

# XML 문서 통합을 위한 상품 표현 모델\*

하상호\* 이강석\* 서진수\*\*

\*순천향대학교 정보기술공학부, \*\*순천향대학교 경영학부

hsh@cse.sch.ac.kr, ksalice@orgio.net, suhksu@asan.sch.ac.kr

## A Product Model for Integrating XML Documents

Sangho Ha\*, Gangseok Lee\*, KunSu Suh\*\*,

\*Dept. of Information Technology, Soonchunhyang University

\*\*Dept. of management, Soonchunhyang University

### 요 약

최근 인터넷을 통한 전자상거래의 규모가 폭발적으로 증가하면서 B2B 거래품목을 합리적으로 표현하고 원활히 교환할 수 있는 표준화된 전자 카탈로그의 개발이 매우 중요한 과제로 떠오르고 있다. 전자 카탈로그에 표현된 상품정보의 내용과 표현은 중요한 표준화 대상이 된다. CommerceNet[1]에서는 전자 카탈로그에 표준화 대상과 범위를 상품분류체계, 게시표준, 전송표준, 포맷표준, 표현 표준, 디렉토리 서비스 등 총 6가지 분야를 제시하고 있다. 본 논문에서는 이들 표준화 대상영역 중에서 상품정보를 게시할 때 여러 가지 이질적인 상품정보를 하나의 표준적 프레임워크로 표현할 수 있는 개념적 모델을 기존의 연구들이 갖는 문제점을 극복할 수 있는 새로운 상품 표현 모델로 제시하고, 이런 모델을 갖는 XML 문서 형태의 상품 정보를 관계데이터베이스에 저장하고 질의하는 시스템을 구현한다.

### 1. 서론

최근 인터넷을 통한 모든 상품들은 전자카탈로그[2] 형태로 표현되고 거래되고 있다. 전자카탈로그는 구매자와 판매자 상호간에 상품 및 서비스에 대한 정보를 교환할 목적으로 네트워크를 통해 상품의 검색, 구매 및 판매업무를 도와주는 시스템으로, 상품과 서비스에 관한 정보를 담고 있는 웹 페이지나 상품 선택을 위한 참조물 또는 그 디렉토리 등을 말한다. 아직까지 전자 카탈로그에 대한 표준화된 상품정보구조와 저장방식이 없기 때문에 업체들간 상품정보의 공유, 비교, 그리고 효율적인 검색에 한계를 드러내고 있다. 이러한 한계를 극복하기 위해서는 거래되는 상품들을 표현할 수 있는 표준 모델을 구축해야 하며, 이를 위해 상품이 갖는 속성들을 유형화 할 필요가 있다. 상품의 속성 정보를 유형화한 기존의 연구들[3,4,5]을 살펴보면, 상품의 속성정보의 유형화 방안들이 각 연구들마다 서로 다르게 제시되어 왔음을 알 수 있다. 이들 각 연구들은 나름대로의 설득력과 직관적인 면이 있지만, 상품의 속성정보의 구분 기준이 명확하지 않고, 모든 이질적 상품들이 공통적으로 가지는 '공통속성정보'와 각 상품마다 고유하고 독특한 정보를 갖는 '고유속성정보'에 대한 구분이 없으며, 상품속성에 대한 분석단위가 존재되어 있다. 따라서, 기존의 연구들이 갖는 문제점들을 극복할 수 있는 새로운 상품 표현모델의 제시가 요구된다.

전자카탈로그는 주문 지불과 같은 비즈니스 프로세스와 통합이 가능하고 B2B 전자상거래의 경우에는 인터넷과 연결되기 때문에 이를 지원할 수 있는 언어로 작성되어야 한다. 따라서, 상품표현의 형식은 웹 문서 기술을 위한 표준이면서, 모든 유

형의 정보에 대한 데이터 기술 언어로서 점차 그 사용 폭을 넓혀가고 있는 XML[6] 언어 스키밍을 기반으로 작성한다. 즉, 상품 표현모델과 상품정보에 대한 표준화된 구조는 XML DTD로 표현한다. XML DTD로 표현된 상품모델들은 XML을 통해 상품 정보가 표현되고 XML 데이터를 데이터베이스에 저장하고 XML 문서 형태의 질의가 가능해야 한다.

본 논문에서는 상품정보를 게시할 때 여러 가지 이질적인 상품정보를 하나의 표준적 프레임워크로 표현할 수 있는 개념적 모델을 기존의 연구들이 갖는 문제점을 극복할 수 있는 새로운 상품 표현 모델로 제시하고, 이러한 모델을 갖는 XML 문서 형태의 상품 정보를 관계데이터베이스에 저장하고 질의하는 시스템을 구현한다.

### 2. XML 문서 통합을 위한 상품 표현 모델

이 장에서는 기존 연구들이 가지는 상품 표현 모델의 문제점들을 극복할 수 있는 상품 표현 모델을 제시한다.

상품 표현 모델의 구축은 상품 속성들을 유형화하는 기준들이 상호 배타적이고 완전해야 하고, 유형화된 상품의 속성정보들을 '공통속성정보'와 '고유속성정보'로 구분하여 표시하되, 모델의 범용성을 높이기 위해 가급적 공통속성정보의 비중을 높여 나가야 하며, 상품의 속성정보는 상품 자체에 속한 것으로 국한하고, 판매업체와 관계된 속성들은 제외해야 한다. 이러한 원칙에 의거하여 상품의 속성정보는 크게 분류체계 정보, 내용 정보, 유형정보의 3 가지 그룹으로 분류한다.

분류체계 정보란 현존하는 전체 상품 리스트를 구성하는 각 대표상품들을 구분하고 분류하는 정보이다. 본 논문은 상품정보를 받아들이고 표현할 수 있는 모델을 제시하는데 그 목적이 있으므로 분류체계 정보를 다루는 것은 본 논문 밖의 문제이

\*본 연구는 정보통신부의 대학 S/W 연구센터지원 사업에 의해 수행된 것임

다.

내용정보는 특정 대표상품을 설명할 수 있는 정보의 집합이다. 내용정보를 구성하는 속성들을 정할 때는 속성 정보의 유연성과 확장성을 동시에 고려해야 한다. 내용정보는 정보제공자에 따라 기본정보, 판매정보, 그리고 부가가치 정보로 구분한다. 기본정보는 상품, 모델명, 가격, 기능 등 상품의 원 생산자 또는 공급자 측으로부터 주어지는 상품의 기본적인 속성정보이다. 판매정보는 판매가나 판매상태와 같이 판촉을 위해 제공하는 속성정보이다. 부가가치 정보는 상품에 대한 소비자, 언론의 평가, 유사상품 및 보완적 상품에 대한 링크 등과 같이 생산자와 판매자 이외의 3자가 제공하는 정보를 말한다.

유형정보는 하나의 대표상품내에서 각 개별상품들을 구분하고 유형화하는 기준을 제공하는 정보로, 주로 소비자의 탐색편의를 도모하기 위해서 제공된다. 유형정보를 도출하기 위해서는 다음과 같이 다섯 가지 원칙을 고려한다. 첫째, 유형화 기준은 중복이 가능하고 배타적이지 않아야 한다. 둘째, 유형화 기준간의 관계는 독립적인 관계를 유지하고, 이들간의 관계(종속, 연관, 인접)는 별도의 정의에 의해 관리한다. 셋째, 유형정보는 소비자의 지식이나 관심의 깊이, 상품별 구매행동 패턴에 따라 다양한 검색이 가능하도록 병렬적으로 제시한다. 넷째, 대표상품내 모든 개별상품이 모든 차원의 유형정보와 관계를 가질 필요는 없으며, 적합성에 따라 선택적으로 관계를 설정한다. 마지막으로, 부분적으로 특정 상품과 상품간의 관계도 가능해야 한다.

위에서 제시된 상품표현 모델의 원칙에 기반하여, XML 문서 통합을 위한 상품표현 모델의 일반적 스키마는 그림 1과 같다.

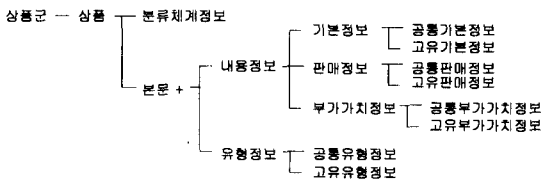


그림 1 XML 문서 통합을 위한 상품표현 모델

그림 1에서 보는 바와 같이 일반적인 상품 표현은 <상품군>을 루트 엘리먼트(Element)로 출발하여 그 아래에는 도서, 컴퓨터 등 각각의 대표상품을 표현하는 <상품> 엘리먼트를 갖는다. <상품> 엘리먼트는 특정 대표상품의 정체성(identity)을 나타내는 <분류체계정보>와 각 개별상품들에 대한 정보를 표현하는 <본문>이 있다. <본문> 엘리먼트에는 대표상품내의 각 개별 상품의 특성 정보를 나타내는 <내용정보>와 동일한 대표상품군내에서 개별상품의 유형을 나타내는 <유형정보>가 있다. 상품의 <내용정보>에는 상품정보를 제공하는 원천(information sources)에 따라 <기본정보>, <판매정보>, <부가가치정보>를 포함한다.

상품표현을 위한 일반적 스키마의 단말 엘리먼트인 <기본정보>, <판매정보>, <부가가치정보>, <유형정보>를 구성하는 하위 엘리먼트의 경우 엘리먼트의 이름과 구조가 대표상품마다 다를 수 있다. 그러나 상품표현모델의 범용성과 모든 개별 대표상품에 대한 표현성을 위해 대표상품들간의 '공통 엘리먼트'와 특정 대표상품에만 존재하는 '고유 엘리먼트'를 정의하고 도출할 필요가 있다. 대표상품들간의 공통엘리먼트는 다음과 같은 조건을 만족해야 한다. 첫째, 모든 대표상품들이 명칭은 다르더라도 본질적으로 같은 내용의 속성을 가진다. 둘째, 소비자의 상품선택 의사결정에 공통적으로 중요한 역할을 한다. 셋째,

해당 엘리먼트의 하부구조가 같다. 따라서 모든 대표상품들이 가지고 있는 속성이지만 소비자의 구매 의사결정에서 중요도가 다르거나, 해당 엘리먼트의 하부구조들이 다른 것은 공통엘리먼트에서 제외한다. 공통엘리먼트에서 제외된 상품표현을 위한 엘리먼트들은 고유엘리먼트가 된다.

### 3. 적용

여기서는 도서와 컴퓨터 상품에 대해 앞에서 제시한 상품표현 모델에 적용해 보고자 한다.

그림 2는 도서와 컴퓨터 상품을 통합한 상품표현 모델을 보여준다. 모든 상품에 공통적으로 나타나는 엘리먼트는 공통 엘리먼트 도출 조건에 따라 도출하여 <공통엘리먼트정보> 하부에 표시하였다. 대표상품들이 명칭은 다르더라도 본질적으로 같은 속성을 갖는 공통 엘리먼트에 대해, 도서의 경우, ( )안에 도서의 개체에 해당하는 엘리먼트의 명칭을 부여하였다. 도서와 컴퓨터 상품이 가지는 고유 엘리먼트정보는 각 상품의 표현 정보에서 공통 엘리먼트정보를 제외한 것으로, <도서/컴퓨터 고유엘리먼트정보> 하부에 표시하였다. 또 반드시 표시되지 않아도 되는 부가적인 엘리먼트들은 모두 '?'를 붙여 표기함으로써 도서와 컴퓨터 상품의 통합 DTD에 유효한 XML문서의 수용범위를 넓히고자 하였다.

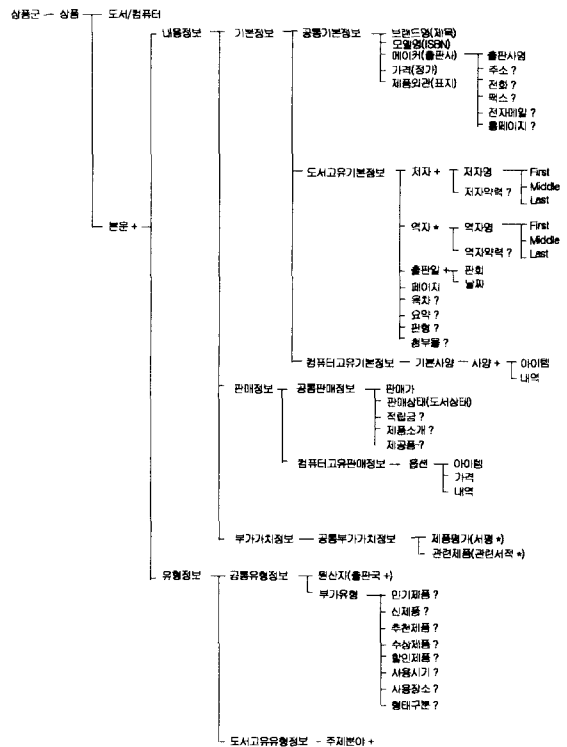


그림 2 도서 및 컴퓨터 통합 상품표현 모델

이를 다른 상품에 확대 적용하는 문제도 공통 엘리먼트는 그대로 사용하면서, 각 상품에 대해 고유 엘리먼트만 도출하면 각 상품에 대한 스키마를 쉽게 표현하고 구성할 수 있다.

4. 구현

여기서는 앞에서 적용한 도서 상품과 컴퓨터에 대해서 XML DTD 기반 상품 통합 구조를 작성하고, 통합 DTD에 따라 작성된 XML 문서를 관계 데이터베이스에 저장하고 질의하는 모듈들을 구현한다.

XML 문서의 저장과 검색 모듈의 전체 시스템 구조는 그림 3과 같다. 시스템은 크게 사용자 부분, 시스템 부분, 그리고 하부 구조 부분의 3부분으로 구성되어 있음을 볼 수 있다. 3부분 모두 별도의 플랫폼으로 독립적으로 구성될 수 있다. 즉, 시스템은 독립된 플랫폼 상에 서버로 구축된다. 사용자는 시스템 서버에 접속하여 시스템에 저장이나 검색을 요청할 수 있으며, 시스템은 다시 데이터베이스 서버에 접속하여 사용자의 요청을 전달하여 수행케 하고, 그 결과를 전달받으며, 다시 이를 가공하여 사용자에게 전달한다.

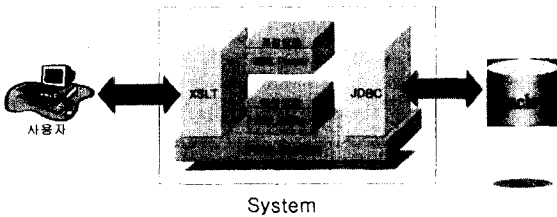


그림 3 전체 시스템 구조도

구현된 XML 문서의 저장과 질의 시스템은 웹과 사용자간의 인터페이스로 Java Servlets를 사용한다. 데이터베이스는 Oracle 8i를 사용하고, 시스템과 하부 데이터베이스와의 연동을 위해서 JDBC를 사용한다. 사용자 질의 결과를 XML 스트림으로 변환하고 사용자에게 볼 수 있는 형태로 변환하기 위해서 Oracle 사에서 제공하는 XDK(XML Developer's Kit)[7]인 SQL Utility와 XSLT(XSL Transformation)를 사용한다. 이 시스템은 UNIX 계열 및 윈도우즈 서버 계열에서 동작이 가능하다.

세부적인 동작 방법은 다음과 같다. XML 문서를 저장하기 위해서는 관계데이터베이스 스키마를 생성해야 한다. 데이터베이스 스키마는 도서와 컴퓨터 상품의 통합 DTD를 이용해서 스키마 생성규칙[8,9]에 따라 생성한다. 스키마가 생성되면, 통합된 DTD에 따라 각각의 상품 정보를 포함하는 XML 문서를 작성한다. 작성된 XML 문서를 저장[8,9]하는 과정은 다음과 같다. 사용자는 웹을 통해 저장할 상품에 대한 XML 문서를 저장 모듈에 전달한다. 저장모듈은 XML Parser를 사용하여 주어진 XML 문서를 파싱하여 유효성을 검증하고, 그 결과로 생성된 DOM 트리를 방문하면서 저장할 데이터를 추출하여 SQL 문을 생성하고, 이를 JDBC를 통해서 Oracle 8i 데이터베이스에 전달하면, Oracle 8i는 SQL 문을 통해 전달받은 데이터를 자신의 데이터베이스에 저장한다. XML 문서를 저장하는 모듈은 URL 정보를 입력받아 저장하는 방법과 에디터를 통해 작성된 XML 문서를 스트림 방식으로 저장하는 방법을 지원한다. XML 문서를 검색[10]하는 과정은 다음과 같다. 사용자가 필요한 정보를 조합하여 질의를 요청하면, 검색모듈은 이를 토대로 SQL 문을 생성하고 이를 JDBC를 통해서 Oracle 8i에 전달한다. Oracle 8i는 전달받은 SQL 문을 수행하여 그 결과를 SQL Utility를 사용하여 XML 스트림으로 생성한다. 마지막으로 XSLT를 사용하여 생성된 XML 스트림을 사용자가 볼 수 있는 형태로 제공하기 위해 스타일 시트를 이용하거나, HTML 문서 형태로

변환한다.

5. 결론

본 논문에서는 XML 문서 통합을 위한 상품표현 모델을 제시하였고, 이 모델을 적용하여 도서와 컴퓨터의 상품을 통합하여 표현하였다. 또한, 제안된 상품표현 모델을 지원하는 XML 기반 DB 모델을 구현하였다. 제안된 상품 표현 모델은 상품의 속성정보들을 개념적으로 상호 배타적이면서 완전하게 표현할 수 있도록 고안되었다. 즉, 상품의 속성정보를 상품간 구분 정보인 분류체계 정보, 상품 자체에 대한 정보인 내용정보, 동일 대표상품내 개별상품간의 구분정보인 유형정보로 분류하였으며, 내용정보는 다시 정보제공의 원천에 입각하여 기본정보, 판매정보, 부가가치정보로 구분함으로써 모든 상품을 하나의 틀 속에서 일관성 있게 표현할 수 있도록 하였다. 또한 모든 이질적 상품들이 공통적으로 가지는 '공통속성정보'와 각 상품마다 고유하고 독특한 정보를 갖는 '고유속성정보'를 고려하였고, 모델의 범용성을 높이기 위해 가급적 공통속성정보의 비중을 높이는 방향으로 DTD 모델을 작성하였다. 제시된 상품 표현 모델은 이론적으로 모든 상품들을 표현할 수 있는 범용적인 틀 (framework)을 갖추었고, 상품표현에 있어 표준화된 프레임워크를 제시하고 있으며, 전자 카탈로그간의 호환성 재고에 기여할 것이다.

향후 연구과제로는 모든 상품의 속성정보를 '공통속성정보'의 틀 속에 편입하여 완전한 의미에서의 범용 상품표현 모델을 개발할 필요가 있다.

참고 문헌

- [1] CommerceNet. "Catalogs for the Digital Marketplace," Research Report, www.commerce.net, march 1997.
- [2] Segev, A., Wan, D. and Beam, C., "Designing Electronic Catalogs for Business Value: Results of the CommerceNet Pilot," CITM working paper, CITM-WP-1005, Haas School of Business, University of California, Berkeley, October 1995.
- [3] 서울대학교 컴퓨터신기술공동연구소, "전자상거래를 위한 디지털 카탈로그 표준 및 서버기술," 1999.
- [4] 한국전산원, "전자 카탈로그 관련 기술 및 사업의 현황분석과 개선방안," 정보화 정책이슈 99-지원-01, 1999. 6
- [5] Price Waterhouse Coopers, "CommerceNet Catalog Interoperability Pilot Evaluation Report," Research Report #99-09, 1999. 3. 16
- [6] 2000 "Extensible Markup Language (XML)," http://www.w3.org/XML/
- [7] Oracle Inc. 1999 Oracle XML Developer's Kit , http://technet.oracle.com/tech/xml/info/htdocs/xdk.htm
- [8] 이상호, 이강석, 백인천 "Oracle8i XDK를 이용한 XML 문서의 저장", 멀티미디어학회, 2000년도 춘계학술대회, 2000. 5.
- [9] 이상호, 이강석, "Oracle8i XDK를 이용한 XML 문서의 질의", 전자공학회, 하계종합학술대회 논문집 제23권 제1호 pp71-74, 2000. 6.