

# XML 뷰 기반의 자료 저장소 래퍼 시스템의 설계 및 구현<sup>†</sup>

박진수<sup>0\*</sup> 최규원\* 정채영\* 김영균\*\* 이미영\*\* 강현석\* 배종민\*  
\*경상대학교 컴퓨터과학과 / 컴퓨터정보통신연구소 \*\*한국전자통신연구원  
{lanez04<sup>1</sup>, jinsilkw, wcdi}@rtp.gsnu.ac.kr {kimyoung, mylee}@etri.re.kr  
{hskang, jmbae}@nongae.gsnu.ac.kr

## Design and Implementation of a Storage Wrapper based on the XML View

Jin-Soo Park<sup>0\*</sup> Kyu-Won Choi\* Chae-Yeong Jeong\* Young-Kyun Kim\*\* Mi-Young Lee\*\*  
Hyun-Syug Kang\* Jong-Min Bae\*

<sup>1</sup>Dept. of Computer Science, Gyeongsang National University /  
Computer Information and Telecommunications Research Institute  
<sup>\*</sup>Electronics and Telecommunications Research Institute

### 요 약

데이터베이스 통합론 중에서 미디어이터 기반의 데이터베이스 통합 방법은 미디어이터가 가상적으로(virtual) 통합된(integrated) 뷰를 유지하도록 하고, 사용자가 통합된 뷰를 통해서 질의를 할 수 있도록 한다. 이를 위해 각 지역 데이터베이스와 직접 대화하는 자료저장소 래퍼(wrapper)가 필요하다. 본 논문에서는 지역 데이터베이스의 내용을 W3C의 표준인 XML Schema로 표현하고, 이를 바탕으로 사용자가 XML 뷰를 정의하는 방법을 제시한다 그리고 이 사용자 정의 XML 뷰 기반의 스키마 관리, 질의어 처리, 연결 관리기로 구성된 래퍼 시스템을 설계, 구현한다.

### 1. 서론

분산된 데이터베이스의 통합에 대한 필요성은 전자상거래, 디지털 라이브러리, 포털서비스, 의학 생물정보 분야 등에서 오래 전부터 제기 되어 왔다. 이는 사용자에게 일고자 하는 정보가 있는 정보원에 대한 위치 결정에 대한 부담을 덜어주고, 다수의 데이터베이스에 대하여 사용자에게 하나의 데이터베이스 관점을 제공함으로써, 연관된 데이터를 다수의 정보원으로부터 검색, 가공하여 하나의 통합된 정보를 제공할 수 있도록 한다.

데이터베이스를 통합하는 방법으로는, 각 지역데이터베이스의 모든 데이터를 하나의 데이터베이스 시스템으로 구축하는 방법(Universal DBMS), 데이터 웨어하우스, 연합형(Federated) 데이터베이스, 메타 검색엔진, 그리고 미디어이터(Mediator) 시스템 등이 있다. 이 중에서 미디어이터 시스템은 데이터베이스의 논리적 통합 방법론 중의 하나로서, 메타 검색과 유사한 기능을 가지나, 모든 종류의 데이터를 지원하면서, 지역 데이터베이스가 구조적, 반구조적, 혹은 비구조적인 데이터를 가지더라도 질의가 가능한 시스템이다.

미디어이터 시스템에서는 데이터베이스를 통합하기 위하여, 각 지역 데이터베이스의 내용에 대해 가상적으로 통합된 뷰를 미디어이터가 유지한다. 그리고 사용자는 통합된 뷰를 통해서 질의한다. 미디어이터가 가상적으로 통합된 뷰를 유지하도록 하고, 사용자가 통합된 뷰를 통해서 질의를 할 수 있도록 하기 위해서는, 각 지역 데이터베이스와 직접 대화하는 자료저장소 래퍼의 도움이 필요하다. 즉, 래퍼는 각 지역 데이터베이스마다 하나씩 존재하여, 지역 데이터베이스에 대한 내용을 미디어이터에게 전달함으로써, 미디어이터가 통합된 뷰를 구성할 수 있도록 정보를 제공하고, 미디어이터로부터 받은 사용자 질의를 지역 데이터베이스에게 전달하여 그 결과를 미디어이터에게 넘겨주는 역할을 한다.

미디어이터 시스템에서 이질적인 정보 시스템들을 통합하기 위해서는 각 지역 데이터베이스의 내용을 표현할 수 있는 하나의 중립적인 모델이 필요한데, XML 기반의 모델이 중요한 대안으로 인정받고 있다. XML 기반의 통합이라 함은, 각 지역 데이터베이스의 내용을 XML 스키마로 표현하고, 사용자 질의는 XML 질의어로 표현되며, 질의 결과는 XML 문서임을 말한다. 본 논문에서는 지역 데이터베이스의 내용을 2001년 5월에 발표된 W3C의 권고안에 따르는 XML Schema[7]로

써 표현한다. 래퍼에 대한 질의어로는 XQuery를 사용하며, 현재 W3C XMLQuery Working Group에 Working Draft로 제안되어 있는 XQuery 1.0[8]을 기반으로 하여, 데이터베이스 통합을 위한 XML 기반의 자료 저장소 래퍼를 설계하고, 관계형 데이터베이스에서 구현한 결과를 제시한다.

### 2. 관련 연구

데이터베이스를 통합하기 위한 방법론 중에서 미디어이터-래퍼를 기반으로 한 시스템으로는 XMF[4], TSIMMIS[5], MIX[6] 등이 있다. XMF는 이질적인 정보 자원들이 제공하는 데이터를 하나의 XML 문서로 통합하여 사용자에게 제공한다. XMF에서 래퍼는 미디어이터로부터 받은 질의를 로컬 시스템에 맞는 로컬 질의로 변환하며 질의어로는 XMF-QL(XMF Query Language)를 사용한다. MIX는 XML 기반의 통합 뷰를 XML DTD로 제공하며, 질의어는 자체 개발된 XMAS 질의어를 사용한다.

한편, XML 뷰 기반으로 질의어 처리가 이루어지는 시스템으로 SilkRoute[1]와 XPERANTO[2]를 들 수 있다. SilkRoute는 자체 정의한 RXL로 XML 뷰를 정의하며, 사용자 질의어인 XML-QL이 RXL과 합성되어 새로운 RXL이 생성된다. 합성된 RXL은 관계형 데이터베이스로부터 데이터를 검색하기 위해 SQL로 변환되어진다. XPERANTO는 관계형 데이터베이스 스키마를 XML Schema로 표현하며, 사용자 입장에서의 XML 뷰를 정의하기 위해 XQuery를 사용한다. 사용자 질의 또한 XQuery이며, 각각의 XQuery는 중간 모델인 XQGM으로 변환되어 합성하게 된다.

### 3. 래퍼 설계 방향

각 지역 데이터베이스들에서 관리되는 데이터에 대하여 모델과는 상관없는 XML 기반의 범용 래퍼를 개발하기 위해서는, 각 지역 데이터베이스의 내용을 하나의 통일된 관점으로 표현하여 관리하고, 각 지역 데이터베이스가 처리할 수 있는 질의어에 맞추도록 사용자 질의어를 번역해야 한다.

먼저, 각 지역 데이터베이스의 내용을 하나의 통일된 관점으로 표현하기 위하여 지역 데이터베이스 내용을 W3C에서 제안한 XML Schema로 표현한다. 이 때, 데이터베이스 질의어 SQL에서 사용자가 정의한 뷰 테이블을 생성하는 기능과 유사한 개념으로, XML 스키마에 대해서 사용자 정의 XML 뷰를 정의하는 방법[1, 2, 3]이 연구된 바 있

<sup>†</sup> 본 연구는 한국전자통신연구원의 위탁 연구 과제인 자료 저장소 wrapper 개발의 일부로 수행된 결과임

는데, 이를 활용하여 사용자는 임의로 정의된 XML 뷰 상에서 질의를 할 수 있도록 한다. 사용자 뷰 정의 언어는 별도로 정의하지 않고, W3C에서 제안된 XML 질의어인 XQuery를 그대로 사용한다.

다음으로, 사용자 질의어 역시 W3C에서 제안된 XML 질의어 XQuery를 사용한다. XQuery로 표현된 질의어는 XML 뷰와 합성되어야 하는데, 이 합성 알고리즘을 개발하고, 합성된 XQuery 질의어를 각 지역 데이터베이스 질의어로 번역하는 질의어 번역기를 개발한다. 그리고 질의어 번역 과정에서 생성되는 XML 템플릿을 기반으로, 지역 데이터베이스로부터 수행된 질의 결과를 XML 문서로 변환하여 사용자에게 전달될 수 있도록 한다.

마지막으로, 미디어이터가 쿼리의 상세한 기능을 알지 못하더라도 그 기능을 사용할 수 있도록 응용 프로그램 인터페이스가 제공되도록 한다.

#### 4. 시스템 구성

그림 1은 설계된 웹 시스템의 구조를 나타낸다. 웹퍼는 각 지역 데이터베이스마다 하나씩 존재하는데, 크게 스키마 관리자(Schema manager), 질의어 관리자(Query manager), 그리고 연결 관리자(Connection manager)로 구성된다.

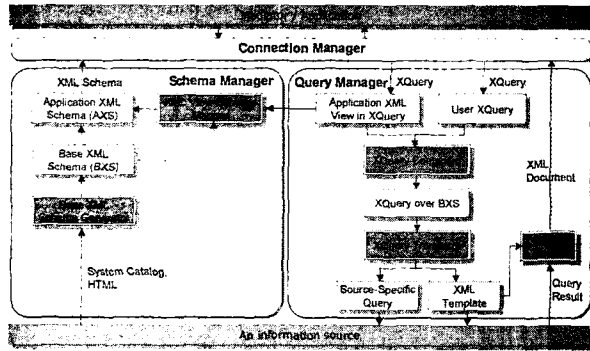


그림 1. 웹퍼 시스템의 구조

#### 4.1 스키마 관리자

스키마 관리기는 지역 데이터베이스 내용에 대한 XML 뷰를 생성, 정의, 등록, 삭제 등의 기능을 담당한다. 기본 XML 스키마 생성기(Base XML Schema Generator)는 지역 데이터베이스의 스키마를 XML 스키마로 자동 생성하는 컴포넌트로서, 지역 데이터베이스의 스키마와 가장 근접한 형태의 XML 스키마를 생성한다. 이 때 생성된 XML 스키마를 기본 XML 스키마(Base XML Schema)라 한다. 기본 XML 스키마를 바탕으로 응용 XML 스키마(Application XML Schema)를 정의할 수 있다. 응용 XML 스키마란, 기본 XML 스키마를 바탕으로 사용자의 관점에서 정의되는 XML 스키마를 말한다. 이는 관계형 데이터베이스 모델에서 뷰 테이블과 유사한 개념이다.

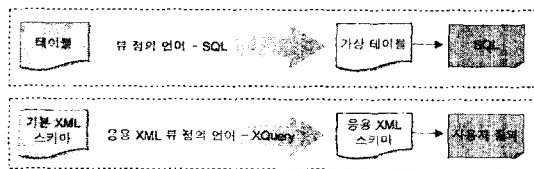


그림 2. 데이터베이스 스키마와 XML 스키마의 관계

응용 XML 스키마는 웹퍼 관리자가 미리 정한 일정한 규칙에 따라서 자동으로 생성, 관리되어서 사용자에게 제공될 수도 있고, 웹퍼 사용자에 의해서 임의로 정의될 수도 있다. 그런데 사용자가 기본 XML 스키마로부터 응용 XML 스키마를 생성하는 것은 간단한 일이 아니다. 따라서 응용 XML 스키마를 쉽게 생성할 수 있는 방법이 제공될 필요가 있다. 본 논문에서 제시된 웹퍼 시스템은 웹퍼 사용자에게 기본 XML 스키마로부터 사용자 정의 XML 뷰를 정의할 수 있는 기능을 제

공한다. 이를 응용 XML 뷰(Application XML view)라 정의한다. 이 때, 사용자 정의 XML 뷰를 정의하는 데 사용되는 언어는 XQuery이다. 그림 2는 데이터베이스 스키마와 XML 스키마의 관계를 나타낸 것이다. 사용자가 XQuery를 사용하여 XML 뷰를 정의하면, XML 뷰-스키마 사상기가 자동으로 응용 XML 스키마를 생성한다. 이 때 기본 XML 뷰와 기본 XML 스키마를 이용하게 되는데, 기본 XML 뷰는 실제 데이터베이스와 기본 XML 스키마상의 대응 관계를 나타내기 위해 자동적으로 생성된 XML 뷰이다. 미디어이터는 통합 스키마를 구축하기 위하여 웹퍼가 제공하는 모든 형태의 스키마를 얻을 수 있으며, 생성된 스키마나 뷰를 등록, 삭제, 변경할 수 있다.

#### 4.2 질의어 관리자

질의어 관리기는 XQuery 합성기(XQuery Composer), XQuery 변환기(XQuery Translator), 그리고 XML 태거(XML Tagger) 세 개의 컴포넌트로 구성된다. XQuery 합성기는 XQuery로 정의된 응용 XML 뷰와 사용자 질의어로 들어온 XQuery를 합성하는 컴포넌트이다. 합성된 결과는 기본 XML 스키마 상에서의 XQuery로 된다. XQuery 변환기는 합성된 결과 새롭게 만들어진 XQuery를 해당 지역 데이터베이스의 질의어로 변환하고 동시에 XML 문서로 변환하기 위한 템플릿을 생성하는 컴포넌트이다. XML 태거는 데이터베이스 검색 결과를 XML 문서로 변환하는 요소이다.

XQuery 질의어를 SQL로 변환하고자 할 때, 먼저 어떤 테이블에서 무엇을 검색할 것인가를 알 수 있어야 하고, 어떤 조건을 가지고 검색 하는지에 대한 정보를 추출하여 SQL로 변환해야 한다. 하지만 사용자 질의어만으로는 이와 같은 정보를 얻을 수가 없다. 왜냐하면 사용자 질의어는 실제 데이터베이스가 아닌 응용 XML 스키마를 보고 만들어진 질의어이기 때문이다. 즉, XQuery 질의어에서 엘리먼트는 사용자가 원하는 관점으로 정의한 뷰에 나타난 엘리먼트 이름이지, 실제 데이터베이스에 있는 테이블 이름이나 컬럼 이름을 가리키는 것은 아니다.

기본 XML 스키마에는 실제 데이터베이스에 대한 정보가 포함되어 있고, 사용자 정의 뷰는 기본 XML 스키마의 경로를 사용하여 정의된다. 사용자 정의 뷰를 바탕으로 만들어지므로 그림 3에서 보는 바와 같이 사용자 질의어와 사용자 정의 뷰를 합성함으로써, 사용자 질의어가 기본 XML 스키마에서 제공되는 정보를 포함하도록 변환될 수 있다. 이렇게 변환된 질의어는 실제 데이터베이스 질의어인 SQL로 쉽게 변환할 수 있다.

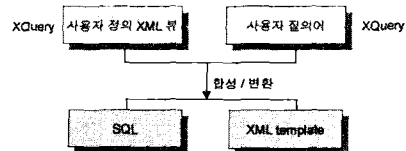


그림 3. XQuery의 합성과 변환

본 연구에서는 그림 4에서와 같이 질의 처리를 위해서, 사용자 정의 XML 뷰에 대해서는 뷰 트리, 사용자 질의어에 대해서는 질의 트리를 구성한다.

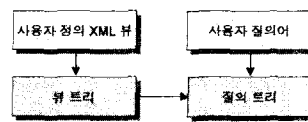


그림 4. 뷰 트리과 질의 트리

사용자 정의 XML 뷰를 트리로 표현함으로써, 사용자 질의어에 내재된 구조적인 정보와 사용자 질의어에 나타나는 임의의 경로식에 적용되어야 하는 조건 정보를 얻을 수 있다. 또한 실제 데이터베이스의 테이블 명과 컬럼 명을 가지고 있기 때문에 SQL 문장 생성을 위한 정보를 얻을 수 있다.

그림 5는 임의의 사용자 정의 XML 뷰에 대한 뷰 트리를 나타낸 것이다. 하나의 완전한 뷰 트리를 구성하기 위해서 모든 노드의 루트가 되는 가상 루트 노드가 추가된다. 그리고 XML 문서의 계층 구조처럼 뷰 트리 역시 엘리먼트의 계층 구조에 따라서 트리가 구성된다. 사용자

질의어에서 나타나는 엘리먼트에 대하여 질의 트리상에 하나의 노드가 생성된다. 엘리먼트의 값으로 나타나는 경로식에 대해서는 SQL 문장의 일부가 생성되면서 동시에 뷰 트리를 참조하여 질의 트리를 기본 XML 스키마상의 구조로 합성시켜서 변환한다. 이는 질의 트리가 XML 문서 생성을 위한 틀이 되는 바탕이 된다. 질의 트리과 뷰 트리를 이용한 번역 과정은 지면 관계상 생략한다.

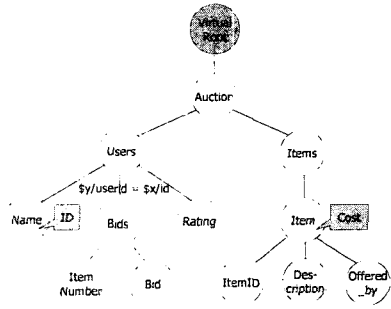


그림 5. 사용자 정의 뷰에 대한 뷰 트리의 예

질의 트리가 완성되면 XML 태거가 실행된다. XML 태거는 질의 트리를 순회하면서 엘리먼트 노드를 만나면 엘리먼트를 구성하고, SQL 질의어가 통합된 노드를 만나면 해당 질의어를 수행시키고, 그 결과를 반환된 결과 집합의 크기만큼 트리의 하위 노드를 순회하면서 XML 문서를 구성하게 된다. XML 문서 결과가 많은 것을 대비하여, XML 문서에 대하여도 데이터베이스 커서 개념을 도입하는데, 본 연구에서는 XQuery에서 최상위 FLWR 질 단위로 커서를 정의 구현하였다.

#### 4.3 연결 관리기

연결 관리기는 웹퍼에 대한 연결 및 연결 정보를 제공하고, 웹퍼의 기능을 사용할 수 있도록 스키마 관리기와 질의어 관리기에 대한 인터페이스를 제공한다.

웹퍼는 자료저장소가 있는 곳에 존재하는 로컬 웹퍼와 미디어터가 있는 머신상에 존재하는 웹퍼 프락시(proxy)로 나누어진다. 웹퍼 프락시는 단순히 웹퍼가 제공하는 인터페이스를 호출하는 형태로 구현되고, 이는 자바 RMI를 통해서 로컬 웹퍼와 연동된다.

#### 5. 구현

데이터베이스 통합을 위한 범용 웹퍼의 설계에 대하여 현재, 관계형 데이터베이스에 대하여만 구현되었다. 그림 6은 미디어터 틀을 실행시킨 모습이다. 사용자는 왼쪽 창에 나타난 기본 XML 스키마를 보고 오른쪽 창을 통해 새로운 응용 뷰를 생성하고 등록한다. 또한 사용자는 등록되어 있는 응용 뷰에 대한 수정이나 삭제가 가능하다.

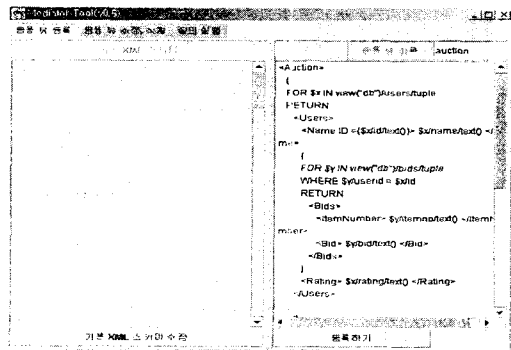


그림 6. 기본 XML 스키마와 응용 뷰 등록

그림 7과 같이 사용자는 등록한 응용 뷰를 통해 생성된 응용 XML

스키마상에서 사용자 질의를 생성하여 질의 결과를 확인할 수 있다. 왼쪽 상단 창에 사용자가 원하는 응용 XML 스키마를 선택하면 이에 해당하는 스키마가 화면에 출력되고, 사용자는 이를 기반으로 질의를 생성한다. 사용자 질의는 관계형 데이터베이스에 접근할 수 있는 SQL로 변환되어 수행되며, 검색된 결과는 XML 형태로 반환되어, 해당 결과가 오른쪽 창에 나타나게 된다. 또한 질의 결과는 관계형 데이터베이스로부터 검색된 모든 내용을 하나의 XML 문서로서 볼 수도 있으며, 질의 결과의 일부분을 차례대로 하나씩 가져와서 볼 수도 있다.

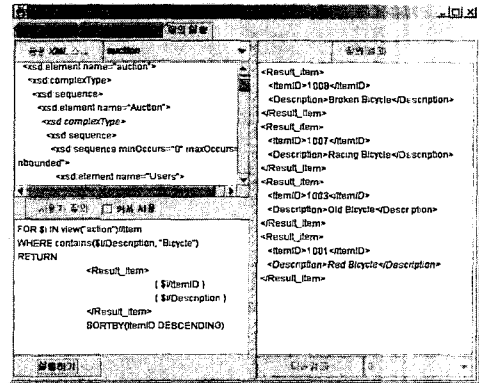


그림 7. 질의어의 생성과 수행 결과

본 도구의 구현에 사용된 언어는 Java이며, 데이터베이스는 Microsoft SQL SERVER 2000을 JDBC는 SQL Server 2000 for JDBC(Version 2.1)를 사용하였다.

#### 6. 결론 및 향후 과제

데이터베이스 통합은 여러 응용 분야에서 필요로 하는 기술이다. 데이터베이스 통합 방법론 중에서, 미디어터 기반의 데이터베이스 통합 방법론은 가장 강력한 기능을 가지고 있다고 할 수 있다. 개발된 웹퍼 시스템은 미디어터가 필요로 하는 일에 대하여 서비스를 제공하는 시스템이다. 본 논문에서는 데이터베이스 통합을 위한 XML 뷰 기반의 범용 웹퍼 시스템의 설계와 관계형 데이터베이스 모델에 대한 웹퍼의 구현을 제시하였다.

XQuery 질의어로부터 SQL로 번역한 결과는 여러 가지 형태로 나올 수 있는데, 그 결과에 따라서 질의 처리의 효율성에 큰 영향을 미치므로, 앞으로 질의어 최적화에 대한 연구가 더 이루어져야 한다.

#### 참고문헌

- [1] M. Fernandez, W. Tan, and D. Suciu, "SilkRoute : Trading between Relations and XML", WWW9, pp. 723-745, 2000.
- [2] M. Carey, D. Florescu, Z. Ives, Y. Lu, J. Shanmugasundaram, E. Shekita, and S. Subramanian, "XPERANTO: Publishing Object-Relational Data as XML", Workshop on the Web and Databases (WebDB), pp. 105-110, May 2000.
- [3] J. Shanmugasundaram, J. Kiernan, E. Shekita, C. Fan, and J. Funderburk, "Querying XML Views of Relational Data", VLDB Conference, pp. 261-270, 2001.
- [4] 이경하, 이상찬, 이규철, "XML 기반의 이질 정보의 통합 방법론", 한국정보과학회, 가을학술발표논문집(I) pp. 96-98, 1999
- [5] Y. Papakonstantinou, A. Gupta, H. Garcia-Molina, and J. D. Ullman, "A Query Translation Scheme for Rapid Implementation of Wrappers", DOOD, pp. 319-344, 1995
- [6] C. Baru, A. Gupta, B. Lud ascher, R. Marciano, Y. Papakonstantinou, P. Velikhov, and V. Chu. "XML based Information Mediation with MIX", SIGMOD, pp. 597-599, 1999.
- [7] W3C, "XML Schema", <http://www.w3.org/0XML/schema>, W3C Recommendation, May 2001.
- [8] W3C, "XQuery 1.0: An XML Query Language", <http://www.w3.org/TR/xquery/>, W3C Working Draft, 2001.