

인터넷 전자문서 형식에 관한 비교 연구

A Comparative Study on the Type of Electronic Document via Internet

조현양, 최성필, 연구개발정보센터
고기형, 송원택, 한국과학기술원 수학과

Cho Hyun-Yang, Choi Sung-Pil, KORDIC
Ko Ki-Hyung, Song Won-Taek, KAIST

최근 급속히 발달한 인터넷을 통해 단순한 전자우편 뿐만 아니라 학술 논문, 저작물 및 보고서, 계약서 등 실제 물리적인 문서를 표현하는 전자 문서의 교환이 빈번하게 이루어지고 있다. 본 연구에서는 현재 사용되고 있는 전자문서 형식들이 가지고 있는 특성과 장단점을 비교 연구하였다. 우선 문서 교환을 목적으로 하는 전자문서 형식이 가져야 할 특성으로 범용성, 신속성, 장치 독립성, 간결성, 확장성 등을 제시하고 이를 기준으로 현재 사용되거나 제안되고 있는 전자문서 형식들을 평가하였다. 특히 DVI, HTML, XML, SGML, PDF, PostScript 등의 문서 형식들을 대상으로 조사하고 평가하였다. 그 결과 연구개발정보센터에서 사용 중인 DVI 문서 형식이 한글 문서를 인터넷 상에 구현하는 가장 효율적인 방법 가운데 하나임이 증명되었다.

1. 서론

미국에서는 식료품과 같은 인터넷 상거래로는 상상하기 어려운 품목까지도 시장애 가지 않고도 직장이나 가정에서 웹을 통하여 주문하고 배달 받을 수 있을 만큼 인터넷이 사람들의 생활습관까지도 바꾸어 놓고 있다. 이제 우리 나라에서도 기업, 기관, 학교 등에서 인터넷을 접하지 않은 사람은 거의 없을 것이다. 대부분 자유롭게 인터넷을 사용할 수 있는 환경이 조성되고 있고 지금의 확산 속도로 볼 때 아마도 1~2년 내에는 각 가정까지도 초고속 망이 상당수 보급되리라 생각된다. 정보서비스를 본래의 사명으로 하는 우리 도서관은 이러한 인터넷의 물결에 직접적인 영향을 받고 있고, 이에 능동

적으로 대처하여 좀 더 나은 서비스를 제공할 수 있어야 한다. 이를 위해서는 보유하고 있는 정보를 컴퓨터와 인터넷을 통해 관리하고 열람이 가능하도록 체계적으로 정리하여 효율적으로 이용자에게 제공할 수 있어야 한다. 이렇게 구축된 전자도서관은 전통적인 의미의 도서관 기능의 수행은 물론, 아울러 사용자 하여금 시간과 공간의 제약을 벗어나 빠른 정보의 획득과 손쉬운 디지털 정보의 복제를 통하여 정보제공자로서의 중요한 임무를 담당하게 될 것이다.

본 논문에서는 가장 보편적인 정보 형태인 문자정보를 중심으로 도서관에서 인터넷 시대에 대처할 수 있는 방안을 제시하고자 한다. 전자

도서관은 문자정보 외에도 이미지, 소리, 동영상 등의 멀티미디어 정보를 제공할 수 있으나 현재로서는 인터넷에 단순한 링크형태로 제공될 수 있을 뿐이고 이들에 대한 검색과 보안 등은 활발히 연구되고 있다. 우리 주변에는 단행본은 물론이고, 학술지, 학위논문, 연구보고서, 특허, 기술규격 및 표준, 신문과 잡지 등 헤아릴 수 없을 만큼 많은 종류의 유용한 문자정보가 존재한다. 문자 정보는 대개 문서라는 형태로 담겨져 있다. 여기서 문서라 함은 단행본과 같은 인쇄매체를 거친 기존 문서와 인쇄매체를 거치지 전 형태로서 컴퓨터 상에 존재하는 전자문서로 나눌 수 있다. 문서의 제작은 대개 워드프로세서로부터 시작된다. 따라서 향후 만들어지는 필사본이나 그림을 제외한 모든 문서는 전자문서로도 존재한다. 도서관의 기능은 크게 정보의 관리 및 보관 측면과 가공 및 서비스 측면의 두 가지로 대별된다. 기존 문서에 대해서는 수서, 대출, 반납 등의 전통적인 관리 방법에다 웹을 통하여서도 서지검색과 조회가 가능하도록 연동해주는 서비스가 부가되던 일단은 인터넷에 대비되었다고 볼 수 있다. 특히 저작권을 보유하고 있는 기존문헌은 이를 스캔하여 원문을 TIFF 등의 형태로 이용자에게 제공할 수도 있다.

디지털 도서관이라고도 불리는 전자도서관의 핵심은 가능한 한 많은 정보의 디지털화에 있으며, 이 작업의 대부분이 자동화되어야만 적은 비용으로 소기의 목적을 달성할 수 있다. 그리고 문자정보의 디지털화는 전자문서 형식을 사용하여 이루어지므로 적절한 전자문서 형식의 선택은 전자도서관의 성패에 결정적인 역할을 한다고 해도 과언이 아니다. 그 이유는 컴퓨터 기술이 하루가 다르게 발전하고 있어 문헌과는 달리 자칫 잘못하면 오랫동안 막대한 예산을 들어 구축한 전자문서가 활용상의 불편함으로 인하여 사장될 수도 있기 때문이다. 전자도서관 체제하에서의 이용자는 항상 새로운 기술의 편리함을 추구하므로 이용상 불편한 정보는 다른 곳에서 비슷한 정보를 좀 더 쉽게 구할 수 있다면 모두 그 곳에 물리게 된다.

전자문서는 크게 세 가지로 나눌 수 있다. 즉 광범위한 공유, 가공 및 사용을 위하여 저장해 두기 위한 형식, 인터넷에서 즉시 서비스가 가능한 형식, 그리고 전자문서를 작성한 워드프로세서의 형식이다. 다음 2 절에서는 저장을 위한 전자문서의 형식에 대하여 고찰해 보고, 그리고 3 절에서는 인터넷 서비스를 염두에 둔 전자문서의 형식에 관하여 다루고자 한다. 마지막으로 전망, 필요 및 비용에 따른 전자문서 선택에 관한 제안은 4 절에서 다루기로 한다.

2. 저장을 위한 전자문서 형식

저장을 위한 전자문서의 목표는 정보의 공유와 재가공에 있다. 따라서 이러한 전자문서는 표준을 따르는 것이 일반적이다. 특정 워드프로세서의 형식을 이러한 용도로 사용하는 것도 고려해 볼 수는 있으나 호환성의 문제와 더불어 컴퓨터 소프트웨어가 급격히 진화하기 때문에 추후 사장되거나 사용범위가 줄어들 가능성이 높아 인트라넷과 같은 특수한 환경을 제외하고 전자도서관에서는 사용하지 않는 것이 바람직하다. 공유와 재가공이 가능한 전자문서는 대체로 내용정보, 구조정보, 양식정보가 결합하여 이루어지는데 대표적인 예로는 TEX, SGML, XML 등을 생각해 볼 수 있다. 여기서 재가공이라 함은 예를 들어 학위논문으로부터 그 해의 학위논문 초록 모음집을 따로 만드는 것과 같이 주어진 전자문서 데이터베이스로부터 필요한 정보를 추출하여 새로운 형태의 문서나 통계자료 등을 만드는 것을 말한다.

2.1. TEX

1970년대 말에 미국이 스탠포드대학교 교수인 Knuth에 의해 만들어진 조판시스템으로서 수식과 도표 처리에 탁월하여 과학기술 문서의 제작에 전세계적으로 널리 사용되어 왔으며, 향후에도 계속 사용될 전망이다. 1980년대 초 Lampert에 의해 만들어진 매크로 팩키지인 LATEX은 TEX의 강력한 조판 컴파일러의 기

능을 충분히 활용하여 TEX을 단순한 조판기
 능을 가진 워드프로세서에서 문서내용을 구조
 정보와 함께 입력하여 양식정보가 담긴 스타일
 파일과 함께 처리하여 원하는 출력을 얻는 오
 늘날의 저장을 위한 전자문서 형식으로 발전시
 켜다. LATEX과 결합된 TEX은 목차와 상호참
 조를 자동으로 갱신하는 기능 뿐 아니라, 색인
 과 참고문헌도 자동으로 처리하는 현재의 워드
 프로세서도 가지고 있지 못한 기능들을 추가함
 으로써 이 후 전자문서 표준규격인 SGML의
 제정에 지대한 영향을 끼치게 된다. 한편 TEX
 은 멀티미디어 관련 기술이 폭발적으로 발전되
 기 전에 개발되었기 때문에 TEX에는 이들의
 처리하는 규격은 정해져 있지 않다. 하지만 현
 재 대부분을 차지하는 문자정보와 삽입된 그림
 위주의 문서에서는 아직도 강력한 도구이다.
 또한 서양에서는 대부분의 자연과학 출판사가
 조판시스템으로 사용하고 있고 저자에게 TEX
 으로 원고를 작성할 것을 권장하는 학술지도
 많이 있다. 그리고 LATEX을 사용하여 작성된
 원고로부터 SGML로의 자동 변환이 가능하고
 실제로 이와 같이 SGML로 보관하여 향후 어
 떤 변화에도 대비하는 출판사도 서양에는 있
 다. 결론적으로 TEX/LATEX의 조합은 문자
 정보 위주의 문서에서는 저작도구로 뿐 아니라
 저장을 위한 전자문서의 형식으로 손색이 없
 다고 할 것이고 향후의 변화에도 적용이 가능하
 다. 또한 3 절에서 다룬 인터넷에 제공 가능한
 전자문서인 DVI나 PDF로 실시간에 추가 비용
 없이 만들어 제공할 수 있다는 이점도 있다.

2.2. SGML

SGML은 1980년 중반 국제표준기구(ISO)가 제
 안한 국제문서규격으로 문서의 구조정보를 기
 술하기 위한 규약을 정한 메타언어이다. SGML(Standard Generalized Markup Language)로 전자문서를 만들려면 문서의 논리구조를 따로 정의한 DTD(Document Type Definition)가 먼저 만들어지는 것이 보통이며, 이 구조에 맞는 내용을 입력하게 된다. SGML 문서로부터의 출력은 좀 더 복잡하다. DTD에

적합한 문서구조에 따라 좋은 모양으로 출력하는
 양식을 작성하여 출력프로그램을 작성하는 것
 이 한가지 방법이나 이것은 구현하는 방법에
 의존하게 되어 정보의 광범위한 공유라는 원래
 의 취지에 벗어나게 된다. 이를 극복하는 한가
 지 방법은 DSSSL(Document Style Semantics
 and Specification Language)이라는 또 하나의
 SGML 문서를 변환하여 조판이 가능한 또 다
 른 SGML문서로 만들어 주는 국제표준규격을
 사용하는 것이다. 이 경우 DTD는 DSSSL규격
 에 맞게 작성되어 있어야 할 필요가 있다. 실
 제로 SGML/DSSSL의 조합으로 작성된 전자
 문서는 추가 비용이 없이 TEX이나 Microsoft
 Word로 원하는 출력물을 얻을 수 있다. 특히
 TEX/LATEX을 사용하여 실시간에 DVI나
 PDF로 만들어 인터넷에 제공하는 것도 가능하
 다. DSSSL 규격을 따르지 않은 SGML문서를
 인터넷에 제공하려면 이를 해석하는 특수한 프
 로그램이 있어야 하고 이는 비용측면 뿐 아니
 라 효율측면에서도 현명하지 못한 선택이 될
 소지가 많다. 결론적으로 SGML은 신중한
 DTD의 작성이 전제된다면 문자와 멀티미디어
 를 포괄하는 정보의 저장을 위한 전자문서로서
 바람직하다고 하겠다. 그러나 다양한 워드프로
 세서 형식의 문서를 SGML로 만드는 데는 현
 재로서는 막대한 시간과 비용을 감수해야 할
 것이고, 규약의 지나친 일반성 때문에 국제무
 역거래 등의 특정 목적을 제외하고는 널리 사
 용되지는 않고 있다는 점을 간과해서는 안 될
 것이다.

2.3. XML

다양한 DTD에 의존한 SGML문서를 모두 웹
 브라우저에서 출력하도록 만드는 일은 불가능
 하다. 설령 가능하다고 해도 SGML문서가 네
 트워의 대역을 너무 차지하게 되어 효율이 떨
 어지게 된다. 그리고 현재 웹 전자문서의 표준
 인 HTML(HyperText Markup Language)은
 문법구조 확장이 한계에 도달한 것처럼 보인
 다. 따라서 이 두 형식의 장점만을 취하여 웹
 에서 제공하는 것을 궁극적인 목표로

W3C(World Wide Web Consortium)에서 제정한 전자문서 표준형식이 XML(Extensible Markup Language)이다. XML문서도 미리 작성된 DTD를 따라 입력되는 것이 보통이다. 그러나 XML의 DTD는 SGML의 DTD와 비교하여 단순화되고 불필요한 사양들이 생략될 수 있다는 것이 다르다. 따라서 XML은 SGML의 부분집합이라고 생각해도 무방하다. 여기에서도 문서의 출력을 구현하는 것이 가장 어려운 문제인데 현재로서는 SGML의 DSSSL과 넷스케이프사의 HTML을 위한 출력 양식인 CSS(Cascading Stylesheet Specification)의 절충형인 XSL(Extensible Style Language)이 제안되고 있으나 XSL에 관한 표준은 아직 정해지지 않고 각 회사마다 규격이 다르다. XML은 W3C 표준인 MathML을 포함할 예정이어서 수식의 처리에도 어려움이 없을 것이다. 또한 레이어의 동작을 제어하는 Dynamic HTML의 기능을 수행한 DOM(Document Object Model)도 아울러 표준화되고 있고, 네트워크와 컴퓨터의 성능이 급격히 향상되는 점을 감안하면 가까운 장래에 저장을 위한 전자문서 뿐 아니라 인터넷을 위한 전자문서의 가장 강력한 표준으로 자리잡을 가능성이 높다. 현재로서도 XML/XSL의 조합으로 작성된 전자문서는 XSL을 변환과정과 TEX 컴파일러를 이용하여 실시간에 DVI 또는 PDF로 만들어 추가 비용 없이 인터넷에 제공하는 것이 가능하다. 결론적으로 XML은 정보의 공유와 재가공이 목적인 전자도서관에서 채택할 수 있는 전자문서 형식 가운데 하나이다. 그러나 XSL의 표준이 명확히 정의되지 아니한 상황이어서 표준이 확정될 때까지는 일단 관망하는 것도 현명한 것처럼 보인다. 더욱이 특정 워드프로세서 형식을 XML로 변환해 주는 도구를 해당 워드프로세서 제작사가 만들 때까지는 변환시 추가 비용을 감수해야만 하는 문제를 안고 있다.

3. 인터넷을 위한 전자문서 형식

인터넷서비스를 위한 전자문서의 목표는 서비스의 효율과 문서의 질을 높이는 데 있다. 특

정 워드프로세서 형식을 이 목적으로 사용하는 것은 일단 파일이 인터넷에서 서비스하기에는 너무 크고, 받는 대로 보여주는 스트림 방식의 처리가 어렵다. 따라서 이러한 전자문서는 정보의 빠른 획득을 원하는 사용자들이 선호하지 않는다. 다만 연구계획서의 작성과 같이 특정 워드프로세서를 보유하고 있는 사용자가 인터넷에서 그 워드프로세서로 만들어진 연구계획서 양식을 받아 재사용하려 할 때는 예외일 수 있으나 이 경우 연구계획서 양식은 인터넷을 위한 전자문서의 개념이라고 보기는 어렵고 일종의 다운로드 받는 자료라고 해야 옳을 것이다. 그리고 전자도서관에서 저자가 작성한 워드프로세서 파일을 그냥 제공하는 것은 정보의 무분별한 재사용을 초래하여 저작권에 문제가 있을 수도 있다는 사실을 고려해야 한다. 인터넷을 위한 전자문서가 소기의 목적을 달성하려면 파일의 크기를 줄이고 플랫폼의 독립을 기하여야 한다. 따라서 내용, 글꼴정보, 위치, 삽입된 기타 오브젝트의 정보만을 전자문서에 담는 것이 보통이다. 이러한 전자문서의 예로는 PDF, DVI 등을 생각해 볼 수 있다. 전자문서에 담기는 것 가운데 글꼴정보는 파일의 크기에 지대한 영향을 미치므로 대부분의 글꼴은 정보의 크기가 큰 글꼴 모양은 담지 않고 글꼴명과 크기, 굵기 등의 간단한 정보만 담고 보여 줄 때는 사용 가능한 글꼴 중 가장 가까운 것으로 나타내게 된다. 특히 글자 수가 많은 한글과 한자의 경우 글꼴의 모양 정보의 분량은 상당하다. 이 점은 대량의 전자문서를 보유할 전자도서관에서는 가장 유의해야 할 사항이다. 왜냐하면 각 전자문서마다 중복해서 들어 있는 글꼴 모양은 저장장치에 상당한 부담일 뿐 아니라 인터넷에서 제공 속도에도 크게 영향을 준다. 따라서 한글 인터넷 전자문서는 글꼴의 모양을 담지 않는 것이 훨씬 유리하고 고정된 폭이 대부분인 한글에서는 글꼴을 대체하는 것이 크게 문제되지 않는다.

3.1. PDF

PDF는 90년대 초 미국의 아도브(Adobe)사가

만든 전자문서 형식이다. 현재는 웹에서 제공하는 전자문서의 형식으로 서양 언어권에서는 많이 채택되고 있지만, 원래의 목적은 DTP을 위해 문서를 상호 교환이 가능하도록 하기 위한 것이었다. DTP를 위해서는 그래픽 패키지를 포함한 많은 응용프로그램과 다양한 글꼴이 사용되기 마련이었고 이러한 문서는 생성한 모든 응용프로그램과 글꼴이 갖추어지지 않은 사람과 교환한다거나 외부 출력소에서 인쇄하는 데 어려움이 있었다. 이러한 어려움을 극복하기 위해 프린터 구동을 위한 아도브 사의 소유 형식인 포스트스크립트를 압축하고 열람의 편의를 위해 하이퍼텍스트 등의 첨가가 가능한 형식으로 발전시킨 것이 PDF형식이다. PDF형식은 아도브 사의 이해관계 상 그들의 포스트스크립트 글꼴형식 단을 사용한다. 바로 이 점이 하나의 글꼴파일의 크기가 엄청난 2바이트 글자 언어권에서는 PDF형식의 문서가 사용되기 어려운 이유이다. 마이크로소프트 윈도우즈와 매킨토시 OS에서는 아도브 글꼴형식에 대항하여 트루타입(True Type)이라는 외각선 글꼴형식이 사용되어 왔다. 따라서 이러한 OS 상에서 PDF문서를 처리하기 위해서는 글꼴을 PDF문서에 내장하거나 사용된 포스트스크립트 글꼴을 미리 사용자 컴퓨터에 가지고 있어야 하는데 글꼴파일의 크기가 비교적 작은 서양 언어권에서는 문제가 되지 않지만 2바이트 글자 언어권에서는 글꼴이 PDF기술에 커다란 부담으로 작용하여 특히 웹에 제공하는 전자문서로서의 기능에 의문을 가지게 한다. 그리고 웹 브라우저와 연동시 필수 기능이라고 할 수 있는 스트림 처리가 PDF문서의 특성상 곤란하다는 것이 또 다른 문제이다. 더욱이 아도브 사 자신도 XML 컨소시엄에서 주도적으로 활동하고 있어 인터넷 전자문서로서 PDF 형식을 고집하지 않을 전망이다. 다만 PDF는 원래의 목적인 출판용으로는 계속 사용될 공산이 크다.

3.2. DVI

DVI는 TEX 문서를 지정된 양식과 함께 TEX

컴파일러를 통과하여 얻는 전자문서 형식이 DVI(DeVice Independent)로서 장치에 무관하다는 명칭이 의미하듯이 내부형식이 PDF에 비해 엄밀하지만 간결하다. DVI로 작성된 문서는 이진(Binary)형식으로 매우 효율적으로 쓰여 삽입된 그림이 없는 경우 압축해도 절반 이하로는 줄지 않는다. DVI는 컴퓨터의 자원이 지금처럼 풍부하지 못하던 시절에 만들어져 저장 공간을 최소화하면서 처리속도를 빠르게 하려는 지혜가 돋보인다. 그리고 DVI는 글꼴의 정보만을 담고 모양자체는 담지 않는다. 따라서 인터넷에 제공하는 전자문서로서 적절하다고 하겠다. 그 동안 DVI의 이러한 장점에도 불구하고 인터넷을 위한 전자문서 형식으로 사용되지 못한 것은 웹 브라우저와 연동하는 출력이 없었기 때문이다. 현재 연구개발정보센터에서 학위논문 서비스에 사용되고 있는 ㈜샘틀로미디어의 TeXplus Viewer는 웹과 스트림 기능을 가지고 연동되면서 아도브 사의 Acrobat Reader에 비해 손색이 없는 기능을 가지고 있다. 2 절에서도 언급한 바와 같이 SGML/DSSSL 문서나 XML/XSL 문서가 실시간에 TEX 컴파일러를 통하여 DVI로 변환될 수 있다. 현재로서는 한글을 포함한 인터넷 전자문서로는 가장 이점이 많은 형식이라고 판단된다. 또한 DVI가 공개된 형식이고 관련 소프트웨어를 개발할 수 있는 기술력을 우리 나라가 보유하고 있어 이번에 연구개발정보센터와 ㈜샘틀로미디어가 공동으로 개발한 서지정보 자동추출기와 같은 도서관 자동화를 위한 다양한 소프트웨어를 국내에서 개발할 수 있다는 것도 DVI의 매력이다.

3.3. 기타

현재 웹의 표준형식인 HTML을 생각해 볼 수 있지만 복잡한 전자문서에 대한 표현 및 향후 확장성에 있어서 한계에 도달해 있으므로 신문, 잡지와 같이 단순한 문서가 아니고는 고려 대상에서 제외된다. 그 밖에도 포스트스크립트 또는 레이저 젯 프린터로 보내지는 형식인 PS 또는 PCL은 전자문서이기에는 글꼴을 비효율

적으로 다루어 파일의 크기가 매우 커지는 단점이 있다. 참고로 압축된 PCL인 PCL6에 기반을 둔 전자문서 형식이 국내에서 개발되어 있다.

4. 결론 및 제안

전자도서관을 새로이 만들 때는 다음과 같은 사항을 염두에 둘 필요가 있다. 첫째, 인터넷에서 제공할 대상 정보가 무엇인가를 고려하여야 한다. 예를 들어 원문이 워드프로세서 등으로 작성된 디지털 파일인지 또는 책자형태의 인쇄물인지를 우선 고려해야 한다. 원문이 디지털 파일인 경우에는 향후 전자문서의 재가공이 필요할 것인지의 여부 또한 고려되어야 한다. 대부분의 전자도서관의 경우에 효율적인 정보 서비스를 위하여 서지정보와 본문 색인 정보만 추출되어 있는 것으로 충분하며, 재가공에 대한 필요성은 비교적 적을 것이다. 따라서 이 경우 인터넷을 통하여 서비스하기 위한 전자문서로의 변환으로 충분하고 저장을 위한 전자문서로의 변환은 필요하지 않다. 그리고 원문이 인쇄물인 경우에는 서지정보를 도서관 측에서 직접 입력하거나 저자에게 입력할 수 있는 인터넷 환경을 제공하여야 할 것이다. 원문을 스캔한 TIFF 파일로부터 서지정보를 인식하는 것은 한글의 경우 인식률이 아직은 그다지 높지 않아 추천할 만한 것이 못된다.

둘째, 정보를 어떻게 인터넷에서 제공할 것인가에 대한 고려이다. 예를 들어 전자문서의 경우에는 서지정보의 검색만 가능하게 할지 또는 본문검색도 가능하게 할지 등을 미리 고려해야 한다. 인터넷을 위한 전자문서인 DVI나 PDF의 경우에 텍스트를 추출하여 색인 엔진을 이용하여 주제어를 추출하는 것은 얼마든지 가능하다.

셋째, 사용자가 어떻게 인터넷에 접속하고 있는가, 즉 사용자의 인터넷 사용환경에 대한 고려이다. 만약 모뎀으로 접속하는 사용자가 주류를 이루면 특히 파일의 크기와 전자문서 열

람기의 크기가 작아야 고객을 생각하는 전자도서관 서비스가 될 것이다.

마지막으로 전자도서관을 구성하는 예산의 제약도 고려해야 할 것이다. SGML이나 XML과 같이 저장을 위한 전자문서로의 변환은 현재까지는 워드프로세서 형식으로부터 요구된 사항을 입력해야 하기 때문에 막대한 비용과 시간을 초래할 것이다. 따라서 원문 워드프로세서 형식을 당분간은 인터넷 전자문서로만 변환하여 서비스하고 추후 워드프로세서 제작사가 적절한 SGML이나 XML로의 변환을 지원할 때까지 보관하는 것이 현명하리라 생각한다.

5. 참고 문헌 및 사이트

- [1] 연구개발정보센터, 인터넷을 통한 복합문서의 전송 및 처리 방안에 관한 연구(2차년도 최종보고서), 1998
- [2] 연구개발정보센터, 인터넷/인트라넷 환경에서의 온라인 문서관리를 위한 문서형식에 관한 연구(1차년도 최종보고서), 1998
- [3] 연구개발정보센터, 인터넷/인트라넷 환경에서의 온라인 문서관리를 위한 MS-Word 형식 문서의 처리에 관한 연구(1차년도 최종보고서), 1998
- [4] 연구개발정보센터, 서지정보 추출 / 자료 가공 자동화 및 통합문서 뷰어의 개발에 관한 연구(1차년도 최종보고서), 1999
- [5] <http://www.tug.org>
- [6] <http://www.oasis-open.org>
- [7] <http://www.w3.org>
- [8] <http://www.adobe.com>
- [9] <http://www.texplus.com>