

모바일 단말기를 위한 인터넷 상품 정보 제공 시스템*

허정수^o 하상호
순천향대학교 정보기술공학부
{majya^o, hsh}@sch.ac.kr

A System for Providing the Internet Product Information to Mobile Devices

Jung-su Hur^o Sangho Ha
School of Information Technology Engineering, Soonchunhyang University

요 약

무선 통신 기술의 발전에 따라, 오늘날은 무선 단말기 상에서 전자상거래가 이루어지는 모바일 커머스 시대이다. 웹상에는 수많은 인터넷 쇼핑몰들이 제공하는 유용한 대량의 상품정보가 존재하나, 이러한 상품 정보가 모바일 단말기의 특성 및 성능 제약으로 인해 모바일 커머스의 콘텐츠로 사용되기에는 적합하지 않다. 웹의 정보를 무선 단말기상에서 이용하기 위한 여러 연구가 진행되고 있으나, 모바일 커머스를 위한 웹 상품 정보 이용에 관한 연구는 아직 미흡하다. 본 논문에서는 인터넷 쇼핑몰의 임의의 HTML 문서에 대해서 상품 항목들을 추출하여 미리 정의된 상품구조의 XML 문서로 포장하여 모바일 폰의 클라이언트에 전달할 수 있는 시스템을 설계하고 구현하며, 이 시스템을 모바일 커머스에 적용한다. 또한, 구현된 시스템에 대한 실행 예를 보인다.

1. 서 론

오늘날 기술의 발전과 함께 많은 모바일 기기들이 보급되고 있다. 모바일 기기들은 개인의 생활에 점점 많은 부분을 차지해가고 있고 무선 네트워크의 발전과 함께 더욱 많은 정보를 제공할 수 있는 환경을 조성하였다. 또한 모바일 기기를 위한 개인 인증 시스템을 도입함으로써 모바일 커머스[1]의 기대를 증대시켰으며, 은행 업무를 모바일 기기로 이용할 수 있게 되었다. 그러나 이러한 기술의 발전이 있고, 발전을 위한 환경이 조성되었으나 사용자가 원하는 대부분의 정보가 웹 환경에 편중되어 모바일 커머스를 위한 상품 콘텐츠가 절대 부족한 형편이다. 문제의 해결을 위해 XML 문서를 이용하여 웹 환경과 모바일 기기 환경 간의 하나의 정보로 여러 플랫폼 적용에 대한 제안이 있었으나, 시스템 교체에 따른 비용 부담으로 인해 널리 사용되지 못했다. 또한, 일부 기업에서는 NTT DoCoMo의 i-모드에서와 같이 별도의 모바일 커머스 사이트를 구축하여 자신의 단말기에 상품 정보를 제공할 수 있으나, 이러한 콘텐츠는 단말기에 중속되며 웹 상품정보와 호환되지 않는다. Nokia의 웹브라우저인 Opera는 모바일 폰을 위한 HTML 브라우저로서, 웹 페이지를 모바일 폰 상에서 디스플레이될 수 있게 한다. 그러나 웹 상품정보가 필터링되지 않고 그대로 단말기에 전달되기 때문에 대량의 데이터 전송에 따른 속도 지연 및 사용자가 웹 페이지를 보기 위하여 많은 화면 스크롤 다운을 해야 하는 문제가 있다.

다른 관련 연구로, 웹의 HTML 문서로부터 개인의 정보를 추출하여 모바일 단말기에 전달하는 연구[2, 3, 4]가 진행되고 있다. [2]에서는 사용자가 관심 있는 웹 사이트를 방문하여 자신의 관심 사항을 반영하여 그 웹 사이트에 대한 scoping과 filtering을 통해서 개인화된 웹

뷰(Web View)를 생성한다. 그러나 웹 페이지로부터 정보 추출 단위는 HTML 문서이며 특정 데이터 항목이 아니다. 여기서 선택된 문서에 대한 하이퍼 링크는 HTML 태그들로 표현된다. [3]에서는 사용자가 관심 있는 웹 페이지 접근 과정을 Smart Bookmark로 기록하고, 해당 웹 페이지에서 추출하고자 하는 항목들을 XPath[5]로 기술함으로써 개인화된 웹 뷰를 생성한다. 그러나 일반 사용자가 XPath를 사용하여 추출하고자 하는 항목들을 서술해야 하는 문제점이 있다. [4]에서는 각 웹 페이지에 대해서 사전에 정의된 도메인 지식에 기반하여 그 페이지로부터 항목을 추출하는데 사용되는 wrapper를 자동으로 생성하고, 사용자가 요청한 검색 결과의 웹 페이지에 해당 wrapper를 적용하여 필요한 항목을 추출한다. 여기서 wrapper는 항목을 추출하는데 필요한 제반의 규칙이나 프로시저를 말한다[6]. 그러나 이 방법은 레이블과 값의 쌍으로 표현된 데이터 항목 추출에만 적용 가능하며, 테이블과 같이 레이블을 갖지 않는 항목에 대해서는 적용되지 않는다. 또한, 웹 페이지가 다른 서술 패턴을 갖는 두 가지 이상의 유형 정보를 포함할 경우에, 그 웹 페이지에 대해서 모두 적용 가능한 패턴을 찾을 수 없다는 문제점도 지적된다.

이 논문에서는 HTML의 임의의 문서에 대해서 항목들을 추출할 수 있는 시스템을 설계하고 구현한다. 시스템의 적용 대상은 모바일 커머스이며, 따라서 사용자에게 모든 상품에 대한 정보를 통일된 형식으로 보여주기 위해서 e-카탈로그[7]의 단순 버전인 3MP 모델[8]의 미리 정의된 상품 구조를 고려한다. 시스템 관리자는 쇼핑몰 사이트로부터 상품 구조에 포함된 항목들을 추출하기 위한 wrapper를 생성한다. 여기서 wrapper는 항목을 추출하는데 필요한 HTML 문서상에서 루트 노드로부터의 항목까지의 HTML 태그들의 시퀀스로 구성된 경로로 표현된다. 항목 추출 과정은 사용자 인터페이스를 통해서 반자동으로 이루어진다. 모바일 폰 사용자가 상품 검색

* 이 논문은 한국과학재단 과제 R05-2004-000-12565-0으로부터 지원되었음

시에 관련된 쇼핑을 사이트에서 검색하여 검색 결과의 웹 페이지에 대해서 해당 wrapper가 적용되어 항목들이 추출되고 추출된 항목들은 미리 정의된 상품구조의 XML 문서로 포장되어 사용자에게 전달되어 디스플레이된다.

2. 시스템 설계

그림 1은 논문에서 고려하는 모바일 단말기를 위한 상품정보제공 시스템에 대한 전체 구조를 보여준다. 시스템은 크게 쇼핑몰 사이트, 모바일 웹서버, 사용자 모바일 단말기의 3가지 요소로 구성된다. 모바일 웹 서버는 래퍼 생성기, PE 경로 추출기, PE 추출기, 상품정보 제공자로 구분된다. 쇼핑몰의 관리자는 래퍼 생성기를 사용하여 쇼핑몰의 정보를 PE 래퍼 저장소에 저장한다. 여기서 래퍼(wrapper)는 해당 HTML 문서로부터 추출하고자 하는 상품 요소에 대한 경로들로 구성된다. 상품정보제공자는 모바일 클라이언트의 사용자로부터 상품검색을 요청받고, 해당 쇼핑몰 사이트를 검색하고, 검색된 HTML 문서와, 해당 래퍼를 저장소로부터 가져와서 PE 추출기에 전달하여 해당 상품요소를 추출하고, 추출된 상품요소를 XML 문서로 취합하고 포장하여 사용자에게 전달한다.

2.1 상품 XML 문서 구조

그림 1에서 상품정보제공자가 추출된 상품요소를 XML 문서로 취합/포장하는데, 여기서 사용되는 XML문서의 구조는 e-카탈로그의 단순 버전인 3MP 모델[8]이다. 그림 2는 설명을 위해서 3MP 모델을 더욱 단순화시킨 버전이다. 이 구조에서 회사와 가격에는 각각 Alias와 Value의 항목을 갖는데, Alias는 상품 특성에 따른 항목의 유의어를 반영하는데 사용된다.

2.2 래퍼와 래퍼 저장소

래퍼는 래퍼 생성기에서 생성되는 XML 문서이며 지역 저장소인 래퍼 저장소에 저장된다. 래퍼는 쇼핑몰에 대한 검색이나 상품을 나타내는 HTML문서에 접근하기 위한 정보와 상품 HTML 문서로부터 추출하고자 하는 상품 요소에 대한 경로 정보들을 포함한다. 그림 3은 래퍼의 구조이다. 하나의 쇼핑몰을 나타내는 SM은 쇼핑몰 정보를 나타내는 SM_Info와 상품요소 경로를 나타내는 SM_PEPATH로 나뉜다. Search_Pattern은 쇼핑몰의 검색 엔진을 사용하기 위한 URL주소 패턴이며, Product_Pattern은 검색된 상품 HTML문서의 주소를 찾기 위한 공통된 URL패턴을 나타낸다. Node는 추출된 상품요소에 대한 문서 경로를 표현한다.

2.3 PE 경로 추출기와 래퍼 생성기

쇼핑몰 웹 서버 관리자는 래퍼 생성기를 호출하여 래퍼를 생성하고 래퍼 저장소에 저장한다. 관리자는 쇼핑몰의 HTML 문서로부터 추출하고자 하는 상품요소에 대한 경로를 웹 브라우저 상에서 point-click의 인터페이스를 통해서 파악한다. 상품요소 경로는 HTML 문서에서 그 요소에 대한 문서 트리 구조상의 경로로 정의된다. 이러한 경로 계산은 PE 경로 추출기에서 이루어진다. 래퍼 생성기는 주어진 HTML 문서에 대해서 추출하고자 하는 모든 요소에 대한 경로들과 그 쇼핑몰에 대한 접근 정보를 포함하여 래퍼를 생성하고 저장소에 저장한다.

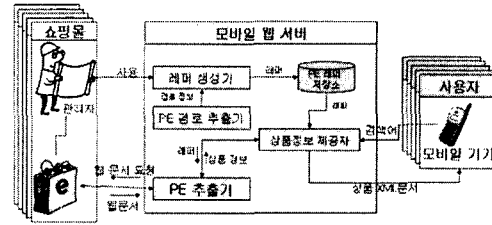


그림 1 시스템 구조

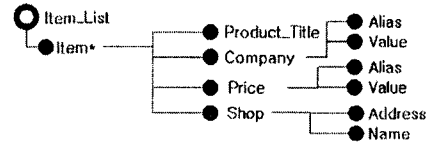


그림 2 상품XML문서 구조

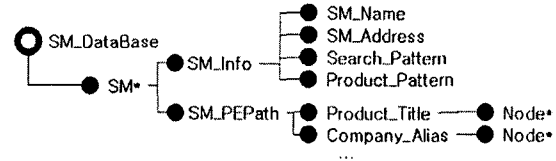


그림 3 래퍼의 구조

2.4 모바일 단말기

사용자의 모바일 단말기에는 사용자 질의를 처리하고, 모바일 웹 서버로부터 검색결과를 전달받은 XML 문서를 디스플레이하는 애플리케이션이 탑재된다. 상품 정보는 계층적 단계적으로 사용자와의 상호작용을 통해서 디스플레이된다.

2.5 PE 추출기

상품정보 제공자로부터 전달받은 검색어와 래퍼를 이용하여 쇼핑몰로부터 요소를 추출하고 XML 문서로 포장한다. 래퍼에 포함되어 있는 쇼핑몰 검색 엔진을 이용하여 해당 쇼핑몰의 상품을 검색어에 대해서 검색한다. 다음에, 검색된 HTML 문서를 파싱하여 래퍼의 상품 URL과 비교하여 검색된 각각의 상품 URL을 추출한다. 추출된 상품의 URL로 HTML문서를 접근한다. 상품을 표현하고 있는 HTML문서를 트리형태로 파싱한다. 트리형태로 변환된 HTML문서는 래퍼의 각 상품 요소 경로를 이용하여 해당 상품 요소를 추출한다. 추출된 개개의 정보를 취합하여 하나의 상품 정보를 생성하고 하나의 쇼핑몰에서 추출된 상품 정보는 다시 취합되어 상품정보 제공자에게 제공된다.

2.6 상품정보 제공자

모바일 단말기 애플리케이션이 사용자 검색어를 매개변수로 하여 상품정보 제공자를 호출한다. 이 모듈은 래퍼 저장소에 속한 각 래퍼에 대해서 사용자 검색어와 함께 매개변수로 하여 PE 추출기를 호출한다. 다음에, PE 추출기로부터 전달받은 XML 문서를 취합하여 한 개의 XML 문서를 생성한 후에 사용자 단말기에 전달한다.

3. 구현

시스템 설계에 의한 간단한 테스트를 위한 시스템을 구현했다. 기본적인 시스템은 Java 플랫폼을 채용하였

다. 그러나 HTML 문서 파싱을 위해 .NET의 Internet SDK[9]를 채용하여, HTML 문서 파싱이 이루어지는 PE 경로 추출기와 PE 추출기는 .NET 플랫폼에서 구현되었으며, PE 추출기는 Java 플랫폼과의 연동을 위해 IIS 서버를 이용하여 웹서비스로 구현되었다. Internet SDK의 mshtml 컴포넌트의 HTMLDocument 객체를 사용하여 HTML 문서를 파싱하여 경로를 추출하고, 경로를 이용하여 상품 정보 요소에 접근한다. 상품 정보 제공자는 Java 플랫폼에서 개발되었으며, 모바일 단말기 애플리케이션은 J2ME를 이용하여 개발되었다. 또한 상품 정보 제공자와 모바일 단말기 간의 통신을 위해 웹서비스를 이용한다. 모바일 단말기는 XML 문서와 웹서비스의 사용을 위해 Sun에서 제공하는 J2ME 개발 툴인 WTK 2.2를 사용하며, 실행은 J2ME 기본 애플레이터를 이용하였다. 상품 정보 제공자는 Apache의 Tomcat 서버와 웹서비스의 사용을 위한 Axis[10]를 이용하여 웹서비스로 구현되었다. 웹서비스의 타입은 .NET 플랫폼과 Java 플랫폼 모두 RPC방식의 Literal형으로 설정했다.

4. 실험 예

쇼핑몰로부터 상품 정보를 클라이언트에게 제공하기 위해 하나 이상의 쇼핑몰의 정보가 등록되어 있어야 한다. 래퍼 등록 작업은 래퍼 생성기를 통해서 이루어진다. 그림 4는 래퍼 생성기의 메인 화면으로, 쇼핑몰의 정보를 입력하고 경로 정보를 추출하는 화면이다.

그림 5는 모바일 단말기의 에뮬레이터로 사용자 클라이언트를 보여준다. 사용자가 모바일 단말기의 프로그램을 이용해 상품 검색을 위한 검색어를 입력하고 검색 버튼을 누르면, 입력된 검색어가 웹서비스에서 제공하는 원격 객체를 이용해 상품 정보 제공자에게 보내진다. 상품 정보 제공자는 사용자로부터 입력된 검색어와 관리자 프로그램으로 생성된 래퍼를 매개변수로 하여 원격 객체인 PE 추출기를 호출한다. PE 추출기는 검색어와 각각의 래퍼를 이용하여 상품을 검색하고 상품 정보 제공자에게 전달한다. 상품 정보 제공자는 하나의 쇼핑몰 단위로 받은 상품 정보를 취합한 XML 문서로 생성하여 모바일 단말기에 전달한다. 모바일 단말기는 전달된 상품 XML 문서를 디스플레이한다. 그림 6은 사용자 단말기 화면에 디스플레이된 검색 결과의 상품 리스트를 보여준다. 그림 7은 사용자가 특정 상품을 선택하였을 때, 그 상품에 대한 상세 사항을 보여준다.

4. 결론

논문에서는 HTML의 임의의 문서에 대해서 항목들을 추출할 수 있는 시스템을 설계하고 구현하였으며, 모바일 커머스에 적용하였다. 개발된 시스템은 개인화(personalization)[11]등 다른 웹 정보 추출을 위한 제반 분야에 적절하게 적용될 수 있다. 또한, wrapper가 XML로 기술되고, 추출된 정보가 XML 문서로 포장되어 클라이언트에 전달되기 때문에 시스템과 사용자 단말기 기종에 종속되지 않게 적용될 수 있다.

앞으로, 상품에 대한 선호도를 명세하는 사용자 프로파일과 확장된 질의어에 기반하여 사용자에게 개인화된 상품정보를 제공하는 기능을 추가하고자 한다. 이러한 기능을 위해서 추출된 항목들로 구성된 XML 문서에 대

한 scoping과 filtering 과정이 필요하다.

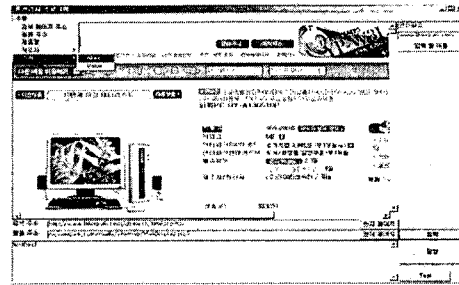


그림 4 래퍼 생성기의 메인 화면

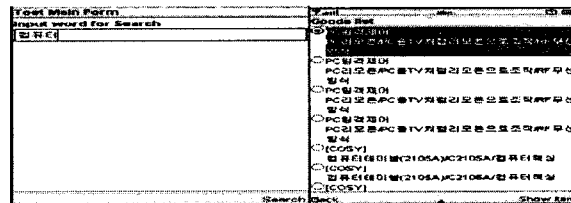


그림 5 검색어 입력

그림 6 검색 결과

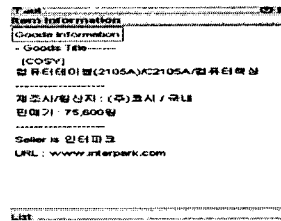


그림 7 상품 디스플레이

5. 참고 문헌

[1] Daniel Bilus, et al., "Adaptive Interfaces for Ubiquitous Web Access", CACM, Vol. 45, No. 5, 2002.
 [2] Zehua Liu, Wee-Keong Ng, Ee-Peng Lim, "Personalized Web Views for Multilingual Web Sources", IEEE INTERNET COMPUTING, p16~p22, July, August 2004
 [3] Juliana Ferie, Bharat Kumer, DanielLieuwen, "WebViews: Accessing Personalized Web Content and Services", ACM 1-58113-348-0, p.576~p.586, May 1-5, 2001
 [4] Jaeyoung Yang, Joongmin Choi, "Knowledge-Based Wrapper Induction for Intelligent Web Information Extraction", Springer, Web Intelligence, No.8, p.153~p.172
 [5] XPath, <http://www.w3.org/TR/xpath>
 [6] N. Kushmerick: Gleaning the Web. IEEE Intelligent Systems 14, 20-22, 1999.
 [7] 하상호, 김경래, XML 기반 상품 카탈로그의 설계 및 적용, 정보처리학회 논문지, 제 9-D권 제 3호, 2002년 6월.
 [8] 하상호, 인터넷 상품정보의 효과적 제공을 위한 모바일 시스템의 설계 및 구현, 정보처리학회 논문지 12-D권 제 3호, 2005년 5월.
 [9] Internet Development SDK, http://msdn.microsoft.com/library/default.asp?url=/workshop/browser/prog_browser_node_entry.asp
 [10] Axis, <http://ws.apache.org/axis/>
 [11] Daniel Bilus, et al., "Adaptive Interfaces for Ubiquitous Web Access", CACM, Vol. 45, No. 5, 2002.