

WEB상의 멀티미디어 문서 관리를 위한 XML문서관리시스템 설계 및 구현

공범용*, 황병곤**, 조세홍***

* 영남이공대학

** 대구대학교 정보통신공학부

*** 한성대학교 멀티미디어정보처리학과

Development of XML Document Management System for Multimedia Documents on the WEB

Beom-Yong Kong*, Byung-Kon Hwang**, Sae-Hong Cho***

* Yeungnam College of Science & Technology

** School of Computer Communication, Daegu Univ

*** School of Information & Computer Engineering, Hansung Univ.

요 약

컴퓨터 기술의 발달은 사회 여러 분야에서 획기적인 변화를 가지고 왔고, 이러한 변화를 주도하는 인터넷, 웹, 멀티미디어, 네트워크 등은 특별한 주목을 받고 있다. 특히, 지식과 정보의 빠르고 정확한 전달은 변화의 핵심을 이루는 요소이다. 지식, 정보화 시대에서 네트워크를 통하여 전달되는 지식과 정보는 멀티미디어 데이터가 그 주종을 이루고 있고, 이러한 멀티미디어 데이터들은 기존의 문자 위주의 데이터 정보에서와는 다른 성질들을 가지고 있기 때문에 새로운 문서 관리 시스템이 요구되어진다. 따라서, 웹상에서는 사용자가 요구하는 멀티미디어 문서를 보다 효과적으로 생성하고 저장, 검색할 수 있는 기술이 요구된다. 본 연구에서는 인터넷 중심의 정보교환 사회에서 생산적이며 효율적인 문서 공유 및 상호교환을 위해 차세대 인터넷문서의 표준인 XML을 기반으로 문서를 교환하며 이를 효과적으로 저장, 검색, 관리할 수 있는 시스템 모델을 제안하고자 한다.

시간적으로 떨어져 있는 정보 제공자와 정보 사용자 간의 원격 상호작용을 통한 접근기회를 용이하게 하여 정보 전달의 능동적인 참여와 아이디어의 공유를 가능하게 함으로서, 여러 사회 분야에서 양적이고 질적인 향상을 가져오는 등의 환경을 근본적으로 변화시키고 있다. 즉 정보통신 기술의 고도화·다양화로 인하여 사회 환경체제에서 시,공간을 초월한 한

1. 서론

인터넷 및 웹으로 대표되는 네트워크의 급속한 확산은 가장 효율적이면서도 효과적인 지식과 정보 전달 방법의 출현을 예고했다. 인터넷은 지리적으로 또는

경, 열린 환경, 사용자와 전달자의 끊임없는 상호작용이 가능한 환경, 정보통신 기술을 최대한 활용할 수 있는 환경, 정보의 탐색·가공·활용을 도와줄 수 있는 환경으로 변화, 개발되어 가고 있는 것이다. 이러한 변화는 산·학·연의 연구개발과 대학을 비롯한 정보통신업체들의 사회적 참여에 따른 확산에 기인한다고 볼 수 있다 [1].

본 연구에서는 인터넷 중심의 정보교환 사회에서 계속 그 중요성이 강조되고 있는 문서 체계에 대하여 논하고자 한다. 현재 인터넷 환경에서는 생산적이며 효율적인 문서 공유 및 상호교환을 위해 여러 연구가 진행되고 있고, 이러한 연구의 중심에는 XML을 이용한 문서 교환의 표준이 그 선두를 달리고 있다. 본 연구는 차세대 인터넷문서의 표준인 XML을 기반으로 문서를 교환하며 이를 효과적으로 저장, 검색, 관리할 수 있는 시스템 모델을 제안하고자 한다. 특히, 멀티미디어 문서를 효율적으로 관리하고 개인 정보를 프리젠테이션 할 수 있는 XML 멀티미디어 문서관리 시스템 설계·구현에 초점을 맞추고자 한다. 본 연구의 검증을 위하여 지식, 정보의 새로운 환경으로 각광을 받고 있는 Cyber 대학의 행정적인 문서관리시스템을 구현하였다.

본 논문의 전체구성은 다음과 같다. 제2장은 XML 및 전자문서에 대한 선행연구 고찰을 하고, 제3장에서는 XML 기반의 문서관리를 위한 시스템 설계를 위하여 문서생성·저장·검색모듈을 설명하며, 제4장에서는 실제 XML 문서관리 시스템 구현을 나타낸다. 마지막 제5장에서는 결론 및 향후 과제 등을 논의하고자 한다.

2. 관련연구: XML

XML(eXtensible Markup Language) 1996년 W3C(World Wide Web Consortium)의 XML Working Group에서 제안한 것으로 웹 상에서 구조화된 문서를 전송 가능하도록 설계된 표준화된 텍스트 형식이다. 1998년 2월 W3C표준에 따르는 XML 초기 사양이 발표되었다. XML(eXtensible Markup Language)의 표준제안 이후 현재까지 그 기능이 확장되고 있고, 기

존의 HTML(Hyper Text Markup Language)이 갖는 단점과 SGML(Standard Generalized Markup Language)이 갖는 단점을 보완하여 작성된 차세대 웹 언어 표준이다. XML 기반의 정보관리는 소프트웨어 및 플랫폼에 종립적인 전자문서관리를 제공함으로써 정보교환이 빈번한 시스템 관리에 적용되고 있으며, 현재 이를 이용한 시스템개발이 활발히 이루어지고 있다. 최근의 인터넷 브라우저인 익스플로러 5.0이상의 Version에서는 XML과 이에 대한 표현양식을 기술한 XSL(eXtensible Style Language)을 지원하고 있어, 인터넷상에서의 XML문서 활용은 더욱더 증가할 것이다. 이에 따라서 XML문서의 장점 및 특성을 응용한 문서생성 및 저장, 검색시스템을 위한 시스템 개발 및 연구가 활발히 이루어지고 있다.

XML 문서관리의 대상은 전자문서이다. 문서정보의 전자적 표현을 전자문서라고 하는데, 기술이 발달함에 따라 정보내용은 텍스트 중심에서 멀티미디어화하고 있다. 기존의 워드프로세서를 이용하여 전자문서를 교환할 경우 특정문서 편집시스템이 앞으로도 계속 사용되리라는 보장 및 시스템 업그레이드에 따른 포맷변환이 필요하다. 또한 특정 시스템을 이용하여 문서정보를 작성, 교환하는 경우 특정 시스템에 의존하게 되므로, 다른 시스템에서 사용하기가 불편하다. 이러한 문제는 문서의 저장형태가 각 시스템마다 독자적인 표현형식을 갖기 때문에 발생한다. 문서의 구조정보는 문서를 이루는 다양한 요소들의 외형과 위치의 차이에 의해 표현되는 데, 이 외형 자체로서는 문서의 정확한 구조적 정보를 전달하지 못한다. 따라서 문서의 외형적 구조보다 논리적 구조를 가지고 정의함으로써, 상호교환 및 공유시 발생하는 문제점을 해결하여 효율적이고 통일적인 문서정보의 관리가 이루어질 수 있는 것이다 [2,3].

문서관리시스템 발전의 근간은 디지털기술에 기반한 데이터베이스 기술의 다양화에 있다. 데이터베이스는 정보를 저장하고 관리하는 체계이다. 디지털 기술의 발달로 정보의 양이 늘어나고, 멀티미디어 기술에 의해 정보의 양식이 다양해짐에 따라 데이터베이스 체계의 구조는 기존의 고정된 길이의 필드

체계에서 가변적 형태로 변화하였고 구조적 연관관계에서 자유로운 연관관계로 발전하였다. 그리고 내용적인 측면에서는 텍스트 처리에서 멀티미디어 데이터의 처리로 변하고 있다. 데이터의 단위화(Encapsulation) 개념과 유전적(Inheritance) 성격을 부여한 객체지향 데이터베이스(Object-Oriented DB) 등 다양한 기술의 등장과 네트워크를 기반으로 한 데이터베이스 개발 등으로 응용의 범위를 넓혀가고 있다.

XML 기술이 많이 활용되고 있는 분야는 Text Processing Application을 들 수 있다. 이는 XML이 문서를 만드는데 필요한 특정한 규칙들을 제공하기 때문이다. 현재 대부분의 검색엔진들은 HTML을 사용하고 있다[4]. HTML은 데이터 포맷하는 방법을 기술하는 언어이므로 데이터간의 관계를 기술하는데는 적절하지 않고, 따라서 검색의 정확성이 떨어진다. 이에 반해 XML은 의미와 구조를 다르므로 검색성을 상당히 향상시킨다. 가장 기대되는 영역은 애플리케이션 개발이다. 데이터교환의 특징과 서버부하를 줄이는 능력은 XML이 분산된 애플리케이션, 특히 데이터가 빈번히 교환되고 업데이트되는 부문에서 우수한 기술이 되게 한다.

데이터를 구조화하는 능력으로 인해 XML 문서로 된 웹 상에서는 검색기능이 향상될 수 있다. XML 문서에서 검색엔진들은 더 이상 태그 속에 있는 정보만을 보는 것이 아니라 태그 자체를 본다. XML은 문서의 내용을 알 수 있게 하는 태그집합을 만들 수 있게 한다. 논리적 계층구조를 형성하는데, 태그집합은 정보에 대한 로드맵(Road Map)으로 기능한다[5]. XML을 이해하는 검색엔진(XML Search Engine)은 콘텐츠 히트(Content hits)보다 마크업 히트(Markup hits)를 더 높게 순위 매김 하게 된다. 또한, XML이 ISO 10646 UNICODE 표준을 지원하므로 XML은 플러그 인에 의존함 없이 다양한 언어들을 해독하고 병주화할 수 있다. 따라서 XML이 지원되는 브라우저는 검색결과를 알파벳 순, 연대순, 주제별, 심지어 파일형태나 언어별로도 디스플레이할 수 있다[6]. 즉 XML기술의 활용은 다음과 같다.

1. 태그와 DTD
2. XML문서형 정의(논리적 구조) 대 XML문서의 설계(정의에 따라 작성된 XML문서)
3. 논리적구조(요소,속성) 대 물리적구조(Entity)
4. XML문서의 특징 및 장점

3. XML 문서관리시스템의 설계 및 구성

본 연구에서 제안하는 문서관리시스템의 구성도는 XML에 기반한 문서관리시스템으로서 문서 작성자가 문서를 생성하고 저장, 검색하는 3개의 병렬 구조로 구성하였다. 또한, 사용자측 구성은 문서작성과 검색을 주요기능으로 하는 일반사용자로 구성되었고, 사용자측 구성을 통해 입력된 정보는 문서관리시스템을 통해 처리된 후 데이터베이스에 저장, 관리된다. 그림 1은 시스템 표준문서를 생성하는 부분과 저장하는 부분, 검색하는 부분 등 세 부분으로 나누어진 문서관리시스템의 전체 구성도를 나타내고 있다.

그림1에서 나타나 있듯이, 문서 생성 모듈은 DTD Generator와 XML Editor/Generator로 이루어져 있으며, DTD에서 엘리먼트 명을 받아 템플릿 XML파일을 생성하며, 사용자 입력 템플릿 폼에서 입력된 데이터를 이용하여 XML문서를 생성한다.

문서저장시스템은 XML Parser, 색인기, 저장모듈로 이루어져 있다. XML Passer는 엘리먼트, 파라미터, 텍스트 등을 분리하여 트리 형태로 나타내며, 구조정보색인기는 트리형태로 재구성된 XML문서에서 객체 엘리먼트, 파라미터, 텍스트 등의 정보를 뽑아낸다. 이렇게 추출된 정보는 XML문서에 대한 인덱스 생성에 활용된다.

문서검색시스템은 질의 입력등 인터페이스를 위한 GUI 모듈과, 통합검색기, 질의처리기 등의 모듈로 세분화 된다. 검색시스템은 데이터베이스에 질의하는 부분과 검색하는 부분으로 나뉘고 각 기능상태별로 질의입력 인터페이스 모듈, 검색기모듈(질의 분석, 단위질의), 질의처리모듈, 질의결과 생성모듈, 단위질의 관리모듈 등 5 가지 모듈로 나눈다

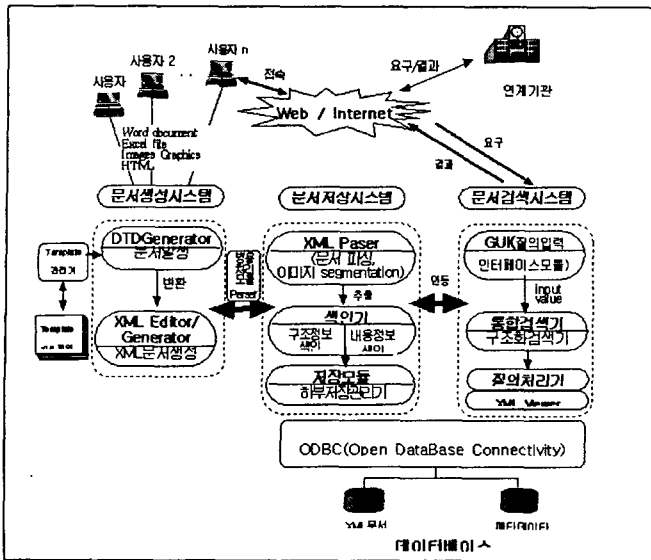


그림 1. XML 문서 관리시스템의 전체 구성도

4. XML 문서관리시스템의 구현

시스템 구현을 위하여 사용한 언어는 Window NT 2000 Server 플랫폼을 기반으로 XML 및 Java를 이용하여 구현하였다. 또한, 문서저장시스템은 Java언어 외에 생성된 XML 문서를 검증한후 DB에 저장하고, 저장할 메타데이터를 추출하기 위해 IBM에서 제공하는 XML4C 파서를 이용하였고, PDF 문서생성을 위해 HANQ 2PDF를 이용하였다. 시스템의 데이터를 관리하는 DB는 MS-SQL 2000 Server을 기반으로 구성하였으며, 사용자는 ASP 3.0을 기반으로 구성되어 제공되는 웹 브라우저를 통해 정보를 입력하거나 제공하도록 하였다.

시스템이 제공하는 스키마는 네 가지 정보로 구성되어 있다. 첫째, 문서를 구분하기 위한 ID정보이다. 등록된 문서에 대해 고유한 ID를 부여하여 문서를 저장 및 관리한다. 둘째, 문서 등록시 문서등록자가 입력하는 정보이다. 작성한 문서파일에서 알 수 없는 등록자 이름, 등록자 이메일, 등록자 패스워드, 문서의 주요 키워드 정보 등 문서관련 정보를 문서 등록시 입력받아 관리한다. 입력된 대부분의 정보는 사용자가 문서에 대한 전체목록 검색시 간략 정보로 제공된다. 셋째, 등록문서에 대한 XML문서에서 추출한 메타데이터 정보이다. 작성한 문서의 내용 중 사

용자의 검색에 자주 사용되는 목차, 요약 등의 정보를 생성된 XML문서에서 추출하여 관리한다. 이러한 정보는 사용자에게 문서에 대한 상세 정보로 제공된다. 넷째, 등록문서에 생성된 문서파일에 대한 정보이다. 문서등록자가 등록한 HWP문서파일 및 등록으로 인해 자동 생성되는 XML문서 파일과 PDF문서파일을 관리한다. 저장된 문서파일은 사용자에게 문서파일 검색 시 인증과 함께 제공된다.

본 연구가 제안한 문서관리시스템의 실증적 검증을 위하여 사이버 대학의 문서관리시스템을 모델로 구현하였다. 문서관리시스템의 구성은 제안한 바와 같이 크게 세 부분으로 나누어지고, 구현한 홈페이지 구성은 그림 2와 같다. 문서를 등록하고, 검색, 열람 및 결재하는 사람도 사용자로서 문서관리시스템은 등록된 사용자만이 사용 가능하므로 사용자의 로그인 기능과 사용자의 인증 기능을 설계하였다.

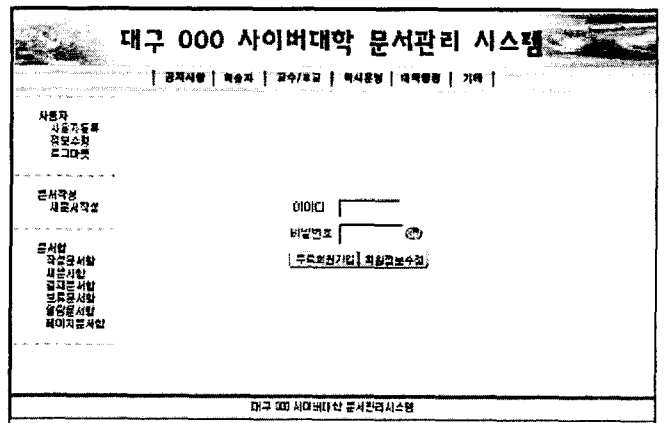


그림 2. XML 문서 관리시스템의 초기화면

XML문서관리시스템 사용자등록 초기화면은 그림 3과 같으며, 전체 문서관리항의 초기화면은 그림 4와 같고, 또한 문서 검색시스템의 초기화면은 그림 5와 같다. 검색된 문서내용 보기화면은 그림 6과 같으며, 수정·삭제할 수 있는 새 문서작성의 초기화면은 그림 7에 나타나 있다.

활용한 웹상의 멀티미디어 문서관리가 다양한 분야에서 널리 활용될 수 있는 가능성을 제공하였다.

향후 연구과제로서는 멀티미디어를 활용한 XML 문서에 대한 정보기술과 인터넷 등이 지속적인 발전을 거듭하면서 XML 문서관리시스템 구축을 위해 더 쉽게 설치, 삭제할 수 있는 종합관리자 모드 개발 등에 대한 기술 개발이 필요할 것이다.

참 고 문 헌

[1] I. F. Cruz, and P. Mahalley, "Temporal Synchronization in Multimedia Presentations," Proc. of the IEEE Int. Conf. Multimedia Computer Systems, Vol.2, pp. 851-856, 1999.

[2] Charles F. Goldfarb and Paul Prescod, The XML handbook, Prentice Hall, 1998.

[3] 이재동, 김재용, XML을 이용한 학생정보관리시스템의 설계 및 구현, 한국컴퓨터교육학회 논문지 제 3권 제2호, 2000.

[4] 정희경, XML가이드 그린 1998.

[5] 김동신, 이용규, XML 문서 접근 권한 관리, 한국멀티미디어학회 추계논문집, 1999.

[6] 최지태, 전영훈, 황대훈, DTD 생성기능을 지원하는 XML 문서 병용 편집기의 설계 및 구현, 한국멀티미디어학회 추계논문집, 1999.