

전문 처리를 위한 웹 기반 정보시스템 구현

김 상 도[†] · 문 병 주[†] · 류 근 호^{††}

요 약

1990년대 웹 개념의 등장으로 인터넷은 개방형 네트워크, 사용자 친근성 및 편의성이라는 특성을 바탕으로 급속히 일반화 되면서 각종 정보를 제공하는 정보 시스템에도 커다란 변화를 가져왔다. 웹은 텍스트 정보뿐만 아니라 이미지, 오디오, 비디오 등 멀티미디어 정보를 쉽게 제공함으로써 기존 정보 시스템의 웹 기반으로의 전환을 가속화 시키게 되었다. 물론 정보 내용도 텍스트 위주의 단순 초록 정보에서 전문(Full-Text) 정보로 급속히 이전되기 시작하였으며, 전문 정보를 처리하는 다양한 문서 포맷들도 등장하기 시작하였다. 하지만, 전문 정보는 본질적으로 대용량의 시스템 메모리, 막대한 처리시간, 보다 많은 전송용량 등을 요구하게 되므로 웹 기반 전문 정보 시스템의 구축시 이러한 요인들을 세밀하게 비교 평가하여야 한다.

따라서 본 연구는 웹 기반 정보 시스템에서 어떻게 전문을 처리할 것인가에 초점을 맞추어 문서의 통합 정보 기능을 제공하는 정보 시스템을 설계 구현한다. 먼저 인터넷상에서 표준 문서형식으로 사용 중이거나 개발중인 문서 포맷들의 특징을 검토하고, 정보 시스템의 관점에서 전문정보를 어떤 문서 포맷으로 처리하는 것이 적합한가를 실험하고 평가한다. 그리고 이를 기반으로 본 연구의 목적인 전문 정보를 PDF 문서로 제공하는 웹 기반 정보 시스템을 설계 구현한다.

Implementation of Web-based Information System for Full-text Processing

Sang-Do Kim[†] · Byung-Joo Moon[†] · Keun-Ho Ryu^{††}

ABSTRACT

As Internet is popularized by the advent of Web concept having characteristics such as open network, user-friendly, and easy-usage, there are many changes in Information systems providing various information. Web is rapidly transferred traditional Information systems to Web-based Information systems, because it provides not only text information but also multimedia information including image, audio, video, and etc. Also, as information contents were changed from text-based simple abstract information to full-text information, there was appeared various document formats processing Full-text information. But, as they naturally demand large system memory, long processing time, broader transmission bandwidth, and etc, estimating of these factors is necessary when constructing information systems.

This paper focuses on how to design and construct information system processing full-text information and providing function of an integrated document. Primarily, we should review standard document format which is used or developed, and any document format is appropriate to process full-text information in review with viewpoint of information system. Also, practically we should construct information system providing full-text information based on PDF document.

† 정 회 원 : 한국전자통신연구원
 †† 종신회원 : 충북대학교 컴퓨터과학과 교수
 논문접수 : 1999년 2월 6일, 심사완료 : 1999년 5월 3일

1. 서 론

정보 검색, 정보 유통 및 정보 표현이라는 정보 제공 체계가 IP(Internet Protocol), HTML(HyperText Markup Language), 그리고 HTTP(HyperText Transport Protocol)의 등장으로 1990년대 초에 획기적인 변화를 맞이하였다. 이미 인터넷은 '정보의 바다'라고 부를 정도로 엄청난 양의 정보가 존재하게 되었으며, 그 이용자 수도 수천만 명에 이른다[19].

인터넷의 급성장의 요인으로 HTML이라는 문서 포맷을 꼽을 수 있다. 즉, HTML은 누구나 손쉽게 배우고 사용할 수 있을 뿐만 아니라 하이퍼텍스트, 하이퍼미디어 기능까지 제공함으로써 인터넷에 보편화의 길을 열게 되었다. 하지만, 최근 들어 HTML의 단순성, 즉 확장성, 구조성, 그리고 데이터 검사 기능 등에서의 단점들이 인터넷 확산의 걸림돌로 작용하고 있다. 특히, 기술문서, 문헌 등을 다루는 기존의 정보 시스템의 웹 기반 전환을 더디게 하는 원인이 되고 있다[7].

이는 정보의 문서 포맷이 텍스트 기반인 경우 HTML 문서로 손쉽게 변환하여 제공할 수 있지만, 이미지, 워드로 작성된 문서의 경우 별도의 방법으로 해결하여야 하기 때문이다. 더욱이 이러한 정보 시스템에서는 수십 내지 수백 페이지에 달하는 엄청난 원문을 매일 신속하게 분류하고 변환하여 구축하여야 한다. 따라서 원문을 처리하는 정보 시스템에서는 다양한 원문 정보를 신속하게 변환하고 저비용으로 구축할 수 있는 방법을 찾는 것이 웹 기반으로 전환하는데 있어 매우 중요한 관건이 되고 있다.

본 연구는 이러한 관점에서 현재 인터넷상에서 표준으로 사용하거나 연구 중인 문서 포맷들의 특징을 검토하고, 정보 시스템의 관점에서 전문 정보를 어떤 문서 포맷으로 처리하는 것이 적합한가를 실험하고 평가한다. 그리고 실질적으로 전문 정보를 PDF 문서로 제공하는 정보 시스템을 설계 구현한다.

본 논문의 구성은 다음과 같다. 제2장에서는 정보 시스템에서 원문 정보를 축적하기 위한 문서 포맷이 어떻게 발전되어 왔는지를 살펴보고, 제3장에서는 정보 시스템에 필요한 각 문서 포맷의 현황 및 장, 단점과 변환 방법을 검토하고 평가한다. 그리고 제4장에서는 원문을 PDF 문서로 처리하는 정보 시스템을 제안하고 구현한다. 제5장에서는 구현한 시스템에 대한 평가와 마지막으로 제6장에서는 결론을 맺는다.

2. 정보 시스템에서의 전문 정보 축적

정보 시스템에서 원문 정보 축적은 1960년대 방대한 양의 문헌을 보유하고 있는 도서관에서 그 기원을 찾을 수 있다. 1966년 미국회도서관에서 MARC(Machine Readable Cataloging)라는 서지 데이터를 코드화하는 방법을 표준화함으로써 도서관 전자화의 기틀이 마련되었다[12]. 그 후 서지데이터에 이어 전문 데이터까지 축적하기 위한 노력들이 이루어지기 시작하였다.

하지만, 전문을 디지털 정보로 축적하는 것은 많은 시간과 인력을 요구하는 작업이다. 즉, 정보를 디지털 정보로 축적하고 관리하는데 드는 인적, 물적 비용이 시스템을 구축하는데 드는 비용보다 훨씬 많이 든다. 특히 원문 전체를 데이터베이스로 구축하기 위해 데이터를 인코딩(Encoding) 하는데 많은 비용과 시간이 소요된다. 이에 전문 정보를 보다 효율적이고 저비용으로 축적할 수 있는 인코딩 방법들의 모색이 필요하다.

이러한 인코딩 방법으로 텍스트 기반의 전문을 축적하기 위한 ASCII-Based, 스캐너를 이용하여 전문을 이미지화하는 Image-Based, 그리고 웹의 등장으로 주목 받기 시작한 Markup-Based 등이 주를 이루게 되었다[3,12]. 하지만, 결과적으로 정보가 생성단계부터 디지털 정보로 이루어지지 않은 이상은 앞의 인코딩 방법들은 많은 시간과 비용이 소요되기 때문에 전문 정보 구축 비용, 정보 관리와 사용자 인터페이스 측면에서 보다 나은 인코딩 방법이 요구된다.

최근 들어 이러한 인코딩 방법으로 PDF(Portable Document Format)라는 새로운 문서 포맷이 널리 이용되기 시작하고 있다. PDF는 대부분의 정보가 워드 프로세서로 작성됨에 따라 새로운 문서 형식으로 부가되기 시작한 것이다. 일반적으로 워드 프로세서로 작성된 정보는 단순한 텍스트뿐만 아니라 도표, 그림이 포함된 복합 문서 형태가 대부분이다. 기존의 호스트 내지 클라이언트/서버 개념의 정보 시스템에서는 워드 프로세서로 작성된 원문을 그대로 저장할 수 있다. 하지만, 웹을 기반으로 하는 정보 시스템에서의 문서 포맷은 웹 환경에 잘 적응하는 표준화된 문서 포맷을 유지하는 것이 중요하다. 이미 웹에는 HTML이라는 표준화된 문서 포맷이 존재하고 있다. 하지만, HTML은 매우 단순하여 복합문서를 처리하는 것이 매우 어려울 뿐만 아니라 막대한 수작업과 시간이 소요된다. 더욱이 복합문서를 이미지화하는 경우에는 많은 데이터 저

상 공간의 요구됨 뿐만 아니라 데이터의 수정이 불가능하여 정보의 재활용 측면에서도 매우 떨어지게 된다. 이러한 두 가지 문서 포맷의 단점을 보완한 것이 PDF이다.

더불어, Markup 기반의 문서 포맷도 기존 HTML의 단점을 보완하기 위한 새로운 문서 포맷에 대한 연구가 활발하게 이루어지고 있다. SGML(Standard Generalized Markup Language), XML(eXtensible Markup Language)이 그 대표적인 주자로 저작 및 변환 툴, 그리고 브라우저에 대한 연구가 활발히 진행되고 있어 이를 활용한 정보 구축이 증가될 것으로 예상된다[7]. 전문 정보의 문서 포맷을 어떤 포맷으로 채택할 것인가는 각 정보 시스템의 특성에 어느 정도 의존하겠지만, 대체적으로 문서 포맷이 정보 구축 및 관리 비용의 절감과 더불어 정보 공유와 전달을 통한 정보 활용도를 높이는 가장 핵심적인 요인이 되므로, 결국 이러한 인코딩 방법 중 정보 공유, 전달 및 활용성을 가장 충족시키는 방법이 널리 이용될 것이다.

3. 전문 구축 포맷 검토 및 실험 평가

앞에서 언급한 바와 같이 기존 정보 시스템이 웹 기반으로 전환하는 경우에 있어서 가장 큰 문제점은 방대한 양의 원문 정보를 어떻게 처리하여야 하는가이다. 기존 호스트 내지 클라이언트/서버 기반의 정보 시스템에선 전문 정보를 축적하기 위한 인코딩 방법으로 주로 ASCII 기반이나 이미지 기반의 문서 포맷이 가장 널리 이용되어 왔다. 물론 최근에는 워드 프로세서가 널리 이용됨에 따라 디지털화된 전문 정보가 생성되고 있어 폐쇄 개념이 강한 기업전산망, 즉 그룹웨어에선 그 자체를 축적하는 경우도 늘고 있다. 하지만, 일반적인 서비스를 기반으로 하는 정보 시스템의 경우 전문 정보를 일정한 문서 포맷으로 통합하여 축적하여야 하므로 ASCII 기반이나 이미지 기반이 중심을 이루고 있다.

이에 반해 웹에서는 HTML과 같은 마크업 언어(Markup Language)가 그 기반을 이루고 있다. 따라서 기존의 정보 시스템이 웹 기반으로 전환하고자 하는 경우 전문 정보를 Markup 기반으로 구축하는 것이 가장 우선적으로 고려될 수 있다. 하지만, 기존 ASCII 기반이나 이미지 기반의 전문 정보를 Markup 기반으로 변환하는 것은 엄청난 인력과 시간이 소요되는 방

대한 작업이나 일예로 최근 미국회도서관을 중심으로 기존 MARC을 이용하여 축적한 원문 정보를 SGML 문서 포맷으로 변환하기 위한 방법[12] 등이 연구되고 있으나 변환에 요구되는 시간과 컴퓨터 자원상의 비용이 엄청난 것으로 추정되고 있다.

따라서 기존 정보 시스템에서 웹 기반의 정보 시스템으로 전환하고자 하는 경우 Markup 기반이 아닌 새로운 문서 포맷이 요구되고 있으며, 이러한 문서 포맷으로 최근 PDF가 주목을 받고 있다. 특히, PDF는 기존 ASCII나 이미지 문서 포맷뿐만 아니라 워드 프로세서로 작성된 원문 처리도 매우 용이하여 웹 기반의 정보 시스템에서 새로운 문서 포맷으로 유용하게 이용할 수 있다. 본 연구에서는 이러한 관점에서 기존 정보 시스템과 웹 기반 정보 시스템에서의 문서 포맷에 대한 장단점을 비교 분석하고 각 포맷별 문서 구축에 대한 장단점을 비교 분석한다.

3.1 기존 정보 시스템에서의 문서 포맷

3.1.1 ASCII-Based

ASCII 기반 시스템은 기존의 모든 텍스트를 키보드 입력하여 ASCII 코드로 저장하는 텍스트 응용 기법을 사용하는 방법과 OCR(Optical Character Recognition) 소프트웨어를 사용하여 ASCII 코드로 변환시키는 방법을 적용한다[10]. ASCII 기반 시스템을 통해 전문에서 사용된 용어를 기준으로 하여 전문 검색과 주제 검색이 가능하며, 기존의 색인보다 효율적인 색인을 자동 생성할 수 있으므로 정보 검색의 속도를 향상될 수 있고, 저장 공간이 절약된다. ASCII 코드의 정보는 텍스트 전체, 또는 각 부분들에 대한 재포맷과 재출력, 출판에 위한 텍스트 편집에 의한 인쇄본, CD-ROM출판, 온라인 서비스 등 다양한 유형으로 활용이 가능하다. 단점이라면 원문이 비 ASCII일 경우에 ASCII 변환이 어렵거나 불완전한 점이며, 텍스트가 보유하고 있는 텍스트 구조를 나타내지 못하기 때문에 텍스트에 대한 구조 탐색이 불가능하다[4].

3.1.2 Image-Based

이미지는 1960년대 후반 마이크로 필름을 사용한 사무자동화에서 출발하여 1980년대 광 디스크(Optical Disk)의 발전으로 이용이 가속화되었다[10,22]. 즉, 종이 문서의 보존, 관리 및 재활용하기 위한 광 파일링 시스템에서 주로 이용되어 왔으며, 웹 사이트인 LG 상

남도서관[26]에서도 잡지류 1,000종과 회의자료 150~200건 등 약 20만 건의 이미지 데이터를 기반으로 서비스를 제공하고 있다[12].

이미지는 저장 매체의 안정성, 관리의 편리성으로 인하여 전문 정보의 무결성(Integrity) 보장, 물리적인 저장 공간의 절약이라는 장점을 가지고 있다. 하지만 상대적으로 이미지 기반이므로 전문 검색과 구조에 기반한 검색이 불가능하며, 대규모의 기억용량이 필요하고 이미지의 삽입, 삭제 등의 조작이 어려워 응용력이 낮다는 단점을 지닌다. 또한, 이미지 처리방법이나 이미지 크기를 줄이기 위한 낮은 해상도를 사용하는 경우 이미지의 질(Quality)에도 문제가 발생하게 된다. 따라서 정보의 검색 및 공유, 전달이 중요시 되는 정보 시스템에서는 비효율적이다[3].

3.2 웹 기반 정보 시스템에서의 문서 포맷

3.2.1 Markup-based

HTML은 웹을 위해 개발된 언어이다[8,14]. 1990년대 웹이 인터넷의 관도를 바꿀 만큼 엄청난 발전을 이룩하였으며, 이의 발전에 가장 큰 공헌을 한 것이 바로 HTML이다. 이는 HTML이 웹에서 하이퍼텍스트, 하이퍼미디어 기능을 지원하며, 누구나 사용할 수 있을 만큼 간단하고, 특별한 데이터 타입을 사용하지 않는 단순한 텍스트이기 때문에 이식성이 높고 사용이 편리하기 때문이다. HTML은 하이퍼 링크를 통해 다른 문서와 연결이 가능하도록 문서의 논리적 구조를 규정하기 위하여 설계된 일종의 SGML의 문서 형식 정의(DTD : Document Type Definition)이다. 웹 브라우저는 HTML로 작성된 문서를 해독하여 보여주는 역할을 하며 웹 브라우저의 발전에 따라 HTML 기능도 점차 확장되면서 최근에 발표된 사양의 경우 특정 브라우저에서만 지원하는 사례가 늘고 있다. 더욱이 HTML의 표준화는 당분간 지속적으로 이루어질 전망이다[15,16,17].

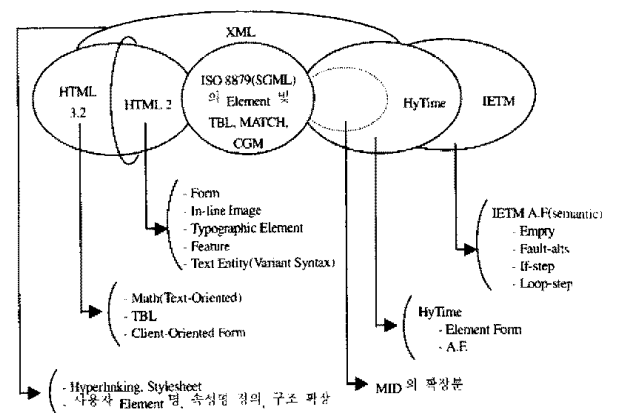
개인 홈페이지 또는 단순한 홈페이지 작업을 위해서는 HTML이 매우 편리하나 대용량의 온라인 출판에 HTML을 이용하기란 매우 불편하다. 이는 HTML의 기능적인 측면에서 살펴 보았을 때 화면상에 보여지는 기능 외에는 별다른 기능을 제공하고 있지 않을 뿐만 아니라 고정된 태그만을 사용하고, 또 페이지 레이아웃 형태를 임의로 지정할 수 없기 때문이다. 이와 같은 단점 중에 많은 것이 CSS(Cascading Style Sheet)

로 해결이 될 수 있지만, 근본적으로 고정적인 태그 셋(Tag set)에서 발생하는 문제나 HTML 문서의 구조적 정보를 담을 수 없다는 것은 여전히 개선할 수 없는 문제이다[11,21].

HTML은 정해진 틀에 맞추어 작성해야 하므로 각 문서가 갖는 논리적 구조를 올바르게 표현하지 못하는 단점을 갖고 있다[6]. 이러한 단점의 해결 방안으로 SGML[13,20]이 표준화되었다. SGML은 1986년 국제표준화기구(ISO)의 표준(ISO 8879)[15]으로 제정되었고, 1992년에 일본공업표준(JIS x4151)[17]으로 제정되었다.

SGML은 한 문서의 논리 구조와 내용을 기술하기 위한 언어이다. SGML을 이용하면 문서 기술에 필요한 태그를 생성할 수 있으며 문서의 내용이나 내용 구조를 정의할 수 있고, 다양한 응용들 사이에서 구조화된 데이터를 상호 교환할 수 있다[1].

SGML은 시스템 또는 플랫폼에 독립적으로 동작하고 문서의 구조를 저장할 수 있기 때문에 문서 구조를 기반으로 한 다양한 응용에 사용할 수 있고, 또 CALS (Continuous Acquisition and Lifecycle Support), EC (Electronic Commerce) 등과 같은 업계의 표준으로 정착하여 많이 사용되고 있다. 하지만 SGML 자체가 워낙 복잡하기 때문에 SGML 전체를 지원하는 소프트웨어의 개발이 용이하지 않다. 또한 SGML은 소프트웨어 산업 전반에 걸쳐 널리 사용될 수 있도록 범용 목적으로 만들어져 있어 화학식의 표현이나 기타 다른 특수 용도의 목적으로 사용하려면 그 목적에 맞는 브라우저를 일일이 개발해야 한다.



(그림 1) Markup-based 문서 포맷간 관계도

웹이 지금까지 발전하는 데에는 누구나 쉽게 만들고 사용할 수 있는 HTML의 단순함이 큰 역할을 하였지

한 사용자의 요구가 다양화되고 이러한 요구를 수용하려는 시도가 바로 XML이 등장하게 된 이유 중 하나이다. XML이 등장하게 된 또 하나의 이유는 SGML이 너무나 복잡하고 어렵다는 것이다. 즉, HTML과 SGML의 장점을 수용하면서 단점을 극복한 것이 XML이다[11].

(그림 1)에서 보는 바와 같이 XML 사양을 SGML의 입장에서 살펴 보면, SGML에서 거의 사용하지 않는 것은 모두 없애고 꼭 필요한 기능만을 수용하였지만 SGML의 중요한 많은 기능이 그대로 남아 있기 때문에 SGML의 서브셋(Subset)이라고 말할 수 있다. 따라서 XML이 SGML의 서브셋이기 때문에 SGML을 XML로 손쉽게 변환할 수 있고, XML을 수정하지 않고 모두 SGML의 응용에서 사용할 수 있다는 특징이 있다.

3.2.2 PDF-based

PDF는 Adobe Systems사가 디지털 서류를 전송하기 위해 개발한 특별한 파일 형식의 하나이다[9]. 문자를 비롯하여 도표, 그림 등 문서에 포함된 내용을 전용 뷰어(Viewer)인 Acrobat Reader를 사용하여 표시, 출력하는 전자문서 배포용 데이터 형식이다. PDF는 국제표준기구(ISO)에서 공식 채택되어 전자출판의 국제 표준으로 널리 사용하고 있는 포스트스크립트 페이지 기술 언어(PostScript)를 이용한 데이터 파일이다. PDF 문서는 정보 전달의 새로운 파일 양식으로 선진국의 산업계는 물론 공공행정 기관에서도 문서 관리 시스템의 표준 파일 양식으로 정착되어 사용하고 있으며, 최근 원문을 제공하고 있는 잡지출판사의 웹 정보 시스템에서 원문의 문서 포맷으로 많이 사용하고 있다.

PDF는 CD-ROM, 인터넷, 인트라넷, 전자메일(E-Mail) 등의 모든 매체에서 텍스트는 물론 컬러 사진, 음성, 동영상까지도 삽입하여 사용할 수 있고, 인쇄된 기록 자료도 자동 문자 인식으로 글꼴, 나란 편집 등 원본의 모양을 유지하면서도 전문 검색이 가능한 파일 양식이다. 또한 PDF는 플랫폼, 매체와 어플리케이션에서 독립, 원본 문서의 페이지 포맷, 멀티미디어, 전문 검색과 하이퍼링크 기능, 기밀 문서 보안 장치, 자동 문서 인식과 컴팩트한 파일 사이즈라는 특징을 가지고 있다.

3.3 문서 구축

정보는 처음 생산될 때 정보 시스템에 적합하게 작

성되길 원한다. 따라서 작성된 정보를 수정하여 활용하게 된다. 즉, 생산되는 정보는 종이 문서, 워드 문서, 웹 문서 등 다양한 형태로 발생하게 된다. 정보 시스템에서는 다양하게 입수되는 문서 포맷을 변환하여 이용자에게는 동일한 문서 포맷으로 제공하는 것이 필요하다. 또한 가급적 표준 문서 포맷을 이용하는 것이 정보 공유와 전달에 중요하다.

향후 XML 관련 기술의 발전으로 XML이 정보 시스템에서의 표준 문서 포맷으로 활용될 경우 매우 효과적으로 전문 정보를 구축할 수 있을 것으로 기대되지만 그 때까지 기다릴 수 없는 정보 시스템 입장에서는 나름대로 효과적인 문서 포맷을 선택하여 활용하여야 하는 어려움이 있다.

이 연구에서는 현재 정보 시스템의 소스가 되는 워드 문서와 종이 문서를 정보 시스템에 적합한 문서 포맷으로 변환하는 문서 변환에 대해 살펴본다.

3.3.1 HTML, XML 문서 변환

최근 문서를 작성하는 툴로는 아래한글, MS-Word 등 상용 워드 프로세서를 주로 사용하고 있다. 따라서 HTML 에디터(Editor)나 HTML DTD 사양에 따라 텍스트 어플리케이션을 이용하여 HTML 문서로 작성하기 보다는 워드 문서를 HTML 문서로 변환하는 방법이 편리하다. 변환은 워드 프로세서에 포함된 HTML 문서 변환 기능을 이용한다. MS Word 97, 아래한글과 같은 워드 프로세서에는 워드 문서를 HTML 문서로 자동 변환해 주는 기능을 갖고 있어 문서를 쉽게 HTML 문서로 변환할 수 있다. 그러나 변환된 HTML 문서는 페이지 레이아웃(Layout)이 원본과 틀리며, 그림이 포함된 복합 문서인 경우 그림이 각각 파일로 구성된다는 문제점이 있다. 필요에 따라 변환된 HTML 문서를 재편집할 수 있으나 많은 시간과 비용이 소요되어 비효율적이다.

이러한 문제점을 해소하기 위해 제안된 것이 SGML, XML이다. 특히, XML은 HTML, SGML의 장점을 수용하고 단점을 보완한 것으로 관심이 높아지고 있다[11]. XML 문서를 만들기 위한 첫 번째 작업은 만들고자 하는 문서의 논리적 구조를 표현할 수 있는 DTD를 만드는 것이다. 이것은 XML 사양에서 제시한 문법에 따라야 한다. 논리적인 구조만을 가지고 있는 XML 문서가 외부로 보여지기 위해서는 포맷팅 처리가 필요하다. 이를 제공하기 위해 SWG(SGML Working

(Group)는 SGML의 포매팅 언어인 DSSSL (Document Style Semantics and Specification Language)을 간소화 해서 사용한다. DSSSL을 이용해서 포매팅을 한다는 것은 DSSSL의 스타일 언어를 이용해서 특정 DTD에 대한 스타일 쉬트의 작성을 의미한다. 아직은 정보 시스템에서 XML 문서 포맷을 활용할 수 있는 틀이 없으며 필요에 따라 어플리케이션을 개발하여야 하는 부담이 있다. 그러나 확장성과 활용성이 높은 XML에 대한 연구가 활발히 진행되고 있어 점차 활용이 증가할 것이다.

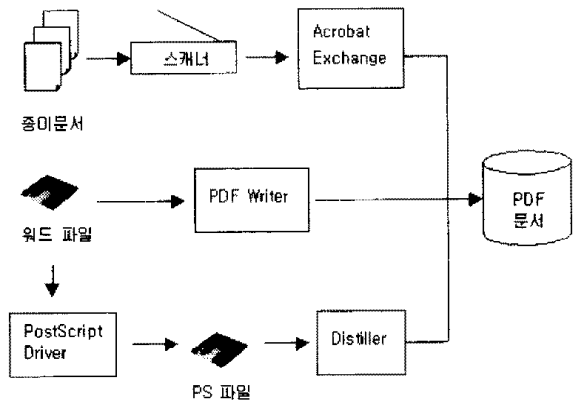
3.3.2 PDF 문서 변환

PDF는 Adobe Systems사의 「Acrobat」[23]이나 대만의 Zeon사에서 개발한 「DocuCom」[24]을 이용하여 작성할 수 있다. 「Acrobat」은 워드 문서를 PDF 문서로 쉽게 변환이 가능하며 다단, 페이지 레이아웃이 원문과 동일한 형태를 유지한다는 특징이 있다. 과정은 먼저 원문을 전자 출판의 표준인 포스트스크립트(Postscript) 형식으로 변환하고 다시 PDF 문서로 변환하게 된다. PDF 문서의 화면 출력은 Acrobat Reader를 이용하게 된다. Acrobat Reader는 웹 브라우저에 플러그인(Plug-in) 되어 이용의 편리성을 높이고 있다.

만일 정보가 디지털화가 되지 않은 종이 문서인 경우에는 위와 같은 방법으로 직접 변환할 수 없다. 종이 문서를 디지털화하여 정보 시스템에서 활용하기 위해서는 이미지 문서 포맷을 이용할 수 있다. 스캐너를 이용하여 종이 문서를 이미지 형식으로 디지털화하여 기계 가독형으로 변환하게 된다. 웹 브라우저에서 이미지 문서를 출력하기 위해서는 이미지 형식에 맞는 뷰어(Viewer)가 필요하다. 이미지 뷰어는 웹 브라우저의 외부 응용(External application)으로 등록하여 이용하게 된다. 이미지 문서 포맷은 「Acrobat Exchange」을 이용하면 PDF 문서로 쉽게 변환이 가능하기 때문에 PDF로 변환하여 웹에서 활용할 수 있다.

(그림 2)는 입력 소스별 PDF 문서 변환 과정을 도식화한 것이다. PDF 문서로의 포맷 변환은 소스 형태에 따라 종이 문서의 변환과 워드 파일 문서의 변환 두 가지로 구분된다. 소스의 형태가 종이 문서인 경우 우선 스캐너로 스캐닝하여 이미지 포맷으로 변환한 후 다시 PDF 포맷으로 변환하는 과정을 거쳐야 한다. 이 경우 PDF 문서는 텍스트 기반이 아니고 이미지 기반이 되어 PDF의 특징인 텍스트 인식이 되지 않는다는

단점이 있다. 텍스트 기반의 PDF 문서로 변환하기 위해서는 이미지 문서를 OCR 소프트웨어를 이용하여 텍스트로 변환하는 방법이 있으나 본 연구에서는 고려하지 않았다.



(그림 2) PDF 포맷 변환 과정

윈도우 기반의 워드 프로세서 파일인 경우에는 프린터로 출력하는 것과 같이 PDF Writer를 프린터로 지정하여 출력하여 쉽게 PDF 포맷으로 변환이 가능하다. 또한 워드 파일을 포스트스크립트 파일로 변환한 후 PDF 포맷으로 방법이 있다.

PDF 문서 변환 툴로는 「Acrobat」과 「DocuCom」이 있다. 「Acrobat」은 Adobe사의 독자적인 폰트인 Type 폰트 [URL 1]를 사용하며, 「DocuCom」은 트루타입 폰트를 사용한다. 이미지 파일과 영문 PDF 문서 생산에는 「Acrobat」이 유용하지만 한글 PDF 문서를 작성하기 위해서는 별도로 개발된 포스트스크립트 드라이버 [25]를 이용하여 포스트스크립트 파일로 변환한 후 다시 「Acrobat Distiller」를 이용해 PDF 문서로 변환하여야 한다. 한글로 작성된 워드 파일의 한글 PDF 문서 변환은 「DocuCom」을 이용하면 쉽게 이루어진다.

3.4 문서 포맷 변환 시험

앞에서는 정보 시스템의 전문 정보 문서 포맷에 대하여 알아 보았다. 여기서는 정보 시스템에서 활용되는 문서 포맷으로 변환한 결과에 대하여 살펴본다. 시험 대상의 소스는 주로 이용하는 워드 프로세서인 MS-Word 97로 작성된 문서를 대상으로 하였다. <표 1>은 시험 문서의 소스 파일 내용을 보여주며 총 65 페이지로 전체 파일 크기가 1,539KB이다. 내용은 텍스트와 도표 13건, 그림 21건이 포함된 복합 문서이다.

모범은 칼라 이미지가 12건, 흑백 이미지가 9건이 포함되어 있다.

<표 1> 소스 파일의 내용 구성

구분		내용
파일 크기		1,539KB
페이지 수		65페이지
도표 수		13 건
그림 수	칼라	12 건
	흑백	9 건

시험 방법은 소스인 워드 파일을 이용하여 틀을 사용하여 직접 HTML과 PDF로 변환과 워드 문서를 이미지로 변환하고 다시 PDF로 변환하는 방식으로 진행하였다. HTML 문서 포맷으로의 변환은 워드 프로세서에 포함된 기능을 이용하였으며 PDF 문서 포맷의 변환은 한글 PDF 문서를 작성할 수 있는 툴인 DocuCom 2.01을 이용하였다. 그리고 이미지 문서 포맷으로의 변환은 소스 내용을 프린터로 출력하여 스캐닝하여 Tiff(Tagged Image File Format) 포맷으로 변환하고 이를 다시 이미지 형태의 PDF 문서로 변환하였다. 그 결과는 <표 2>와 같다.

<표 2> 문서 포맷별 시험 변환 결과

문서포맷	파일크기	페이지당 크기	비 고
워드	1,519KB	23.4KB	• MS-Word 97
HTML	858KB	13.2KB	• HTML 파일 : 1 • 이미지 파일 : 21개 (GIF 포맷)
PDF 텍스트	891KB	13.7KB	• 텍스트 인식 가능
이미지	1,761KB	27.1KB	• CCITT G4 Tiff 포맷 • 200DPI
	2,730KB	42.0KB	• 300DPI
PDF 이미지	1,756KB	27.0KB	• 텍스트 인식 불가능 • 200DPI
	2,725KB	42.0KB	• 300DPI

각 문서 포맷별 변환 결과 파일 크기에서 HTML과 PDF 문서가 원본의 1/2 수준으로 줄어들고, 200DPI로 스캐닝한 이미지포맷의 경우 비슷한 크기를 보였다. 변환된 HTML 문서는 앞에서 문제점으로 검토한 바와

같이 원본과 편집 체계가 다르게 나타났으며, 문서일 상자에 포함된 내용 일부가 변환이 되지 않은 것이 확인되었다. 또한, 수식과 이미지는 별도의 Gif(Graphics Interchange Format) 이미지로 변환되었다.

이미지 문서의 가장 큰 단점은 해상도이다. 스캐닝 해상도를 높일수록 출력 해상도를 높일 수 있으나, 파일의 크기가 커지는 것을 알 수 있다. 이는 이미지 문서 포맷으로 원문 정보를 구축할 때 적절한 스캐닝 해상도를 선택할 필요성을 보여준다. 본 연구에서는 스캐닝 해상도를 200DPI로 선택하였다. 200DPI로 스캐닝할 때 화면과 프린터 출력시 내용을 확인하는데 크게 부리가 없으면서 파일의 크기도 적당하다고 판단하였다. 스캐닝한 이미지 포맷을 변환한 PDF 이미지 문서를 PDF 리더로 화면과 프린터 출력을 시험한 결과 이미지 뷰어의 출력물과 해상도에서 비슷한 결과를 보이면서 사용자 편의성을 높일 수 있다는 결과를 얻었다.

시험 결과 가장 우수한 문서 포맷으로 결과를 얻은 것은 역시 텍스트 기반의 PDF 문서이다. 워드 파일에서 PDF 문서로 쉽게 변환이 가능하고 원본과 동일한 편집 체계를 유지하면서 하이퍼 링크, 북마크(Bookmark), 섬네일(Thumbnail) 등 다양한 부가 기능을 포함하여 사용자 편의성을 높이면서 파일 크기도 작아 디스크 스페이스도 절약된다는 결과를 얻었다.

4. 웹 기반 정보 시스템 설계 및 구현

본 연구의 시험 시스템으로 한국전자통신연구원의 연구문서 관리 시스템을 대상으로 하였다. 연구문서 관리 시스템은 연구 활동 중 발생한 신제품 개발 정보, 아이디어나 Know-how 등을 정리한 중요한 지적 재산 정보를 관리하는 정보 시스템이다. 연구문서의 서지사항, 요약 정보뿐만 아니라 전문을 데이터베이스화하여 웹 환경으로 정보를 제공하고 있다.

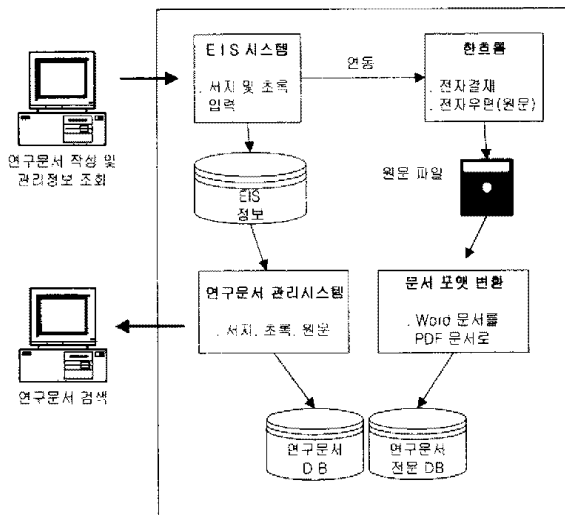
기존 연구문서 관리 시스템의 전문 정보는 소스가 종이 문서로 되어 있어 광 파일링 시스템을 이용하여 전문 정보를 이미지 포맷으로 축적하여 왔다[2]. 종이 문서로 구성된 원문은 스캐너를 이용하여 200 DPI 해상도로 스캐닝하여 Tiff 파일로 저장하였다. Tiff 파일로 저장된 전문 정보는 이용자 검색 결과인 요약 정보와 링크되어 이용자 요구에 따라 자체 개발한 Tiff 이미지 뷰어(TicView 1.0)[2]로 제공된다.

4.1 시스템 설계

새로운 연구문서 관리 시스템의 설계는 두 가지 관점에 주안점을 두었다. 첫번째 관점은 시스템의 외부 환경 변화에 따른 관련되는 타 시스템과의 통합화이다. 두 번째 관점은 원문 정보의 정보 시스템에서의 원문 정보에 대한 문서 포맷의 통합이다.

연구원내에는 전자결재 시스템인 '한흐름'과 업무 전산화를 위한 클라이언트/서버 기술로 개발된 EIS (ETRI Information System)와 연동하는 연구원 종합 정보시스템[5]이 운영되고 있다. 새 시스템에서는 EIS와 '한흐름'과의 통합 환경을 구축하여 연구문서 처리 단계를 줄여 신속하고 효과적인 연구문서 관리 및 이용 체계를 구축하도록 설계하였다.

통합 환경으로 다양한 관리 기능이 있는 EIS와 전자 결재가 가능한 '한흐름'을 활용함으로써 연구문서의 관리 기능이 강화되고 연구문서의 서지 및 요약 정보와 원문 파일 전송이 자동적으로 이루어져 연구문서 획득 단계가 신속하고 자동적으로 이루어지게 된다. 시스템 통합 구성도는 (그림 3)과 같다.



(그림 3) 시스템 통합 구성도

(그림 3)은 새로운 연구문서 관리 시스템의 통합 환경 구성도로 연구 문서 작성에서 이용까지의 자료 흐름 과정을 간략히 보여준다. 연구부서에서 연구문서가 작성되면 우선 EIS 시스템을 이용하여 연구문서의 서지사항과 요약 정보를 입력한다. 이 정보는 EIS 데이터베이스에 수록되어 개인별, 부서별 관리 정보로 활용된다. EIS 정보를 '한흐름'과 연동하여 첨부 문서로 원문 파

일을 포함한 후 전자결재 기능을 거쳐 결재를 마친 연구부서의 등록 절차를 마치게 된다.

결재된 연구문서의 서지 및 요약 정보는 연구문서 관리 시스템의 입력 정보로 받아들여 종합관리번호 부여, 색인어 추출, 원문 파일 연결 과정을 거쳐 정보 시스템에서 활용하기 위해 데이터베이스화 과정을 거치게 된다. '한흐름'의 첨부 문서로 수신된 원문 파일은 정보 시스템에서 원문 정보로 이용하기 위해서 정보 시스템의 원문 문서 포맷인 PDF 문서로 변환하여 원문 데이터베이스에 저장한다.

'한흐름', EIS 시스템과의 연동으로 연구문서 작성에서부터 이용자 제공까지 신속하게 이루어지게 되었으며 EIS를 통한 연구문서 통계 정보와 연구문서 관리 시스템의 정보 공유와 전달 기능이 통합되어 종합적인 관리 및 이용 체계가 이루어졌다.

4.2 원문 문서 포맷 통합

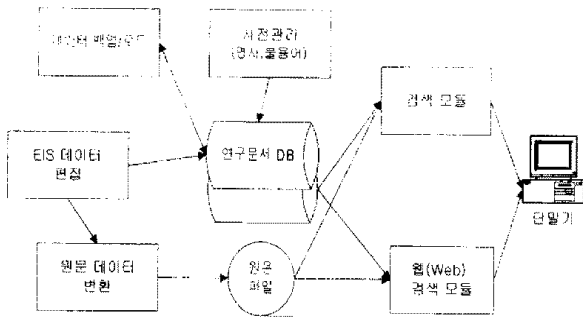
정보 시스템에서는 문서 포맷의 통일이 필요하다. 문서 포맷의 통일은 이용자에게 정보 이용의 혼란을 줄이고 편의성을 높여주며 정보를 효율적으로 관리할 수 있다. 그런데 새로운 통합 환경이 구축되면 정보 시스템의 정보 소스가 종이 문서에서 워드 파일로 바뀌게 된다. 소스의 형태가 바뀌게 됨으로써 기존 문서 포맷과의 호환성이 유지되는 새로운 문서 포맷의 선택이 요구된다.

새로운 문서 포맷으로는 앞에서 검토한 바와 같이 기존 문서 포맷과 호환성을 유지할 수 있고 신속한 원문 정보 데이터베이스 구축이 가능하면서 북마크, 섬네일, 하이퍼링크 등 다양한 이용자 편의 기능을 포함하고 있는 PDF 포맷을 정보 시스템의 표준 포맷으로 통일하였다. PDF 포맷으로 통일함은 기존 이미지 포맷을 PDF 포맷으로 변환하여야 하는 문제점이 있으나 워드 파일을 이미지 포맷으로 변환하는 것 보다는 이용자 편의성, 정보의 재활용성, 향후 웹 문서 포맷과의 호환성 유지에 바람직하다고 판단하였다.

4.3 시스템 환경

정보 시스템을 구현하는 서버측 환경으로 유닉스(Unix) 시스템인 Hp D270 서버를 이용하였으며, 데이터베이스 시스템으로 Sybase 11.0을 이용하였다. 웹 CGI 프로그래밍은 C 언어를 사용하였으며, 이용자 질의에 대한 검색 수행을 담당하는 IR(Information Re-

trivial) 엔진을 데이터베이스 시스템을 이용하여 자체 개발하여 구현하였다.



(그림 4) 연구문서 관리 시스템 서버 모듈 개념도

전문 정보 처리 환경은 Windows NT Workstation 버전 4.0 OS 환경에 PDF 변환 툴인 DocuCom 2.01을 탑재한 펜티엄 Pro 200MHz PC를 이용하였다. 종이 문서 처리를 위해 고속 스캐너인 Canon DR-3020을 이용하였으며, 이미지 문서의 PDF 변환 변환은 Acrobat Exchange를 이용하였다. 전문 처리 환경을 Windows NT 환경으로 구현한 이유는 동일한 PC 환경에서 Windows 95 환경과 Windows NT 환경에서 PDF 문서 변환을 시험한 결과 Windows NT 환경에서 변환 속도가 30~50% 향상되었기 때문이다. 특히, 이미지, 복잡한 도표가 많은 문서의 경우 매우 효과적이었다.

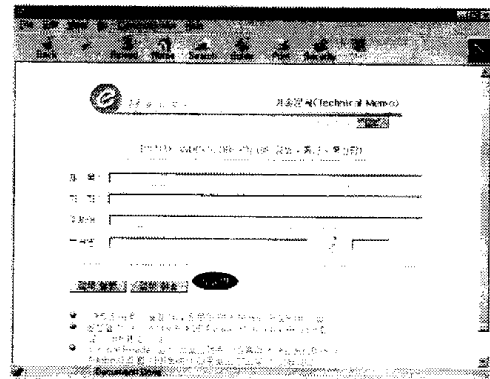
4.4 시스템 구현

정보 시스템은 관리 기능과 이용자 검색 기능을 구현하였다. 관리 기능은 데이터 입력, 색인어 추출, 데이터 유지/관리, 원문 정보 구축 등 주로 연구문서 관리자가 이용하는 기능이다. 이 관리 기능은 EIS와 연계하여 클라이언트/서버 환경으로 구현하였다. 이용자 검색 기능은 이용자의 편의성, 확장성을 고려하여 웹 환경으로 구현하였다. (그림 4)는 연구문서 관리 시스템의 서버 모듈 개념도이다.

데이터베이스는 제목, 저자, 주제어 등에서 추출한 색인 데이터베이스와 저지 사항과 요약 정보가 수록된 마스터 데이터베이스로 구성하였다. 그리고 원문은 파일 시스템으로 구축하였으며, 마스터 데이터베이스의 원문 정보 경로(Path) 필드에 원문 위치 정보로 링크되도록 구성하였다.

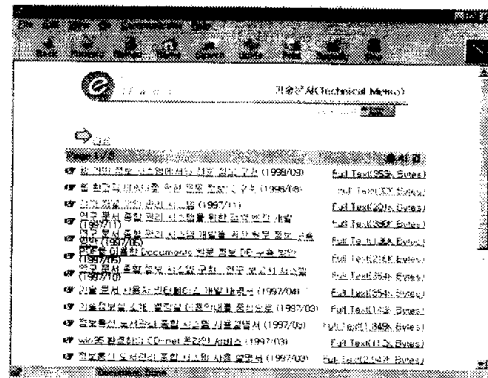
이용자 검색 모듈은 일반적인 웹 기반의 정보 검색과 유사한 기능으로 구현하였다. 이용자 검색 기능은

입력된 질의에 대한 검색을 실행하여 간략 정보와 제공하고 원하는 정보의 저지 및 요약 정보 그리고 원문 정보를 제공하는 3단계로 구성되어 있다. (그림 5)는 검색 화면으로 이용자 질의에 대한 검색을 수행하여 간략 정보를 보여준다. 필드내 질의어에 대한 연산은 AND, OR 부울리언 연산을 수행하며 필드와 필드간 연산 AND 연산을 수행한다.

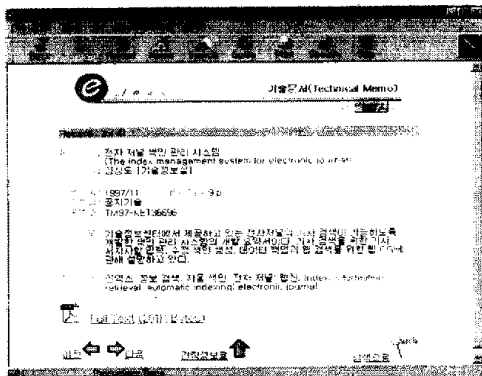


(그림 5) 검색화면

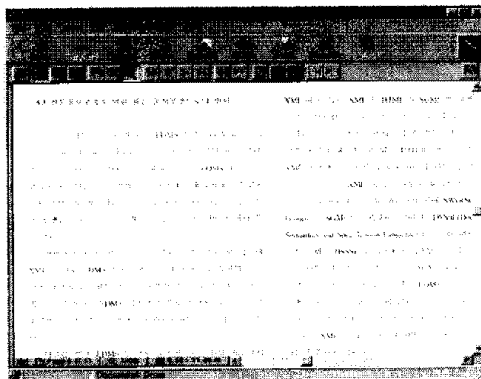
검색된 결과는 연구문서의 제목, 요약과 원문의 문서 포맷을 상징하는 이미지로 표시하여 요약 정보나 원문과 링크되도록 구성하여 이용자에게 편의성을 제공하였다. 요약 정보는 텍스트로 구성하여 HTML로 제공하며 원문은 PDF 리더(Reader)를 이용해 볼 수 있다. (그림 6) (a)는 연구문서의 제목과 작성 년월을 나타내는 간략 정보로 직접 원문을 연결할 수 있다. (그림) (b)는 문헌의 요약을 나타내며 검색된 연구 문서의 전, 후 내용을 볼 수 있도록 하여 이용자 편의성을 높였다. 문헌의 전, 후 이동은 질의어 재검색 방법으로 해결하였다. (그림 5) (c)는 문헌의 원문을 보여준다.



(a) 간략 정보 보기



(b) 문헌 요약 보기



(c) 문헌 원문 보기

(그림 5) 검색된 문헌의 간략(a), 요약(b), 원문(c) 예

5. 시스템 평가

연구문서 관리 시스템의 평가는 원문 정보를 PDF로 구축한 것에 대해 자체 업무 수행 관점에서 분석하였다. 원문 정보 구축 측면에서 보면 첫째, 신속한 원문 정보를 구축으로 시간과 비용이 절감되었다는 점이 가장 큰 장점이다. 워드 문서를 HTML로 변환하여 구축하는 경우에 비하여 수백~수십배 빠른 처리가 이루어졌다. 이는 HTML로 변환할 경우에 비해 재편집할 필요가 없기 때문이다. 둘째, 원문 문서 포맷이 인터넷에서 널리 활용되고 있는 문서 포맷으로 통일했다는 것이다. 기존 시스템에서 운영되던 이미지 문서 포맷과 워드 문서를 PDF로 전환하였다. 셋째, 전문 정보를 구축할 때 대량의 정보로 인하여 대용량의 저장 공간이 요구되는데 고도로 압축된 텍스트 파일로 저장되는 PDF 문서 포맷을 사용하여 저장 공간을 절약할 수 있었다. 시험 결과 <표 2>에서 보듯이 50% 이상 저장 공간을 절약하는 결과를 얻었다. 넷째, 연구원 내 그룹웨어 연계하여 시스템의 통

합화를 이루었다는 것이다. 연구부서에서 작성된 연구문서를 전자결재를 거쳐 자동적으로 접수되어 데이터베이스를 구축하는 종합적인 관리 체제를 구축하였다.

정보 공유와 전달 측면에서 보면 인터넷에서 널리 이용하고 있는 문서 포맷을 이용함으로써 이용자의 거부감을 최소화하고 편의성을 높인 것이다. 문서 포맷 별로 브라우저를 설치하지 않고 공개 소프트웨어인 Acrobat Reader나 DocuReader 하나로 해결하였다.

문제점이라면 첫째, 웹 브라우저에 PDF 리더를 설치하여야 한다는 것이다. PDF 리더로 유명한 Acrobat Reader가 비록 공개 소프트웨어이고 널리 이용되고는 있지만 이용자에게 부담을 주는 것은 사실이다. 둘째, 향후 인터넷의 표준 문서 포맷이 될 SGML과 XML과의 관계가 어떻게 될 것인가 하는 것이다. 이점은 PDF가 인터넷에서 널리 사용하고 있으며 PDF를 개발한 Adobe사도 XML에 참여하고 있으며 Adobe에서 개발한 그래픽 문서 포맷인 PGML을 W3C에 제안하고 있다는 점에서 XML에 포함되거나 호환성을 유지할 것으로 기대하고 있다.

6. 결 론

지금까지 정보 시스템에 있어서 원문정보를 처리할 수 있는 다양한 문서 포맷의 특징과 장단점을 비교 분석하여 보았다. 특히, 기존 정보 시스템이 웹 기반으로 전환함에 있어 가장 큰 걸림돌이 되고 있는 원문 정보의 처리 및 구축에 있어서 보다 효율적이고 경제적인 방법을 제시하였다.

원문 정보는 이미지나 워드 등 다양한 문서 포맷으로 작성된다. 웹 기반의 정보 시스템에 있어서 이러한 원문 정보를 신속하고, 효율적으로 처리할 수 있는 문서 포맷을 선택하는 것이 매우 중요한 관점이 되고 있다.

이러한 관점을 갖고 현재 운용되고 있는 정보 시스템을 사례로 다양한 문서 포맷으로 원문 정보를 처리하여 그 장단점을 분석하였다. 여기서 분석 팩터로 정보 전송 속도와 연관된 파일 크기, 원문정보 처리의 신속성과 관련된 처리시간 및 원문정보 관리와 연관된 파일 수를 중심으로 고려하였다. 하지만, 실질적으로 웹 기반의 정보 시스템을 운영하는데 있어서는 이외에 전문 검색(Full-Text Retrieval), GUI(Graphic User Interface) 등 다양한 팩터들을 고려하여야 할 것이며, 이러한 팩터들에 대한 분석은 웹 기반의 정보 시스템들을 운영하면서 더욱 연구되어야 할 것으로 본다.

다음의 작업까지의 분석은 신문 정보의 처리라는 측면에서만 고려하여 각 문서 포맷을 비교 분석하였지만, 실질적으로 웹 기반의 각 문서 포맷들은 단순성, 보안성, 정보의 재활용성 측면에서 고유의 장점들을 가지고 있어 나름대로 자신의 영역을 확보하고 발전해 나갈 것이다. 즉, HTML은 HTML 나름대로 자신의 영역을 확보하고 발전해 나갈 것이며, SGML, XML도 나름대로의 영역을 개척하여 사용될 것이다.

따라서 이 연구는 웹 기반의 정보 시스템의 문서 포맷으로 어떤 포맷이 가장 효율적인가를 밝히기 보다 정보를 구축하고자 하는 사람들이 자신의 정보에 맞는 문서 포맷을 선택하는데 있어 판단의 근거 자료로 제시하였다.

참 고 문 헌

[1] 根岸正光, 石塚英弘(공저), 김성혁(역), SGML의 기본과 이해, 성안당, 1996.
 [2] 기민호 외, 정보통신 종합정보센터 운영, 한국전자통신연구원, 1997.
 [3] 김성혁, "인코딩 포맷 및 방법론", 1995년 UNION DB 전문가초청세미나 발표자료집, pp.1-36, 1995.
 [4] 김성혁 외, "가상도서관 모델 및 구현", 정보통신부 지원 초고속 정보 통신 응용 기술개발 1차년도 연구개발 결과보고서, pp.66-68, 1996.
 [5] 김영희, 전병운, "그룹웨어 기반의 연구소 종합정보시스템 구현", 20주년기념 ETRI 논문집, pp.283-287, 1997.
 [6] 김태규 외, "유니코드 한자 지원 문법 지시적 SGML 편집기의 설계 및 구현", 1998년도 한국정보과학회 봄 학술발표논문집(B), Vol.25, No.1, pp.464-466, 1998.
 [7] 문병주 외, "차세대 웹 문서 XML의 기본 개념", 주간기술동향, 847호, pp.1-13, 1998.
 [8] 박재현, "웹과DB 연동 기술의 모든 것", 마이크로소프트웨어, 통권167호, pp.240-248, 1998.
 [9] 白旗保則, "ハブリック技術の最新動向 - DTPによる出版から PDF による電子出版へ-", 情報科学と技術, 47巻1號, pp.29-34, 1997.
 [10] 유사라, "하이퍼미디어 도서관 정보시스템", 한국도서관협회, 1997.
 [11] 이강찬, 이원석, "XML(eXtensible Markup Language) Introduction", 충남대학교 컴퓨터공학과 데

이타베이스연구실, 1998.
 [12] 최석두, "전자도서관의 개념과 발전추세", 도서관문화, Vol.38, No.2, pp.38-57, 1997.
 [13] Goldfarb, C.F. *The SGML Handbook*, Oxford university Press, 1990.
 [14] Ian S. Graham. *The HTML Sourcebook*, John Wiley & Sons, Inc., 1995.
 [15] *ISO 8879 : Information processing - Text and Office systems - Standard Generalized Markup Language(SGML)*, ISO, 1986.
 [16] *ISO/IEC 10179 : Information technology - Text and Office Systems - Document Style Semantics and Specification Language(DSSSL)*, ISO, 1996.
 [17] JIS X 4151, 문서기술언어 SGML, 일본규격협회, 1992.
 [18] Jonathan B. Postel, "Simple Mail Transfer Protocol," RFC 821, 1982.
 [19] Juliana Nelson, "Global Web Site Analysis Market Review And Forecast 1997~2002," IDC Market Research, Mar. 1998.
 [20] Martie Colby & David S. Jackson(지음), 김숙자(역), SGML의 모든 것, 성안당, 1997.
 [21] Richard Light, 채규혁(역), 차세대 웹의 혁명 XML, 도서출판 대림, 1998.
 [22] William B. Frakes, Ricardo Baeza Yates(지음), 류근호, 김진호(공역), 정보검색, 시그마프레스, pp.125-152, 1995.
 [23] Adobe Acrobat, <http://www.adobe.com/>
 [24] DocuCom 2.0 --- A Powerful PDF Tools for Electronic Publishing, <http://www.zeon.co.tw/>
 [25] Welcom to SNI Korea, <http://www.snikorea.co.kr/>
 [26] LG 상남도서관, <http://www.lg.co.kr>



김 상 도

e-mail : kds@etri.re.kr
 1987년 한남대학교 수학과 졸업 (학사)
 1997년 3월~현재 충북대학교 대학원 전자계산학과 석사과정
 1999년 현재 한국전자통신연구원 선임기술원
 관심분야 : 정보 검색, 데이터베이스, 지식관리시스템 등



문 병 주

e-mail : bjmoon@infocetri.re.kr
1980년 부산대학교 전자공학과 졸업(학사)
1998년 3월~현재 충북대학교 대학원 전자계산학과 석사과정

1999년 현재 한국전자통신연구원 기술정보센터 선임연구원

관심분야 : 데이터베이스, 지식관리, EDI 등



류 근 호

e-mail : kuryu@dblalab.chungbuk.ac.kr
1976년 숭실대학교 전산학과 졸업(이학사)
1980년 연세대학교 대학원 전산전공(공학석사)
1988년 연세대학교 대학원 전산전공(공학박사)

1976년~1986년 육군 군수 지원사 전산실(ROTC장교), 한국전자통신연구소(연구원), 한국방송통신대학교 전산학과(조교수)근무

1989년~1991년 Univ. of Arizona. Research Staf(TempIS 연구원, Temporal DB)

1986년~현재 충북대학교 컴퓨터과학과 교수

관심분야 : 시간 데이터베이스, 시공간 데이터베이스, 지식기반 정보검색, 객체 및 지식베이스 시스템