

XML을 이용한 웹기반 상점 설계 및 구현

배수선, 박승수
이화여자대학교 컴퓨터학과

Design and Implementation of a Web-based Shop using XML

Su Sun Bae, Seung Soo Park

Dept. of computer Science and Engineering, Ewha Womans University

요 약

기존 웹상 HTML 사용의 경우 문서 브라우징의 역할밖에 못하여 그 자체로는 의미가 없었기에 실행 가능한 프로그램을 따로 설계하여야 했다. SGML의 서브셋인 XML을 사용할 경우 단순한 문서가 아닌 객체로 취급되어 구조화된 데이터를 표현하고 전달하고 교환할 수 있다. 이러한 이유로 JAVA와 XML을 이용한 융통성있는 웹 어플리케이션 개발이 큰 관심거리에 속하고 연구가 이루어지고 있다.

본 논문에서는 웹 콘텐츠를 XML로 작성하고 JAVA를 기반으로 XMLDOM을 이용하여 그 객체를 언어와 간단한 인터넷기반 마켓에 적용시킨다. 또한 그 자료를 DATAMINING에 활용하는 시스템을 설계, 구현하였다. 서버에 별도의 데이터베이스를 만들고 그것을 클라이언트에 연결하는 복잡한 방법을 XML 객체를 사용하는 방법으로 간단하게 대체하였다. 데이터는 XML객체로 이루어져 상점거래가 이루어지고, 이것은 고객분석 및 상품분석을 위한 DATAMINING에 사용할 수 있게 해준다. 이로써 단순한 XML이 여러 다양한 데이터베이스로써 역할을 하게 된다.

1. 서론

정보화의 홍수시대에서 정보를 어디서, 어떻게 얻느냐는 더 이상 중요한 것이 아니다. 인터넷, 월드 와이드 웹등의 발달로 정보의 대량화, 대중화가 이미 이루어져 수많은 정보가 산재해 있으나 중복된 경우가 흔하며 오염된 데이터가 많다. 그렇기에 웹콘텐츠 등을 이용한 일반 검색의 경우 정확하고 유익한 원하는 정보를 얻기가 힘든 경우가 많다. 이러한 단점을 해결하기 위해서 구조화된 데이터를 표현하고 전달할 수 있는 XML의 연구가 활발히 진행되고 있다. 이를 사용함으로써 의미 있는 웹콘텐츠를 만들 수 있고 단지 문서가 아니라 데이터로써 활용이 가능한 것이다. 또한 웹콘텐츠를 대상으로 하는 데이터마이닝에 이 XML을 이용할 경우 기존 HTML에 비해서 손쉽게 자료를 얻어 기타 여러 데이터베이스처럼 활용 가능하다. 이러한 이유로 XML 데이터를 활용하는 시스템을 설계, 구현하였다.

XML을 이용하여 상점거래를 하는 상점을 만들고, 여기서 생성된 상품거래 내역은 데이터마이닝을 위한 데이터베이스로써 활용한다. 서버에 별도의 데이터베이스를 두고 클라이언트에 연결하는 기존방법을 간단히 XML로 작성하고 JAVA기반

DOM을 이용해 객체를 얻어오는 방법으로 대체한다. 이렇게 생성된 객체들은 또한 데이터마이닝을 위한 데이터베이스로써 활용하므로써 단순한 데이터를 여러 유용한 데이터로 활용한다.

2. XSED 설계

2.1 XML(eXtensible Markup Language)

XML은 구조화된 데이터를 표현하고 전달하고 교환할 수 있는 텍스트 기반 포맷이다. 이는 SGML을 웹에 최적화시킨 SGML의 서브셋으로 W3C에 의해 정의된 것이다. 표준 HTTP 프로토콜을 통해 데이터를 전송할 수 있으며, 웹기반 어플리케이션을 통해 데이터를 표현하고 교환하기 위한 중요한 발전중의 하나이다. 이 XML을 이용하여 상품리스트 및 고객거래내역 데이터를 작성, 변경, 삭제하며, 데이터마이닝을 위한 데이터로도 사용된다.

2.2 DOM(Document Object Model)

DOM은 프로그램에 의한 문서구조, 내용, 스타일 시트를 창조, 변경하기 위한 프로그램 인터페이스이다. 이는 HTML, XML, CSS 스타일시트를 대상으로 하며 플랫폼, 프로그래밍

* 본 연구는 KOSEPI(98-0102-01-01-3)의 지원에 의해서 수행되었음.

언어에 독립적이다. 이러한 특징을 이용하여 JAVA를 기반으로 XML 객체를 뽑아 내오는데 DOM을 사용하여 데이터로써 사용한다.

2.3 데이터마이닝

데이터마이닝은 대용량의 데이터베이스에서 정보 다이아몬드를 탐색하는 것이다. XSED에서 특정 데이터 베이스가 없더라도 쉽게 액세스 할 수 있는 XML을 이용하여 데이터를 얻어 데이터마이닝 데이터로 활용한다.

2.4 전체 시스템의 구성

웹컨텐츠의 데이터 활용을 위해 전자상거래 및 검색엔진등 XML이 활발히 연구되고 있으나, 이들은 어느 한 분야로써의 데이터 이용에 국한된 것이다. 이를 인터넷마켓의 상품거래를 위한 데이터로써도 활용하고 또한 이를 통해 생성된 것을 데이터마이닝 데이터로써도 활용함으로써 그 이용가치를 높이고자 하였다.

XSED에서는 각종 서비스를 위한 여러 데이터 소스 및 메인 프레임은 미리 서버에 위치해 두었고, 클라이언트인 데스크탑에서는 인터페이스가 위치, 그들 중간에서 서비스를 해주는 웹 서버가 있다. 인터페이스에서는 로그인후 인증단계를 거쳐 고객에게는 상품거래란으로 관리자에게는 관리 및 분석란으로 나누어 원하는 정보를 입력할 수 있다. 여기서 받은 질의를 통해 웹서버에서는 DOM을 이용해 XML로부터 객체를 얻어 오거나 생성, 변경 등을 한후 그 결과를 HTML로 내보내어 준다. 관리자가 상품분석을 요구한 경우는 상품거래XML을 대상으로 데이터마이닝한 결과를 애플릿으로 볼 수 있다.

XSED의 전체적인 구성은 다음 그림 1과 같다.

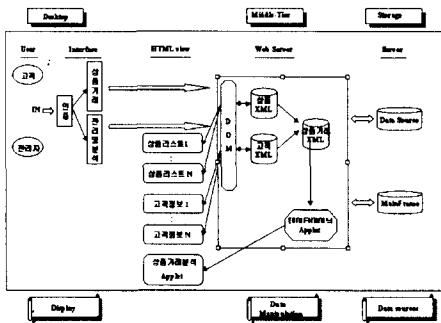


그림 1 전체 시스템 구성도

3. XSED의 구현

XSED의 구현을 위하여 데이터베이스의 역할을 하는 데이터로써 XML 문서를 작성하였으며 이를 객체로 사용하기 위

하여 JAVA와 DOM을 이용하였다. 여기서 만들어진 XML들은 IE 5.0 내 HTML보기로 이루어진다. 데이터마이닝 도구는 JAVA환경 애플릿으로 작성되었다. JAVA 개발도구는 JDK 1.1.7을 사용, XML parser로 IBM사의 xml4j_1_1_14.jar를 사용하며, IE 5.0내 JAVA 스크립트 XMLDOM 객체를 이용한다.

3.1 XML 데이터 작성

XSED에서 사용되는 데이터들은 XML로 작성되어 그 객체를 얻어 사용하는데 크게 상품카달로그에 대한 자료와 고객에 대한 자료, 데이터마이닝을 위한 자료로 나눌 수 있다. 상품에 대한 XML은 미리 상점관리자가 작성하며, 후에 이를 변경하거나 삭제할 수 있다. 고객은 이 상품리스트를 여러 가지 형태로 보고 상품에 대한 정보를 얻거나 구매를 할 수 있다. 고객에 대한 XML은 고객이 새로 계정을 신청할 때 한 객체씩 삽입되며, 다시 로그인 할 경우는 변경되는 형태이다. 상품을 거래한 후 그에 따른 상품거래내역 역시 마찬가지로 사이사이 변경된다. 상점관리자가 상품분석 및 고객분석을 원할 시 데이터마이닝을 위한 데이터베이스 역할을 하는 XML을 형성하여 이를 처리하여 사용한다.

XSED에서의 일부 XML 다음 그림 2와 같다.

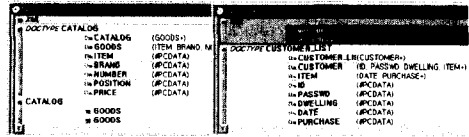


그림 2. XML Sources

3.2 사용자 인터페이스 작성

본 시스템의 인터페이스는 인터넷의 IE5.0이상 브라우저에서 사용가능하며 HTML 보기로 이루어진다. 상점고객의 경우 사용자 확인 및 정보변경, 상품검색 및 상품거래를 위해 사용되며. 상점관리자의 경우 사용자 확인 및 상품리스트 관리, 상품과 고객 분석을 위한 인터페이스로써 사용된다.

3.3 DOM 객체를 사용한 XML 객체 얻어오기

일반 상품리스트 및 고객정보를 위한 XML은 자체 XML을 지원하는 IE5.0 브라우저를 이용한다. 여기서 JAVA스크립트를 작성하여 XMLDOM을 통해 이미 작성되어 있는 XML 객체를 얻어오거나, 변경 및 수정, 삭제를 할 수 있다.

데이터마이닝을 위한 XML은 상품거래에서 생성된 고객 및 상품 XML을 고객관리자가 분석을 원할 때 적당한 형태로 변경된 XML을 사용한다. 이를 위해서 데이터마이닝도구는 애플릿으로 작성되었고, 이는 IBM XML parser를 통하여 JAVA

환경에서 DOM을 이용하여 XML객체를 얻어올수 있다. 상품리스트의 일부는 그림 3과 같다

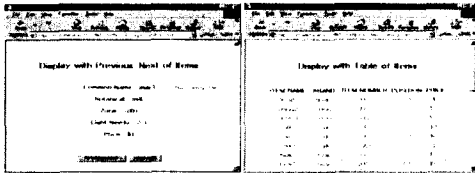


그림 3. DOM 이용해 표현한 상품리스트

3.4 XML 객체를 사용한 상점거래

상품에 관한 리스트들은 상점관리자들에 의해 미리 작성되어져 있는 상태이고 고객들이 사용자 확인후 여러 가지 형태로 그 상품을 볼 수가 있다. 똑같은 XML 데이터를 사용하는데, 이 데이터를 얻어와서 어떻게 배치를 달리 하느냐에 따라 그 보기가 달라지는 것으로 사용자가 편한 것을 선택하여 볼 수 있다. 상품검색을 통한 보기도 가능하며 이를 통해 구매 또는 취소할 수 있다. 이때 각 고객의 상품거래 XML이 추가 또는 변경, 삭제되는 것이다.

3.5 XML 객체를 이용한 데이터마이닝

상점관리자가 거래된 상품이나 고객들의 취향을 분석하려고 할 때 사용되는 것으로 이미 거래된 내역의 XML 데이터의 일부 수정된 XML데이터를 사용한다. 이때 애플릿으로 작성된 데이터마이닝 도구가 수행되며, 연관관계 룰을 사용한다. 이를 통해 상품들간 상관관계, 고객이 주로 이용하는 것 등을 분석할 수 있다.

마켓거래후 생성된 데이터마이닝을 위한 XML은 그림 4와같다.

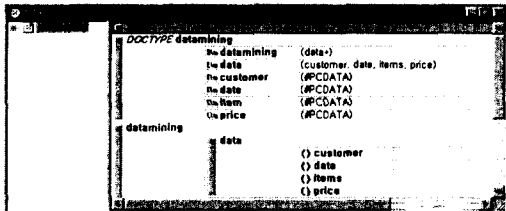


그림 4. 데이터마이닝을 위한 XML

3.6 구현예

XSED에서는 고객일 경우 인증 후 상품거래를, 관리자인 경우 관리 및 분석을 할 수 있다. 고객들은 원하는 상품리스트를 여러 가지 형태의 리스트로 보고 구매를 할 수 있고, 관리자는 고객들이 일정한 시간동안 그러한 거래를 한 후 갱신된 고객 및 상품거래 XML을 이용해 상점에서 거래된 상품들간의 관계 분석을 위해 데이터마이닝을 할 수 있다. 예를 들어 그 상점에서 함께 팔릴 가능성이 높은 제품들을 비슷한 위치에 재배열하기 위해서 팔린 상품들간 관계를 조사한다고 할 때,

마시는 차를 대상으로 연관관계 룰을 이용해 데이터마이닝한 결과로 그림 5와 같다. 커피와 차, 커피와 실탕이 함께 팔린 사례가 많은 것으로 알 수 있다.

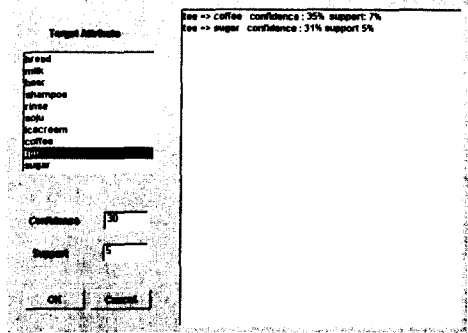


그림 5. 데이터마이닝한 결과의 예

4. 결론 및 향후과제

본 논문에서는 XML의 다양한 활용을 위해 비교적 간단한 XML을 이용해 상품거래에서의 데이터로써 사용 및 상품분석을 위한 데이터마이닝의 데이터로써 사용을 하였다. 일반 서버에 여러 데이터베이스를 생성하고 항상 데몬을 띄워야 연결이 가능한 일반 방법보다 훨씬 데이터생성 및 데이터추출방법도 쉬우며 효율적으로 데이터를 사용할 수 있다.

본 연구를 통해 얻을 수 있는 기대효과는 합부로 접근, 혹은 연결하기 힘든 데이터베이스를 XML로 교체함으로써 간편한 데이터 사용을 할 수 있다는 것이며, 여기서 생성된 웹컨텐츠를 바로 데이터마이닝에 연결할 수 있다는 것이다. 하지만 XML 만으로는 유익한 데이터만 될 뿐 기능적인 측면을 보완하기 위해서는 에이전트 기법을 도입하여야 할 것이며 데이터마이닝도 여러 가지 기법을 사용할 수 있도록 하여야 할 것이다.

5. 참고문헌

- [1] Alex Homer, "XML IES", WROX, pp 184-247, 1999
- [2] " Extensible Markup Language (XML) 1.0 ", http://XML.t2000.co.kr/kor/XML/spec_korean.html, 1999
- [3] Frank Bounmphrey, et al. ,"XML application", pp 185-281, ADTEC, 1999
- [4] Rakesh Agrawal et al., "Fast Discovery of Association Rules", ADVANCES IN KNOWLEDGE DISCOVERY AND DATA MINING, pp 307-328, 1999
- [5] William J. Pardi, "XML in Action Web Technology", Microsoft, pp 76-136, 1999
- [6] Yuhichi Nakamura, "A Framework for Representing Agent Interaction Protocols Based on XML", IBM Research Tokyo Research Laboratory, 1998