

XML 기반 범용 Wrapper의 설계

황명진, 정채영, 배종민, 강현석
경상대학교 컴퓨터과학과 / 전산개발연구소

Design of A XML-Based General Wrapper

Myung-Jin Hwang, Chai-Young Jeong,
Jong-Min Bae, Hyun-syug Kang
Dept. of Computer Science, Gyeongsang National University

요약

인터넷 환경에서 이 기종 데이터베이스의 통합은 중요한 연구 과제 중의 하나로서, 그를 위한 자료저장소 Wrapper에 대한 관심이 높아지고 있다. 한편, XML은 이 기종 시스템들간에 효과적으로 데이터를 교환하는 도구로서 최근 많은 DBMS들이 XML을 지원하고 있다. 그러나 그들은 각기 자신의 데이터 모델만을 지원하기 때문에 상호 호환성이 떨어진다.

본 논문은 인터넷 환경에서 분산된 이질의 데이터베이스들을 효과적으로 통합할 수 있도록 일관된 변환 모듈을 사용하여 데이터 모델에 독립적으로 동작할 수 있는 XML 기반의 범용 Wrapper를 제안한다.

1. 서론

인터넷 환경에서 기존의 정보 시스템들은 각자 고유의 데이터 모델을 사용하여 자신의 데이터를 표현한다. 이에 따라 사용자들은 찾고자 하는 정보의 위치와 정보원의 사용 방법을 알기 위해 많은 노력을 기울여야 한다. 예를 들어, 유전자 관련 데이터베이스는 유형이 대략 10여가지가 있는데, 각 유형에 따라 수십개의 데이터베이스가 개발되어 있다. 따라서 생명공학자가 활용할 수 있는 데이터베이스의 수는 수백 내지 수천 개에 이르며, 이러한 데이터베이스들의 통합은 생물정보학에서 중요한 연구주제가 되었다.[3].

이러한 이질적인 정보 시스템들을 통합하기 위해서는 하나의 중립적인 데이터 모델이 필요한데, 최근 XML 기반의 모델이 중요한 대안으로 떠오르고 있다. XML은 W3C에서 웹을 기반으로 하는 구조화된 문서를 기술하는 방법에 대한 표준안으로 제시되었다.

이에 따라, 기존의 정보 시스템들이 갖는 데이터

모델을 XML 데이터 모델로 변환해야 할 필요가 생겨났고, 통합하고자 하는 정보 시스템간의 연동을 담당하는 Wrapper가 필요하게 되었다.

본 논문에서는 이를 위해 일관된 변환 모듈을 사용하여 데이터 모델에 독립적인 XML 기반의 범용 Wrapper를 제안한다. 제안하는 XML 기반의 범용 Wrapper는 사용자의 질의를 해당 정보 시스템 상에서 사용 가능한 질의어로 변환하는 기능을 가지며, 질의의 결과를 XML 문서로 변환시키는 기능을 수행한다. 그리고 XML 문서를 해당 정보 시스템의 저장구조로 사상시키는 기능도 수행한다.

2장에서는 관련 연구를 알아보고 3장에서는 크게 3개의 부시스템으로 구성된 범용 Wrapper의 구성을 알아본다. 그리고 4장에서는 결론 및 향후 계획을 논의한다.

2. 관련 연구

XML 문서를 저장하는 XML 데이터베이스의 경우, 데이터베이스 스키마를 기술하기 위해 DTD를 사용할 수 있다. 그러나 DTD는 XML 문서 형식과 다르고 타입을 지원하지 않는 등 스키마 언어로서

의 기능을 효과적으로 수행하기가 힘들다. 이에 따라 XML Schema[6,7], DCD[8], XML-Data[9], SOX[10], DSD[11], DDML[12] 등의 다양한 XML 스키마 언어가 출현하였다. 범용 Wrapper의 개발을 위해서는 이러한 XML 스키마 언어를 XML 문서와 각종 데이터베이스 모델 사이를 사상하는 중간 표현으로 활용할 수 있다. 우리는 이중 최근 많은 지지를 받고 있는 W3C 제안의 XML Schema를 사용한다.

국내외에서 최근들어 여러 가지 형태의 자료저장소 Wrapper의 개발이 보고되고 있다. 대표적인 것으로 외국의 DB2XML과 국내의 OODB2XML을 들 수 있다. DB2XML[2]은 RDBMS에 저장되어 있는 정보나 질의 결과를 XML 문서로 변환하는 시스템이다. 그리고 OODB2XML[1]은 XML 질의어를 객체지향 데이터베이스 질의어로 변환하고, 질의 결과를 XML 문서로 변환한다.

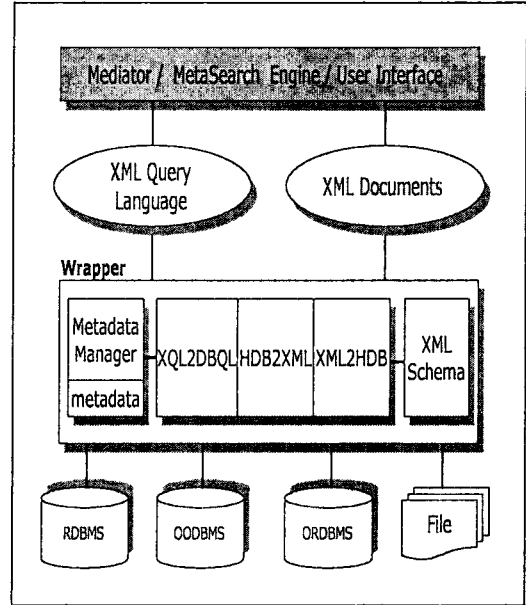
위의 연구결과들 외에도 XML 문서를 기존의 데이터 모델의 저장 구조로 사상시키는 연구[3], 그리고 질의처리 결과를 XML 문서로 보여주는 연구[4] 등이 활발히 이루어지고 있다.

그런데, 이들은 대부분 자신이 사용하는 각 DBMS의 데이터 모델에 맞게 개발되어 있어 DBMS가 달라질 경우, 그대로 사용할 수 없게 된다. 따라서 본 논문에서는 데이터 모델에 독립된 형태로 다양한 DBMS에 사용할 수 있는 범용 Wrapper를 제안한다.

3. 범용 데이터 저장소 Wrapper의 구성

범용 Wrapper의 개략적인 구조는 아래 <그림 1>과 같다. 그림에서 보는 것과 같이 범용 Wrapper의 핵심 구성 요소는 3개의 부시스템들이다. 즉, (1) XML 질의어와 자료저장소 질의어 사이를 변환하는 XQL2DBQL(XML Query Language To Data Base Query Language), (2) DB의 내용이나 질의 결과를 XML로 변환하는 HDB2XML(Heterogeneous DB TO XML), (3) XML 문서를 DB 스키마로 사상하는 XML2HDB(XML TO Heterogeneous DB)이다.

그리고 제안하는 범용 Wrapper의 사용자는 데이터베이스 통합을 위한 Mediator나 메타 검색엔진이 될 수 있으며, 질의는 XML 질의어인 XQL을 사용한다. 그리고 일반 사용자도 손쉬운 사용자 인터페이스를 통해 범용 Wrapper을 이용하여 XML 문서를 검색하거나 관리할 수 있다.

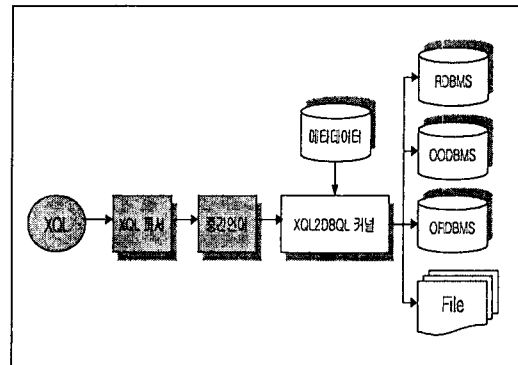


<그림 1> XML기반의 범용저장소 Wrapper구성도

여기서는 이중 핵심이 되는 3개의 부시스템들을 중심으로 알아본다.

3.1 XQL2DBQL의 구조

XQL2DBQL은 XQL 질의어를 목적 데이터베이스 질의어로 변환하는 모듈로서, 그 구조는 <그림 2>와 같다.



<그림 2> XQL2DBQL 구조

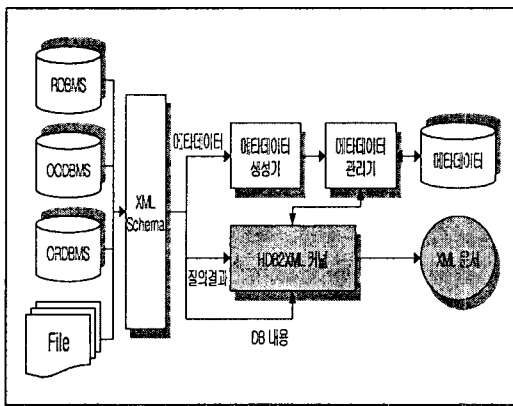
XQL로 기술된 XML 질의어는 XML 파서에 의해서 파서 트리 형태로 표현되는데, 범용 질의어 변환기(XQL2DBQL 커널)는 이 파서 트리를 입력으로 받아서 메타데이터를 바탕으로 각각의 목적 데이터베이스에 맞게 목적 데이터베이스 질의어를 생성한다.

XML 질의어를 목적 데이터베이스 질의어로 변환하는 기술의 핵심은 변환된 질의어 처리의 효율성, 변환의 완전성, 그리고 범용성에 있다. 질의어 변환시, 데이터의 저장구조에 크게 의존하는 경우에는 변환된 질의어 처리의 효율성은 뛰어나지만 범용성이 떨어진다. 반면에, 저장 구조에 대한 최소한의 정보만을 활용하면 범용성은 높으나 질의어 처리의 효율성과 완전성은 감소한다. 따라서 저장구조에 대한 적절한 정보를 갖는 것이 중요하다.

우리는 이를 위해 XML Schema를 이용하며, 이때 데이터베이스 스키마 생성을 위한 여러 가지 저장에 관련된 제약이나 결정 사항들을 관리함으로써 효과적으로 질의어를 변환할 수 있도록 한다.

3.2 HDB2XML의 구조

HDB2XML은 데이터베이스의 내용이나 질의 결과를 XML 문서로 변환하는 모듈로서, 그 내부 구조는 <그림 3>과 같다.



<그림 3> HDB2XML 구조

HDB2XML은 데이터베이스로부터 스키마 정보를 추출하여 메타데이터 DB를 구축하고, 이를 바탕으로 데이터베이스 내용이나 질의 결과를 XML 문서로 변환한다.

XML 문서를 DB에 사상시키는 방법은 DTD나 XML Schema와 같은 스키마가 명시적으로 주어진 경우와 이들이 명시적으로 주어지지 않은 경우로 나눌 수 있는데, 전자의 경우가 상대적으로 사상하기가 용이하다. 왜냐하면, 후자의 경우에는 완전히 확정된 스키마를 갖지 않고 줄거리만 어느 정도 정해진 반 구조적(Semistructured) 문서를 다루어야 하기 때문이다.

우선 DTD 기반의 XML 문서 사상 기법의 경우, 관계형 모델에서는 값의 집합을 표현하는 문제와

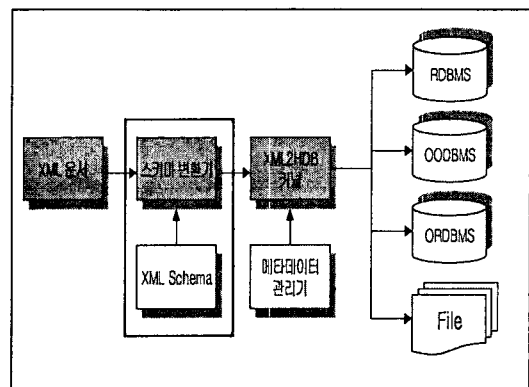
재귀적(Recursive) 관계에 있는 엘리먼트를 처리해야 하는 어려움이 있다. 따라서 객체지향 모델을 활용하는 기술이 많이 발표되고 있는데, 본 논문에서는 이에 대해 사상 결과의 효율성이 데이터 모델에 크게 종속되는 문제를 극복할 수 있도록 XML Schema 언어를 이용해 중간 구조로 우선 사상하는 방안을 사용한다.

한편, 일반적으로 반 구조적(Semistructured) 문서에는 스키마가 어느 정도 암시되어 있다고 볼 수 있는데, 반 구조적 문서에 대한 데이터베이스 관리 기술은 그 응용 범위가 넓은 관계로 세계적으로 활발히 연구되고 있다. 그러나 국내에서는 아직 상대적으로 연구 결과가 미미한 편이다. 특히 반 구조적 문서로부터 스키마를 지능적으로 추출하는 기술은 데이터베이스의 효율과 질에 미치는 영향이 매우 크다. 우리는 이를 위해 Lore[13]에서와 같이 OEM 데이터 모델을 사용하는 방법을 사용한다.

결과적으로, 우리는 현재 두 가지 경우에 대해 각기 다른 방식으로 접근하고 있다. 하지만, 앞으로 이를 체계적으로 통합해서 다루는 방법을 고안할 예정이다. 그런데, 이러한 구조적 또는 반 구조적 XML 문서를 통합된 DB 스키마로 사상하는 기술을 만족할 수 있을 정도로 개발하기 위해서는 많은 시간이 소요될 것으로 전망된다.

3.3 XML2HDB의 구조

XML2HDB는 XML 문서를 목적 데이터베이스 스키마로 사상하는 모듈로서, 그 구조는 <그림 4>와 같다.



<그림 4> XML2HDB 구조

저장될 XML 문서는 XML Schema 언어로 기술된 XML 스키마로 변환되고, 저장 구조 사상은

목적 시스템에 대한 메타데이터를 바탕으로 XML 스키마를 목적 데이터베이스의 저장 구조로 사상한다.

앞에서 언급한 바와 같이 DB 검색 결과를 XML 문서로 제공하는 기능은 최근의 DBMS들에서 거의 대부분 옵션으로 제공하고 있을 정도로 매우 보편화되고 있다. 뿐만 아니라 최근에는 DB에 저장된 내용을 XML 문서로 직접 제공하는 시스템도 나타나고 있다. 그러나, 이들의 경우 각 DBMS들이 별도의 시스템으로 이들을 가능하게 하기 때문에 서로간의 호환성이 없다.

그래서 본 제안 시스템에서는 이러한 비호환성을 극복하기 위해, 동일한 XQL 인터페이스를 통해 입력된 질의 요구를 XML Schema 언어로 기술된 공통된 개념적 수준의 내부 데이터베이스 스키마로 우선 사상시킨 다음, 이 공통 스키마를 각각의 상이한 DBMS들에 사상시킴으로써 외부에서 볼 때 일관된 사용이 가능하게 하면서도 다양한 DBMS들을 사용할 수 있게 한다.

현재, 이 분야의 기술은 특정 데이터 모델에 대해서는 안정되고 있다고 판단되기 때문에 기존의 각 데이터 모델에 적용된 개별 기술들을 효과적으로 통합하기 위하여 내부적으로 공통된 틀을 사용하려고 하는 것이다.

4. 결론 및 향후 계획

데이터베이스 통합은 세계적으로 중요한 연구 주제이며, 데이터베이스 통합을 위한 기본적인 도구로 XML이 부상하고 있다. 즉, XML은 데이터베이스 통합을 위한 기본적인 프로토콜이 될 것이 확실하다. 이미, 새로 개발되는 DBMS들은 거의 대부분 질의처리 결과를 XML 문서로 제공하는 기능을 가지고 있다. 따라서 기존의 데이터베이스는 물론이고 새로 개발되는 데이터베이스에 대하여 XML 질의어 처리, XML 문서의 저장 구조, 데이터베이스의 내용 및 질의어 처리 결과를 XML 문서로 변환하는 요구가 매우 많을 것으로 전망된다.

본 논문에서는 이러한 XML을 기반으로 이질의 데이터베이스를 인터넷 환경에서 효과적으로 통합하는 방법을 제공할 수 있는 범용 Wrapper를 개발하는 초기 단계의 연구 상태를 기술하였다. 이러한 범용 Wrapper는 XML 질의어 처리 기능, XML 문서의 저장 기능, 데이터베이스의 내용 및 질의어 처리 결과를 XML 문서로 변환하는 기능 등을 가지고 있으며, 이를 통해 이질의 데이터베이스들을

효과적으로 통합하는 방법을 실현함으로써 기존의 Wrapper 보다도 한층 더 강력한 기능을 발휘할 수 있을 것이다.

앞으로 제안된 범용 Wrapper를 구현하여 이와 같은 사항들을 실증적으로 점검할 것이다.

참고문헌

- [1] 김태현, 김경일, 이강찬, 이규철, "XML 기반의 정보통합을 위한 OODB2XML 래퍼의 설계 및 구현," '99 한국정보과학회 학술발표논문집, pp. - , 1999. 10.
- [2] <http://www.informatik.fh-wicsbaden.de/~turau/DB2XML/>.
- [3] 배종민, 김현주, 정채영, 박은경, "XML DTD로부터 관계형 테이블로의 사상 구조 설계".
- [4] 이경하, 조정수, 이강찬, 이규철, "XML 기반 분산 이질 정보 자원의 통합 프레임워크".
- [5] 이경하, 이강찬, 이규철, "XML 기반의 이질 정보의 통합 방법론".
- [6] <http://www.w3.org/TR/xmlschema-1>.
- [7] <http://www.w3.org/TR/xmlschema-2>.
- [8] <http://www.w3.org/TR/NOTE-dcd>.
- [9] <http://www.w3.org/TR/1998/NOTE-XML-data>.
- [10] <http://www.w3.org/TR/NOTE-SOX>.
- [11] <http://www.brics.dk/DSD/>.
- [12] <http://www.w3.org/TR/NOTE-ddml>.
- [13] J. McHugh, S. Abiteboul, R. Goldman, D.Quass, and J. Widom, "Lore : A Database Management System for Semistructured Data,"