

XML 저장 관리 시스템에 관한 연구

A Study for XML Storage Management System

정지문, 우성구, 이창용*, 최성
남서울대학교 컴퓨터학과

Ji-Moon Jung, Sung-Gu Woo, Chang-Yong Lee*, Sung Choi
Dep. of Computer Science NamSeoul University

요 약

대용량의 XML 문서를 관리하기 위해서는 전형적인 XML 저장관리시스템이 필요하다. 본 연구에서는 XML을 활용한 저장관리시스템의 구조 및 구성모듈과 시스템 구현시 고려되어야 할 기능과 핵심 기술을 살펴본다. 또한 지금까지 개발된 국내외 XML 저장관리시스템의 특성을 연구하였다.

1. 서론

XML은 차세대 웹 문서 포맷으로 부각되고 있는 것으로 W3C에서 제안된 국제 표준의 전자문서 메타 언어이다. XML은 웹에서 구조화된 문서를 전송 가능하도록 설계된 표준화된 텍스트 형식으로, 문서를 구성하는 각 요소들의 독립성을 보장하게 함으로서 문서의 호환성, 내용의 독립성, 요소 변경의 용이성 등의 특성을 제공한다.

이에 따라 HTML의 새로운 대안으로 떠오르고 있는 XML이 현재 인터넷 사용자 및 관련업계의 주요 관심사가 되고 있으며, 인터넷 웹 문서뿐만 아니라 전자도서관, CSCW(Computer Supported Cooperative Work) 그리고 CALS(Commerce At the Light Speed)를 포함한 다양한 분야에서 XML을 활용하고자 폭 넓은 연구를 하고 있다.

이처럼 XML이 광범위한 응용분야에서 채택되는 이유는 다음과 같다.

첫째, 시스템에서 독립적인 문서 교환 및 처리가 가능하기 때문에 이질적인 네트워크 및 분산 환경 하에서의 전자 상거래 및 EDI의 정보 기술 언어로 활용할 수 있다.

둘째, 정보 생성에 드는 비용과 시간을 줄일 수 있기 때문이다. 하나의 단위로 참조될 수 있는 문자들의 집합인 엔터티를 정의한다. 이와 같은 엔터티 기능은 동일한 정보를 다양한 목적에 따라 효율적으로 재구성, 재사용할 수 있게 하며, 데이터베이스 시스템을 저

장소로 이용하여 사용자 요구에 의한 문서를 동적으로 생성할 수 있다.

셋째, 문서의 논리적인 마크업과 포매팅 정보를 분리시킴으로써 특정 문서 집합에 대한 표준화된 출력과 동일 문서에 대한 다수의 출력 포맷을 지정할 수 있다.

이와 같이 전세계적으로 XML에 대한 관심이 고조되고 실제 많은 분야에서 활용되고 있기 때문에 향후 XML 문서의 확산이 더욱 가속화될 것이며, XML 문서의 생명주기 전 과정에서 정보의 생산성, 재사용성, 지속성, 이식성 등과 같은 XML 문서 사용의 장점들을 얻기 위해서는 XML 문서들을 저장, 관리 및 검색할 수 있는 XML 저장관리시스템이 필수적으로 요구되고 있다.

2. 본론

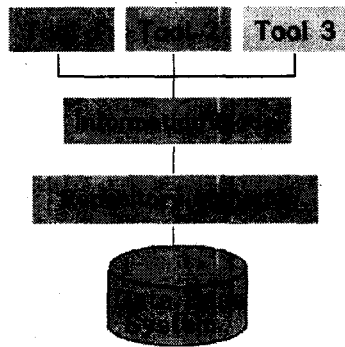
I. 일반적인 저장관리시스템

일반적으로 저장관리시스템은 소프트웨어, 문서, 지도, 전자 회로 등과 같은 특정 분야의 정보를 관리하는 시스템이다. 예를 들어, 소프트웨어 개발을 지원하는 저장관리시스템은 데이터베이스 설명, 형식 정의, 제어, 문서들, 인터페이스 정의,

소스 코드, 도움 문서, 실행 화일, 아이콘, make file들의 내용을 저장하고 관리한다. 저장관리시스템과 DBMS의 가장 두드러진 차이점은 내장된 정보모델을 갖는다는 것이다.

이는 저장하고자 하는 정보의 특징을 잃어버리지 않고 데이터베이스에 저장하기 위해 반드시 필요하다. 또한 데이터베이스에서 제

공하는 기능 이외에 저장하고자 하는 데이터의 특성에 맞는 객체관리(Object Management), 버전관리(Version Management), 관계관리(Relationship Management), 동적확장(Dynamic extensibility), 통지(Notification)서비스 등 다양한 기능을 제공한다는 점이다. <그림 1>은 전형적인 저장관리시스템의 구조를 보여주고 있다.



<그림 1> 일반적인 저장관리시스템의 구조

소프트웨어 툴을 사용하여 생성된 정보는 데이터베이스에 저장하고자 할 때 저장하고자 하는 정보의 특징을 살려주기 위해 설계한 정보 모델로 변환을 하고 저장관리자(Repository Manager)를 통해 데이터베이스에 저장한다. 저장관리자(Repository Manager)는 정보들을 저장, 관리 및 조작할 수 있는 서비스들을 제공한다.

II. XML을 활용한 저장관리시스템의 특징

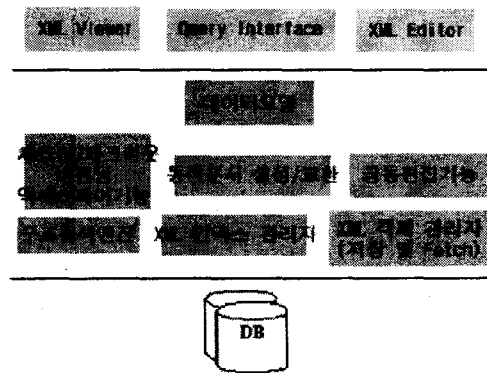
일반적인 저장관리시스템에 비해 XML 저장관리시스템은 XML이 갖고 있는 다음과 같은 특성을 고려하여야 한다. XML 문서는 복잡한 구조와 다양한 미디어를 포함할 수 있다. 이러한 XML 문서는 구조 정보를 이용하여 문서를 효율적으로 관리하기 때문에 데이터베이스에 XML 문서, DTD, 구조정보 및 다양한 미디어를 저장, 관리해야 된다. 즉, XML 문서들이 포함하고 있는 많은 특성들을 손실 없이 데이터베이스에 표현할 수 있도록 스키마 설계가 잘 이루어져야 한다.

또한 기존의 정보검색시스템(IRS: Information Retrieval System)은 문서의 논리적 구조에 따른 정보검색을 거의 이용하지 못하고 있다. 그러나 XML 문서의 특성을 볼 때

다음과 같은 검색 조건을 줄 수 있어야 하며 이것을 혼합하여 사용할 수 있어야 한다.

- 첫째 XML 문서에 대한 내용질의
- 둘째 XML 문서가 갖는 논리적인 계층 구조를 이용한 질의
- 셋째 엘리먼트가 갖는 속성에 대한 질의

<그림 2>는 XML 저장관리시스템 구현 시 일반적으로 고려해야 될 기능들은 보여주고 있다. 각 기능들에 대한 설명은 아래와 같다.



<그림 2> XML 저장관리시스템의 주요기능

① 데이터 모델링 기능

XML 문서를 저장하기 위한 데이터 모델은 저장 시스템에 삽입하기 전의 문서와 저장된 문서를 다시 복원했을 때의 데이터 손실을 방지하기 위해 XML의 모든 특성들을 고려해야 한다. 또한 문서 내에 포함된 멀티미디어 정보와 XML 문서의 구조정보의 저장을 고려해야 한다.

② XML 객체 관리자

XML 저장관리시스템에서 가장 핵심적인 기능을 담당한다. XML 객체관리자에서는 위에서 설계한 데이터 모델을 바탕으로 실제 XML 문서를 저장하기 위한 스키마를 생성하고, 생성되어 있는 스키마에 XML 문서, 구조정보, 멀티미디어 등을 저장하며, 저장된 XML 문서에서 사용자가 원하는 전체 문서 또는 문서의 일부분을 꺼내는 일을 담당한다.

③ XML 인덱스 관리자

XML 저장시스템에 저장되는 문서 및 기존에 저장된 문서의 수정/삭제 시 색인을 생성하고 구축된 색인 정보들을 동적으로 갱신하는 기능 등을 제공한다.

④ 구조 검색 엔진

저장된 XML 문서 중 사용자에게 접근 권한이 허용된 문서들에 대한 구조, 속성, 엘리먼트, 키워드에 대한 검색을 처리하는 기능을 담당한다. 사용자는 정보 검색 시 검색 범위를 문서의 일부 특정 부분으로 한정시켜 질의를 할 수 있으므로 검색 결과를 향상시킬 수 있다.

⑤ 라이브러리서비스 기능

문서의 생명 주기 제어 기능, 사용자 권한에 의한 액세스 제어 기능, 다수 저자의 동일 문서에 대한 공동 저작 시 발생하는 문서의 불일치를 방지하는 check-in/check-out 기능, 문서의 생명 주기 동안의 변경 사항을 관리하는 버전 기능 등을 제공해야 한다.

⑥ 공동 편집 기능

XML 문서들이 주어졌을 때, 몇몇 사용자들이 종종 동시에 전체 문서의 서로 다른 부분들을 접근한다. 그러므로 사용자는 저장소로부터 문서의 일부분을 추출할 수 있는 또는 다양한 툴(XML Editor, Viewer 등)에 부분을 전송할 수 있고 처리한 후에 재통합하는 문서 단편서버(document fragment server)와 같은 서비스들이 필요하다.

⑦ 동적 문서 생성/교환 기능

사용자의 질의 또는 문서 생성 템플릿에 의해 저장소에 저장된 객체들을 재구성하여 동적으로 문서를 생성하는 기능 및 문서의 내용을 유지할 수 있는 부분 교환 기능을 제공해야 한다.

III. XML 저장관리시스템의 핵심기술

① 데이터 모델링

XML 문서 관리 시스템에서 가장 핵심적인 역할을 수행하는 XML 객체 관리자를 개발함에 있어서 가장 먼저 선행되어야 하는 것이 데이터 모델링이다.

XML 문서 모델링에 대해서 기존에 연구되었던 내용들에 대해서 살펴보면 DBMS 활용에 따른 데이터 모델링에서 관계형 모델과 객체지향 모델로 나눌 수 있고, XML 문서의 저장 방식에 따른 데이터 모델링으로는 분할 저장 모델, 비분할 저장 모델, 혼합 모델로 나눌 수 있다. 각각의 특징을 보면 다음과 같다.

② DBMS 활용에 따른 데이터 모델링

(i) 관계형 모델

관계형 모델은 현재 가장 많이 사용하고 있는 관계형 데이터베이스를 기반으로 하고 있기 때문에 쉽게 접근할 수가 있어 사용자들의 전반적인 확산이 빠르고 쉽게 상용화할 수 있으며 내용 검색 등에 용이한 장점이 있다. 그러나, 관계형 데이터베이스의 특성상 문서의 구조에 대한 충분한 정보를 유지하기 위해 필요로 하는 테이블과 튜플의 수가 기하급수적으로 늘어날 수 밖에 없으며, 이에 따른 JOIN 연산으로 인한 시스템의 성능이 저하되고, 상속의 개념이나 복잡한 하이퍼미디어 객체의 표현, 엘리먼트의 속성인 애트리뷰트 등의 표현이 불가능하다.

(ii) 객체지향 모델

객체지향 모델은 객체지향 데이터베이스에서 스키마를 설계할 때 사용되는 클래스를 이용하여 각 요소들의 구조를 정의할 수 있다. 이 모델은 데이터베이스에서 지원하는 객체지향 개념을 이용할 수 있기 때문에 상속과 같은 객체지향 특성을 이용할 수 있으며, 엘리먼트 간의 전후종속 관계를 클래스에 기반한 객체들간의 링크로 나타낼 수 있기 때문에 구조적인 문서를 모델링하는데 적합하다 할 수 있다.

③ 저장방식에 따른 데이터 모델링

(i) 분할 저장 모델

분할 저장 모델은 XML 인스턴스를 엘리먼트 별로 나누어서 저장하는 방식을 말하며, 문서의 실제 내용을 가지고 있는 각각의 단말 엘리먼트안에 문서의 내용이 쪼개져서 저장된다. 따라서, 문서의 구조적 정보나 일부 내용들이 수정되었을 때 관계되는 노드들만 수정하면 되므로 문서의 편집 및 관리가 쉽고, 동일한 내용을 갖는 노드들을 공유할 수 있다는 장점이 있지만, 문서의 내용을 추출하고자 할 때 각 단말 노드들을 순회하며 통합하는 과정에서 시스템의 성능을 저하시키는 문제가 발생한다.

(ii) 비분할 저장 모델

비분할 저장 모델은 분할 저장 모델과 같이 문서의 내용을 나누어서 저장하지 않고 문서 전체를 BLOB 형태로 저장한 다음, 각각의 단말 노드는 오프셋 정보를 가지고 접근하는 방식이다. 이는 문서를 한꺼번에 저장하였기 때문에 통합 과정이 필요 없어 문서 참조를 빨

리 할 수 있지만, 내용의 일부만이 수정되었을 때도 문서 전체를 재구성해야 한다는 큰 단점이 있다.

(iii) 혼합모델

혼합모델은 분할 저장 모델과 비분할 저장 모델을 혼용하여 사용하는 모델로 각각의 모델에서 단점을 보완하고자 상대 모델의 특성을 일부 포함하였다. 하지만 혼합 모델의 단점인 저장 공간이 많이 소모된다는 점과 문서의 일부를 가져오려고 했을 때에는 어쩔 수 없이 통합 과정이 필요하게 되는 문제점이 있다.

IV. XML 저장관리시스템 개발 현황

지금까지 ObjectDesign사의 eXcelon, Oracle 8i의 인터넷 파일 시스템(iFS) 등과 같은 XML 저장관리시스템이 출시되고 있으며, 국내적으로는 케이오테크, 테크노2000프로젝트등 국내 XML 전문업체에서 XML 저장관리시스템을 잇따라 선보이고 있다.

ObjectDesign사의 eXcelon은 XML 문서를 파싱하고 파싱된 형태로 문서를 저장한다. 또한 XML 화일 뿐만 아니라 텍스트를 포함한 MIME 데이터, multimedia, HTML등 여러 종류의 문서를 저장하고 XML 문서에 질의하기 위해 XQL을 지원한다. eXcelon에 저장된 XML 데이터를 조작하기 위해 DOM(Document Object Model)을 사용한다. eXcelon 클라이언트는 COM 객체를 통하여 이용가능하고 XML 데이터는 여러 머신에 분산되어 저장될 수 있다.

오라클 인트라넷용 데이터베이스의 새로운 버전인 Oracle 8i의 인터넷화일시스템(iFS)은 XML을 지원한다. 오라클에서 개발한 XML 파서는 Java 언어로 작성되었고 데이터베이스에 들어오거나 나가는 XML문서를 자동적으로 프로세싱하고 XML 문서를 각 컴포넌트로 나누어 저장한다. 구조정보검색을 위한 ConText 엔진은 XML 태그에 기반한 검색을 통해 검색이 문서의 특정 섹션, 혹은 극히 짧은 문단 안에서 이루어지도록 한다.

테크노2000프로젝트는 XML 문서를 효과적으로 관리할 수 있는 XML 저장관리시스템인 XDMS2000을 개발하고 있다. 국산 DB인 UniSQL 기반으로 개발된 XDMS2000은 대용

량 XML 문서에 대한 구조검색과 전문검색을 지원하며 검색결과 순위표시 기능, 요약정보 미리보기 기능, 링크에 대한 질의 기능 등을 갖추고 있다.

3. 결 론

본 연구에서는 대용량의 XML 문서를 관리하고 저장할 수 있는 XML을 활용한 저장관리시스템의 특징, 기술, 개발 시 문제점을 살펴보았다. 또한 국내의 개발된 XML 저장관리시스템의 특성을 조사하고 분석한 결과를 제시하였다. 이러한 분석 결과는 본 인이 직접 제품을 사용해보지 않고 개발회사의 소개책자, 논문, 잡지 등에 기반하여 도출되었기 때문에 약간의 오류가 있을 수 있다는 것을 밝혀두는 바이다.

결론적으로 XML 저장관리시스템은 XML로 작성된 문서의 구조 정보와 대용량의 멀티미디어 데이터를 손실없이 저장, 검색, 관리할 수 있게 해 주는 시스템으로 향후 XML 문서가 점차 보편화되면 문서의 다양한 활용을 위한 가장 중요한 기반이 되는 기술이다.

따라서 정부, 관련연구기관 및 산업체 등에서 XML 저장관리시스템에 대한 많은 관심과 지원이 적극적으로 이루어졌으면 한다.

< 참고문헌 >

[1] 연제원, 장동준, 김용훈, 이강찬, 이규철, "효율적인 검색 지원 SGML 저장 관리기의 설계 및 구현", '99 한국 데이터베이스 학술대회 논문집 15권 1호, pp136-143, 1999

[2] 이원석, 대량의 구조화 문서 관리를 위한 SGML 저장 관리기의 설계 및 구현, 석사 학위 논문, 충남대학교(1998).

[3] 유재수외8인, "전자도서관 표준 문서 관리를 위한 XML 저장관리기 개발", 시스템매뉴얼, 케이오테크, 1999.

[4] W3C, "Extensible Markup Language(XML) 1.0",

<http://www.w3c.org/TR/1998/REC>

-xml-19980210, 1998.

[5] ORACLE Technical Data Sheet, <http://www.oracle.com>

[6] Techno2000 Products Information, <http://xml.t2000.co.kr>