

XML기반의 가상 스키마를 이용한 정보 통합 검색시스템 설계

문석재^o 정계동 국윤규 최영근
광운대학교 컴퓨터 과학과

{msj1568^o, gdchung, ykkook ygchoi}@kw.ac.kr

Design of Virtual Schema using Information Integration Search System based on XML

Seok-Jae Moon^o, Kye-Dong, Jung, Youn-Gyou Kook, Young-Keun Choi
Dept. of Computer Science, KwangWoon University

요 약

XML을 기반으로 한 시스템간의 통합하는 과정에서 발생하는 데이터 정의 이질성 문제, 데이터 표현의 이질성 문제, 유사표준의 중복 개발 등이 야기되므로, 이에 대한 대비책이 필요하다. 따라서 분산된 이 기종 시스템을 마다 각각 구조화된 데이터베이스를 통합하는 과정에 발생하는 시스템간의 불일치 문제를 해결할 수 있는 방법이 메타데이터를 이용하는 것이다. 본 논문에서는 어질적이면서도 분산되어 있는 정보 저장소로부터 메타데이터를 수집하고, 이를 통합된 메타데이터 베이스로 구축하여, 분산된 데이터베이스에 대한 검색정보 및 데이터 트랜잭션을 할 수 있도록 제안한다. 정보 통합 방법으로는 메타데이터를 기반으로 한 가상 스키마를 이용한다. 스키마 이질성과 데이터 이질성 해결하기 위한 방법으로 메타데이터 가상 스키마를 설계 하였다. 메타데이터 가상 스키마 기반으로 한 정보 통합은 XML 뷰어 언어인 XDR을 기반으로 한다. 이는 XMVS 템플릿으로 분해하여, 통합 처리 할 수 있는 XML기반의 가상 스키마를 이용한 정보 통합 검색시스템을 설계 제안한다.

1. 서 론

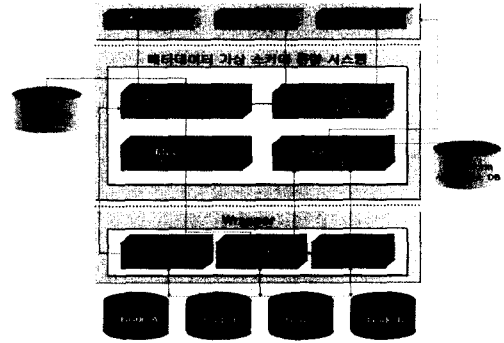
최근 XML기술을 적용하는 사례가 증가하고 있으나, 산재된 데이터를 통합할 때, 데이터의 이질적인 문제 때문에, 각각의 데이터 데이터베이스 스키마를 전역스키마로 통합하는 방법이 제안되었고, 웹과 같은 동적 환경에도 지능형 중계기, 도메인 컨트롤러, 에이전트, 그리고 고도의 질의어 등을 이용한 방법들이 제안되었다.[5][4] 이러한 방법은 인터넷 기반에서 다양한 정보를 실시간으로 사용할 수 있다는 측면에서는 아주 흥미로운 일이다. 하지만, 이질적인 정보 통합은 데이터베이스 분야에서는 아주 중요한 문제이다. 이러한, 이질적인 데이터베이스의 정보 통합 과정에서 의미적인 통합 연구가 진행 중이다. 의미적 이질성을 스키마 이질성과 데이터 이질성으로 나눌 수 있다. 데이터 정보를 관리해주는 데이터를 메타데이터라 하고, 메타데이터를 관리해주는 저장소를 메타데이터 레퍼지토리라 한다. 메타데이터는 데이터 요소로 구성된 메타레지스터로 데이터 표준의 역할을 담당하며, 새로운 시스템 구축 시 데이터베이스 스키마 적용의 기준이 될 수 있다. 정보를 접근하는 방법은 다양해지고 있으나, 현재 대부분의 메타데이터 레퍼지토리는 특정 소프트웨어나 DBMS에 의존적이기 때문에, 메타데이터의 정보획득은 물론, 특정 소프트웨어에 의존적일 수밖에 없게 된다. 본 논문에서는 XML을 기반으로 한 메타데이터 가상 스키마를 설계 하므로써, 웹상에서의 XML로 구조화된 문서로 전송 가능하고, 여러 장소에서 생성된 메타데이터 및 데이터 정보 교환이 가능하게 한다. [1][2] 또한, 제안된 메타데이터 가상 스키마 모델을 기반으로 하여 메타데이터 정보 추출의 범위를 확대하여 메타데이터 가상 스키마 뷰어 모듈로 테이블간의 관계 정보를 표현 할 수 있도록 확장한다.[5] 그리고, XMVS는 메타데이터 가상 스키마를 사용하여, 분산된 DB를 통합하기 위한 MDR(메타데이터 레지스터)를 XMVS(XML 메타데이터 가상 스키마)로 정의한다. 본 논문의 구성은 2장에서는 XML기반의 메타데이터 가상 스키마를 이용한 정보 통합 검색 시스템을 설계하고, 3장은 적용 사례를, 마지막 4장은 결론 및 향후 연구에 대하여 기술한다.

2. XML기반의 메타데이터 가상 스키마를 이용한 정보 통합 검색 시스템

분산 환경에서의 정보 통합 방법으로, 메타데이터 가상스키마를 기반으로 한 메타데이터 정보 추출의 대한 범위를 확대하고, 메타데이터를

XML뷰어 모듈로 테이블간의 관계정보 표현을 확장할수 있도록, 메타데이터 가상스키마를 이용한 XML기반의 정보 통합 검색 시스템 구조를 설명한다.

2.1 전체 시스템 구성



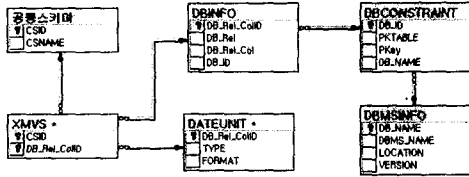
(그림1) 전체시스템 구성도

(그림1)은 분산환경에서 메타데이터 가상스키마를 이용한 XML기반의 정보통합 검색시스템을 설계한 것이다.

- 전체적인 동작 과정은 다음과 같다.
- (1) 클라이언트가 정보 검색을 질의 한다.
 - (2) 분산 환경의 지역 DB들의 스키마 정보, 구조정보, 위치정보를 미리 메타데이터 가상 스키마에 수집하고, 수집된 메타데이터를 사용하여, 지역 DB의 맞는 XMVS템플릿을 생성한다.
 - (3) 분산 환경이 지역 DB들의 맞게 생성된 XMVS템플릿을 Wrapper에 전송한다.
 - (4) Wrapper에서 XMVS템플릿으로 분산 환경의 지역 DB들의 데이터를 검색하여, XML 문서로 결과를 추출한다.
 - (5) XML 문서 수집기에서 검색하여, 추출된 XML 문서들을 수집하여 통합한다.
 - (6) 통합한 XML문서들 XSL 형식에 맞추어 클라이언트에게 검색된 내용을 보여준다.

2.1.1 메타데이터 가상 스키마 매핑 및 관리

지역 DB의 스키마 정보를 수집하여, 메타데이터의 매핑 및 통합하여 관리한다. 지역 DB스키마 정보가 전송되어지면, 스키마 매핑기에 의해 (그림2)인 메타데이터 가상 스키마 테이블에 수집되어 진다. 수집되어진 지역 DB스키마 정보들은 분산된 지역 DB들을 데이터를 검색할 때, XMVS템플릿에 적용되어진다.



(그림2) 메타데이터 가상 스키마 테이블 구조

- (1) 공통스키마 - 지역DB스키마에서 공통적으로 추출한 XML스키마 정보를 저장한 테이블.
 - CSID - 스키마 ID.
 - CSNAME - 지역 DB스키마에서 공통적인 스키마를 추출하여 통합된 스키마 명.
- (2) DBINFO - 지역DB스키마 정보를 추출하여 저장한 테이블.
 - DB_Rel_ColID - 지역 DB테이블이 컬럼ID.
 - DB_Rel - 지역 DB의 테이블명.
 - DB_Rel_Col - 지역 DB 테이블의 컬럼명.
 - DB_ID - 지역 DB의 제약사항 정보키 ID.
- (3) XMVS - 지역 DB스키마 정보의 스키마 이질성을 해결하기 위해 공통스키마 정보와 DBINFO 정보를 매핑한 테이블.
 - CSID - 공통스키마 ID.
 - DB_RelColID - 지역 DB테이블 컬럼 ID.
- (4) DBCONSTRAINT - 지역 DB스키마의 제약사항을 저장한 테이블.
 - DB_ID - 지역 DB의 제약사항 정보키 ID.
 - PKTABLE - 지역 DB스키마의 주기가 있는 테이블명.
 - PKey - 지역 DB의 주기 필드명.
 - DB_NAME - DB명.
- (5) DBMSINFO - 지역 DB의 서버정보 테이블.
 - DB_NAME - DB명.
 - DBMS_NAME - DBMS명.
 - LOCATION - 원격서버정보(IP)
 - VERSION - 버전 정보.
- (6) DATEUNIT - 지역 DB스키마 정보의 데이터 이질성을 해결하기 위한 정보 테이블.
 - DB_RelColID - 지역 DB테이블이 컬럼ID.
 - TYPE - 지역 DB테이블의 컬럼들의 DB타입.
 - FORMAT - 지역 DB스키마의 날짜, 단위등의 양식.

2.1.2 메타데이터 질의어 처리기

클라이언트에서 검색 질의한 정보를 기반으로 메타데이터 가상 스키마로부터 메타데이터를 추출한다. 추출된 메타데이터는 XMVS템플릿에 적용된다. (그림3)은 XMVS템플릿에 적용하기에 필요한 메타데이터를 추출하기 위한 질의문과 결과이다.

```
SELECT DBCONSTRAINT.*, DBINFO.*, XMVS.*, 공통스키마.CSNAME
FROM 공통스키마 INNER JOIN
XMVS ON 공통스키마.CSID = XMVS.CSID INNER JOIN
DBINFO ON XMVS.DB_Rel_ColID = DBINFO.DB_Rel_ColID INNER JOIN
DBCONSTRAINT ON DBINFO.DB_ID = DBCONSTRAINT.DB_ID
```

DB_ID	PKTABLE	PKey	DB_NAME	DB_Rel_ColID	DB_Rel	DB_Rel_Col	DB_ID	CSID	DB_Rel_ColID	CSNAME
C02	mes	the_id	BaroBook D6		mes	C02	CS1	D6		도서명
C04	authors	au_id	BaroBook D16		authors	au_name	C04	CS2	D16	저자
C01	publishers	pub_id	BaroBook D1		publishers	pub_name	C01	CS3	D1	출판사명
C02	mes	the_id	BaroBook D11		mes	pubdate	C02	CS4	D11	출판일
C02	mes	the_id	BaroBook D9		mes	price	C02	CS5	D9	가격
C02	mes	the_id	BaroBook D10		mes	advance	C02	CS6	D10	요약

(그림3) 메타데이터 가상 스키마 질의문 및 결과

2.1.3 XMVS 템플릿 생성 및 관리

XMVS템플릿은 sql:relation, sql:field, sql:limit-field, sql:limit-value, sql:datatype, sql:relationship key-relation으로 구성된다. 이들은 element, ElementType, attribute를 데이터베이스의 특정 테이블이나 컬럼명으로 매핑하는데 사용한다. 요소 ElementType, Attribute를 데이터베이스의 특정 테이블이나 컬럼으로 매핑하는데 사용한다.

- sql:relation : 지역 DB의 테이블명.
- sql:field : 지역 DB의 컬럼명.
- sql:limit-field, sql:limit-value : 제한된 조인을 위한 컬럼명과 값.
- sql:datatype : 컬럼의 타입.
- sql:relationship key-relation foreign-relation : 지역 DB스키마 테이블의관계 제약사항을 고려한 주기, 외래키.

```
<?xml version="1.0" ?>
<Schema xmlns="urn:schemas-microsoft-com:xml-data"
xmlns:sql="urn:schemas-microsoft-com:xml-sql">
<ElementType name="Pub" sql:relation="publishers">
<AttributeType name="Name"/>
<AttributeType name="Type"/>
<attribute type="Name" sql:field="pub_name" />
<attribute type="Type" sql:limit-field="pubdate"
sql:limit-value="1991-06-12"
sql:datatype="datetime"
sql:field="title"
sql:relation="titles"
sql:foreign-key="pub_id"
foreign-relation="titles" foreign-key="pub_id"/>
</attribute>
</ElementType>
</Schema>
```

(그림4) XMVS템플릿 뷰

(그림4)에서는 스키마 이질성과 데이터 이질성을 해결하여 생성된 XMVS템플릿 뷰이다.

2.1.4 XML문서 수집기

Wrapper에서 검색 추출된 XML문서는 XML문서 수집기에 의하여 각 지역 DB들로부터 수집된다. 그리고 XML문서 수집기에서는 최종적으로 문서들을 통합하여, 클라이언트에게 보여준다.

2.2 Wrapper

2.2.1 지역 스키마 관리

각 지역DB의 스키마 정보를 통합하기 위해 지역 스키마 관리에서 (그림5)와 같이 수집하여 전송한다.

```
<?xml version="1.0" encoding="utf-8" ?>
<SOAP:Envelope xmlns:SOAP="urn:schemas-xmlsoap-org:soap-v1">
<SOAP:Body>
<local:SchemaInfo>
<RemoteInfo>
<SchemaInfo>MS-SQL2000</DBMSINFO>
<DBNAME>BaroBook</DBNAME>
<LOCATION>128.134.47.6</LOCATION>
</RemoteInfo>
</local:SchemaInfo>
</Tables>
<Table Name="pub_info">
<Table Name="publishers">
<field Name="pub_id" Type="string" PrimaryKey="True" />
<field Name="pub_name" Type="string" PrimaryKey="False" />
<field Name="dtu" Type="string" PrimaryKey="False" />
<field Name="state" Type="string" PrimaryKey="False" />
<field Name="country" Type="string" PrimaryKey="False" />
</Table>
<Table Name="titles">
<Table Name="titleauthor">
<Table Name="authors">
</Tables>
</local:SchemaInfo>
</SOAP:Body>
</SOAP:Envelope>
```

(그림5) 지역 DB스키마

지역 DB스키마는 DB원격정보와 테이블명, 필드명, 타입, 제약사항 등을 수집되어 전송되어진다.

2.2.2 지역 XMVS 템플릿 처리

(그림4)인 XMVS 템플릿 뷰가 지역 XMVS 템플릿 처리기에 보내져, http://128.134.47.6/BaroBook/schema/BaroBook.xdr/Pub?root=BaroBook 프로토콜로 지역DB에 접근하여, 결과를 추출한다.

2.2.3 지역 질의어 처리

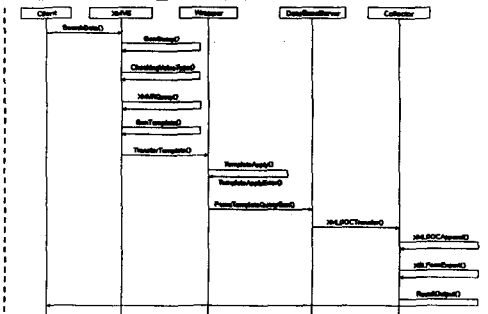
각 지역DB에 XMVS템플릿이 적용이 되지 않아 검색질의가 되지 않는 경우, 전송된 XMVS템플릿을 지역 질의어 처리에서 파싱하여, 각 지역 DB에 맞는 질의를 생성하여 검색을 하게 된다. (그림6)은 질의문이 생성된결과이다.

```
Select pub_name, title From titles
Inner Join publishers On titles.pub_id = publishers.pub_id
Where pubdate='1991-06-12'
```

(그림6) 지역 SQL질의문

2.3 XML기반의 메타데이터 가상 스키마를 이용한 정보통합 검색시스템 설계

(그림7)은 정보통합검색을 위한 순서도이다. Client에서 조회 검색을 하면, 메타데이터 가상스키마에서 검색조건을 기반으로 메타데이터 질의하여, XMVSTemplate로부터 생성한 후, Wrapper에게 전송하여, XMVSTemplate로부터 검색할 데이터를 추출한다. 이때, 지역DB 특성에 의해 검색을 실행시, XMVSTemplate를 파싱하여, 지역DB 특성에 맞는 질의를 생성한 후 검색한다. 검색된 XML문서형식의 결과를 XML문서수집기인 Collector에서 수집하여, 클라이언트에게 검색 결과를 보여준다.



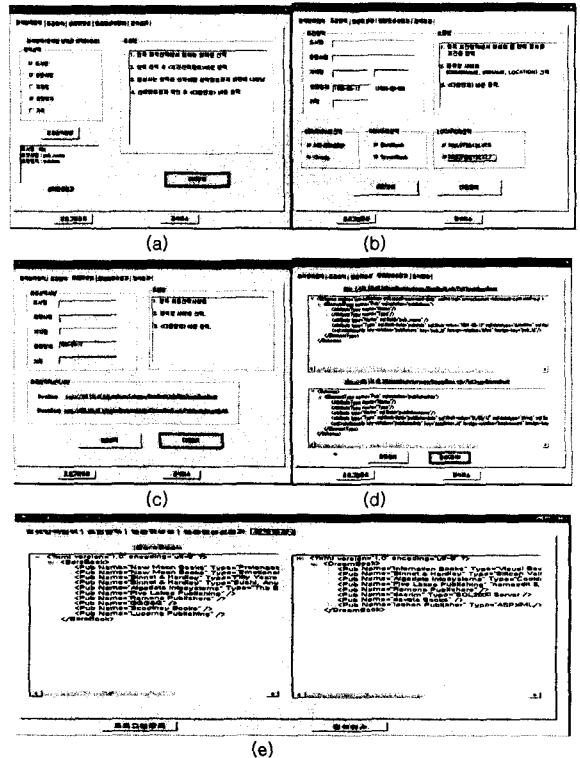
(그림7) 정보통합검색 흐름도

메소드는 다음과 같다.

- (1) SearchData() - 클라이언트가 검색질의 할 내용을 입력한다.
- (2) GenQuery() - 입력된 검색질의 데이터를 기반으로, 질의문을 생성한다.
- (3) CheckingValueType() - 질의 쿼리에 대한 타입 체크.
- (4) XMVSTemplate() - 생성된 쿼리문으로 메타데이터를 추출한다.
- (5) GenTemplate() - 쿼리문에 의해 추출된 메타데이터를 기반으로, 각 시스템에게 전송할 XMDR 템플릿을 생성한다.
- (6) TransferTemplate() - 생성된 XMDR템플릿을 각 시스템에게 전송한다.
- (7) TemplateApply() - 각 시스템에게 전송된 XMDR템플릿으로 DB에 데이터를 검색하여, XML문서형태로 추출한다.
- (8) TemplateApplyError() - 각 시스템의 특성에 의해 템플릿뷰로 DB의 데이터를 검색하지 못할때, 에러를 발생한다.
- (9) ParseTemplateQueryGen() - XMDR템플릿으로 각 시스템 DB에 검색질의시 에러가 발생된 것은 XMDR템플릿을 파싱하여, 각 시스템 DB의 특성에 맞는 DB질의문을 생성하여, 다시 DB의 데이터를 검색하여 XML문서로 추출한다.
- (10) XMLDOCTransfer() - 각 시스템들의 DB에서 XML문서형태로 추출된 것을 전송한다.
- (11) XMLDOCAppend() - 각 시스템에서 전송된 XML문서들을 통합한다.
- (12) XSLFormExport() - 보여줄 클라이언트에 맞는 Form형식을 추출한다.
- (13) ResultOutput() - 통합된 XML문서를 추출한 Form형식에 맞춰 클라이언트에게 보여준다.

3. 적용 사례

(그림8)은 적용 사례로서 도서상품검색에 대한 검색 결과이다. 이는 SQL2000 Server에서 제공하는 XML뷰 스키마 언어인 XDR을 사용했고, 도서상품검색을 예로 하였다.



(그림8) 적용사례 - 도서상품검색

- (a) 사용자가 도서명, 출판사명, 출판일자에 대한 조회 선택.
- (b) 출판일자에 대한 조건 입력 및 원격지 정보 선택.
- (c) 조회할 내용과 원격지 정보를 보여줌.
- (d) 선택한 원격지에 전송할 템플릿 생성 결과.
- (e) 템플릿에 의한 검색 결과.

4. 결론 및 향후 연구

분산된 각 시스템들의 기종이 달라 구조화된 데이터베이스를 통합하는 과정에 발생하는 스키마이질성, 데이터질성 문제의 해결 방안을 제안했다. 또한 통합 방법으로 메타데이터를 기반으로 한 메타데이터 정보 추출의 대한 범위를 확대하고, 메타데이터 뷰어 모듈에서 테이블간의 관계 정보를 표현할수 있도록 확장하였다. 키의 제약사항과 컬럼의 타입을 고려하여, 데이터 베이스 일관성을 유지하였다. 그러나, 데이터들의 일관성 문제를 통합하는 과정에 있어서는 데이터들의 변경에 대한 문제나 데이터베이스의 스키마 구조가 변경되었을 때, 좀더 일관성 있고, 유연하게 대처하는 방안을 해결하기 위한 연구가 필요하다.

참고문헌

- [1]David Wang, "Automated Semantic Correlation between Multiple Schema Information Exchang," M.I.T MM May 2000]
- [2]Daniela Florescu, Alin Deutsch, Mary Fernandez, AlonLevy, David Maier, DanSuciu, "Querying XML data," Data Engineering Bulletin, Vol.22, No.3, pp27-34, 1999.
- [3]Kevin Williams, et al, "Professional XML DataBase." Wrox, 2001, pp743-797]
- [4]Mike Uschold and Michael Gruninger, Ontologies : " Principles, Methods and Applications," Knowledge Engineering Review, 1996.]
- [5]백두권, 최요한, 박성공, 이정욱, 정동원 "MDR과 온톨로지를 결합한 3계층 정보통합 시스템" 정보처리학회논문지 D 제10-D권 제2호 2003.4]