

# XML을 위한 데이터 모델 (A Data Model for XML)

이대우, 최옥, 김영찬

한밭대학교, 정보통신·컴퓨터공학부

(DaeWoo LEE, Ok CHOI, YoungChan KIM

Division of Information Communication and Computer Engineering,  
Hanbat National University)

## 요 약

본 논문에서는 XML(eXtensible Markup Language)을 위한 데이터 모델을 제시하기 위해 현재 광범위하게 사용되고 있는 관계형 데이터베이스(relational database) 개념을 적용한다. 관계형 데이터베이스는 업무 요구사항의 분석으로부터 고수준의 개념적 데이터 모델을 사용해서 데이터베이스에 대한 개념적 스키마(conceptual schema)를 생성한 다음, 고수준 개념적 데이터 모델을 구현 데이터 모델로 변환하여 논리적 데이터베이스 스키마를 생성한다. 이때, 고수준 개념적 데이터 모델링의 대표적인 방법으로 ER모델을 사용하고, 구현 데이터 모델로 관계 모델(relation model)을 사용한다.

XML은 문서의 논리적 구조를 정의하는 DTD와 XML Schema 등을 갖는다. XML의 DTD와 정보 모델링 기법인 E/R 모델은 모두 작은 세계(real world)를 모델링하는 도구들이라고 할 수 있다. 본 논문에서는 XML의 DTD와 E/R 모델의 구성요소들을 분석하여 서로 사상(mapping)시키는 'XML을 위한 데이터 모델(A Data Model for XML)'을 제시한다. 최종적으로 제시된 XML을 위한 데이터 모델에 의해 작은 세계(real world)를 모델링하는 XML DTD에서 정의한 구조(XML Structure)를 따르는 유효한 XML문서(validate XML document)들은 관계형 데이터베이스에 저장할 수 있게 된다.

## 본 문

XML은 E-Business와 응용프로그램들 사이의 정보 교환을 위한 표준으로 자리잡고 있다. 또한, XML은 플랫폼 독립적인 방식으로 다른 형태의 데이터를 통합하기 위한 강력한 환경을 제공하기 때문에 향후 정보 교환의 표준으로서 지배적으로 사용될 것이라 기대한다. XML에서는 작은 세계를 XML문서로 표현하게 되는데, XML문서의 구조를 정의하는 DTD를 정의하여 유효한 문서를 작성하게 된다. 이때, XML의 DTD는 XML문서의 스키마로 간주할 수 있다.

현재 비즈니스 데이터들은 대부분 관계형 데이터베이스에 저장되어 있는데, 이러한 관계형 데이터를 XML으로 표현하기를 원하는 이유는 다른 시스템과의 비즈니스 공유(sharing business data), 비-호환적인 시스템들간의 상호 운용성(interoperability), XML을 사용하는

응용프로그램들을 위한 Legacy Data의 표현, B2B(Business-to-Business) 트랜잭션들 등을 예로 들 수 있다.

이러한 비즈니스들은 XML을 사용하는 다른 비즈니스와 워크 플로우 컴포넌트들 사이에 통신할 필요가 있다. 그러나, 비즈니스 데이터의 대부분은 관계형(relational)과 객체 관계형(object-relational)의 테이블의 열들과 컬럼들에 존재한다. 또한, 현재 상용 DBMS들은 우수한 질의 능력, 확장성, 이용성 등을 제공하기 때문에 계속 존재할 것이다.

XML의 스키마는 데이터가 마크업 되는 방식으로서 문서 안에서 어떤 요소형과 특성, 값들을 사용할 수 있는 지에 대한 규정이라고 할 수 있다. 즉, 하나의 XML문서 안에 어떤 것들을 담을 수 있는지에 대한 규칙들의 집합이며, XML Vocabularies 또는 XML Dictionaries라고도 불린다.

1998년 2월 XML 1.0 recommendation은 W3C working group에 의해 표준으로 승인되었는데, XML 문서 구조를 기술하는 메커니즘인 DTD(Document Type Definition)는 XML을 위한 모델 언어 또는 스키마이다. XML 1.0에서는 XML 문서 타입을 정의하기 위한 메커니즘을 제공하지만, XML 개발자들은 DTD를 구현한 경험에 의해 DTD의 수많은 단점과 내용 모델을 정의하는 더욱 포괄적이고 엄격한 방법이 필요하다는 것을 알게 되었다. 그래서, 많은 그룹들이 DTD를 보완한 XML-Data, DCD(Document Content Description), SOX(Schema for Object-oriented XML), DDML(Document Definition Markup Language), DT4DTD 등의 XML 스키마의 제안들이 W3C에 제출된 스키마 관련 노트(note)들로 제안되었고, 그 결과 2001년 5월 1일에 W3C에서 표준으로 승인된 XML Schema이다.

XML문서의 스키마인 DTD(Document Type Definition)의 구성 요소는 element, attribute, entity, notation등으로 구성한다. XML의 엘리먼트 콘텐츠 모델의 정의는 XML 구성 요소의 순서(sequence)를 쉼표(,)에 의해 표현하고, 구성 요소의 선택(choice)을 세로막대(|)으로 표현한다. 엘리먼트의 카디널리티(cardinality)은 Optional(?), Zero or More(\*), One or More(+) 기호로 정의한다. 결과적으로, XML DTD는 EBNF 표기에 의한 정규 표현식(regular expression)으로 정의된다. 관계형 데이터베이스 설계단계에서 업무 요구사항을 수집하고 분석한 후, 고수준 개념적 데이터 모델을 사용하여 데이터베이스에 대한 개념 스키마를 생성한다. 이 단계를 개념적 데이터베이스 설계 단계라고 하는데, 이때, 고수준 개념적 데이터 모델링의 대표적인 방법으로 ER모델의 정보 모델링 기법을 사용한다. ER 모델은 엔티티(entity type), 관계(relationship type), 애트리뷰트(attribute)들로서 데이터를 기술한다. ER 모델에 의해 업무 요구사항들은 ER 다이어그램이라는 그래픽 표기로 데이터베이스 응용을 위한 스키마를 어떻게 나타내는지 보여준다.

ER모델의 구성요소는 엔티티 타입(entity type), 애트리뷰트(attribute), 관계 타입(relationship type)으로 데이터를 기술한다. 일반적으로 관계 타입은 관계 인스턴스에 참여하고 있는 엔티티들의 가능한 조합을 제한하는 제약조건을 가진다. 관계 제약조건은 카디널리티 비율(cardinality ratio)와 참여 제약조건(participation constraint)라는 2가지 유형으로 구분할 수 있는데, 카디널리티 비율은 엔티티가 참여할 수 있는 관계 인스턴스의 수를 지정한다. 일반적으로 이진 관계 타입에 대한 카디널리티 비율은 1:1, 1:N, M:N이다. 참여 제약조건은 한 엔티티의 존재가 관계 타입을 통해 연관되어 있는 다른 엔티티에 의존하는지의 여부를 명시한다. 참여 제약조건에는 전체 참여(mandatory)와 부분 참여(optional)의 2가지 유형이 있다. 전체 참여는 엔티티들의 전체집합에 속하는 모든 엔티티가 관계 타입에 의해서 다른 엔티티에 연관되어야 한다는 것을 의미한다. 때로 전체 참여를 존재 종속성(existence dependancy)이라고도 한다. 부분 참여는 엔티티들의 전체집합이 아닌 부분집합이 관계 타입에 의해서 다른 엔티티에 연관되어 있다는 것을 의미한다. 관계 타입의 카디널리티 비율과 참여 제약조건을 구조적 제약 조건(structural constraint)라고 한다. 이러한 제약

조건들은 관계가 표현하는 실세계의 상황으로부터 결정된다.

XML문서의 구조를 정의하는 DTD와 ER 모델의 구성요소를 분석하여 사상시킴으로서 ER모델에 대한 DTD 모델을 생성할 수 있다. 이렇게 생성된 DTD는 작은 세계(real world)를 모델링하는 정보 모델링 기법으로 간주되고, 이러한 DTD에 의해 생성된 유효한 XML문서는 다른 시스템들과 비즈니스 데이터를 공유할 수 있게 된다. 최종적으로 제시된 'XML을 위한 데이터 모델'에 의해 작은 세계(real world)을 모델링하는 XML DTD 모델에서 정의한 구조(XML structure)를 따르는 유효한 XML문서(validate XML document)들은 임의의 관계형 데이터베이스에 저장할 수 있게 된다.