

# 인터넷 원거리출판의 응용과 PDF의 인쇄활용에 관한연구

유영수, 강영립, 김병현, 이광숙\*

중부대학교 인쇄공학과  
중부대학교 광고홍보학과\*

## A Study on the online of PDF Electronic Documents System

### Abstract

PDF(Portable Document Format) is a file format that Adobe advances postscript technique and use in managing document information or electric publishing(internet, CD-ROM, DVD). PDF is a devised document type for being able to read and print anywhere, independent of OS, printer type, resolution, and the kind of computer etc. Because this includes a compressing function, it transfers document through a small size of file in internet or intranet. In addition, that is a file format has various advantages- sharing of information and transferring documents in on line or off line environment.

In this paper, we developed electronic document system using PDF format. Electronic document system consists of filter, automatic indexing, special searching system and web server. The information used in this paper is database made using Zwon's DocuCom.

The filter recognizes various kinds of document structure. And according to property of document, it produces ASCII output. In addition to processing various formats of document, the filter can extract keywords in documents of MS WORD, Excel, PowerPoint, PDF, CAD etc. This filter uses the structure of window printer drive and can extract the information for text, page, font type and size from relevant document. The automatic indexing recognizes the formatted tag of document from ASCII text produced by filter and extracts adequate keyword to structure and property of document.

PDF electronic document system proposed in this paper can be used in Internet, PC communication. Users can choose and read electronic documents by two ways. First, users can choose and read relevant books using PDF electronic document homepage. Second, users can use PDF integrated-search system. User can search after inputting keyword

and choose reference field and type of data.

But, now, PDF products of Adobe can't support the Korean character. If this problem is resolved, we think that PDF applications system looks active. Although there is limited function in case of using Zwon DocuCom used in this study, we think that there isn't a great deal of difficulty in electronic document and building digital database.

## 1. 서론

다양한 디지털정보와 멀티미디어정보를 포함하고 있는 전자문헌들을 인터넷상에서 우리는 자주 접할 수 있다. 아무리 훌륭한 정보시스템과 네트워크 환경을 갖추어도 운용할 데이터 및 정보가 없다면 무용지물이다. 그리고 데이터의 생산과 축적기간은 시스템 구축기간보다 월등히 많은 시간과 인력이 투입되며 또한 표준성과 범용성이 결여된 데이터는 재활용할 수 없다는 이유 등으로 더욱 많은 연구가 이루어져야 할 것이다. 미래의 디지털도서관에 있어서 정보의 유통 및 공유라는 기본목적을 충족하기 위한 중요한 요소 중의 하나는 정보자료의 형식이라 할 수 있을 것이다(Schamber 1996, Collier 1997). 이러한 전자문헌의 포맷은 현재 여러 종류의 형식이 사용되고 있으며, 그 포맷 형식에 따라 제공하고 있는 기능 및 특징 또한 다양하다. 이 같은 전자문헌의 형식은 장차 디지털도서관에서 사용되어질 디지털 문헌의 포맷이 될 것으로 예상된다. 최근에는 PDF, SGML, HTML, XML, VRML 등의 포맷이 널리 사용되고 있다. 본 연구에서는 이 중에서 PDF를 중심으로 한 웹을 통한 원거리출판, 전자문헌, 데이터베이스의 구축에 관하여 살펴보고자 한다.

PDF(Portable Document Format)는 Adobe에서 일반 및 인터넷에서 통용되는 공통표준문서로 제안한 문서 파일 형식이다. PDF는 인터넷 이전에 이미 컴퓨터의 기종, 운영체제, 프린터 종류 및 해상도 등에 제한없이 어디서나 읽고 출력할 수 있도록 고안된 문서 형태이다. PDF는 인터넷에서 종이 인쇄물과 같은 역할을 하면서 전자문서 전자출판을 서비스 할 수 있는 솔루션이라 할 수 있다. PDF 문서는 다양한 문서의 포맷을 여러 시스템 환경에서도 제공하고 공유하고자 하는 경우, 문서 자료의 압축된 자료로 전송하고 보여 주고자 하는 경우, 원격지로 특정 문서를 주고받으며 원고 교정 및 , 고품질의 인쇄를 원하는 경우에 사용된다.

PDF 포맷으로 구성된 문서를 읽거나 출력을 하려면 Acrobat Reader와 같은 프로그램이 필요하다. PDF를 열람할 수 있는 방법은 2가지이다. 첫째, Adobe Acrobat Reader를 통하여 온라인(인터넷/인트라넷)상에서 웹 브라우저(익스플로러, 넷스케이프)에 플러그인(plug-in)하여 PDF를 열람하는 방법이다. 둘째, Acrobat Reader 자체 뷰

어(viewer) 프로그램을 통하여 PDF를 볼 수 있는 방법이다. PDF는 일반 HTML과 마찬가지로 전체 PDF 파일을 전송비율만큼 자료를 보여 주기 때문에 사용자는 전송 중에도 현재의 PDF 전송율만큼 자료를 볼 수 있다.

외국에서는 PDF를 다양하게 활용하고 있다. 예를 들어 2001년 4월 15일 하루에 Adobe의 웹에서 무료로 배포되는 Acrobat Reader는 90,000 카피가 다운로드 되었으며 전 세계적으로 약5천만 카피가 유통되는 것으로 추정된다. 현재 우리 나라에서는 Adobe의 Acrobat PDF의 경우 몇 개월 전까지만 해도 한글 처리가 원활하게 지원되지 못하기 때문에 널리 사용되지 못하였으나, 한국어도비와 한빛소프트에서 공동개발한 PDF Invader(Mac)와, 조광 프리프레스의 crackerjack 3.0(Mac,Pc)등은 P/S용 한글을 포함하고있어 널리 사용되고있다.

본 연구의 목적은 PDF 포맷을 이용 원격지로 특정 문서를 보내고 교정을 주고받으며, 고품질의 인쇄를 하는 경우와, 전자문헌시스템을 구축하는 것이다.

본 연구에서는 우리말의 지원이 가능한 제품으로 한국어도비와 한빛소프트에서 공동개발한 PDF Invader(Mac)와, 유니덱스와 한글과컴퓨터사에서 공동 개발한 EzPDF solution, 조광 프리프레스에서 판매하는 Crackerjack 3.0/ Quite A Box Of Tricks/ Imposing plus,(Mac,Pc) JPD internet HaanQ PDF Solution V3.0, Zwon사의 DocuCom을 이용하여 고급인쇄, 각신문사PDF서비스, 전자도서, 전자잡지, 학위논문, 약보 등 다양한 유형의 자료를 대상으로 원격지로 특정 문서를 보내고 교정을 주고받으며, 고품질의 인쇄를 하고, 데이터베이스를 구축하였다. 또한 PDF 본문정보에서 키워드를 지능적으로 추출하기 위하여 필터와 자동색인기를 이용하였으며 이용자는 PDF 전자문헌시스템에 접속하여 전문검색시스템, 분류 등의 방법을 이용하여 해당 문헌을 검색하고 읽을 수 있다.

## 1.2 연구의 방법 및 범위

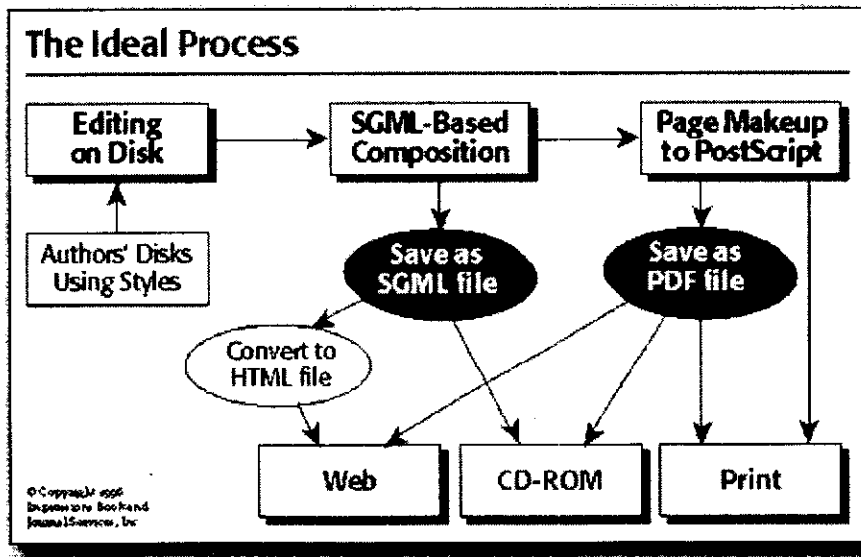
본 연구에서는 현재 기술을 이용한 전자출판을 위한 방법과 여러 유형의 데이터 포맷을 고찰하였다. 그리고 PDF 포맷의 역사, 특징, 관련 시스템, 활용에 관하여 살펴 보았으며 우리나라의 환경에 적합한 전자도서의 구조와 다양한 유형의 전자문헌으로 이루어진 전자문헌시스템인 PDF문서를 원격지로 문서를 보내고 교정을 주고받으며, 고품질의 인쇄를 하는 방법을 연구하였다.

PDF 전자문헌시스템은 서치 정보와 PDF 포맷으로 이루어진 본문정보를 갖고 있다. 필터는 다양한 형태의 문헌 구조를 인식하고 문헌의 특성에 따라 최종적으로 ASCII 출력물을 생산한다. 자동 색인기는 필터에 의해 생성된 ASCII 텍스트에서 문헌의

포맷된 태그를 인식하고 문헌의 구조와 특성에 적합한 키워드를 추출한다. 이용자는 인터넷, PC통신을 이용하여 PDF 전자문헌시스템에 접속하여 키워드, 분류 등을 이용하여 해당 문헌을 검색하고 읽을 수 있다.

PDF 전자문헌시스템은 인터넷, PC 통신에서 운영되며 이용자는 두 가지 방법에 의하여 전자문헌을 선택하고 읽을 수 있다. 첫째, PDF 전자문헌 홈페이지를 이용하여 특정 자료를 선택하고 해당 문헌을 읽을 수 있다. 둘째, PDF 통합검색시스템을 이용한 방법이다. 이용자는 키워드를 입력하여 탐색하며 원하는 검색 필드와 자료의 유형을 선택할 수 있다. 2001년 6월 현재 홈페이지 제작틀인 나모웹에디터로 유명한 나모 인터랙티브사에서 나모딕서치를 개발하여 PDF 전용서치엔진 프로그램을 개발시판하고 있기도 하다.

## 2. PDF의 개념과 전자출판



<그림 1> 전자출판의 처리 과정

### 3. PDF 시스템

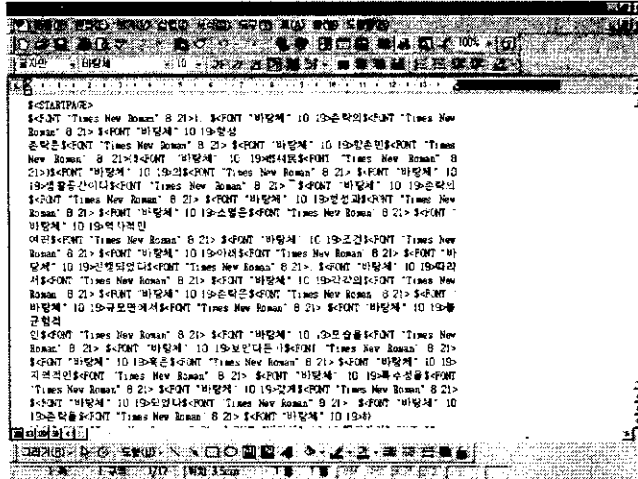
분류	시스템	기능
보기 도구	Acrobat Reader Acrobat Exchange	PDF Viewer 기능과 검색 텍스트 추출, 프린팅 기능
제작 도구	PDFWriter PDFDistiller Acrobat Exchange  Acrobat Capture	PDF 파일 제작기능 PS 파일 PDF 변환기능 PDF 문서 편집(전자출판 요 소, Link, Bookmark, Actions Edit 등) 문서 스캔, OCR, PDF변환
2 검색 도구	Acrobat Catalog	검색 인덱스 제작 기능

<표 7> Acrobat의 시스템 구성

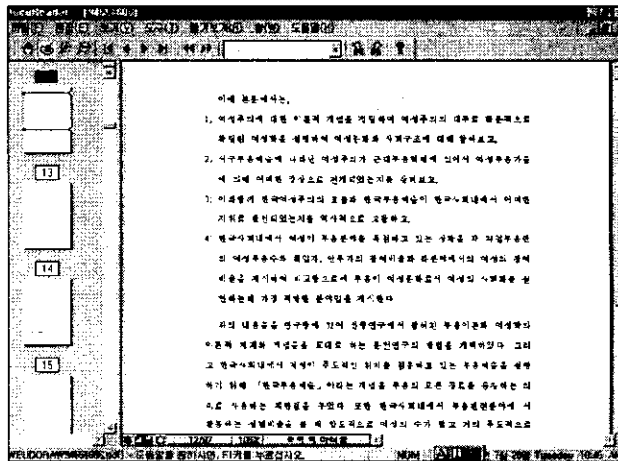
시스템	주요 기능
DocuDriver	다른 응용 프로그램의 출력을 포스트스크립트 파일로 생성
DocuMaker	포스트스크립트 파일을 PDF 문서로 변환
DocuPlus	PDF 문서의 검색, 주석 첨부, 문서 편집에 사용
DocuReader	브라우저로서 모든 PDF 문서의 검색에 기울이면 포스트스크립트 보기 엔진

<표 8> DocuCom의 시스템 구성

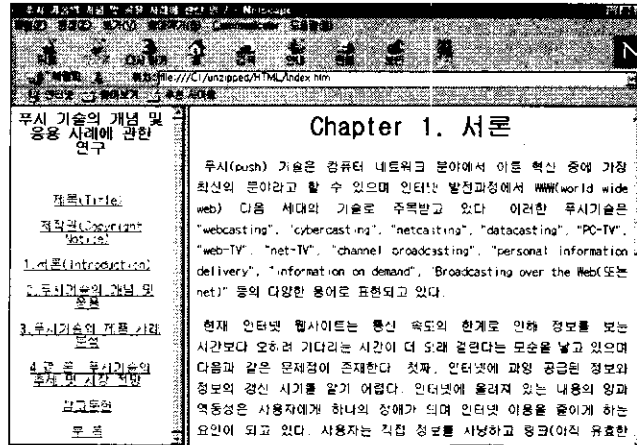




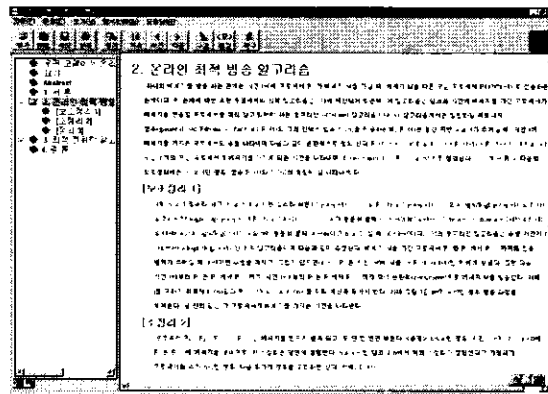
<그림 3> 필터의 처리 사례



<그림 4> PDF 썸네일 화면 사례



<그림 5> HTML 화면 사례



<그림 6> SGML 화면 사례

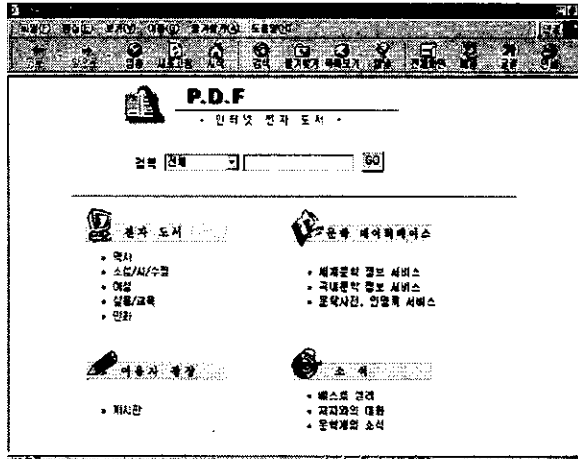
### 4.3.3 PDF 전자문헌시스템의 운영

본 연구에서 구축된 PDF 전자문헌시스템은 인터넷에서 운영되며 이용자는 두 가지 방법에 의하여 전자문헌을 선택하고 읽을 수 있다.

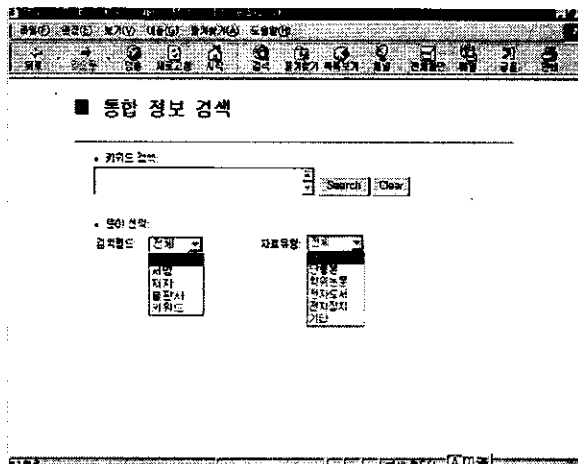
첫째, PDF 전자문헌 홈페이지를 이용한 방법이다. 이용자는 분야별로 구분된 메뉴를 이용하여 문헌을 선택하고 해당 본문을 읽을 수 있다. 또한 이용자는 전자문헌의 유형별로 구분하고 키워드를 입력하여 특정 전자문헌을 선택하고 읽을 수도 있다.

둘째, PDF 통합검색시스템을 이용한 방법이다. 이용자는 원하는 검색 필드(저자, 출판사 등)와 자료의 유형(전체, 전자도서 등)을 구분할 수 있으며 키워드를 입력하여 특정 도서를 검색할 수 있다. 탐색결과에 대한 간략 메뉴, 상세 메뉴의 예는 다음과 같다.

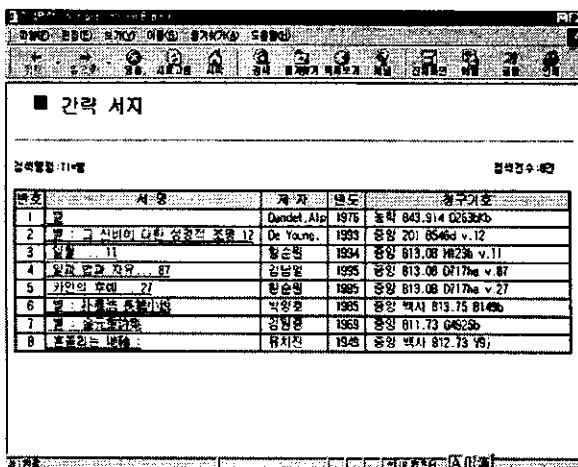




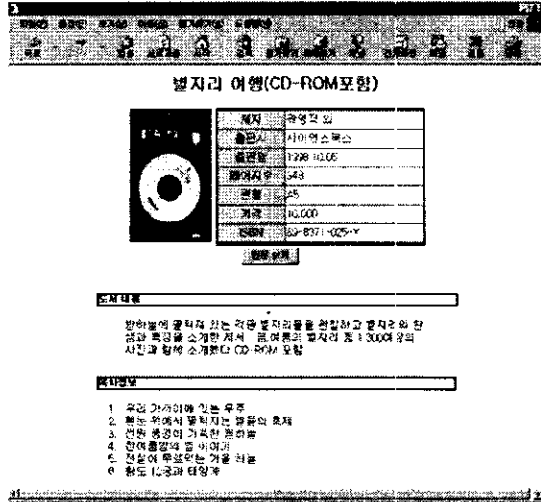
<그림 7> 전자문헌 홈페이지 화면



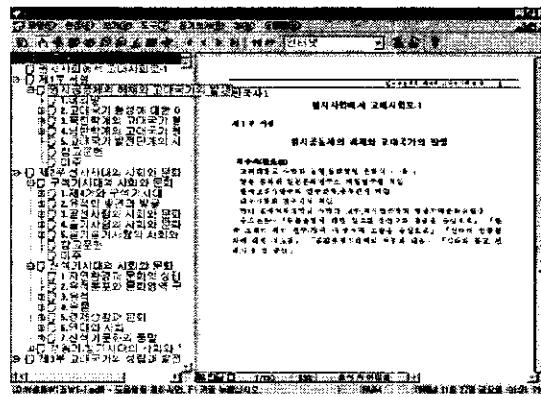
<그림 8> 전자문헌 통합검색시스템의 화면



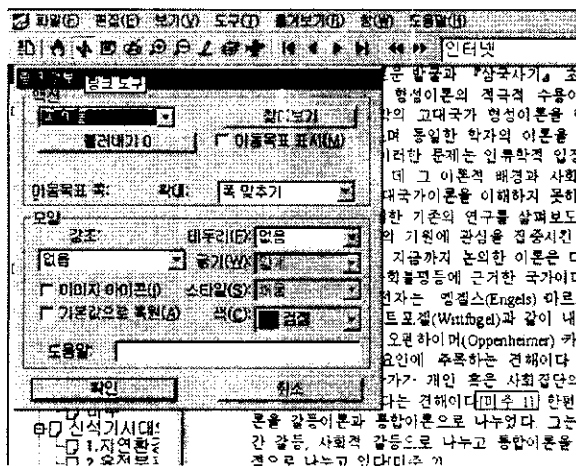
<그림 9> 탐색 결과의 간략 메뉴 화면



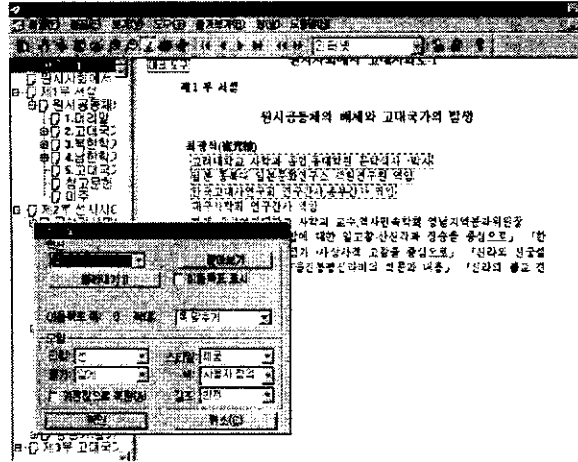
<그림 10> 탐색 결과의 상세 메뉴 화면



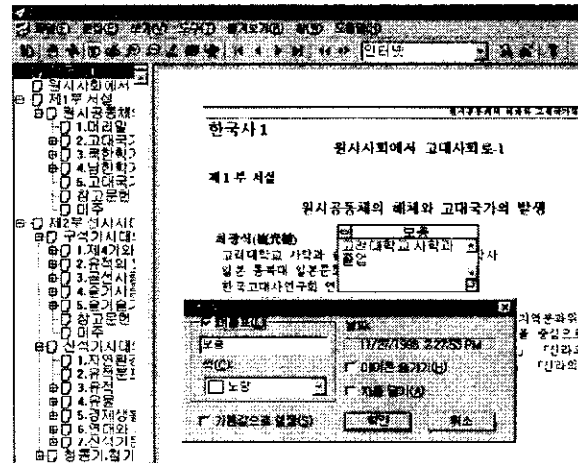
<그림 11> 북마크 기능의 사례



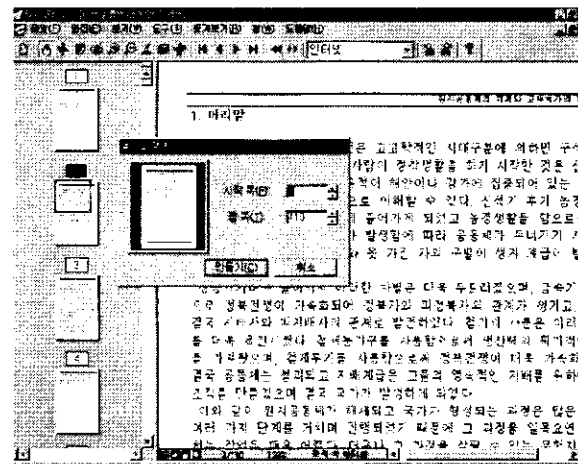
<그림 12> 링크 기능의 사례



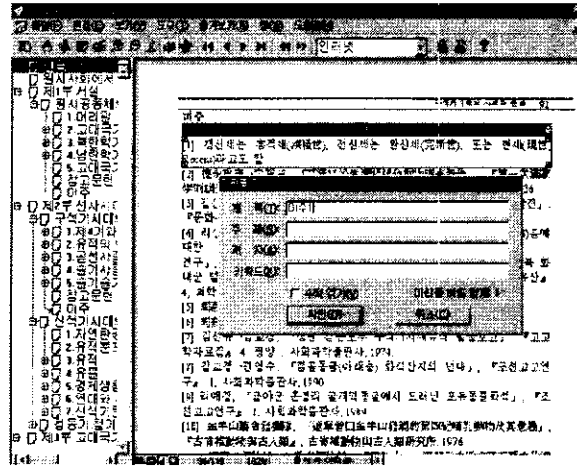
<그림 13> 마커 기능의 사례



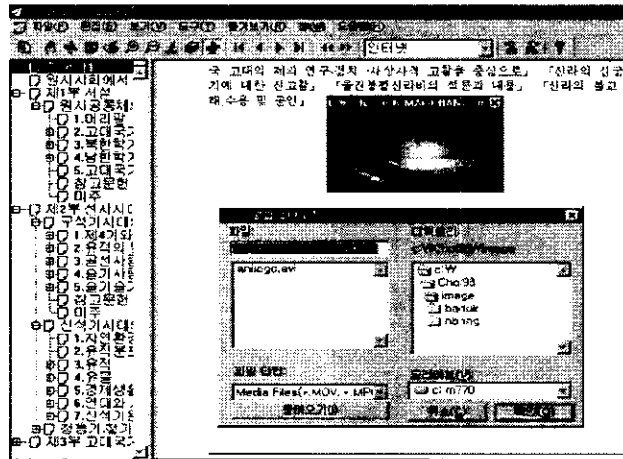
<그림 14> 노트 기능의 사례



<그림 15> 썸네일 기능의 사례



<그림 16> 아티클 기능의 사례



<그림 17> 미디어 기능의 사례

사용 기관	시스템 유형
기업, 공공기관, 개인,  인터넷, 인트라넷	도서, 잡지, 학위논문
	기업 전자문서 시스템, 도면 관리 시스템, 마케팅 자료 인사관리(입사지원서, 주민등본, 자격증,
	디지털도서관, 제품 카탈로그 및 매뉴얼, 기업 전자 사보, 지도, 학습지

<표 4> PDF의 활용 시스템 사례

### 3. 결 론

본 연구에서는 PDF 포맷을 이용한 PDF인쇄활용시스템을 개발하였다. PDF인쇄활용 시스템은 서지 정보와 PDF 포맷으로 이루어진 본문정보를 갖고 있다. 본문정보와 인쇄활용 실험은 한빛소프트사의 PDF Invader와PDF파일 생성조광프리프레스사에서 판매하는 Crackerjack 3.0으로 출력인쇄실험하고, 대만의 Zwon사에 의하여 개발된 DocuCom을 이용하여 데이터베이스를 구축하였다.

필터는 다양한 형태의 문헌 구조를 인식하고 문헌의 특성에 따라 최종적으로 ASCII 출력물을 생산한다. 필터는 다양한 형태의 포맷으로 이루어진 문헌을 처리하며, 윈도우 95, 98,2000/서버, NT, Mac 환경에서 MS WORD, Excel, PowerPoint, PDF, CA D 등의 여러 유형의 문헌에서 키워드를 추출할 수 있다. 이 필터는 윈도우 프린터 드라이브 구조를 이용하며 해당 문헌으로부터 텍스트, 페이지, 폰트 유형 및 크기에 관한 정보를 추출할 수 있다. 자동 색인기는 필터에 의해 생성된 ASCII 텍스트에서 문헌의 포맷된 태그를 인식하고 문헌의 구조와 특성에 적합한 키워드를 추출한다.

본 연구에서 구축된 PDF 전자문헌시스템은 인터넷, PC통신에서 운영되며 이용자는 두 가지 방법에 의하여 전자문헌을 선택하고 읽을 수 있다. 첫째, PDF 전자문헌 홈페이지를 이용하여 특정 자료를 선택하여 해당 문헌을 읽을 수 있다. 둘째, PDF 통합검색시스템을 이용한 방법이다. 이용자는 키워드를 입력하여 탐색하며 원하는 검색 필드와 자료의 유형을 선택할 수 있다.

몇 개월 전부터 Adobe의 PDF 제품은 한글을 지원하고 있다. 이 문제가 해결되어 PDF 관련 응용시스템은 활성화될 것으로 보인다.

본 연구에서 사용한 한빛소프트사의 PDF Invader로 PDF파일생성하고 조광프리프레스사에서 판매하는 Crackerjack 3.0으로 출력 인쇄실험 했다. 현재 한빛소프트사의 PDF Invader는 출력소 인쇄소 기획실에 널리 사용되고 있다.

조광프리프레스는 Crackerjack 3.0로 출력서비스와 고급인쇄를 2년 전부터 서비스하고 있다.

### <참고 문헌>

- 권영하 1997. "PDF를 이용한 디지털 도서관 구축 방안", 디지털도서관 5 : 86-95.
- 김용, 남궁황 1998. "전자문헌개발도구에 관한 고찰 : SGML, HTML과 PDF를 중심으로", 정보관리연구. 29(1) : 1-19.
- 김두식 1993. 전자출판론 I, 서울 : 타래.
- 김무상 1992. "DTP와 전자출판," 전자공학회지. 19(7) : 626 - 631.
- 김성백 1991. 윤곽선 한글 폰트의 생성과 한글 수용을 위한 포스트스크립트의 개선에 관한 연구. 서울 : 서울대학교 석사논문.
- 김영상, 이종철 1992. "Postscript 언어의 사전 객체에 대한 기억 장소 관리 전략에 관한 연구", 경북대전자기술연구지 13 : 41-52.
- 박홍서 1997. "전자문서의 표준-SGML", 디지털도서관 5 : 72-82.
- 여미라, 한윤섭 1990. "Postscript의 구성과 활용", 전기과학회지 39(7) : 60-70
- 오무석 1997. "소규모 도서관의 디지털 라이브러리 구축을 위한 소고," 정보관리연구. 28(4) 16-23.
- 원세연 1996. "전자도서관과 SGML," 과학기술정보 워크샵 회의록, 40-52.
- 유사라 1996. "멀티미디어 정보 관련 기술과 표준안에 대한 고찰," 정보관리학회지. 13(2) : 39-71.
- 이수연 1992. "전자출판의 발전상 및 해결과제," 전자공학회지. 19(7) : 567 - 577.
- 이수연 1987. "LBP와 Postcript", 정보과학회지 20 : 71-78.
- 임순범 1991. "윤곽선 글자꼴의 처리 기술 및 활용 추세", 전자공학회지 18(11) : 69-76.
- 임순범 1992. "전자출판에서의 서체 및 출력기술," 전자공학회지. 19(7) : 578 - 583.
- 정동주 1995. 한글 출력을 위한 범용 컬러 postscript 프린터 드라이버의 설계. 서울 : 송실대학교 석사학위 논문.
- 조왕근 1997. SGML DTD 및 데이터베이스 구축에 관한 연구. 서울 : 한양대학교 교

육대학원 석사학위논문.

- 최원태 1998. "A Digital Library Prototype for Access to Diverse Collections," 한국문헌 정보학회지, 32(2) : 295-307.
- 최원태, 임상원 1998. "PDF를 이용한 학위논문 데이터베이스 구축", 한국정보관리학회논문집 15 : 93-96.
- 石田晴久 외 1987. "Postscript (1)", NIKEI BYTE 1987 4 : 175-189.
- 石田晴久 외 1987. "Postscript (2)", NIKEI BYTE 1987 5 : 211-221.
- 石田晴久 외 1987. "Postscript (3)", NIKEI BYTE 1987 6 : 179-189.
- Adobe 1995. "Adobe offers a new approach with 'Capture'", Electronic Documents 4(4) : 11-12.
- Keyhani, Andrea. 1994. "OCLC Electronic Publishing: Creating New Pathways to Information," OCLC Newsletter. 301 : 16.
- Barrertt, Jon. 1987. " Kirk Reistroffer : Designing a Raster Image Processor", BYTE : 171-180.
- Brian, E. and Aale C. Waldt, 1995, The SGML Implementation Guide : A Blue Print for SGML Migration, Berlin : Springer-Verlag.
- Collier, M. W. 1997. "A model for the electronic university library", Towards a worldwide library: a ten year forecast. Proceedings of the 19th International Essen Symposium 23-26 September 1996. Edited by Ahmed H. Helal and Joachim W. Weiss. Essen, Germany, Universitätsbibliothek Essen, 1997 : 180-190
- Rawlins, Gregory J. E. (1993). "Publishing over the Next Decade," Journal of the American Society for Information Science. 44(8) : 474-479.
- Jacso, P. 1996. "A plug for plug-ins, ActiveX controls: add-ons enhance the 3.0 web browsers' capabilities enormously", Information Today 13 (10) : 26-27, 32.
- Lemay, Laura, Teach Yourself : Web Publishing. Indianapolis, IN: Sams Publishing, 1995
- Lovegrove, W. S. and Brailsford, D. F. 1995. "Document analysis of PDF files: methods, results and implications", Electronic Publishing 8(2/3) : 207-220.
- Lunin, L. F. 1995. "Content and context: imaging's buzz words for the '90's", Information Today 12(7) : 20-22.
- Lyon, J. and Brailsford, D. 1996. "Using Acrobat in journal publishing", Digital Publishing Technologies 1(10) : 17-18.
- Mostert, P. and Fransen, P. 1997. "Technical and functional aspects of electronic journal systems developed in TULIP and EES projects", Library Acquisitions : Practice and Theory 2(3) : 347-353.