

XML문서를 이용한 사용자 프로필 생성 시스템 설계 및 구현

권순덕^o 고재진 정승훈 양상석
울산대학교 컴퓨터·정보통신 공학부
{winder2000^o, jjkoh, noru37, winli}@mail.ulsan.ac.kr

Design and Implementation User Profile Creation System using XML Document

Sundeok Kwon^o, Jaejin Koh, Seunghoon Jung, Sangseok Yang
Dept. of Computer Engineering & Information Technology, University of Ulsan

요 약

본 논문은 다양한 인터넷 콘텐츠를 제공하는 시스템에서 XML문서를 이용한 사용자 프로필 생성 시스템을 설계하고 구현하였다. 본 시스템은 구매와 같은 하나의 트랜잭션이 발생하기 전에 사용자 구매 패턴을 분석하기 위해서 필요한 프로필을 정의하는 모듈과 이를 적용한 사용자 프로필 생성 모듈로 구성되어 있다. 프로필 정의 단계에서는 XML의 DTD를 이용해서 사용자 프로필 규칙을 정의하고 저장한다. 그리고 사용자 프로필 생성 모듈에서는 트랜잭션이 일어날 때 발생하는 XML문서를 파싱해서 XPath형태로 저장되어 있는 규칙 리스트에 적용한 후 필요한 XML문서의 요소 값들을 이용해서 사용자 프로필을 생성한다. 또한 추가적인 장점으로는 하나의 복잡한 모듈이나 패키지뿐만 아니라 제공되어지는 규칙을 사용하지 않고 관리자나 사이트를 운영하는 운영자가 직접 규칙을 정의하고 규칙에 대한 설명을 통해서 자신의 환경에 맞는 사용자 프로필에 대한 정보 추출 및 CRM(Customer Relationship Management)을 지원 할 수 있다.

1. 서 론

웹에서 제공하는 일반적인 정보 검색 시스템은 질의에 적합한 문서들을 검색하여 사용자에게 제공함으로써 대용량의 데이터로부터 주어진 시간 내에 원하는 정보를 발견할 수 있도록 도와준다[1]. 또한 정보 검색 시스템에서는 순위 결정 방법(Ranking)을 적용하여 유사도(Similarity)를 계산하고 계산된 유사도에 따라 문서에 순위를 부여하여 사용자에게 높은 순위의 문서를 우선적으로 검토함으로써 필요한 정보를 얻는데 소모되는 시간을 최소화할 수 있다[2]. 또한 정보 선택 제공 시스템(Selective Dissemination of Information)은 분산되어 있는 데이터를 적합한 사용자에게 제공하기 위해서 단순 키워드 매칭(keyword matching)[3] 보다 더 효율적인 사용자 프로필을 이용하고 있다. 또한 전자상거래에서도 주어진 질의를 이용한 검색보다는 사용자의 프로필(Profile)을 이용한 검색과 정보 제공이 더 효율적이다[4]. 새로운 트랜잭션(transaction)이 발생할 때마다 각 사용자에게 적합한 프로필을 작성하거나 추가해서 제공하고자 하는 정보를 원하는 사용자에게 전송한다. 또한 각 사용자의 프로필을 이용해 관리자 관점에서 필요로 하는 정보를 구하거나 이용할 수 있다. 이 논문에서는 XML문서를 기반으로 하는 사용자 프로필을 형성하고 이용하는 방법을 제시하고자 한다. 기존의 사용자 질의(Query) 문을 파싱해서 사용자 프로필을 생성하는 시스템[5]과는 달리 새로운 XML문서가 발생하는 시점에서 XPath형태의 규칙(Role)을 이용해서 사용자의 프로필을 생성한다. 또한 관리자 관점에서 필요한 정보를 구하기 위해서 규칙을 생성할 수 있는 툴(Tool)을 제공함으로써 사용자의 프로필을 이용한 통계 및 다양한 정보를 구할 수도 있다.

2. 관련 연구

2.1 XML

월드 와이드 웹 컨소시엄(W3C, World Wide Web Consortium)은 1990년 중반에 SGML의 유연성과 HTML의 단순성을 결합한 마크업 언어인 XML(eXtensible Markup Language)를 만들었다.

2.2 XML 스키마

자기 기술적(self-describing)이며 확장 가능한 언어인 XML의 장점을 극대화하기 위해서는 어떤 형태의 XML 문서에서도 허용될 수 있는 내용을 정의하고 기술하는 방법이 필요하다. 과거에는 XML 문서의 적합성을 검사하기 위해 DTD를 이용했지만, XML 데이터를 다루기에는 많은 부분에서 한계를 차지하고 있었다. XML Schema는 DTD와는 달리 XML 문법 자체에서 허용된 문서 구조들을 기술할 수 있게 함으로써, XML 문서를 구축하는 절차를 좀 더 견고하게 해준다. XML 스키마는 마크업은 어떻게 사용되고 또 XML 문서 구조의 컴포넌트 부분은 어떻게 서로 잘 맞게 되는지 기술하게 된다. 또, XML 문서의 구조(structure), 의미론(semantics) 그리고 콘텐츠(contents)에 대한 정의도 수용할 수 있어서 DTD보다 훨씬 강력하고 유연한 메커니즘을 제공하고 있다[6].

2.3 XPath

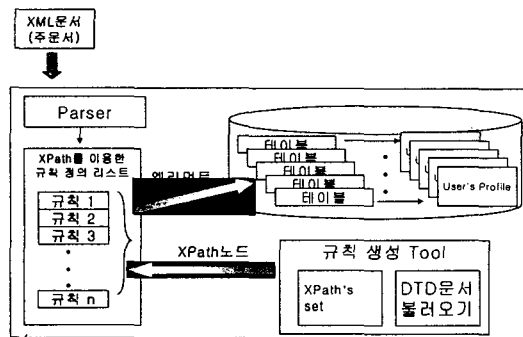
XML은 문서의 요소들을 노드(node) 개념으로 접근하여 소스 트리의 정확한 위치를 지정해주기 위한 경로 지정 문법으로 XML과 별개의 문법체계를 가지고 있다. XPath 단독으로 사용하지 않으며, XSL, XSLT에서 결과트리를 만들기 위한 경로 지정에 사용된다. 또한 XPath는 XSLT, XLink, XPointer에 의해 사용될 수 있는 XML 문서의 내용을 질의하고 위치를 부여하기 위한 일반 문법과 의미를 제공하기 위해 만들어진 언어로서 XSLT 1.0과 함께 1999년 11월 W3C에서 XPath 1.0으로 권고(Recommendation)되었다.

2.4 개인화(personalization)

개인화 전략은 개인별로 맞춤 서비스를 제공하기 위한 핵심적인 전략중의 하나이다. 콘텐츠(content)를 제공하는 사이트에서 각 사용자의 개인화를 표현하기 위해서 사용자 프로필을 이용한다. 이러한 각 사용자마다의 프로필을 이용함으로써 고객들에게 차별화된 서비스를 제공해 고객들과 친밀한 관계를 유지하여 충성도(loyalty)를 높일 수 있다[7].

3. XML문서를 이용한 사용자 profile 생성과 추천 시스템

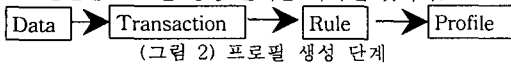
본 논문에서 제안하고자 하는 것은 특정 웹사이트(전자 상거래와 같은)의 관리자가 필요로 하는 규칙을 생성하는 방법과 생성된 규칙을 근거로 특정 XML문서(주문서)가 발생하면 사용자 프로필과 CRM(customer relationship management)에 필요한 정보를 추출할 수 있는 시스템을 설계 및 구현하고자 한다. 개략적인 설명은 다음과 같다. 우선 관리적 차원에서 필요한 규칙을 관리자나 사용자가 직접 정의할 수 있다. 이 때 생성된 규칙은 시스템내의 XPath(XML Path Language)의 형태로 리스트(List)라는 자료구조를 이용해서 저장한다. 그리고 특정 XML문서가 만들어지면 시스템 내에 정의된 규칙 리스트(List)에 따라 데이터베이스 내에 개별적인 테이블로 생성, 저장된다. 이런 방법으로 저장된 테이블은 규칙에서 정해진 엘리먼트(Element)의 값을 저장한다. 또한 엘리먼트를 기준으로 인덱싱(indexing)되기 때문에 검색 시간을 단축시킬 수 있다. 다음 (그림 1)은 본 논문에서 제안하는 XML문서를 이용한 사용자 profile 생성과 추천 시스템의 개념도이다.



(그림 1) 전체 시스템 흐름도

3.1 사용자 프로필(users profile)

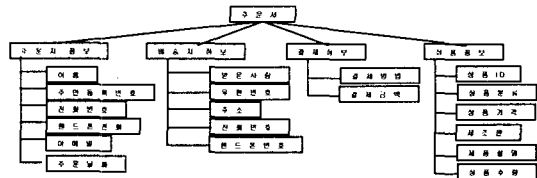
사용자의 프로필을 생성하는 단계는 크게 두 단계로 나뉘어진다. 첫 번째 단계에서는 주어진 데이터를 수집하는 것이고, 두 번째는 주어진 데이터를 이용해서 새로운 사실과 규칙을 만드는 것이다. 예를 들어 전자상거래 시스템에서는 고객의 이름, 주소, 생년월일 등과 같은 데이터를 주어진 데이터 혹은 사실(fact)라고 하고, 특정 기간동안의 거래 내역등과 같은 데이터를 새로운 사실 혹은 트랜잭션(transaction) 데이터라고 말한다 [7]. 또한 어떤 트랜잭션 데이터를 이용해서 사용자의 행동(behavior)에 따른 정보를 수집하고 저장한 후 이를 이용해서 새로운 데이터를 생성할 수 있는데 이를 사용자 프로필을 이용한 정보 수집이라고 말한다. 예를 들어, "구매자가 8월 한달 동안 가장 많이 구매한 품목이 맥주이다."와 같은 구매 트랜잭션 데이터는 "구매자의 8월 한달 동안의 구매 선호 품목이 맥주"라는 사용자 프로필로 표현 될 수 있다. 이러한 사용자 프로필을 통해서 각 사용자의 소비 패턴 분석, 개인별 선호도 그리고 다양한 정보를 관리자 입장에서 손쉽게 구할 수 있다. (그림 2)는 간단한 프로필 생성 단계를 나타낸 것이다.



(그림 2) 프로필 생성 단계

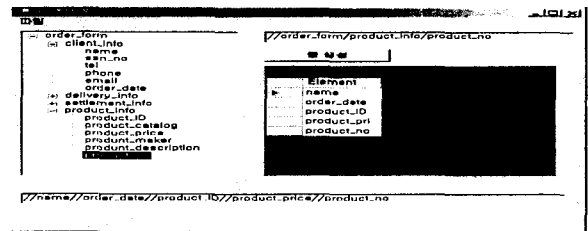
사용자 프로필을 생성하는 방법으론 XML의 DTD(Data Type

Definition)를 이용해 시스템에서 필요로 한 정보를 트랜잭션(Transaction)이 발생하기 전에 GUI 응용프로그램을 통해 새로운 규칙을 생성할 수 있는 규칙 생성 단계와 생성된 규칙을 통해서 사용자 프로필을 생성하는 프로필 생성 단계로 나누어진다. 또한 새로운 규칙이 필요한 경우 주문이 발생하기 전에 규칙을 생성해 둬으로써 차후에 발생하는 모든 사용자의 행동에 대해서 규칙을 적용할 수 있다. 이는 관리자 측면에서 필요한 규칙을 언제든지 간단한 방법을 통해서 생성할 수 있다는 장점을 가지고 있다. (그림 3)은 본 논문에서 사용되는 주문서의 구조를 표현한 것이다.



(그림 3) 주문서 구조

위에서 제시한 규칙(사용자의 특정 기간동안의 구매에 대한 정보)을 구하기 위해서는 주문서 DTD(Data Type Definition)에서 주문자 정보(이름, 주문 날짜), 상품 정보(상품ID, 상품분류, 상품 수량)등이 필요하다. 선택된 요소(Element) 집합은 XPath의 형태로 규칙 리스트(List)에 저장된다. 이렇게 생성된 규칙 리스트는 사용자의 새로운 주문이 발생할 때마다 주문서의 내용을 파싱(parsing)하고 사용자의 구매 정보를 파악하는데 사용된다. (그림4)는 규칙을 생성하는 GUI기반의 응용 프로그램을 보여주고 있다.



(그림 4) 규칙을 생성하는 응용 프로그램

위의 방법과 같이 생성된 규칙은 규칙을 식별하기 위해서 규칙 식별자와 함께 데이터베이스에 저장된다. 또한 규칙을 만드는 과정에서 규칙에 대한 설명과 규칙에 대한 이름을 부여함으로써 사용자의 프로필에 대한 상세 정보를 제공할 수 있다.

3.2 사용자 프로필 데이터베이스

사용자에게 적용된 규칙과 규칙의 정보는 데이터베이스에 저장된다. (표 1)에서는 구매 정보에 대한 사용자에게 적용된 규칙 정보를 보여준다. 사용자 프로필에는 구체적인 규칙의 정보를 기록하지 않고 규칙을 식별하기 위한 규칙 식별자만을 기록함으로써 사용자 프로필의 크기를 줄일 수 있다. 그리고 새로운 규칙을 적용할 때마다 새로운 행이 추가됨으로써 여러 개의 규칙을 추가적으로 적용시킬 수 있다.

(표 1) 사용자 프로필 구성 예

사용자 프로필		
사용자 ID	Rule_ID	적용여부
권순덕	Purchase_info	true

각 규칙에 대한 상세 정보는 표2와 같은 구성되어 규칙에서 필요한 요소에 대한 정보를 가지고 있다. (표 2)에서 가지고 있는 정보를 이용해서 XML문서(주문서와 같은)가 발생할 때 각 문서에 대해 엘리먼트의 값을 추출한 후 데이터베이스내에 규칙의 이름과 동일한 테이블을 생성 한다.

(표 2) 규칙의 정보

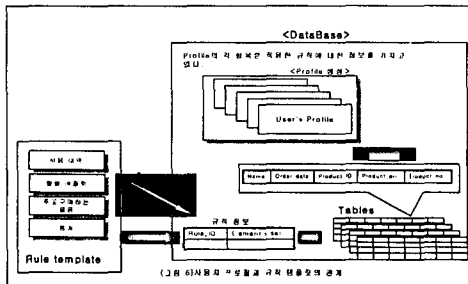
Rule_ID	Rule_Description	Element 정보 (규칙을 적용하기 위해서 필요한 요소)
Purchase_info	구매자의 구매 정보	Name, order_date, product_ID, product_pri, product_no

이와 같이 생성된 테이블은 사용자의 "구매정보"를 분석하는데 필요한 요소만을 가지고 있다. (그림5)에는 규칙에 따라 생성되는 데이터베이스내의 purchase_info 테이블의 구조를 나타낸 것이다.

Name	order_date	product_ID	product_pri	Product_no
권순덕	03/07/12	sa-100s	520,000	1
이건태	03/07/14	stv-120a	640,000	2

(그림 5) purchase_info 테이블

데이터베이스내의 테이블을 생성하기 위해서 규칙 템플릿을 이용한다. 먼저 테이블을 생성하기 위해서 테이블의 이름은 각 규칙의 이름을 따라서 테이블의 이름을 정하고, 각 테이블이 가지고 있는 필드(field)는 규칙 템플릿에서 사용한 각 요소(주문서에서 사용되는 엘리먼트)를 이용해서 테이블을 생성한다. 따라서 각 테이블은 하나의 트랜잭션(transaction)에 대해 규칙을 적용한 값이 들어간다. (그림6)은 사용자 프로필과 규칙 템플릿의 관계를 도식적으로 표현한 것이다.



(그림 6) 사용자 프로필과 규칙 템플릿의 관계

3.3 사용자 프로필을 이용한 필터링(Filtering)

트랜잭션(Transaction)이후에 생성되는 XML문서(예를 들어, 주문서와 같은 문서)는 XML 파서(parser)를 통해서 필터링 작업과 규칙을 기반으로 추출한 요소의 값을 데이터베이스에 저장하는 프로세스를 수행한다. 이번 논문에서는 이벤트 기반(event-based)의 XML 파서인 SAX를 사용하고 있다[9].

규칙을 작성할 때 생성된 요소의 집합을 기반으로 파싱한 XML문서의 요소 값을 데이터베이스에 저장한 후, 테이블에 저장되어 있는 내용을 XSLT(Extensible Style Language for Transformation)라는 간단한 언어를 이용해서 스타일 시트를 작성한다. 규칙에 대한 설명도 함께 표현 가능하다.(그림 7)는 스타일 시트에 적용한 사용자 프로필을 보여주고 있다.

Rule_description		필별 매출 현황			
user_ID	Name	order_date	product_ID	product_pri	Product_no
kwonsd	권순덕	2003/07/12	sa-100s	520,000	1
leekt	이건태	2003/07/14	stv-120a	640,000	2

(그림 7) 스타일시트가 적용된 사용자 프로필

4. 결론 및 향후 과제

본 논문에서는 사용자의 정보나 사용자의 프로필을 이용해 보다 효율적인 서비스를 제공하고자하는 시스템에서 활용될 수 있는 프로필과 규칙을 생성하는 방법에 대해 논의하였다. 또한 기존의 프로필 생성 방법과는 다른 방법으로 트랜잭션(Transaction)이 발생하는 시점에서 사용자의 프로필을 생성, 추가하고, 이를 관리할 수 있는 방법을 제시하였다. 그리고 프로필 생성과 규칙을 정의하기 위해서 플랫폼-독립적인 XML문서 형식을 이용하여 확장성과 다른 시스템과의 연계를 용이하게 하였다. 또한 하나의 복잡한 모듈이나 패키지로만 생성할 수 있는 규칙을 관리하나 사이트를 운영하는 운영자가 직접 규칙을 정의하거나 규칙에 대한 설명, 재 정의를 함으로써 다양한 시스템에서 필요로 하는 사용자 프로필에 대한 정보 추출 및 CRM(customer relationship management)을 할 수 있다. 앞으로 본 논문에서 제시하는 방법은 고객과의 지속적인 관계를 유지하기 위한 개인화 전략, 고객의 프로필 생성을 필요로 하는 마케팅 분야에서 응용할 수 있을 것으로 기대된다.

5. 참고 문헌

- [1] Salton, G., Historic note: The past thirty years in information retrieval, Journal of the American Society for Information Science, Vol. 38, No. 5, pp. 375-380, 1987.
- [2] Callan, J. and Connell, M. Query-based sampling of text databases, Technical Report IR-180 Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts. 1999.
- [4] Mehmet Altinel, Michael J. Franklin, "Efficient Filtering of XML Document for Selective Dissemination of Information", Proceedings of the 26th VLDB Conference, Cairo, Egypt, 2000.
- [5] Ugur Cetintemel, "Self-Adaptive User Profiles for Large-Scales Data Delivery", IEEE, 2000.
- [6] W3C, XML Schema Part I: Structures, W3C Recommendation, <http://www.w3c.org/TR/xmlschema-1>, Nov. 2001.
- [7] G. Rossi, D. Schwabe, D., Guimarães, R. M., "Designing Personalized Web Applications", Prof. of the International World Wide Web Conference (WWW), Amsterdam, 2001.
- [8] G. Adomavicius and A. Tuzhilin, Using Data Mining Methods to Build Customer Profiles. IEEE, February, 2001.
- [9] Megginson Technologies, "SAX 1.0: a free API for event-based XML Parsing", <http://www.megginson.com/SAX/index.html>, May, 1998.