[1. 보고서번호] KSRI-91-52-IR [2. ISBN//가격] [3. 주제분야] 0502 [4. 서명//부제] 광기술을 이용한 보고서 관리시스템 구축 연구 [5. 저자//공저] 박신석 이상현 김익철 김혜진 김충렬 [6. 기관//부서] // 기술정보-SRD실 [7. 출판지//사] [8. 발행일] 1991.3 [9. 면수] 106p. [10. 보충사항] [11. 배포구분] 01 [12. 참고문헌] [13. 자료내역] 1차년도 보고서 [14. 초록 해제] 본 연구의 목적은 연구보고서의 출판, 저장, 이용, 보급 건공정의 완전한 자동화 시스템 구축이다. 연구보고서, 논문 그리고 기타의 연구결과로서 발생하는 출판물들은 차후 연구활동 참고용으로뿐만 아니라 연구활동 기록으로서, 또 이를 통한 연구활동 경제성 분석과 연구소 정책수립에의 기본자료로서 중요한 가치를 갖는다. [15. 위탁기관] [16. 키워드] 광기술 / 전문데이터베이스 / 하이퍼텍스트 / 표준화 / 멀티미디어 / 기술보고서			

그림 6. 보고서 입력 전용화면(영문요약문)

제어번호: KSRI00000015	상태: 수정	갱신일: 92.01.25	데이터: 영문
구분: 보고서	언어: kor	입력일: 92.01.25	
[4. 서명//부제] Study on Construction of Technical Reports Management System Using Optical Technology [5. 저자//공저] Park, Shin Sok Lee, Sang Hun Kim, Ic Cheol Kim, Hye Jin Chung Ryul Kim [6. 기관//부서] // Office of Tech Info & SRD [7. 출판지//사] [8. 발행일] 1991, Mar. [10. 보충사항] [12. 참고문헌] [13. 자료내역] 1st Report [14. 초록 해제] This project is aimed at building a computerized management system which integrates publishing, storage, usage, and distribution of technical reports. Various kinds of printed materials resulting from research activities(e.g., reports, papers, memoranda, and so on) are valued not only as records and reference for future researches but also as fundamental data for policy-making and analyzing the adequacy of projects of research institution. Currently, such technical reports also serve as basic media for information flow within Korea. In a wider sense, therefore, the construction of an effective technical reports management system can enhance the contribution of research activities. Up until now, technical reports have been distributed mostly in printed formats and microfilms have been used for long-term storage of these documents digital technologies made it possible to design a new integrated technical reports management system. [15. 위탁기관] [16. 키워드] Optical technology Hypertext Image Database Technical Report Automatic Indexing Multimedia			

입력이 완료된 후 시스템은 내부에 저장된 대응표에 의하여 전용화면의 필드들을 MARC의 태그로 변환하고, 국·영문간의 관계를 880과 ▼6을 사용, 연계시켜 하나의 MARC레코드를 완성시킨다.(표 2 참조)

시스템 사용자는 필요시 시스템 내부에 파라메타로 저장되어 있는 대응분법을 수정하거

나, 디폴트로 제공되는 필드들을 추가, 삭제, 변경할 수 있으며 보고서 전용 입력화면의 사용 도중에도 필요한 필드를 MARC태그 형식으로 추가시켜 사용할 수 있다. 일단 저장된 서지정보 레코드는 사용된 입력방식에 관계없이 MARC형식 또는 서지통제양식으로 자유롭게 지정하여 출력시킬 수 있다.

표2. 서지통제양식과 MARC 대비표

서지통제 양식항목	MARC 표시 및 식별	비고
수행기관 보고서번호	088▼a	
수행기관	710▼a	
수행부서	710▼b	
발행일	260▼c	
제 목	245▼a	
부 제	245▼b	
저 자	100▼a	
공저자	700▼a	
주제분야	072▼a	*
위탁기관	536▼a	*
계약번호	536▼b	*
자료내역(보고서유형)	513▼a	*
자료내역(수목기간)	513▼b	*
보충사항	500▼a	
초 록	520▼a	
주요 참고문헌, 관련문헌	510▼a	*
키워드	653▼a	
배포구분	506▼a	
해제기간	506▼b	**
면 수	300▼a	
ISBN	020▼a	
가 격	020▼c	

* : USMARC 참조

** : 본 시스템에서 임의 부여

4.2.2 영문 OCR을 이용한 입력기능

영문서지정보를 입력할 수 있는 또 하나의 방법은 영문 서지통제양식의 이미지를 스캐너를 사용하여 입력한 후 이를 영문 OCR을 통하여 문자정보로 변환 사용하는 방법이다.

본 시스템에서 영문 OCR기능을 위하여 구축한 환경은 다음과 같다.

스캐너 :

제작사 및 모델명 : Hewlett Packard Scanjet+
입력속도 : 20장/분
용지크기 : A4
기록밀도 : 300DPI
ADF용량 : 30장

OCR 소프트웨어 :

제작사 및 모델명 : Caere Omnipage/386
판독속도 : 480장/일(A4 40줄기준)
인식률 : 92-97%
최소단위 : 8호활자
요구환경 : IBM PC 386/SX 이상
메모리 4MB 이상
하드디스크 : 10MB이상
폰트 : 영어(명조, 고딕 등 인쇄체)
불어, 독어, 이탈리아(선택사양)

다음은 서지통제양식의 원문과 이를 OCR 처리 후에 출력한 교정용 데이터의 예이다.

OCR 사용시 발견되는 오류에는 <그림7>에서 볼 수 있듯이 다음의 두가지 경우가 있다.

- 1) 문자인식 불가능: 원문상태의 불량 또는

OCR 제품의 인식범위를 벗어난 문자체나 도형문자로 인하여 인식하지 못한 경우로서 공급회사가 발표한 에러률의 범위에 속하는 경우이다. 여기에는 **로 표시되고 있다.

예) "2. PERFORMING I" ← "2. PERFORMING LAB"을 잘못 인식함

2) 타문자로 오인 : OCR 프로그램이 원문의 문자를 전혀 다른 문자로 오인한 경우로서 글자 사이의 간격, 얼룩, 인쇄 상태 등 여러가지 원인이 있을 수 있으나 공급회사가 발표한 에러률의 범위 밖에 속한다.

예) "14. KEYU/ORDS" ← "14. KEYWORDS"를 잘못 인식함

따라서 교정담당자의 면밀한 대조작업이 없을 경우 발표된 에러율을 훨씬 초과하는 오자가 발생하게 되지만 일단 대조작업을 거칠 경우에는 키보드 입력보다 월등히 데이터 오류가 적으며, 전체적으로 소요되는 시간도 키보드 입력에 비하여 약 1/4정도에 불과하였다.

일단 오류수정을 거친 서지정보는 텍스트 화일로 저장된 후, 워드프로세서를 사용하여 저자, 서명등의 필드를 식별할 수 있는 태그를 부여하는 편집과정과 이 화일을 서지정보 모듈로의 도입과정을 차례로 거침으로써 키보드로 입력된 영문서지정보와 동일하게 처리될 수 있다.(그림 9 참조)

4.2.3 한글자동색인

입력된 서지정보는 단어별 형태소 분석을 거쳐 키워드를 추출하는 한글자동색인 과정을

그림 7. 영문서지통제양식 원문

BIBLIOGRAPHIC DATA SHEET		1. REPORT NO.	2. PERFORMING LAB. Office of Tech Info & SRD	3. REPORT DATE Mar. 1991
4. TITLESUBTITLE Study on Construction of Technical Reports Management System Using Optical Technology				7. SUBJECT CATEGORY 0502
5. AUTHOR(S) Shin Sok Park, Sang Hun Lee, Ic Cheol Kim, Hye Jin Kim, Chung Ryul Kim.			8. PERFORMING ORGANIZATION REPORT NO. KSRI-91-52-IR	
6. PERFORMING ORGANIZATION NAME			9. CONTRACT OR GRANT NO.	
			10. TYPE OF REPORT IST REPORT	
11. SPONSORING ORGANIZATION				
12. SUPPLEMENTARY NOTES				
13. ABSTRACT This project is aimed at building a computerized management system which integrates publishing, storage, usage, and distribution of technical reports. Various kinds of printed materials resulting from research activities(e.g., reports, papers, memoranda, and so on) are valued not only as records and reference for future researches but also as fundamental data for policy-making and analyzing the adequacy of projects of research institution. Currently, such technical reports also serve as basic media for information flow within Korea. In a wider sense, therefore, the construction of an effective technical reports management system can enhance the contribution of research activities. Up until now, technical reports have been distributed mostly in printed formats and microfilms have been used for long-term storage of these documents digital technologies made it possible to design a new integrated technical reports management system.				
14. KEYWORDS Optical technology ; Hypertext, Image Database, Technical Report Automatic Indexing ; Multimedia ;				
15. CLASSIFICATION 01		16. SCHEDULE OF DECLASSIFICATION		17. NO. OF PAGES 106
				18. PRICE

그림 8. OCR을 이용하여 만든 텍스트 파일 예

1. REPORT NO. 2. PERFORMING I-B

BIBLIOGRAPHIC DATA SHEET
Info & SRD

Office of 1

Study on Construction of Technical Reports Management
System Using Optical Technology

5. AIJ-)-R(S)

shin Sok Park, Sang Hun Lee, Ic Cheol Kim,
Hye Jin Kim, Chung Ryul Kim.

8. PE-MING ORGANIZ-TION REPORT N-).

KSRI-91-52-IR

6. PERFORMIN-, ORGNIZATION NAME 9. (-IR-CT OR (;RANT N-).

10. TYPE OF REPORT 1st Report

11. S-11-; ORG-NIZATION

1-. SUPPLEI-IARY N-IES

18. AB'-IR-T

This project is aimed at building a computerized management
system which integrates publishing, storage, usage, and distribut-
ion of technical reports.

Various kinds of printed materials resulting from research
activities(e.g., reports, papers, memoranda, and so on) are valued
not only as records and reference for future researches but also
as fundamental data for policy-making and analyzing the adequacy of
projects of research institution. Currently, such technical reports
also serve as basic media for information flow within Korea. In a
wider sense, therefore, the construction of an effective techical
reports management system can enhance the contribution of research
activities.

Up until now, technical reports have been distributed mostly in
printed formats and microfilms have been used for long-term storage
of these documents digital technologies made it possible to design
a new integrated technical reports management system.

14. KEYU/ORDS

Optical technology;Hypertext; Image Database;Technical Report
Automatic Indexing;Multimedia;

15 . CIASS I F I C-1 ICN

0 1

1 16. -ILE OF I-CIASSIFICA I-N 17. N-). OF P-ES 106

그림 9. 영문요약문을 서지데이터베이스로 입력하기 위한 텍스트화일 예.

#1# KSRI-91-52-IR
 #6# // Office of Tech Info & SRD
 #8# 1991, Mar.
 #4# Study on Construction of Technical Reports Management System Using Optical Technology
 #5# Park, Shin Sok || Lee, Sang Hun || Kim, Ic Cheol || Kim, Hye Jin || Kim, Chung Ryul
 #13# 1st Report
 #14# This project is aimed at building a computerized management system which integrates publishing, storage, usage, and distribution of technical reports.
 Various kinds of printed materials resulting from research activities (e.g., reports, papers, memoranda, and so on) are valued not only as records and reference for future researches but also as fundamental data for policy-making and analyzing the adequacy of projects of research institution. Currently, such technical reports also serve as basic media for information flow within Korea. In a wider sense, therefore, the construction of an effective technical reports management system can enhance the contribution of research activities.
 Up until now, technical reports have been distributed mostly in printed formats and microfilms have been used for long-term storage of these documents digital technologies made it possible to design a new integrated technical reports management system.
 #16# Optical technology || Hypertext || Image Database || Technical Report || Automatic Indexing || Multimedia

통하여 데이터베이스에 추가된다. 여기에 사용되는 참조사전과 테이블은 다음과 같다.

- 동의어 테이블-시소러스와 유사한 기능으로서, 동의어들의 대표어를 선택, 사용 가능하게 한다. 예를 들면, 영어의 “be, is, am, are, was, were”의 경우도 동사 ‘be’로 대치시키는 용도로 사용 가능하다.
- 불용어 테이블-검색에서 의미어로 쓰일 수 없으면서 출현빈도가 높은 단어들을 제거하여 색인화일의 크기를 적정수준으로 유지하는데 사용한다.

예를 들면, 1) “우리, 이, 그들, 등, 것, 상, 수”와 같은 대명사나 불완전 명사 따위의 기능어(functional word) 2) “연구, 고찰, 문제, 큰, 높은” 등과 같이 내용구성에 관계가 없는 명사, 형용사, 부사 3) “본, 나타난, 있다, 되는” 등과 같은 동사와 그 변화형들이 여기에 속한다.

- 구(phrase) 테이블-각 단어보다는 서로 연관되어 특별한 의미를 가지는 단어들로 구성된다.

그 형태로는 1) 완전한 형태의 숙어 2) 이웃하는 단어간의 거리를 지정해 준다.

· 접사(접두사, 접미사) 테이블-텍스트는 문장으로 구성되며, 문장은 어절로, 어절은 단어로, 단어는 여러 형태소로 구성된다.

형태소는 크기가 가장 작고 독립된 의미를 지닌 언어표현으로서 어휘적 형태소와 문법적 형태소로 나눌 수 있다. 어휘적 형태소는 의미를 규정하는데 결정적 역할을 하고, 문법적 형태소는 형태소간의 문법적 관계와 다른 형태소 및 구성 성분의 의미를 명백히 해 주는 것이다.

이 테이블을 사용하여 어근과 접두사, 접미사 등의 접사를 최장 일치법에 의하여 분리시킴으로써 형태소를 획득하여 자동색인에 이용할 수 있다.

· 명사사전 - 명사, 복합명사를 분리하는데 사용된다.

· 조사 테이블 - 각 형태소의 접속 유형, 불규칙 음운현상 등을 해석하는데 사용된다.

· 동사사전 - 동사별 활용 예를 코드화하여 수록.

시스템은 텍스트의 한라인씩을 읽어 들여 3바이트 조합형으로 코드변환을 실시한 후 색인생성과정을 수행한다. 조합형 코드의 사용이유는 한글의 자소단위처리를 위해서이며, 3바이트 코드 사용이유는 패턴 검색과 크기 비교를 효과적으로 처리할 수 있고 2바이트 보

다는 처리가 쉽기 때문이다.

그밖에 한글, 한자, 일어, 로마자 데이터의 처리에는 다음의 원칙들이 적용되었다.

- 1) 한글, 한자, 일어 이외의 언어 즉 영어, 독어, 불어 등은 단어별로 분리하여 키워드로 삼되, 불용어 사전에 등록되어 있는 것은 제외한다.
- 2) 일어문자 (히라가나, 카타카나)는 영어에 준해 처리하되 한자에 붙어있는 조사 (예를 들면, -が, -は)는 분리하여 제거한다.
- 3) 한글, 한자는 명사사전, 조사사전, 동사사전을 참조하여 복합명사, 조사, 동사 등을 분리하여 명사, 복합명사만을 추출한다.

다음은 한글자동색인을 이용하여 서명 키워드를 추출하는 예이다.

(서명) 광기술을 이용한
 보고서관리시스템
 구축연구.

(형태소 분석 결과) 광기술 + 올
 이용 + 하 + L
 보고서
 구축+연구

(키워드 추출) 광기술
 보고서
 시스템
 관리시스템

* 키워드 추출시 '이용', '관리', '구축', '연구'는 불용어사전에 등록되어 있으므로

제외되지만 '관리시스템'의 경우는 복합명사로서 불용어가 아닌 '시스템'을 포함하고 있으므로 키워드로 추출되었다.

4.3 원문 이미지정보처리기능

원문 이미지정보 모듈에서는 이미지관리정보와 이미지정보를 관리한다. 이미지관리정보는 보고서별 원문이미지가 저장된 광디스크 번호, 보고서 종류별 이미지 존재 여부 등을 지시하는 정보로서 하드디스크에 저장된다. 이미지정보는 보고서 원문을 비트이미지로 저장한 것으로서 광디스크에 저장된다.

이미지정보는 스캐너, 펌웨어인 이미지처리용 보드, 그리고 전용 소프트웨어를 이용하여 처리된다.

4.3.1 이미지 관리정보

이미지관리정보는 검색된 서지정보가 지시하는 원문 이미지정보로의 연계를 위한 중간단계의 정보를 관리한다. 서지정보와 이미지정보의 연계키는 보고서 번호로서 서지정보 레코드내에 수록되어 있으므로 직접 연결될 수도 있으나 다음과 같은 이유에서 중간관리 단계가 필요하였다.

1) 연계키는 이미지정보가 수록되어 있는 디스크내에서의 논리적 주소가 되어야 한다. 그러나 서지정보의 입력과 이미지정보의 입력이 별도로 행해지는 경우, 예를 들어, 외부용역을 이용하여 이미지정보를 입력할 경우 서지정보레코드 부분에서는 이미지정보의 논리적 주소를 알 수 없다. 또한 이미지 정보의 추가, 수정의 작업결과 발생하는 논리적주소의 변동상황을 매번 서지정보에 반영시키는 것은 비효율적이다.

2) 이미지는 데이터 압축기법을 사용할 경우 A4 용지 한매당 20-30 킬로 바이트 정도를 차지한다. 따라서 광디스크 한면(300MB 정도)에 약 10,000 페이지, 다시 말하면, 100 페이지 정도의 보고서 100 권 정도가 수록될 수 있다.

수천종 이상의 보고서가 입력되면 수십개의 광디스크가 필요하게 되며 시스템은 어느 광디스크의 어느 면에 필요한 정보가 있는가를 확인하여야 한다.

본시스템에서는 이와 같이 이미지정보의 소장위치를 지시하는 관리정보를 다음과 같은 구조로 설계하였다.

표3. 이미지관리정보화일 구조

필드명	필드 타입	필드 길이	참조사항
보고서번호	문자	20	
광디스크번호	숫자	5	
최초입력일	문자	8	YYYYMMDD
최종갱신일	문자	8	YYYYMMDD
최종사용일	문자	8	YYYYMMDD

4.3.2 원문 이미지정보

4.3.2.1 이미지 화일 형식

이미지 화일을 저장하는 방식은 TIFF (Tagged Image File Format), PCX, GIF (Graphic Interchange Format) 등이 현재 많이 사용되고 있다. 그 중 TIFF와 GIF는 미국 및 유럽의 여러 회사가 컨소시움을 형성하여 제정한 특수표준화일형식이고 압축시 효율이 좋기 때문에 사용이 권고되고 있다. [35]

이미지 압축방식으로는 CCITT의 Group3, Group 4, TIFF Type 2, PCX의 Packed Byte 방식등이 있는데 압축비율은 이미지화일의 저장경제성에 중요한 항목이므로 DOS에서 일반적으로 사용되는 형식인 TIFF와 PCX를 대상으로 화일의 크기와 압축율을 조사했다. A4 규격 (219mm X 297mm)의 보고서 50페이지를 스캐너로 읽어 각 저장방식별로 그 크기를 비교한 결과 TIFF와 IMG 포맷은 1 페이지 당 평균 16K바이트, PCX는 평균 82 KB를 차지하였다. 이러한 화일 크기는 보고서에 수록되는 데이터에 따라 어느 정도 가감되겠지만 PCX 포맷이 다른 포맷보다 약 5배 정도 더 큰 것으로 확인되었다. (표 4 참조)

따라서 TIFF 포맷을 사용하면 평균 30:1,

PCX 포맷을 사용하면 평균 6:1 정도로 원문 이미지를 압축하여 저장할 수 있으므로 TIFF 포맷이 경제적임을 알 수 있다. 그리고 TIFF 포맷은 사무용DTP(Pagemaker, 장원금제등)의 대부분이 채택하고 있는 형식으로 향후 DTP시스템과 융합시 인터페이스를 제공할 수 있을 것으로 판단됨에 따라 본 시스템에서는 TIFF포맷을 채택하였다.

4.3.2.2 이미지 입력방식

본 시스템에서는 원문이미지 입력을 위하여 펌웨어 방식을 사용하였다. 다음은 이 모듈에서 사용한 스캐너와 이미지처리보드의 사양이다.

스캐너 :

제 작 사 : Fujitsu

모 델 명 : M3096E

입력속도 : 25장/분

용지크기 : B4

기록밀도 : 200, 300, 400DPI

ADF용량 : 50장

35. Clifford A. Lynch, op.cit., p.582.

표4. 이미지화일 형식 및 압축방식별 화일 크기 비교

	화일형식/압축방식별 화일 크기(50페이지 : 단위 바이트)					
	TIFF G4	TIFF G3	TIFF TYPE 2	TIFF RAW	PCX PACK-BYTE	IMG G4
총 계	834514	833898	833984	834088	4110810	834154
평 균	16690	16679	16677	16681	82215	16683

이미지처리보드:

제작사: KOFAX

모델명: KF-9250(2.5MB)

처리속도: 40장/분

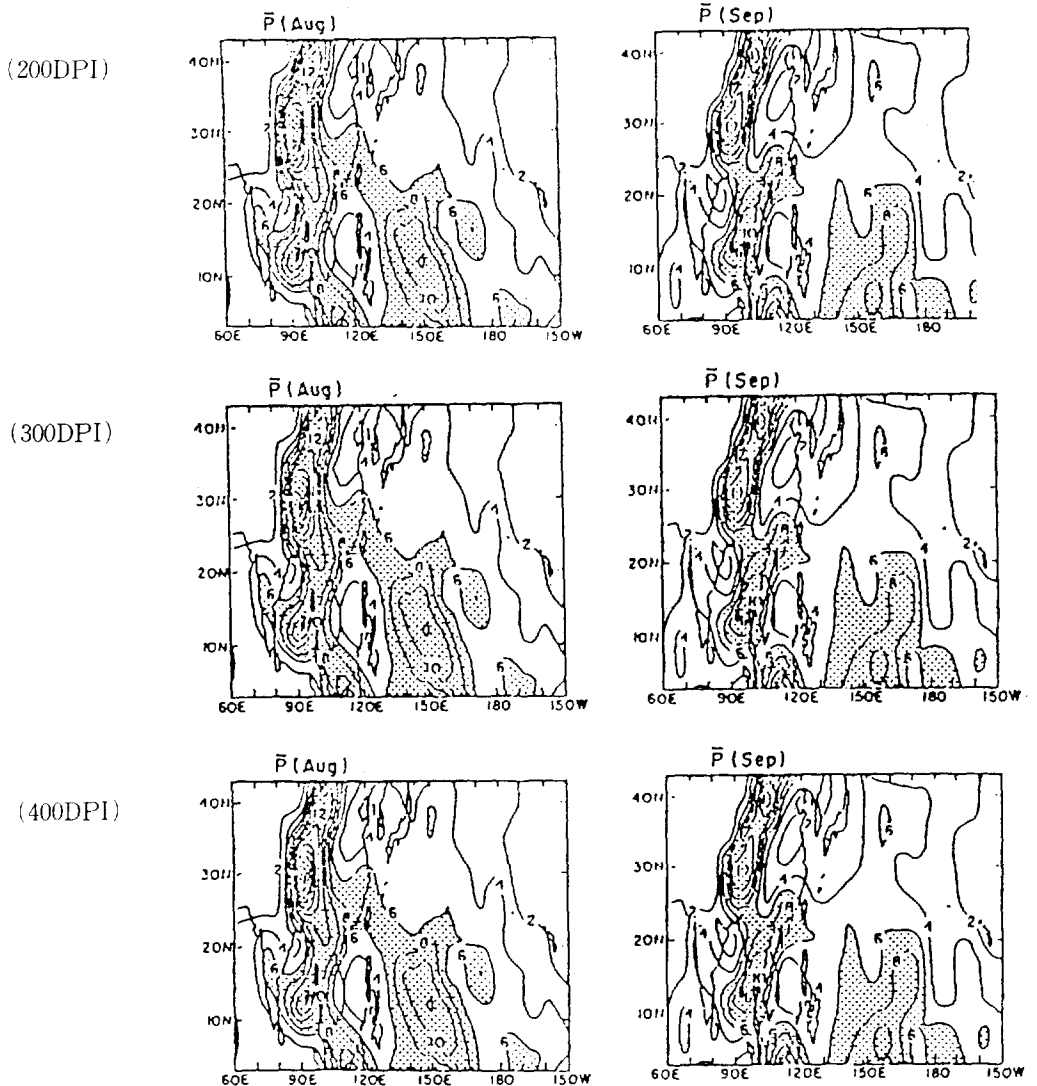
화일형식: TIFF G3/G4, PCX, IMG

화상형식: 흑백 문서형식, 흑백 사진형식(256 grey)

스캐너는 200-400 DPI의 기록밀도 및 음영 수준 256단계까지의 사진형식을 지원할 수 있으나 저장용량의 절감과 속도향상을 위하여 적절한 DPI를 결정하고자 200, 300, 400 DPI로 입력 및 출력을 통하여 해상도를 실험하였다.(그림 10 참조)

그 결과 보고서의 수록내용이 문자가 주인

그림 10. 보고서 원문을 해상도를 달리하여 스캔하여 프린트한 결과



경우는 200DPI로도 무난하고 상세한 도형을 포함하는 경우도 300DPI로 충분하다는 결론을 얻었다. 이들 결과를 마이크로펄름 자료와 대비한 결과도 동일했다. 이러한 구성에 의하여 실용성 있는 이미지 압축, 복원, 회전처리 능력을 사용할 수 있게 되었으며, TIFF G4 포맷 뿐만 아니라 KF-9250에 내장된 소프트웨어 라이브러리를 이용하여 PCX, IMG형식의 화일들도 처리가능하게 되었다.

4.3.2.3 이미지 저장방식

원문 이미지는 과제별로 부여된 별도의 디렉토리 아래에서 한 페이지의 이미지가 한 화일 방식으로 저장되며, 한 디렉토리 아래에서는 보고서의 성격에 따라 다음과 같은 방식으

로 화일명을 부여하도록 하였다.

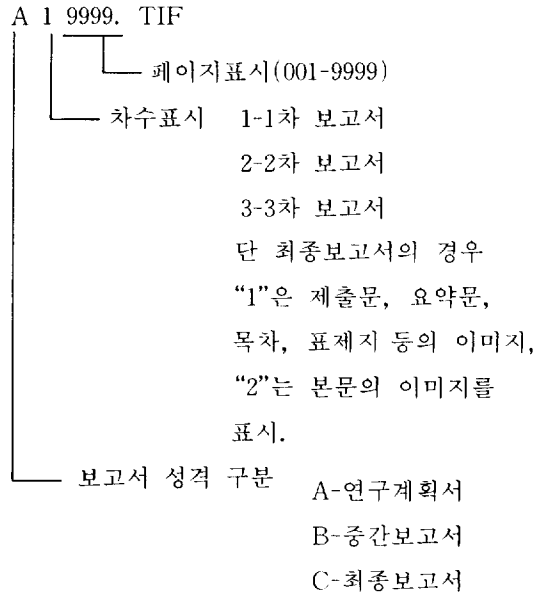


그림 11. 보고서 원문 입력화면

보고서 원문 입력 초기값			
광 드라이브	E	광디스크 번호	110
보고서 번호		KSRI-91-52-IR	
광기술을 이용한 보고서관리시스템 구축연구 / 박신석			확인
보고서 종류	차수	<input type="checkbox"/> 평가서 ?	취소
연구계획서 중간보고서 최종보고서 요약 최종보고서 본문 논문	1		
스캔 시작 페이지 번호		1	
스캔 페이지 증가		<input checked="" type="checkbox"/> +1	<input type="checkbox"/> +2
		<input type="checkbox"/> -1	<input type="checkbox"/> -2

이 방법에 의하여 한 연구과제에 관련되어 생산된 계획서, 중간보고서, 최종보고서를 일괄 처리하거나 특정 성격의 보고서를 지시하여 별도 처리할 수 있으며, 최종 보고서의 경우는 차수표시 "2"의 화일명을 사용하여 본문의 특정 페이지에 직접 접근할 수 있다.

예를 들면, 보고서번호 "KSRI-91-52-1R"은 KSRI9152.1R 디렉토리에 원문의 이미지가 저장되며, 이 디렉토리내의 C10005.TIF 화일에 저장된 목차 이미지에서 확인된 본문 50 페이지의 내용출력을 요구할 경우 시스템은 C20050.TIF 에서 가져온 이미지를 화면에 출력시켜 준다.(그림 11 참조)

원문 이미지를 입력하기 위해서는 원문의 크기, 상태, 포함된 그림이나 사진의 해상력 등에 따라 스캐너의 파라메타들을 새로 설정하거나 변경할 수 있어야 하며, 이미지가 한

페이지씩 입력될 때마다 즉시 화면으로 자동 출력하여 입력상태를 감시하고자 할 경우에는 화면출력을 위한 파라메타가 설정되어야 한다.(그림 12,13 참조)

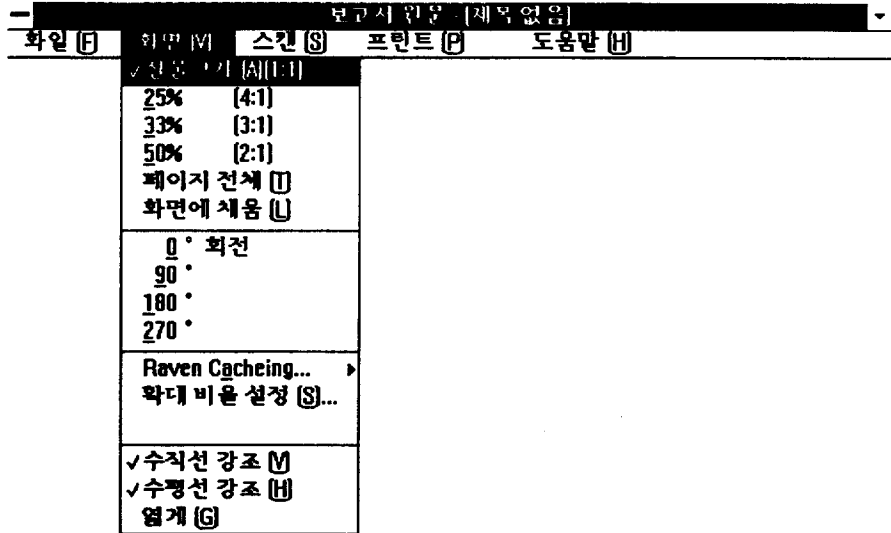
4.3.3 서지정보와 원문정보의 연결검색

보고서 정보검색은 우선 보고서 서지정보를 검색한 후 해당 보고서에 관련된 원문정보를 골라 화면에 보여주는 방식을 취한다. 다음은 이 과정을 순차적으로 설명한 것이다.

- 1) 서지정보검색은 단행본, 도서정보검색과 유사한 방식으로 구현하되, 이미지정보와 한 화면에서 운영되기 위하여 MS WINDOWS 환경을 사용하였다. 이용자는 서지정보검색용 메뉴의 저자, 서명, 주제어 및 검색어구 항목에 검색용 키워

그림 12. 스캐너 파라메타 설정화면

그림 13. 화면출력 파라메타 설정화면



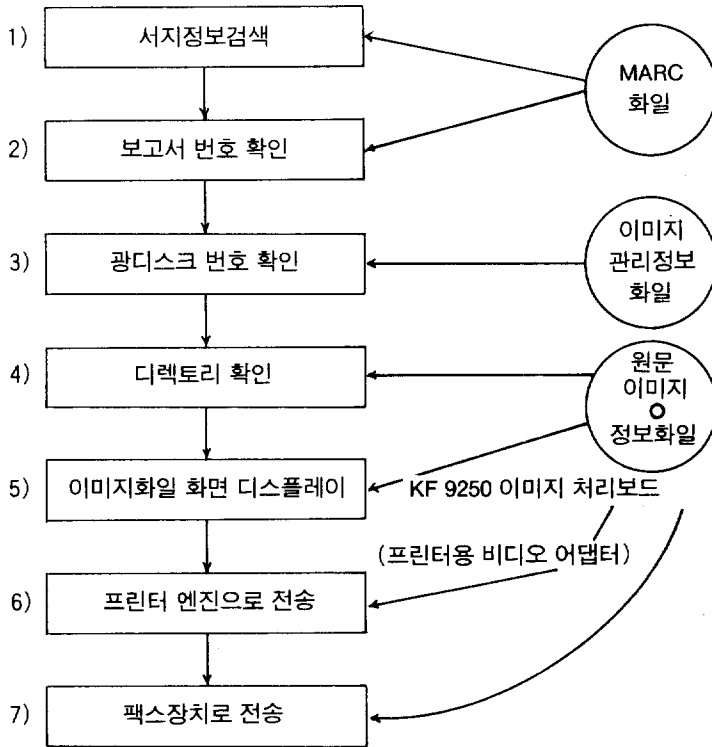
드를 입력하거나, ISBN, 보고서번호 등을 입력하여 원하는 보고서를 검색한 후 그 이미지를 요구한다.

- 2) 시스템은 서지정보에서 확인된 보고서번호를 연계키로 사용하여 이미지관리 정보를 확인한다.
- 3) 이미지관리정보에서 확인된 특정 광디스크가 드라이브에 들어있지 않을 경우 해당 광디스크를 넣으라는 메시지가 화면에 나타난다.
- 4) 관리정보와 일치하는 광디스크가 확인되면 시스템은 해당 디렉토리를 확인한 후 필요한 화일들을 열어놓고 이용자 명령을 기다린다.
- 5) 이용자는 화면조작 메뉴를 사용하여 화면에 나타난 원문이미지를 확대, 축소, 반전, 회전 또는 복수의 화면을 동시에

출력하는 등 원하는 형태로 조작하여 상세하게 열람할 수 있다.

- 6) 이용자는 화면에 나타난 보고서 원문정보의 전부 혹은 일부 페이지를 프린터로 출력할 수 있다. 현재 프린터 출력은 KOFAX사의 KF-9250 이미지처리보드에 연결된 프린터용 비디오어댑터를 통해 프린터 엔진으로 직접 이미지를 보내는 방식을 사용하여 프린트 속도를 높이고 있다. HP LaserJet III Si 프린터를 이용한 원문 이미지의 출력 속도는 A4 용지를 기준으로 할 때 평균 20 PPM 정도이다.
- 7) 이용자는 화면에 나타난 보고서 원문정보의 전부 혹은 일부 페이지를 컴퓨터에 내장된 팩스시스템을 통해 원거리 목적지에 있는 팩스장치로 전송할 수 있다.

그림 14. 시스템 기능 흐름도



현재 상용 팩스장치가 CCITT G3포맷을 사용하고 있으므로 G3포맷으로 전송한다. 만약 차후 G4팩스장치가 널리 보급 되면 G4포맷으로 원문이미지를 전송할 수 있을 것이다. 현재 원문이미지 전송용으로 시험중인 팩스시스템은 HAYES사의 JT9600 이다.

JT9600이 컴퓨터에 내장되면 원문이미지 처리모듈에서 자동적으로 이를 감지하여 팩스기능을 사용할 수 있도록 해줄 것이다.

5. 결론

과학기술투자 증가에 비례하여 국내보고서 생산량도 급증하고 있으나 보고서의 서지적 특성을 잘 반영하고, 최근 실용화되고 있는 디지털이미지 처리기술을 적절히 구사한 경제적인 보고서관리시스템 개발사례는 발표되지 않고 있다.

이에따라 디지털이미지처리방식의 아나로그 이미지처리방식에 대한 상대적 강점과 발전 잠재력을 인식한 기관들에서는 사무자동화용

으로 공급되고 있는 문서관리용 광화일링시스템을 보고서관리업무에 도입하는 사례가 늘고 있다.

그러나 이러한 시스템을 보고서관리업무에 도입하는 데는 다음과 같은 문제점들이 있다.

- 1) 보고서정보의 보존 및 이용기관인 도서관으로서는 불요불급한 기능인 문서작성 및 이미지편집기능의 부착으로 시스템의 규모 및 구입가격이 필요이상으로 클 뿐만 아니라, 필요한 기능의 전문성은 상대적으로 취약하다.
- 2) 레코드구조면에서 제한적인 입력 및 탐색항목설정과 키워드 설정방식으로 인하여 레코드구조가 경직되어 다양한 서지정보를 처리할 수가 없으며 검색 기능 역시 매우 제한적일 수 밖에 없다.
- 3) 단행본, 연속간행물등과 보고서의 서지정보처리 표준이 상이함에 따라 한 도서관내에서 두개의 데이터베이스가 각기 별도의 기준에 따라 작성되고 운영되는 것은 이용자교육, 시스템유지보수 및 운영전반에서 비효율적임이 틀림없다.
- 4) 서지데이터의 입력, 대량의 데이터의 소급변환등 도서관업무에서 발생하는 업무의 특성이 고려되어 있지 않다.
- 5) 이와 같이 표준화되지 않고 서지적으로 불충분한 보고서정보가 각 기관에서 생산된다면 보고서를 위한 국가종합목록데이터베이스 구축을 위해서는 또다른 변환작업이나 최악의 경우 재입력등의 사태가 필연적이다.

이에 따라 본 연구에서는 다음과 같은 특징

을 갖는 보고서관리시스템을 개발하였다.

- 1) 보고서정보의 보존 및 이용에 중점을 두어 시스템의 규모를 경제적인 수준으로 축소하고, 기능을 전문화하였으며 최소한의 기본적인 표준 이외에는 국내시장에서 도입가능한 하드웨어와 소프트웨어로 사용, 교체, 보완할 수 있는 개방형시스템을 구축하였다.
- 2) MARC의 개념과 도서관자동화시스템의 편목, 이용자목록검색 등의 기능을 그대로 도입함으로써 서지정보검색에 적합한 유연성을 확보하고, 도서관 자동화시스템과 완벽히 통합 또는 연계되어 운영될 수 있는 시스템을 구축 하였다.
- 3) 표시기호 880과 식별기호 ▼6 및 ANSI Z39.18에 입각한 서지통제양식 개념 등 보고서정보처리의 특성을 반영시킬 뿐만 아니라, 영문 OCR에 의한 영문서지통제정보의 신속한 소급변환과 이미지정보의 일괄적인 소급변환시의 신속한 처리방식을 구사하여 데이터베이스 구축에서 가장 많은 시간이 요구되는 소급변환부분을 효율화하는 등 도서관 업무의 특성을 시스템에 반영하였다.

이러한 특성에 의하여 이 시스템은 보고서관리시스템의 표준모델로 활용될 수 있다고 판단되며, 여기에 적용된 각종 표준들을 국가 보고서관리시스템에서 채택할 경우 국가차원의 보고서 종합목록/원문데이터베이스의 구축을 효율적으로 실현할 수 있을 것이다.