

XML과 관계형 데이터 모델간의 스키마 변환 방법에 관한 비교론적 고찰

허보진*, 김형석*, 김창석*

*공주대학교 대학원 멀티미디어학과

e-mail: {bobe04, replay97, csk}@kongju.ac.kr

A Comparative Study on Schema Conversion Methods Between XML and Relational Data Models

Bo Jin Heo*, Hyeong Seok Kim*, Chang Suk Kim*

*Dept of Multimedia, Kongju National University

요 약

XML이 인터넷상에서 비즈니스 데이터 교환과 저장의 표준으로 자리 잡아감에 따라 많은 데이터베이스 연구자들이 XML과 관계형 모델 간의 스키마 변환 방법에 관한 연구를 수행하게 되었다. 지금까지 두 모델간의 변환 방법이 국내외에서 발표되고 있으나 이들 연구들에 대한 비교 및 고찰에 대한 체계적인 연구가 부족하였다. 본 논문에서는 XML과 관계형 모델 간의 스키마 변환 방법에 관한 국내외 연구들을 분류 정리하여 비교·고찰함으로써 향후 XML 스키마와 관계형 데이터베이스 간의 데이터 교환 및 저장에 관한 연구의 사전(事前) 연구가 되도록 하였다.

1. 서론

최근 XML이 인터넷상에서 비즈니스 데이터 교환과 저장의 표준으로 자리 잡아가고 있다. 그러나 아직 대부분의 비즈니스 데이터들은 관계형 데이터베이스나 객체-관계 데이터베이스에 저장하여 사용하고 있다. XML 전용 데이터베이스 시스템이 널리 사용되지 않은 현 시점에서는 XML 데이터를 관계형 데이터베이스 시스템에 저장하고거나, 이미 저장된 관계형 데이터베이스 시스템에 저장된 데이터를 XML로 출력하는 것이 필수적이다. 그래서 많은 데이터베이스 연구자들이 XML 데이터와 관계형 모델 간의 스키마 변환 방법에 관한 연구를 수행하게 되었다. 지금까지 두 모델간의 변환 방법이 국내외에서 발표되고 있으나 각 연구기관마다 독자적으로 연구가 진행되어 연구의 주관점이 모두 다르게 진행되고 있다.

지금까지 XML과 관계형 데이터 모델 간의 데이

터 변환에 대한 연구들을 수집·분석하여 이들 연구들에 대한 비교 및 고찰에 대한 체계적인 연구가 부족하였다. 본 논문에서는 XML과 관계형 모델 간의 스키마 변환 방법에 관한 국내외 연구들을 분류 정리하여 비교·고찰함으로써 향후 XML 스키마와 관계형 데이터베이스 간의 데이터 교환 및 저장에 관한 연구의 사전(事前) 연구가 되도록 하였다.

우선 계층적 구조 형태인 XML을 플랫폼(flat) 파일 구조인 관계형 데이터 모델로 변환하는 방법을 살펴본다. 이때 두 표현 방법의 차이를 어떻게 극복하는지를 중점적으로 분석한다. 다음으로 이미 구축된 관계형 데이터를 XML로 출력하기 위해 플랫폼 파일 구조를 계층적 특성을 가지는 XML 스키마 구조를 변환하는 방법을 분석한다. 그래서 이들 방법들의 접근 방법을 의미적인 측면과 구조적인 측면으로 구분하여 고찰해 본다.

2. XML에서 RDB 변환 방법들

XML 스키마에서 관계형 데이터베이스 모델링까지의 과정을 자동화시키고자 하는 많은 연구가 진행되고 있지만, XML은 계층적 구조를 가지고 있으므로 이를 2차원적인 테이블의 집합인 관계형 정보로 표현하기에는 한계가 있다. 따라서 계층적 구조를 가진 DTD를 관계형 데이터베이스에 저장하기 위한 모델링으로 변환하기 위한 작업이 필요하다.

지금까지 DTD의 구조적인 측면은 관계형 모델로 변환이 되었는데 의미적인 측면은 제대로 전달되지 않았다. 여기서는 계층적 구조를 가진 XML 문서를 플랫폼한 구조의 관계형 데이터 베이스에 저장하는 작업을 살펴본다.

2.1 XML 문서의 공통구조를 이용한 방법[5]

[5]에서의 연구는 XML문서의 구조 정보를 담고 있는 DTD를 이용하여 동일한 DTD를 사용하는 XML문서들의 공통스키마를 추출하는 기법을 사용하여 관계 데이터베이스 시스템에 XML 문서를 저장하기 위한 릴레이션 스키마 추출기법을 제안하였다.

- 기존 스키마 추출 기법의 단점인 릴레이션 테이블에 다수의 널(NULL)값이 존재하는 단점을 극복해 릴레이션 테이블 수를 줄일 수 있는 기법을 제안하였다.
- DTD 설계자가 상위 노드에 중요한 데이터를 '+'와 'n'연산자가 아닌 '*와 '?' 연산자로 설계한다면, DTD 그래프의 CO(Cardinality Operator) 적용에 있어 Level을 고려하여야 하며 이러한 과정은 또한 기존의 방법의 단점들을 포함 할 수 있다.
- DTD를 이용한 절대 스키마 추출기법
 - 1) Cardinality Operator = CO라 약칭함
CO를 이용해 각 노드간의 발생 빈도 범위를 설정하여 DTD그래프에 적용함으로써 XML문서들의 공통스키마를 추출한다.
 - 2) XML 문서의 DTD 속성 선언부에는 각각의 속성 값들의 필수 존재 여부를 결정해주는 속성 기본 파라미터 값을 적용한다.
 - 3) ENTITY나 NOTATION같은 DTD 구성 요소는 고려하지 않은 상태에서 절대 릴레이션 스키마 추출과정은 아래와 같다.
 - ① DTD 그래프 생성은 DTD의 유효성을 검증하여 초기화된 DTD를 통해 각각의 노드정보를 가져온다.
 - ② 합 연산자 그래프 생성은 DTD내의 CO정보를

추출하여 합 연산자를 생성한다.

③ 절대 스키마 생성은 CO 정보를 통해 절대스키마를 생성할 노드를 분리 한다.

④ 비 절대 스키마 생성은 절대 스키마 외의 나머지 노드들을 따로 생성한다.

⑤ 테이블 노드 생성은 절대 스키마에서 '+'연산자를 가지는 노드는 하나의 테이블로 매핑이 되며, 부모와 자식 테이블 노드와는 foreign key로 참조된다. 테이블 노드의 자식 노드들은 릴레이션 테이블의 속성 값이 된다. 또한, 속성 기본 파라미터 값이 'IDREFS'인 노드 또한 하나의 테이블로 매핑되며 부모 노드와 외래키로 참조된다. 각각의 테이블노드에 대한 primary key 필드를 생성한다. ID 속성이 존재하면 사용하며, 없는 경우는 인적으로 생성한다. '@' 표기는 속성 값을 표현한다. 공유되는 엘리먼트를 위해 릴레이션 테이블에 parent field와 root field 생성한다.

절대 릴레이션 스키마는 NULL값을 포함하지 않으며 전체 XML 문서들에 공통적으로 존재하는 노드들만으로 구성된 것이다.

4) 비 절대 릴레이션 스키마 추출

비 절대 스키마에 포함되는 데이터들은 절대 스키마 노드들의 자식 노드들로 구성된다.

2.2 XML Schema를 위한 관계형 스키마 자동생성기의 개관[6]

이 연구는 XML Schema가 시스템에 입력되면 타입 분리가 XML Schema 내의 구성요소들을 분류하여, 엘리먼트 처리기, 데이터 타입 변환기, 스키마 그래프 생성기 등을 거쳐서 최종적으로 관계형 스키마 생성기에 의하여 관계형 스키마를 생성하게 한다.

XML Schema로부터 관계형 테이블들을 생성하기 위해 필요한 정보들로 구성된 스키마 그래프와 타입그래프를 제시하고 XML Schema의 데이터 타입, 상속, 다형성 등과 같은 다양한 각각의 요소들에 대해서 관계형 스키마를 어떻게 생성할 것인가를 제시하였으며, 시스템의 성능을 향상시킬 수 있는 휴리스틱 기법들과 생성된 관계형 스키마의 이용을 돕기 위한 테이블 참조 매뉴얼을 제시하였다.

2.3 CPI[3] 알고리즘

CPI(Constraints-preserving Inlining) 알고리즘은 DTD에서 발견할 수 있는 의미적인 제약조건들을 추출하여 관계형 스키마로 변환할 때 이 제약조건들

을 반영하는 것이다. 제약조건들은 다음과 같다.

① Cardinality Constraints

DTD 선언시 레코드에 일어날 수 있는 4가지 카디널리티 관계를 살펴보면 다음과 같다.

(element와 sub-element 관계)

- (0,1)은 element와 sub-element가 0 또는 1의 매핑 관계를 나타내고 있다(?).
- (1,1)은 element와 sub-element가 1 또는 1의 매핑 관계를 나타내고 있다.
- (0,N)은 element와 sub-element가 0 또는 N번의 매핑 관계를 나타내고 있다(*).
- (1,N)은 element와 sub-element가 1 또는 N번의 매핑 관계를 나타내고 있다(+).

그 외 Inclusion Dependencies(INDs), Equality-Generating Dependencies(EGDs), Tuple-Generating Dependencies(TGDs) 등이 있다.

2.4 XML DTD 그래프 모델을 이용한 연구[2]

본 논문에서의 연구는 XML의 문서 타입 정의(DTD)를 분석하여 이를 DTD 그래프로 표현한다. 이 그래프를 확장하여 정규 간선과 백 간선을 가지는 엘레먼트 그래프로 변환하여 관계 데이터베이스로 변환 한다.

3. RDB에서 XML로 변환 방법들

관계형 모델에서 XML스키마 변환방법의 주된 논쟁거리는 단지 관계형 모델 스키마의 구조를 그대로 XML스키마로 변환하는 것이 아니라, 관계형 모델 스키마를 분석하여 xml 스키마에 의미적 제약조건(semantic constraints)을 부여하여 정보의 손실 없이 변환하는 것이다.

3.1 NeT(Nesting-based Translation Algorithm)[1]

NeT[?]는 플랫폼 릴레이션 스키마로부터 계층화된 XML 스키마로 이끌어 내는 알고리즘이다. 관계형 스키마의 테이블들은 XML DTD의 엘레먼트로, 관계형 스키마 컬럼들은 XML DTD의 애트리뷰트로 변환하는 간단한 구조이지만, XML에서 제공되는 정규표현 연산자(e.g., * or +)를 표현하기 위해 릴레이션을 'nest operator'로 내포하여 XML 스키마의 특성인 계층화로 표현하는 알고리즘이다.

NeT 알고리즘은 다음과 같다. 관계스키마 R이 적용된 각 테이블 t_i 는 nesting이 성공적으로 안 될 때까지 'nest operator'를 반복시킨다. 선택되었던 기준에 의거하는 가장 내포된 테이블을 선택한다.

nesting된 컬럼들 $\{c_1, \dots, c_{k-1}\}$ 위에서 성공했던 곳에 $t'_i(c_1, \dots, c_{k-1}, c_k, \dots, c_n)$ 로서 이 테이블을 보여준다.

만일 $k=1$ 이면 FT 변환을 따른다. 만일 $k>1$ 이면 각 컬럼 $c_i(1 \leq i \leq k-1)$ 일 때 스키마 R이 널을 허용하면 엘레먼트로 변환시 컬럼 c_i 는 *를 붙여 주고, 널을 허용하지 않으면 c_i+ 를 붙여 준다. 각 컬럼 $c_j(k \leq j \leq n)$ 일 때 스키마 R이 널을 허용하면 엘레먼트로 변환시 컬럼 c_j 는 ?를 붙여 주고, 널을 허용하지 않으면 c_j 로 표현한다.

3.2 CoT 알고리즘[1]

NeT 알고리즘은 단지 단일 테이블에 적용되는 XML 변환 알고리즘이지만 CoT(Constraints-based Translation Algorithm)은 변환시 다수 테이블들 간의 의존되어있는 관계와 내부적으로 연결되어 있는 테이블들을 계층적이고 일관성 있는 부모-자식구조의 최종 XML 스키마로 병합하는 내용을 고려한 XML 변환 알고리즘이다. CoT의 주된 아이디어는 각각 컬럼 X 와 Y를 포함한 구별되는 두 테이블 s 와 t 에서 우리는 $\alpha \subseteq X$ 와 $\beta \subseteq Y$ 에서 $s[\alpha] \subseteq t[\beta]$ 를 제약적인 외래 키를 가졌다고 가정한다. 또한 s에 대한 키는 $K_s \subseteq X$ 의 관계가 있다고 가정한다. 그리고 나서 s와 t사이에 구별되는 카디널리티 관계는 다음의 조합에 의해 관계모델이 표현되어 진다. 1) α 는 unique/not-unique 2) α 는 nullable/non-nullable. 그런 다음, 두 테이블 s, t의 외래키의 의미적 작업은 다음과 같다.

1. 만약 a가 값이 없지 않다면
 - a) 만약 a가 unique 하고 s와t사이의 관계가 1:1관계가 있다면 아래와 같이 기록되어 진다
<!ELEMENT t(Y,s?)>
 - b) 만약 a가 non-unique하고 s와 t 사이의 관계가 1:n관계가 있다면 아래와 같이 기록되어 진다.
<!ELEMENT t(Y, s*)>

2. 만약 s가 t의 서브 엘레먼트로서 나타내어진다면, s를 위한 키는 K_s 에서 $(K_s - a)$ 로 바뀔 것이다.

이와 같은 방법으로 다수 테이블들의 내부적으로 연결되어진 구조를 계속해서 살펴나가면 최종적인 계층적이고 일관성 있는 XML 스키마를 구해 낼 수 있다.

3.3 Shanmugasundaram 알고리즘[2]

이것은 SQL에 XML Constructor를 만들어 질의를 수행하여 XML 문서가 생성되도록 했다. 여기서

XML Constructor는 XML문서의 프래그먼트를 생성하도록 설계되어 있다. 이것은 스키마 변환을 하지 않지만, 질의를 수행할 때 마다 XML 문서 구조와 관계형 스키마를 고려해야 하는 번거로움이 있다.

[4]는 질의를 수행할 때 XML 문서 구조와 관계형 스키마를 지정하지 않고도 데이터 손실이 없는 변환방법을 제안하였다.

4. 비교 및 고찰

4.1 구조적 측면

XML과 관계형 데이터 모델 간의 데이터 변환에 대한 연구들을 살펴보면 초기에는 XML과 관계형 스키마의 구조적인 사실만 고려하여 변환을 시도하였으나, 최근에는 제약조건이나 의미적인 정보를 추출하여 변환의 정확성을 높여가고 있음을 알 수 있다.

Shanmugasundaram은 Inline 알고리즘을 제안하여 계층적인 구조의 XML과 플랫폼 관계형 모델간의 변환에 중요한 변환 개념을 제공하였다.

4.2 의미적 측면

Lee[?]는 XML과 관계형 데이터 모델 간의 데이터 변환에 의미적 제약조건을 처음으로 도입하였다. 기존의 연구가 변환에 구조적인 측면만 고려한데 비해, Lee는 관계형 모델의 특성인 카디널리티 제약조건, Inclusion Dependencies, Equality-Generating Dependencies, Tuple-Generating dependencies 등을 XML DTD에서 추출하여 변환에 이용하였다.

그래서 XML DTD의 계층적인 성질을 관계형 스키마 변환에 반영하여 좀더 의미적이고 정확한 변환이 되도록 하였다.

4.3 휴리스틱 측면

아직까지 위에서 제안한 제약조건 이외에 저장된 데이터를 마이닝하거나 학습(learning)기법을 이용한 휴리스틱 방법은 제안된 적이 없다. 향후 이런 연구도 필요할 것으로 보인다.

5. 결론

지금까지 XML과 관계형 데이터 모델 간의 데이터 변환에 대한 연구들을 수집·분석하여 이들 연구들에 대한 비교 및 고찰에 대한 체계적인 연구가 부족하였다. 본 논문에서는 XML과 관계형 모델 간의 변환 방법에 관한 국내외 연구들을 분류 정리하여

비교·고찰하였다.

먼저 계층적 구조 형태인 XML을 플랫폼(flat) 파일 구조인 관계형 데이터 모델로 변환하는 방법을 살펴본 후, 두 표현 방법의 차이를 어떻게 극복하는지를 중점적으로 분석하였다. 이때 모델의 차이를 극복하기 위해 구조적인 정보에 의미적인 제약 조건을 도출하여 변환에 이용하고 있음을 보였다. 다음으로 이미 구축된 관계형 데이터를 XML로 출력하기 위해 플랫폼 파일 구조를 계층적 특성을 가지는 XML 스키마 구조를 변환하는 방법을 분석하였다.

이 연구에서 모델의 변환을 충실하게 하기 위해서 구조적 요소뿐만 아니라, 의미적인 요소를 도출하여 변환에 이용하고 있음을 알 수 있다. 본 연구가 XML과 관계형 데이터베이스 변환 연구 및 도구 개발에 활용되기를 기대한다.

감사의 글

본 연구는 정보통신부 2002년 기초기술연구지원사업(C1-2002-024-099-3)과 한국과학재단 목적기초연구(R01-2002-000-00068-0) 지원으로 수행되었음.

참고문헌

- [1] Dongwon Lee, "Schema Conversion Methods between XML and Relational Models," Knowledge Transformation for the Semantic Web, 2003.
- [2] Jayavel Shanmugasundaram, Bridging Relational Technology and XML, Ph.D. dissertation, Univ. of Wisconsin-Madison, 2001.
- [3] CPI: Constraints-Preserving Inlining Algorithm for Mapping XML DTD to Relational Schema, Dongwon Lee, Wesley W. Chu, In J. Data and Knowledge Engineering (DKE), Vol. 39, No 1, pp. 3 - 25, October 2001
- [4] 정유나, 황인준, "관계형 Schema 기반의 XML 문서작성기 설계," 2001 한국정보과학회 봄 학술발표논문집, Vol. 28, No. 1, 2001. pp. 226-228
- [5] 안성은, 이정선, 최황규, "XML 문서의 공통구조를 이용한 효율적인 릴레이션 스키마 추출기법," 2002 한국정보과학회 봄 학술발표논문집, Vol. 29, No. 1, 2002. pp. 49-51.
- [6] 김정섭, 박창원, 정진완, "XML 스키마를 위한 관계형 스키마 자동생성기의 개관," 2002 한국정보과학회 봄 학술발표논문집, Vol. 29, No. 1, 2002. pp. 10-12.