

표준 문서 관리에 XML 기술의 적용

유정연*, 나재열*, 권석훈*, 구경철**, 이준섭**, 박기식**이규철*,
*충남대학교 컴퓨터공학과
*한국전자통신연구원 정보기술본부 PEC
e-mail : jyyou@ce.cnu.ac.kr

An application of XML Technology to a Standard Document Management

JeongYoun Yu*, JaeYoul Na*, SeokHun Kwon*,
KyoungChul Koo**, JunSeob Lee**, KiShik Park**, KyuChul Lee*
*Dept. of Computer Engineering, Chungnam National University
**Protocol Engineering Center
Electronics and Telecommunications Research Institute

요 약

급속한 과학 기술의 발달로 인해 상호간의 정보교환의 요구는 증가하게 되었으나, 서로 다른 시스템 환경으로 인해 정보 교환에 많은 문제점이 발생하였다. XML 기반의 정보 교환은 이를 위한 해결하기 위한 방안이며, 여러 연구자들이 공동으로 의견을 교환하여 작성해야 하는 표준 문서의 관리에 XML 을 적용하면 매우 효과적이다. 본 논문에서는 보다 생산적이며 효율적인 표준 관련 문서의 공유 및 상호 교환을 위해 차세대 인터넷 문서의 표준인 XML 을 기반으로 문서를 교환하며 이를 효과적으로 저장, 검색, 관리할 수 있는 시스템 모델을 제안하였다. 또한, 표준 제정 과정에서 이루어지는 문서의 편집, 수정, 삭제의 기능을 사용자의 권한에 따라 제한함으로써 효율적인 정보 교환 기능을 제안하였다.

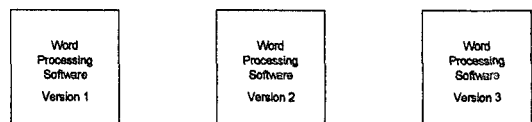
1. 서 론

하드웨어, 운영체제, 네트워크 등 시스템의 발달과 함께 이를 기반으로 한 다양한 형태의 컴퓨터 응용 개발은 개인과 개인, 기업과 기업, 그리고 사회 전반에 걸친 업무 관리 및 개발 환경에 많은 영향을 미치고 있다. 이러한 변화는 과학 기술은 물론, 업무 처리에 관련된 정보의 급속한 증가를 초래하였으며 이로 인해 발생하는 각 관련 정보들간의 효과적인 정보 교환 기능을 요구하게 되었다.

그러나, 기존의 시스템 환경으로는 이러한 업무 형태의 변화를 제대로 지원하지 못하는 문제가 있다. 즉, [그림 1-1] 과 같이 처리하는 운영 환경, 응용 프로그램 등 업무 환경이 서로 달라 정보의 교환 시 응용 프로그램의 변경, 데이터의 변환 등 다양한 문제점을 가지게 된다. 이러한 문제점들은 현 업무의 개발

에 많은 장애 요인으로 작용한다.

이러한 문제점을 해결하기 위해서는 소프트웨어와 플랫폼에 중립적인 문서 표준이 필요하다. XML (extensible Markup Language)은 인터넷 전자 문서의 표준으로 1998년 2월에 W3C (World Wide Web Consortium) 에서 표준으로 제정된 것으로, 웹 상에서 구조화된 문서를 전송 가능하도록 설계된 전자 문서 형식이다.



[그림 1-1] 응용프로그램에 의존적인 문서 환경

이러한 문제점을 해결하기 위해서는 소프트웨어와 플랫폼에 중립적인 문서 표준이 필요하다. XML (extensible Markup Language)은 인터넷 전자 문서의 표

본 논문은 한국전자통신연구원의 표준 관련 문서 교환을 위한 XML-DTD 및 관련 Repository 기술 연구의 위탁과제로 수행하였음

준으로 1998년 2월에 W3C (World Wide Web Consortium)에서 표준으로 제정된 것으로, 웹 상에서 구조화된 문서를 전송 가능하도록 설계된 전자 문서 형식이다. 이는 인터넷에서 기존에 사용하던 HTML의 한계를 극복하고 SGML의 복잡성을 해결한 표준으로 소프트웨어 및 플랫폼 중립적인 전자문서를 효과적으로 검색, 관리 및 공유 할 수 있는 유일한 방안으로 제시되고 있다.

위와 같은 시스템 환경과 사회적 흐름은 새로운 시스템의 개발 대신 기존의 시스템 환경을 이용하며, 이들 간의 문서를 보다 효과적으로 교환하기 위한 방안으로 XML 기반의 문서 처리에 대한 요구가 증가하고 있다.

여러 형태의 문서들 중 시스템 개발을 위해 기초가 되는 표준 문서에 대한 처리는 중요하다. 이는 표준화 제정 과정에서 표준 관련 문서들을 상호간에 서로 교환하여 표준화 작업을 수행하므로, 사용자의 목적에 따라 다양한 형태의 정보를 효과적으로 전달하는 것이 요구된다.

그러므로 본 논문에서는 인터넷 중심의 정보 교환 사회에서 보다 생산적이며 효율적인 표준 관련 문서의 공유 및 상호 교환을 위해 차세대 인터넷 문서의 표준인 XML을 기반으로 문서를 교환하며 이를 효과적으로 저장, 검색, 관리할 수 있는 시스템 모델을 제안하였다. 또한, 표준 제정 과정에서 이루어지는 문서의 편집, 수정, 삭제의 기능을 사용자의 권한에 따라 제한함으로써 효율적인 정보 교환 기능을 제안하였다.

본 논문의 전체 구성은 다음과 같다. 제 2장은 국내·외적으로 표준 문서와 관련하여 XML 기반의 문서 교환을 위해 이루어지고 있는 관련 기술들을 살펴보고 있다. 제 3장에서는 표준 문서를 효율적으로 관리 및 교환하기 위한 시스템 모델을 제안하였다. 제 4장에서는 앞으로의 연구 방향을 제안하였다.

2. 관련 기술 동향

2.1 RFC(Request For Comments)

RFC는 설명요청,자료요청이란 뜻이며 1969년 이후 인터넷에 관련된 여러 가지 내용들을 기록해 놓은 일련의 문서들을 지칭하는 용어로 인터넷과 TCP/IP 사양서를 말한다. 이는 문서번호와 함께 표기를 하며 인터넷 통신규약과 여러 가지 실험 결과들을 자료로 만들어 놓은 것으로, 인터넷 관련 문서를 체계적으로 관리함으로써 인터넷 관련 정보들의 오류들을 최소화하기 위한 목적을 가지고 있다.

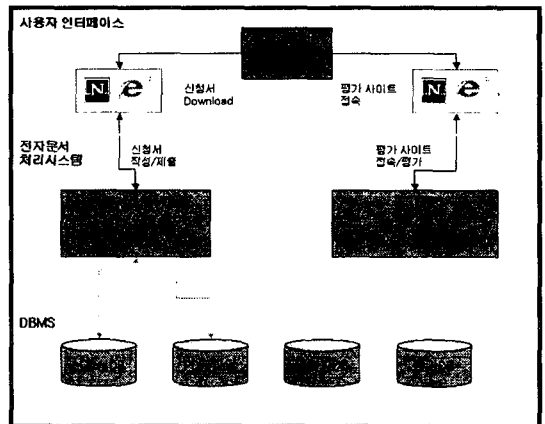
Invisible Worlds, Inc of the Marshall T.Rose는 XML을 기반으로 하여 I-D(Internet-Draft)와 RFC series를 위한 문서 작성 정의를 제안하였다. 이를 위해 Invisible Worlds Inc는 XML 기반의 메타데이터 관리를 위한 통합 솔루션인 Blocks 프로토콜을 만들어왔으며, 실제계의 어플리케이션에 Blocks의 파워와 확장성을 가져

올 정보 관리 툴을 제공하고 있다.

2.2 한국 과학 재단

현재까지 국내에서는 표준 문서 교환을 위해 XML을 사용하는 사례는 거의 드물다. 그러나, 이와 유사한 사례로는 최근에 한국 과학 재단에서 전자 문서의 유통 방안 연구로서, 연구 과제를 위해 제안서의 제출에서 평가까지 처리되는 여러 가지 과정들을 웹을 기반으로 교환하고 처리하기 위해 기초가 되는 파일럿 프로젝트를 수행하였다.

즉, 웹을 통하여 연구 제안자들로부터 접수 받은 워드 프로세서 형태의 제안서를 다양한 플랫폼에 중립적이며 차세대 웹 표준인 XML 문서로 변환하여 저장 및 관리하고, 제안서에 대한 평가를 위하여 저장된 문서에 대한 검색 및 문서 처리 작업등을 하였다.



[그림 2-1] 한국 과학 재단 수행 시스템 모델

3. 표준 문서 관리 시스템 모델

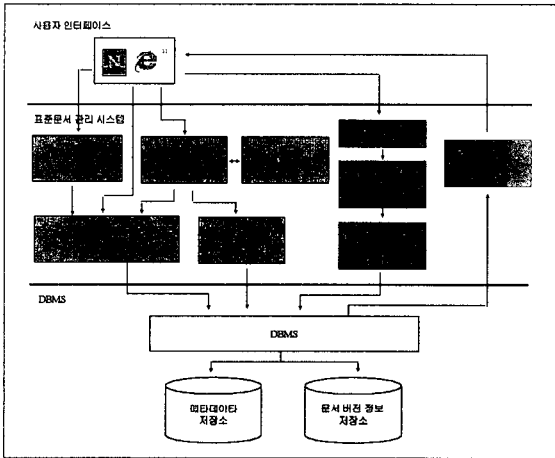
3.1 전체 시스템

본 논문에서 제안하고자 하는 시스템 모델은 첫째, XML 기반의 문서 관리를 위해 새로운 툴의 적용에 대한 부담감을 줄이기 위해 기존에 사용해 왔던 시스템 환경을 기반으로 XML 기반의 문서 관리 기능을 제공하는 특징을 가지고 있다. 즉, 사용자는 현재 이용되고 있는 HWP나 Word 형태의 워드 프로세서를 이용하여 문서를 작성하면, 표준 문서 관리 시스템은 작성된 문서를 분석하여 필요한 XML 문서의 형태로 변환한 후, 관리하였다. 이러한 형태의 시스템 변화는 현재 국내·외적으로 큰 관심의 대상이 되고 있고 이를 위한 연구 역시 진행되고 있다.

둘째, 표준 문서의 효과적인 교환을 위해 사용자의 권한에 따라 여러 가지 형태로 변환하여 제공하는 특징을 가지고 있다. 특히, 표준 문서 제정 과정에서는 사용자에게 따른 권한 제공 서비스가 중요 시 되고 있

다. 이에, 본 연구 모델에서는 사용자의 권한에 따라 수정, 편집이 가능한 문서 형태와 수정이 불가능한 PDF 형태의 문서를 제공하였다. 또한, 기존의 메뉴 일 방식을 이용한 문서의 변환이 아니라, 보다 자동화된 문서 처리 기능을 제공하여 기존의 시스템과는 다른 생산적인 시스템 모델을 제안하였다.

표준 문서 교환을 위한 전체 시스템 구조는 [그림 3-1]과 같다.



[그림 3-1] 전체 시스템 구조

표준 문서 교환을 위한 관리시스템의 전체 시스템 구조는 사용자 인터페이스, 표준 문서 관리 시스템, DBMS로 구성된다. 사용자 인터페이스는 관리 시스템에 접근이 쉬운 웹을 기반으로 제공한다. 표준 문서 관리 시스템은 표준 관련 문서를 데이터베이스에 변환, 검증 및 저장하는 저장 관리 시스템과 저장된 표준 문서를 사용자의 질의에 맞게 검색하여 전달하는 검색 관리 시스템으로 나뉘어진다. DBMS는 표준 문서 관리 시스템을 통해 제공되는 표준 문서 메타데이터와 문서에 대한 버전 정보를 관리하는 저장소로 나누어서 관리하게 된다.

가. 표준 문서 저장 관리기

· XML 생성기

XML 생성기는 두 단계의 변환과정을 통해 XML 형태의 문서를 생성한다. 즉, 워드 프로세서에서 작성한 문서를 서식 형태의 문서로 변환한 후, 변환된 서식 문서를 분석하여 XML 형태의 문서를 생성하는 단계로 이루어진다.

· XML 검증기

생성된 XML 문서에 오류가 없는지 또는 DTD에 맞게 정의가 되었는지를 체크하는 기능을 제공한다.

· PDF 생성기

PDF 문서 변환 방법은 크게 워드 프로세서 형태의 문서를 이용한 변환 방법과 생성된 XML 형태의 문서를 이용한 변환 방법으로 나눌 수 있다. 워드 프로세서 형태의 문서를 기반으로 변환할 경우 원래 작성한 문서 형태를 그대로 보존할 수 있으나, 워드 프로세서에서 PDF 변환 기능을 제공하는 API나 툴들의 기능을 이용하여야 한다. XML 형태의 문서를 변환할 경우 현재 이러한 기능을 제공하는 툴이 연구 진행 중에 있으며, 또한, 원래 작성된 표준 문서와 동일하게 표현하는 것이 현재의 기술로는 아직 어렵다. 이에, 본 연구의 PDF 생성기에서는 워드 프로세서의 문서를 PDF로 변환하기 위해 현재 제공되고 있는 워드 프로세서의 툴이나 API를 이용하여 서비스를 제공한다.

· 메타데이터 관리기

메타데이터 관리기는 메타데이터 추출기와 메타데이터 저장기로 나뉘어진다. 메타데이터 추출기는 XML 생성기를 통해 제공되는 XML 형태의 문서를 분석한 후, 필요한 정보를 추출하는 기능을 제공한다. 추출된 정보는 사용자에게 제공되는 질의 정보 및 결과 처리에 중요한 역할을 한다. 메타데이터 저장기는 메타데이터 추출기를 통해 전달된 메타데이터를 정의된 스키마에 따라 데이터베이스에 저장하는 기능을 제공한다.

· 문서 저장기

문서 저장기는 웹을 통해 전달되는 XML DTD와 워드 프로세서 문서, 그리고 XML 생성기나 PDF 생성기를 통해 생성된 문서들을 문서들의 버전에 따라 문서 버전 정보 저장소에 저장하는 기능을 제공한다.

나. 표준 문서 검색 관리기

· 사용자 인증기

사용자의 권한을 인증한 후, 표준 문서의 접근 권한을 부여하는 기능을 제공한다. 본 연구에서는 사용자에 따라 수정 및 편집 가능한 워드 프로세서 형태의 문서와 편집이 불가능한 PDF 형태의 문서를 제공한다.

· 질의 분석기

사용자는 질의 목적 및 대상에 따라 키워드, 문장, 연산 등 다양한 형태의 질의를 시스템에게 제공하며, 시스템은 사용자의 목적에 적합한 질의 결과를 제공하기 위해 질의를 분석하여 처리하는 기능을 제공해야 한다. 질의 분석기는 표준 문서 교환을 위한 검색 인터페이스를 통해 전달된 질의 정보를 분석한 후, 시스템에 적합한 질의 언어로 변환하는 기능을 제공한다.

· 질의 처리기

데이터베이스에 정의된 스키마에 따라 질의 분석기를 통해 전달 받은 질의를 처리를 수행하며

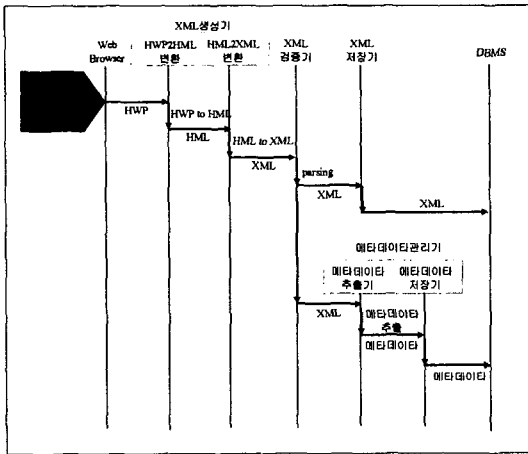
통해 데이터베이스에 저장된 데이터에 대한 질의 처리를 수행한다.

결과 처리기

사용자는 다양한 형태의 질의를 전달하는 반면 다양한 형태의 질의 결과를 요구한다. 결과 처리기는 요구한 질의 결과 형태에 따라 질의 처리기를 통해 처리된 질의 결과를 처리하는 기능을 제공한다.

3.2 시스템 기능 설계

가. 저장 관리 시스템 기능 설계



[그림 3-2] 저장 관리 시스템 흐름도

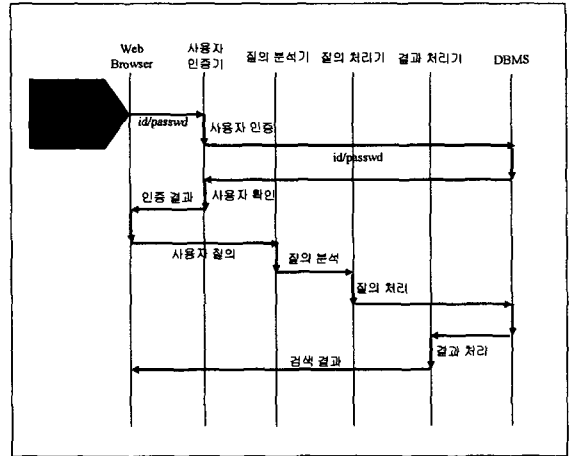
저장 관리 시스템은 웹을 기반으로 한 표준 문서 공유 사이트를 통해 작성된 워드 프로세서 문서를 전달 받는다. 전달된 워드 프로세서 문서는 HWP2XML 변환기를 통해 HML 형태의 중간단계를 거쳐 XML 형태의 문서로 생성된 후, XML 검증기를 통해 오류를 검사한다.

검증된 XML 문서는 XML 저장기와 메타데이터 추출기를 통해 나누어 처리된다. XML 저장기는 생성된 XML 문서를 저장하는 것으로 문서 버전 정보 관리를 통해 이전 XML 문서에 대한 버전 정보를 관리 한 후 데이터베이스에 저장한다. 메타데이터 관리는 생성된 XML 문서를 분석하여 필요한 메타데이터 정보를 추출한 후, 데이터베이스에 정의한 스키마에 저장한다.

나. 검색 관리 시스템 기능 설계

사용자는 인증 인터페이스를 통해 ID/Passwd를 입력하며, 입력된 ID/Passwd는 사용자 인증기를 통해 사용자의 권한을 체크하며, 사용자에게 주어진 권한에 따라 다양한 질의 인터페이스를 제공한다. 질의 인터페이스를 통해 입력된 질의 정보는 질의 분석기

를 통해 시스템에 적합한 질의로 분석한 후, 데이터베이스에 저장된 정보에 대해 질의 처리기를 통해 최적화된 질의 처리를 수행한다. 처리된 질의 결과는 결과 처리기를 통해 표준 분과별로 제안된 표준 문서에 대한 간략한 정보와 문서 파일을 사용자가 원하는 형태의 결과로 보여준다.



[그림 3-3] 검색 관리 시스템 흐름도

4. 결론

본 논문에서는 현재 국내·외적으로 표준 문서를 위한 관련 기술 연구 사례들을 조사하였다. 또한, 현재의 시스템을 기반으로 보다 효과적인 표준 문서를 관리하기 위한 표준 문서 관리 시스템 모델을 제시하였다.

본 논문에서 제시한 시스템 모델은 표준 문서의 효율적인 관리를 위해 XML 형태의 문서 교환을 제공하며, 사용자에게 익숙한 시스템을 제공하면서 새로운 문서 변환에 대한 투명성을 제공함으로써 이제까지 제안되어 왔던 시스템과는 달리 보다 생산적이며 효율적인 기능을 제공한다.

향후, 표준 문서에 대한 DTD를 설계하고 본 논문에서 제안한 시스템 모델을 기반으로 각 모듈의 구현할 예정이다.

참고문헌

[1] Bray, T., et al., "Extensible Markup Language(XML)," PR-xml,971208, W3C, <http://www.w3.org/TR>, 1997
 [2] invisible World Inc, <http://www.invisible.net/developers>
 [3] RFC Editor, <http://www.rfc-editor.org/>
 [4] TetraSix, Majix, <http://www.tetrasix.com/English/majix.htm>
 [5] HWPML(한글 마크업 언어) 매뉴얼 97 기능 강화판, 한글과 컴퓨터
 [6] TTA 단체 표준 문서, 한국 정보 통신 기술 협회, <http://www.ttr.or.kr>