

다형구조 검색질의를 지원하는 XML데이터 시스템의 설계 및 개발

신주열, 변정용
동국대학교 전자계산학과

e-mail: rome@wonhyo.dongguk.ac.kr

A Design and Development of XML Data System for Hybrid Retrieval

JooYoul Shin, JeongYong Byun
Dept of Computer Science, Dongguk University

요 약

XML 데이터 검색 질의처리를 위해서 개개의 데이터베이스 시스템은 여러 형태로 대안을 제시하고 있다. 그런데 데이터베이스관리 시스템별로 XML 질의처리방식과 함수처리 방식이 달라서 다형 구조의 데이터를 통합하여 검색하는데 어려움이 있으며 여기에 관계형 데이터베이스시스템과 XML 스키마 결합의 불일치 문제점도 해결 과제로 남아 있다.본 논문에서는 이런 문제점을 해결하는 방안으로 XML 데이터 검색 질의처리 시스템을 제안한다. 또한 다형적 구조의 환경에서 XML 데이터 시스템과 구조적인 데이터 시스템의 연계방안과 검색질의의 시스템에 관해서도 논의한다.

1. 서론

최근 XML을 지원하기 위한 시스템 구축에 관한 논의가 활발히 일어나고 있다. XML은 인터넷을 통해 구조적인 데이터를 교환하는 표준이 되고 있다. 대형 전자상거래 사이트와 같은 경우 다수 XML 데이터를 검색하는 기능이 요구된다.

관계형 데이터베이스 시스템에서는 XML질의를 지원하기 위해 XML문서 엘리먼트데이터를 관계 테이블의 컬럼에 매핑하여[1] 컬럼정보에 기초해 질의하며, Lore[2]과 같은 XML 전용 데이터베이스 시스템은 XML데이터만을 질의하는데 최적화 방안이 제시되고 있다.

근래들어 XML데이터베이스 시스템과 별도로 기존의 구조적 데이터베이스시스템 환경과 XML데이터처리를 위한 시스템을 통합, 관리하기 위한 시스템에 대한 필요성이 제기된다. 전자상거래 사이트의 경우 다형적이고 이질적인 데이터 서버의 데이터를 통합하고 질의할 필요가 있다. 특히 XML 데이터와 구조적인 데이터를 통합적으로 검색 질의하기 위한

방안이 제시 되어야 한다. 다형적인 시스템을 통합하는 시스템은 구조적인 데이터 처리시스템과 XML 데이터 처리시스템들 사이에서 중립적 데이터 처리 및 질의수행이 가능하며 XML스키마 처리도 자연스럽게 할 수 있다는 장점이 있다. 본 논문에서는 이에 다형적인 구조의 데이터 처리 환경에서 구조적인 데이터와 XML 데이터를 검색 질의하기 위한 방안을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 XML데이터 처리 방식에 관해 논하며 3장에서는 XML데이터 서버 시스템의 전체 구조와 구현내용에 관해서 설명한다. 4장에서 결론을 맺는다.

2. 관련 연구

XML데이터의 질의처리를 위해서 관계형 데이터베이스 시스템의 경우 SQL의 뷰 개념을 사용해서 처리한다. 그런데 뷰 개념으로 구조적인 데이터와 XML 데이터를 처리할 때 XML 데이터의 완전한 통합이 어렵고, 구조정의가 용이하지 못하다. 또한

문서의 엘리먼트의 구조가 변경 될 때마다 관계 테이블 구조를 재구성해야 하는 문제점이 있다. 상용의 관계형 데이터베이스 시스템의 경우 XML 데이터와 구조적 데이터를 함께 질의하기 위한 방안을 제시하고 있다.[3][4] 그러나 다형적이고 상이한 시스템이 섞여있는 환경에서는 XML질의를 중립적으로 처리하는 것이 쉽지 않다. 해당 관계형 시스템을 유지하는 상태에서 XML스키마를 이에 적용할 때 시스템마다 상이한 XML 질의 처리어 구조로 구성되어 있기 때문에 시스템간 호환이 되지 않는 단점이 있다.

3. 다형 구조 검색질의의 시스템의 설계 및 구현

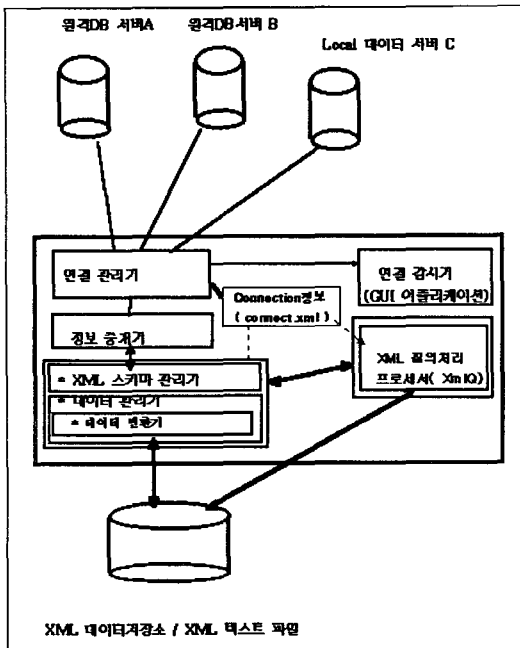


그림 1. 다형적 구조환경에서 검색질의를 지원하는 시스템

다형적 구조의 환경에서 검색질의를 지원하는 시스템의 경우 해당 XML스키마에 기반한 질의뿐 아니라 관계 스키마나 다른 스키마를 지원하는 시스템과 XML데이터를 통합 형태로 중립적 질의를 수행할 수 있어야 한다. XML 데이터를 질의와 데이터를 관리하는 프로그램, 외부 데이터와 연결하기 위한 연결관리 프로그램, XML 데이터 변환기 프로그램 등이 필요하다. 이에 따라 구성한 XML 데이터 서버의 구조는 그림1과 같다.

3.1 연결 관리기

구조적인 데이터와 반구조적 데이터를 얻기 위해 해당 데이터베이스 서버로 연결하는 작업을 관리하게 된다. 구조적 데이터의 경우 관계형 데이터베이스 시스템은 JDBC드라이버를 사용하며 객체지향형 데이터베이스 시스템의 경우 해당 데이터베이스 시스템별로 자바 데이터베이스 드라이버를 사용하는 것으로 가정한다. 관계형 데이터베이스 시스템의 JDBC 연결 루틴과 객체지향 데이터베이스 시스템은 상이한 메소드를 사용하기 때문에 이들을 통합 관리하는 상위 래퍼가 필요하다. 본 시스템의 경우 JDBC 연결루틴과 유사한 형태로 래퍼를 구성하였다. 객체지향형 데이터베이스 시스템의 접속 메소드 정의는 ODMG 홈페이지[5]를 기준으로 래퍼를 구성할 수 있다.

3.2 정보 중계기

구조적인 데이터를 XML 데이터로 매핑하기 위해 필요하게 된다. 관계형 데이터베이스 시스템의 경우에 테이블들에 대한 정보, 컬럼 정보, 데이터베이스 정보를 추출하기 위한 영역(변수)을 할당하게 되는 데 아래 소스는 실제 모듈의 루틴이다

```
public void DbParse(ResultSet rs,String tablename)
throws SAXException, SQLException, IOException
{
    ResultSetMetaData rsmd=rs.getMetaData();
    int numCols= rsmd.getColumnCount();
    while(rs.next()){
        // SAX Document Handler 저장
        for(int i=0;i< numCols ; i++){
            // rs.getXXX메소드로 받은 값을 다음 메소드로 넘김
            Columns_to_SAX_이벤트화(rsmd, rs, i );
        }
    }
    protected void Columns_to_SAX_이벤트화
    (ResultSetMetaData rsmd, ResultSet rs,
    int columnIndex)
    { String columValue=rs.getString(columnindex);
    char[] elementName=columValue.toCharArray();
    ...
    _document.startElement(...);
    }
}
```

3.3 XML 스키마 관리기 / 데이터 관리기

XML 스키마 관리기는 XML 데이터에 관한 DTD와 스키마를 가지는 모듈이며, 데이터 관리기는 XML 질의처리가 요청하는 질의 데이터를 처리한다. 즉 XML형태로 변환 가능한 외부 데이터 서버에서 정보를 받거나 함께 있는 데이터 변환기로 데이터를 보내서 XML 저장소의 데이터를 외부 데이터로 변환한다. 구조적인 데이터 시스템의 데이터나 XML텍스트 데이터를 엘리먼트와 애트리뷰트로 저장하는데 이를 위해서 JDOM파서와 Xerces를 사용해서 SAX와 DOM 파싱을 병행했다. JDOM파서는 엘리먼트를 DOM객체의 트리로 저장하는데, 일반적으로 DOM파서가 객체를 트리화 할 때 메모리 사용의 효율을 개선한 특징이 있다.

3.4 데이터 변환기

데이터 변환기는 XML 텍스트 파일형태로 저장된 데이터를 구조적인 데이터로 변환하거나, XML 데이터베이스를 XML 문서로 변환하는 작업을 한다.

3.5 연결 정보

다형적인 환경에서 XML을 처리하기위한 원거리 데이터 시스템의 기본적인 정보를 XML형태의 정보 파일로 저장하게 되고 여기에는 데이터베이스 시스템 연결에 관한 기본적인 정보와 데이터베이스 시스템의 테이블 정보 등이 담기게 된다.질의처리프로세서와 데이터관리기 프로그램과 XML 스키마 관리프로그램은 이를 참고하게 된다.

```
<?xml version="1.0" ?>
<db type="rdb" connect="remote">
  <alias>oracle_test</alias>
  <uri>jdbc:oracle:thin:dblab/dblab2@sunduk.dongguk.ac.kr:1521:EDU</uri>
  <user>dblab</user>
  <passwd>dblab2</passwd>
  <table>...</table>
  <driver>oracle.jdbc.driver.OracleDriver</driver>
</db>
<db type="rdb" connect="remote">
  <alias>mybookdb</alias> <uri>...</uri>...
</db>
```

그림2. 원격 데이터베이스 및 데이터 정보(예)

그림2는 본 시스템에서 테스트용으로 사용한 연결 정보이다. 다형적인 데이터 베이스 환경을 지원하고 XML data로 변환하기 위해 <db>엘리먼트하부에 <alias>엘리먼트를 사용하게 된다.<alias>엘리먼트는 XML 데이터를 질의할 때 사용자로 하여금 중립적인 질의를 할 수 있도록 하며, 데이터 연결 과정에 상관없이 질의과정을 수행하기 위해 필요하다.

3.6 연결 감시기

외부 데이터와의 연결을 감시는 작업을 한다.GUI 어플리케이션 형태나 자바 서블릿을 통해 개발이 가능하다. 현재는 자바 스윙라이브러리를 이용해 어플리케이션을 개발 중이며 차후 브라우저를 이용한 감시 기능을 가능하게 하기 위해 서블릿이나 CGI 형태로 구현 확장도 가능하다.

3.7 XML질의처리 프로세서(XmlQ)와 XML질의언어

본 논문에서 제안하는 XmlQ 질의어는 기본적으로 Lorel 질의어의 문법을 참조하되 SQL과 유사한 언어로 설계한 것이 특징이다. 현재까지는 select - from- where 구문만 지원하며 group by / having 은 구현되지 않았다. 표1에 나타난 Lore데이터베이스 시스템의 Lorel 질의어[6]의 경우 XML 엘리먼트의 질의를 수행하는데 있어서 SQL에 비해 문법의 엄격성이 낮는데 이 경우 XML 데이터를 질의할 때 있어서 유연성을 가지게 된다. 본 XML 질의처리어 설계에 있어 기타 다른 언어들도 있지만 다수의 데이터베이스 시스템 개발자가 SQL에 익숙한 것을 고려서 SQL과 유사한 형태의 언어로 설계되었다.

표1 . 질의 형태 비교

질의어	질의 형태 (예시)
Lorel	select xmlDB명 select - from select - where select-from-where
SQL	select- from - [where]
XmlQ language	select- from - [where]

XmlQ 언어의 경우도 select-where 구문만으로 xml을 질의하는 구문으로 고려할 수 있지만 다형적 환경을 지원하는 XML 데이터 서버라는 측면에서 원거리 시스템의 데이터를 정의하는 부분이 필요하고 이를 from절에 기술하도록 정의하였기에 from

절은 select문 뒤에 오도록 했다. 표2를 보면 from 절에서 외부 데이터 소스를 정의할 때 SQL의 @을 사용해 구문을 서술한다.

표2 . Xml질의어의 from절 사용

질의어 종류	질의형태 (예)
SQL 질의어	select 필드명 from 참조테이블 @다른데이터베이스
XmlQ 질의어	select 엘리먼트명 from dburi: 참조엘리먼트 @외부데이터_별명

XmlQ 언어는 Lorel 검색질의의 구문과 유사하지만 구문에서 SQL의 select-from-[where]구조를 필히 준수하도록 설계되었다. 본 시스템의 테스트를 위해 사용한 데이터는 아래 표3과 그림3과 같다

표3. 관계형 테이블 데이터

title	subtitle	author	year
Modern...	-	Fred	1999
XML ...	Building...	S. Mag	1999
Data on...	From Rel...	S.Abiteboul	1998

```
<bib> <book>
<title>Xml By Example </title>
<author>S. Mag</author>
<year> 1999 </year></book> ...
</bib>
```

그림3. XML 데이터

예제 데이터를 토대로 검색질의를 수행한 결과는 그림 4와 같다. XML 데이터와 관계테이블 형태로 구성된 데이터를 엘리먼트형태의 스키마로 구성하여 질의한 결과이다.

```
XmlQ>select mybook.title
from dburi:mybook@mybookdb,
xmluri:/home/doc/bib.xml
where mybook.author=bib.book.author

result query=>
mybook.title
Modern Database Management
mybook.title
Xml By Example
```

그림 4. 검색 결과

4. 결론

본 논문에서는 XML데이터와 관련된 다형구조의 환경을 지원하는 XML데이터 검색질의의 시스템을 설계 및 개발 방안에 관해 논해 보았다.다형구조의 환경에서 통합된 검색 질의를 가능케 하려면 시스템은 특정 시스템에 종속적이지 않고 다수의 데이터 시스템을 질의하는데 적용이 가능해야 한다. 또한 XML을 포함한 질의라는 측면에서 질의어도 다형적인 구조를 고려한 형태로 구성해야 한다. 본 논문에서는 이에 제안한 시스템 구조에 기초해 구현을 통해 검증해 보았다. 향후 연구방향은 질의처리기의 성능향상 방안에 대한 연구가 필요하다.

참고문헌

[1] Alin Deutsch , Mary Fernandez and Dan Suciu , "Storing Semistructured Data with STORED" , SIGMOD Record, Volume 28, Issue 2, pp431-442 , June 1999

[2] Lore 데이터베이스 시스템 홈페이지
<http://www-db.stanford.edu/lore/pubs/index.html>

[3] DB2 UDB XML Extender 웹사이트
<http://www-4.ibm.com/software/data/db2/extenders/xmlext/>

[4] Oracle XSQL 웹사이트
<http://technet.oracle.com/tech/xml/info/index2.htm?Info&htdocs/relational/index.htm>

[5] ODMG 3.0 스펙문서
<http://www.odmg.org/files/odmg30.zip>

[6] S.Abiteboul, D. Quass, J. McHugh, J. Widom, and J. Wiener. "The Lorel Query Language for Semistructured Data", International Journal on Digital Libraries, 1(1), pp68-88, April 1997.

[7] Serge Abiteboul , Peter Buneman and Dan Suciu, "Data on the Web (From Relations to Semistructured Data and XML)",pp167-210, Morgan Kaufmann Publishers