

객체지향 데이터베이스의 XML로의 표현

윤정희^o 박창원 정진완
한국과학기술원 전자전산학과 전산학전공
{cotoy, cwpark, chungcw}@islab.kaist.ac.kr

Representing Object-Oriented Databases in XML

Jung-Hee Yun^o Chang-Won Park Chin-Wan Chung
Division of Computer Science,
Department of Electrical Engineering & Computer Science, KAIST

요 약

데이터 변환은 데이터 재사용, 데이터 교환 및 통합 등에 사용되는 중요한 기술이다. 본 논문에서는 이러한 데이터 변환 기술 중 객체지향 데이터베이스에 저장된 데이터를 XML 문서로 변환시키는 기술을 제시한다. 먼저 객체지향 데이터베이스의 스키마와 데이터, DTD와 XML 문서를 정의한 뒤 이 정의를 기반으로 스키마를 DTD로 변환하는 알고리즘과 데이터를 XML 문서로 변환하는 알고리즘을 제시한다. 그리고 변환에 의한 결과 XML 문서가 잘 구성된 문서이고 결과 DTD에 대해서 유효한 XML 문서임을 보임으로써 제시한 변환 기술의 유용성을 보인다.

1. 서론

최근 웹은 많은 데이터를 얻을 수 있는 곳으로 다양한 사용자들에게 점점 친숙해지고 있으며 그 규모도 날이 증가하고 있다. 그런데 웹 상에서 데이터를 얻을 수 있을 뿐만 아니라 웹을 이용해서 데이터를 표현할 수도 있다. 따라서 특정 데이터 소스에 보유하고 있는 데이터를 웹 상에서 볼 수 있는 데이터로 표현할 수 있는 기술이 필요하며 이와 관련하여 여러 데이터 소스들에 대한 래퍼(wrapper)와 중재기(mediator), 그리고 질의 언어를 제공하여 데이터를 HTML 형식으로 제공하는 기술을 제시한 연구 결과가 있었다[1].

그러나 현재는 웹 데이터들이 HTML 형식으로 표현되고 있지만 데이터 교환 및 통합을 쉽게 하고자하는 목적에서 XML(eXtensible Markup Language)[2]이 새롭게 등장하였고 차세대 인터넷 문서 표준으로 부각되고 있다. 이러한 배경에서 HTML 형식의 데이터들을 XML 형식으로 변환시켜서 제공하는 것과 관련된 연구도 이루어졌다[3].

XML에서의 태그는 데이터 자체를 표현하는 것으로 XML 문서에 대한 질의어가 가능하다. 따라서 웹 상에서 제공되는 XML 문서에 대해 데이터베이스 시스템과 관련된 기술을 적용시키려는 노력이 있어 왔다[4].

또 이와는 반대로 데이터베이스 시스템에 저장된 방대한 데이터를 어떻게 웹 상의 데이터로 표현할 것인가에 대한 연구도 필요하다. 이러한 배경에서 본 논문에서는

데이터 변환 기술의 하나로 객체지향 데이터베이스(OODB)에 저장된 데이터를 XML 문서로 변환시키는 기술을 제안하고자 한다. 즉 OODB에 저장된 데이터를 XML 문서로 제공하여 많은 사용자들이 접근할 수 있고 또 다른 목적에서 재사용 할 수 있도록 하는 기반을 마련하는 것이다.

그리고 변환에 의한 결과 XML 문서가 잘 구성된(well-formed) XML 문서이고 스키마를 변환한 결과 DTD에 대해서 유효한(valid) XML 문서임을 보임으로써 변환 기술의 유용성을 제시하겠다.

2. 관련연구

대부분의 중재기 시스템에서 이질적인 데이터들간의 교환과 통합 기술을 요구하는데 이때 데이터 변환 기술이 필요하다. 이러한 중재기 시스템으로 STRUDEL 시스템[5]과 MIX(Mediation of Information using XML) 중재기 시스템[6]을 들 수 있다. MIX 중재기 시스템의 경우 데이터 모델로 XML을 사용하여 웹 전체를 커다란 분산 데이터베이스 시스템으로 보기 때문에 여러 데이터 소스들과 XML 간의 래퍼가 필요하다. 즉 여러 가지 데이터 포맷들과 XML 간의 데이터 변환 기술이 사용되는 것이다.

임의의 데이터 포맷간의 데이터 변환이 가능하도록 하는 목적에서 만들어진 시스템으로 YAT 시스템[7]이 있는데 트리 형태의 YAT 모델을 정의하고 변환 언어 YATL(YAT Language)를 사용한다. 이질적인 데이터들의 통합을 필요로 하는 어플리케이션이 증가하면서 데이터

이 연구는 과학기술부의 핵심 S/W 기술개발사업에서 지원받았음.

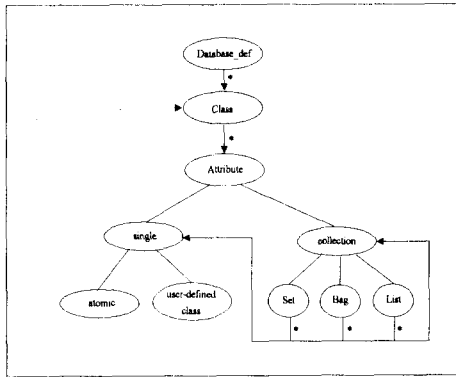
변환 기술의 필요성이 높아졌는데 기존의 데이터 변환 기술이 재사용이 불가능한 경우가 대부분이라는 것에 착안하여 YAT 시스템의 경우는 YAT 데이터 모델에 대한 래퍼가 구현된 모든 이질적인 데이터 포맷간의 데이터 변환이 가능하도록 하였다.

본 논문에서의 연구와 가장 유사한 연구로 관계형 데이터베이스(RDB)에 저장된 데이터를 XML 문서로 변환하는 기술에 대한 연구를 들 수 있는데 이 분야에 대한 연구 결과들은 많으며 그 중 한 가지로 DB2XML[8]이라는 것이 있다. 현재 웹 페이지들 중 많은 부분이 기존의 데이터베이스에 저장된 데이터로부터 생성되었다는 점에 착안하여 RDB에 저장된 데이터를 XML 문서로 자동 변환 시켜주는 툴을 만든 것이다. DB2XML은 RDB에 저장된 데이터에 대해서 XML 문서와 DTD를 제공하는데 관계형 모델의 각 구성 요소에 대한 매핑(mapping) 규칙을 제시하여 그 규칙에 의한 데이터 변환을 제시하였다.

3. 객체지향 데이터베이스의 XML로의 변환 기법

3.1 OODB의 스키마, 데이터, DTD, XML 문서의 정의

OODB의 스키마 정보는 데이터베이스 이름과 여러 클래스들로 구성되며 각 클래스는 애트리뷰트들로 구성되고 애트리뷰트는 자신의 타입을 정보로 가진다. 이러한 OODB의 스키마 구조를 <그림 1>과 같이 표현할 수 있다.



<그림 1> OODB의 스키마 구조도

OODB에 저장된 데이터는 객체들의 모임으로 볼 수 있는데 각 객체는 자신이 속한 유일한 클래스가 있다.

이러한 특징의 OODB 스키마와 데이터를 <표 1>과 <표 2>로 정의하였다.

```

database_def:=(name, class;...;class)
class:=(name:sname, attribute;...;attribute)
attribute:=(name, single)|(name, collection)
single:=atomic | user_defined
collection:= set(single) | set(collection)
user_defined:= name
    
```

<표 1> OODB의 스키마 정의

```

database:=(class_extnt;...;class_extnt)
class_extnt:=(object;...;object)
object:=(oid, attr_value;...;attr_value)
attr_value:= single_value | collection_value
    
```

<표 2> OODB의 데이터 정의

본 논문에서 제시하는 데이터 변환 기법에 의한 결과 XML 문서는 다음과 같은 조건을 만족한다.

- 속성은 ID, HREF 속성만 포함한다.
- 엘리먼트의 내용으로 데이터와 다른 엘리먼트의 혼합이 되는 경우는 없다.
- 엔터티(entity)는 포함하지 않는다.

<표 3>과 <표 4>는 DTD와 XML 문서를 정의한 것이다.

```

DTD := {<n : attr(n) : type(n)>}
n ∈ NAME
attr(n) ∈ {ID, HREF, null}
type(n) : regular expression over NAME
           or PCDATA or EMPTY
    
```

<표 3> DTD 정의

```

XML_DOC:=element
element:=start_tag(name):attribute
          <content> end_tag
attribute:=null | (attr_name : data)
content:=empty | pcdat | element;...;element
    
```

<표 4> XML 문서 정의

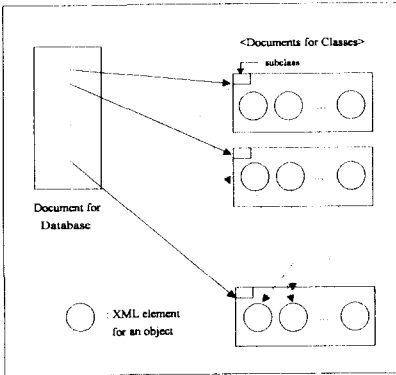
3.2 변환 알고리즘

OODB의 XML 문서로의 변환은 크게 스키마를 DTD로 변환하는 부분과 데이터를 XML 문서로 변환하는 두 부분으로 나뉘어진다. 데이터 변환에서 가장 먼저 고려할 부분으로 어떤 단위로 하나의 XML 문서를 만들 것인가의 문제인데 전체 데이터베이스, 클래스별, 객체별 등을 생각할 수 있는데 여기서는 클래스별로 하나의 XML 문서를 만드는 방법을 사용한다. 그리고 클래스별로 만든 XML 문서들에 대한 링크를 모두 모아서 하나의 데이터베이스를 표현하는 XML 문서로 나타내도록 하였다. <그림 2>는 결과 XML 문서의 구조를 표현한 것이다.

스키마 정보를 DTD로 변환하는 알고리즘은 3.1에서 정의한 database_def 타입을 입력으로 하고 DTD 타입을 결과로 한다. 먼저 전체 데이터베이스를 표현하는 DTD를 만들고 각 클래스에 대해서 해당하는 DTD를 만들어서 포함시킨다. 그리고 애트리뷰트에 대한 DTD를 포함시킨다.

데이터를 XML 문서로 변환하는 알고리즘은 database 타입을 입력으로 하고 XML_DOC 타입의 집합을 결과로 한

다. 알고리즘의 전체 구조는 스키마를 DTD로 변환하는 알고리즘과 유사하다.



<그림 2> 결과 XML 문서간의 관계

결과 XML 문서간의 관계 그림에서 알 수 있듯이 변환 알고리즘은 데이터베이스로부터 시작하여 클래스, 객체, 애트리뷰트 순서의 하위요들의 재귀적 호출로 구성된다. 전체적인 변환 알고리즘의 구조는 <그림 1>에서 표현된 OODB 스키마 구조도와 유사하다. 그리고 <표 5>에서 간단한 OODB 스키마의 DTD로의 변환예를 보여준다.

```

class PERSON          <!ELEMENT PERSON (NAME, AGE)>
{
  <!ATTLIST PERSON ID ID (#REQUIRED)>
  char NAME[30];      <!ELEMENT NAME (#PCDATA)>
  short AGE;         <!ELEMENT AGE (#PCDATA)>
}
    
```

<표 5> OODB 스키마의 DTD로의 변환 예

3.3 잘 구성된 XML 문서와 유효한 XML 문서

잘 구성된 XML 문서라는 것은 결과 XML 문서의 구조가 올바르다는 것을 나타낸다. 앞에서 정의한 XML_DOC 타입이 잘 구성된 XML 문서를 나타내는 것이므로 변환 알고리즘의 결과가 XML_DOC 타입임을 보이는 것으로 결과가 항상 잘 구성된 XML 문서가 됨을 증명할 수 있다.

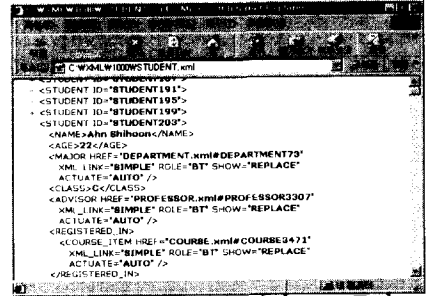
스키마가 주어지고 그 스키마를 만족하는 데이터가 주어졌을 때 스키마를 변환시킨 DTD에 대해서 데이터를 변환시킨 결과 XML 문서는 항상 유효한 문서가 됨을 증명할 수 있다. 결과가 유효한 XML 문서가 되므로 유효한 XML 문서를 필요로 하는 어플리케이션에 결과 문서를 제공할 수 있게 된다.

3.4 실험 결과

변환 알고리즘을 구현하여 OODB에 저장된 데이터를 XML 문서로 변환시킨 결과를 인터넷 익스플로러 5.0으로 확인한 결과를 <그림 3>에서 보여준다.

4. 결론

본 논문에서는 데이터 변환 기법의 하나로 객체지향



<그림 3> IE로 본 결과 XML 문서

데이터베이스에 저장된 데이터를 XML 문서로 변환하는 기법을 제안하였다. OODB 스키마, 데이터, DTD, XML 문서를 정의하고 그 정의에 기반 하여 스키마를 DTD로 변환하는 알고리즘과 데이터를 XML 문서로 변환하는 알고리즘을 만들었다. 그리고 결과 XML 문서가 잘 구성된 문서이고 결과 DTD에 대하여 유효한 XML 문서임을 보였다. 본 논문에서 제시한 데이터 변환 기술은 나아가 다른 포맷들에 대한 데이터 변환 기술의 개발이나 XML 문서를 OODB 시스템의 구조에 맞게 변환, 저장하는 기술 개발 등의 바탕이 될 것이다.

5. 참고 문헌

- [1]M. Fernandez, D. Florescu, A. Levy, and D. Suciu. A Query Language for a Web-Site Management System. In SIGMOD Record, Sep.1997.
- [2]World Wide Web Consortium(W3C). Extensible Markup Language(XML) 1.0, 1998. <http://www.w3.org/TR/1998/REC-xml-19980210>.
- [3]A. Sahuguet and F. Azavant. Looking at the Web through XML glasses. In International Conference on Cooperative Information Systems, 1999.
- [4]Jayavel Shanmugasundaram, Kristin Tufte, Gang He, Chun Zhang, David DeWitt, and Jffrey Naughton. Relational Databases for Querying XML Documents : Limitations and Opportunities. In Proceedings of the 25th VLDB Conference, 1999.
- [5]M. Fernandez, D. Florescu, A. Levy, and D. Suciu. A Query Language for a Web-Site Management System. In SIGMOD Record, September 1997.
- [6]Y. Papakonstantinou and P. Velikhov. Enhancing Semistructured Data Mediators with Document Type Definitions. In IEEE International Conference on Data Engineering, 1999.
- [7]S. Cluet, C. Delobel, J. Simeon, and K. Smaga. Your mediators need data conversion!. In Proceedings of ACM SIGMOD Conference, 1998.
- [8] DB2XML home page : <http://www.informatik.fh-wiesbaden.de/~turau/DB2XML/index.html>.