

# 웹마이닝을 이용한 M-Commerce 추천시스템 설계 및 구현

이경호<sup>+</sup>, 윤창현<sup>++</sup>, 박두순<sup>++</sup>

## 요 약

추천 시스템들은 고객들이 E-Commerce 사이트에서 상품들을 사는 것을 도와주기 위해서 지속적인 증가추세로 사용되었다. 추천시스템들은 다양한 고객들의 선호도에 따라 유사성과 비유사성에 대한 정보의 기초위에서 고객들의 잠재적인 관심 항목들에 대해 개인의 취향에 맞게 추천하는 기술들을 제공한다. 그러나 추천시스템에 많은 관심을 가짐에도 불구하고 그들의 성능에 대한 공개된 기술이나 정보는 매우 제한적이다. 본 논문에서는, 과거 고객들의 구매행동, 고객정보, 데이터마이닝의 연관규칙을 이용한 M-Commerce 추천시스템을 설계하고 구현하였다.

## A Design and Implementation of the M-Commerce Recommendation System using Web Mining

Kyong-Ho Lee<sup>+</sup>, Chang-Hyun Yoon<sup>++</sup>, Doo-Soon Park<sup>++</sup>

## ABSTRACT

Recommender systems are being used by an ever-increasing number of E-Commerce sites to help consumers find products to purchase. Recommender Systems offer a technology that allows personalized recommendations of items of potential interest to users based on information about similarities and dissimilarities among different user' tastes. However, despite enormous interest in recommender systems, both the number of available published techniques and information about their performance are limited. In this paper, we design and implement an M-Commerce recommendation systems using the past buying behavior of the consumer, consumer information, and association rule mining.

## 1. 서 론

초기 인터넷 쇼핑물은 전자상거래 (E.C. : Electronic Commerce)의 발전과 더불어 눈부신 성장과 발전을 가져오게 되었다. 이러한 성장과

발전에는 기술적인 부분이 뒷받침해 주고 있다. 초창기 웹은 정적인 정보를 교환하는 수단에 지나지 않았고, 이러한 정적인 정보교환은 웹상에서의 커뮤니케이션에 있어서 한계를 가져오게 되

+ 준회원 : 순천향대학교 전산학과

++ 정회원 : 순천향대학교 정보기술공학부 교수

논문접수: 2003년 6월 9일, 심사완료: 2003sys 6월 27일

본 연구는 정보통신부 ITRC 사업에 의해 수행되었음.

있고, 이는 동적 콘텐츠의 등장을 야기하게 되었다. 즉, 웹상에 동적인 기능을 제공할 수 있는 기술이 등장하게 된 것이다. 그 대표적인 기술로 CGI(Common Gateway Interface)를 들 수 있고, 이러한 기술들을 바탕으로 인터넷 쇼핑몰과 같은 동적인 콘텐츠를 제공하는 수많은 사이트들이 생겨나게 되었다. 인터넷 쇼핑몰은 웹기술의 발전과 함께 기술적인 성장의 급격한 발달로 외형적인 성장이 지속되고 있지만 전체 상점거래액에서 차지하는 비율은 미미한 수준이다. 이는 가격경쟁의 악화와 수익기반 확보 및 소수 메이저 업체간의 양극화 현상 등 구조적 취약성의 문제로 인한 결과이다.

대부분의 인터넷 쇼핑몰들은 단순히 상품판매에만 의존하는 실태이다. 이러한 인터넷 쇼핑몰들은 고객의 수요증가 및 다양성으로 인한 질적 향상 및 차별화된 서비스를 필요로 하게 된다. 따라서 고객들의 구매패턴 분석, 로그기록 분석, 및 연관규칙을 이용한 데이터 마이닝 기법을 통한 상품추천엔진 개발은 단순한 상품판매가 아닌 각각의 고객에게 상품을 자동으로 추천해주는 차별화된 서비스를 제공하는 상품추천 시스템을 설계하고 구현하였다. 이 방법은 고객들의 재구매율 증진 및 고객의 효율적인 관리 뿐만 아니라 향후 인터넷 쇼핑몰의 구조적인 문제의 해결방안에 있어 하나의 해결책이 될 것으로 본다. 또한, 이를 무선단말기에도 서비스가 제공되기 위해 M-Commerce 상의 서비스 향상을 위해 웹상의 정보를 무선단말기로의 서비스를 위한 변환기 개발을 설계하였다.

## 2. 연구배경

상품추천엔진 개발을 위해서는 다량의 고객정보를 수집해야 하는데, 이는 웹로그 수집 모듈개발과 고객의 행위정보 수집 모듈개발로써 고객의 행동패턴정보를 수집하는데 사용되었다. 또한 고객의 등록정보 및 캠페인 수행정보를 통한 고객정보수집도 이루어졌다. 이렇게 수집된 다량의 정보들은 상품추천엔진의 전처리 과정에서 대부분 수행되게 된다. 즉, 웹로그 수집 모듈 등은 전

처리 과정에서 데이터를 추출, 변환, 정제하는 일련의 과정중 일부분이라고 생각하면 될 것이다. 이와같이 상품추천엔진에서 유용하게 사용될 데이터들은 전처리 과정을 통해 데이터베이스에 저장되게 되고, 이러한 데이터들과 고객의 구매데이터 및 등록 데이터 등을 근거로 상품추천 DB가 구축되게 되는 것이다.

이러한 상품추천엔진 개발과 상품추천서비스에 대한 연구[1,2,3,4]는 매우 활발하게 진행되었으며 본 논문에서는 상품추천엔진 개발과 상품추천 서비스를 무선단말기에서 제공하기 위하여 필요한 내용을 간단하게 기술한다.

### 2.1. 로그분석

사용자가 웹사이트를 이용하게 되면 이에 대한 기록이 로그라는 형태로 흔적이 남게 된다. 로그분석이란 로그를 기반으로 다양한 정보를 추출해내는 것이라 할 수 있다. 이러한 로그분석은 온라인 비즈니스가 치열해지면서 경쟁에서 생존하기 위한 기본적인 방법 중의 한 가지로써 방문객 행위 분석이라 할 수 있다. 이 때 중요한 점은 범용 Log Software를 이용한 단순한 페이지 View를 측정하는 것만으로는 온라인 방문객 행위 분석을 수행하였다고 볼 수 없다는 것이다.

로그분석은 크게 좁은 의미의 로그분석과 넓은 의미의 로그분석으로 나눌 수 있다. 먼저 좁은 의미의 로그분석은 분석 대상을 Web Log에 한정 짓고 이에 대한 분석을 실시하는 것을 의미한다. 기본적인 페이지 뷰부터 방문자의 클릭스트림 및 방문자 수, 참조 페이지 등을 조사함으로써 현재 대부분의 로그분석 툴들이 분석하는 값들을 말한다. 다음으로 좁은 의미의 로그분석에 반해, 넓은 의미의 로그분석은 단지 로그 데이터 뿐만 아니라, 웹사이트에서 보유하고 있는 고객등록정보(Master DB, Demographic DB), 구매정보(Transaction DB)등과 같은 각종 Legacy 정보들을 같이 연계시켜서 실시하는 분석을 의미한다.

특히 거래 정보를 기반으로 하는 고객 분석은 데이터 충실도 측면에서 순수 로그분석이 보여줄

수 없는 고객의 측면을 다양한 각도에서 보여줄 수 있다는 특징이 있다. 또한 Web Log DB가 기존의 Master DB와 연동되는 경우 인구 통계학적으로 Web Usage 현황을 파악할 수 있으며 Case Based Reasoning과 같은 각종 데이터마이닝 기법과의 연계를 통해 각 개인에게 맞는 콘텐츠 및 상품 등을 One to One Real Time으로 추천할 수 있는 기반이 마련되게 된다[6].

## 2.2. 웹로그 및 방문자의 행위패턴 수집방법

웹상에서 이루어지는 E-Customer들의 행위기록은 기본적으로 웹서버에 의해 자동적으로 기록된다. 이러한 기록은 E-Customer가 웹페이지를 액세스할 때마다 기록되는 것으로, 사용자의 IP와 액세스한 파일, 액세스한 시간 등의 정보가 남게 된다.

일반적으로 사용자의 요구로, 웹서버는 해당 웹페이지와 관련된 여러 파일등에 접근하게 된다. 따라서 E-Customer가 요청하는 특정 웹페이지 뿐만 아니라 해당 웹 페이지와 관련된 이미지 파일, 이미지 데이터, 모든 연관 파일 등에 대한 정보를 웹로그 파일에 기록하여 저장되는 것이다. 이렇게 로그파일은 웹서버에 의해 수집되며, 웹서버에 따라 조금씩 다르게 나타난다. 이처럼 웹서버에 저장하는 로그 데이터 포맷이 서로 맞지 않아 초창기에는 로그분석에 어려움이 많았다. 즉, 하나의 로그분석 프로그램이 서로 다른 웹서버에 의해 기록된 로그파일을 분석할 수 없기 때문이다. 그래서 등장하게 된 것이 Common Log Format(CLF)이다. 이러한 CLF을 사용할 경우 공통적인 형식으로 로그 데이터를 저장하기 때문에 웹서버가 다르다 하더라도 로그데이터 형식은 유사하게 저장되어 위와 같은 문제는 해결할 수 있게 된다[7].

## 2.3. 웹마이닝

웹마이닝은 웹에서 발생하는 모든 데이터를 분석 대상으로 삼는다. 이러한 데이터로는 서버 접속 로그 데이터(server access log data), 사용자

등록 정보(user registration data 또는 profile), 사용자 세션(session), 또는 트랜잭션(transaction), ERP 데이터(enterprise resource planning data)가 있다.

사용자의 웹 사이트 이용 패턴을 분석하기 위해서는 사용자가 웹을 사용할 때마다 사용자가 웹 서버로 보낸 요구를 기록해야 한다. 이러한 웹로그에서 패턴 발견을 위해 데이터를 추출, 변환, 정제하는 일련의 과정을 전처리 과정이라 할 수 있다. 정확한 결과를 얻기 위해서는 전처리 과정이 매우 중요하다[6].

## 2.4. 연관규칙

연관규칙은 데이터마이닝의 핵심기술로써 상품 혹은 서비스간의 관계를 살펴보고 이로부터 유용한 규칙을 찾아내고자 할 때 이용될 수 있는 기법이다. 연관성측정은 어떤 특정 문제에 대해 아직은 일어나지 않는 답을 얻고자하는 예측의 문제나 고객들의 특정목적에 따라 분류하는 문제가 아니라, 상품 혹은 서비스의 거래 기록데이터로부터 상품간의 연관성 정도를 측정하여 연관성이 많은 상품들을 그룹화하는 클러스터링의 일종이라 할 수 있다. 또한 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 장바구니분석이라고도 한다.

연관성측정에서의 연관규칙은 “상품 A가 구매되어진 경우는 상품 B도 구매된다”라고 해석된다. 연관성규칙기법을 이용할 수 있는 데이터는 판매시점에서 기록되어진 거래와 품목에 관한 정보를 담고 있어야 하며 데이터의 형태는 결과변수(targets)를 갖지 않는 Unsupervised data이며 특별히 각 고객들이 누구인지에 대한 구분이나 고객들에 관한 성별, 나이 등의 인구 통계학적인 자료를 비롯한 기타 정보들을 필요로 하지는 않는다. 측정의 기본은 얼마나 자주 구매되었는가 하는 빈도이다[8].

## 2.5. 개인화된 추천 시스템

일반적으로 개인화된 정보를 추천하는 기법으

로 내용기반 추천기법(content based recommendation technique), 협업필터링 기법(collaborative filtering technique), 규칙기반 필터링 기법(rules based filtering technique) 그리고 선호도기반(preference scoring based technique) 기법을 들 수 있다. 내용기반 추천기법은 제품의 속성과 고객 관심의 매칭을 통한 정보를 추천하며, 문서내 단어의 중요도에 의해 문서를 표현한다. 협업필터링 기법은 고객 사이에서 발견된 상관관계를 이용하여 비슷한 취향을 가진 고객들간의 비교를 통해 새로운 고객이 흥미를 가지거나 구매할 가능성이 높은 정보를 예측하고 추천한다. 규칙기반 필터링 기법은 고객의 프로파일이나 행위에 근거한 조건문 형식의 규칙을 이용하여 개인화된 추천을 제공하는 방법으로 가장 일반적으로 전통적인 방법이다. 선호도기반 기법은 고객의 초기 프로파일, 구매행위, 그리고 인터넷 쇼핑몰에서의 다양한 관심표명 행위를 바탕으로 정보의 선호도를 결정하여 개인화된 정보를 추천한다. 선호도를 계산하기 위해 사용되는 고객 관련정보는 규칙기반 필터링 기법과 동일하지만 복잡한 절차없이 선호도의 크기에 따라 고객에게 개인화된 추천을 할 수 있어 보다 간편하다[9].

2.6. HTML을 WML로 변환하는 기술

<표 1>은 기존의 HTML 페이지를 WML로 변환하는 Converter들이다. 하지만 다양한 멀티미디어 요소가 포함되어진 HTML 웹페이지는 여러 기능에서 아직 완벽한 변환이 불가능함을 <표 1>에서 보여주고 있다[10,11].

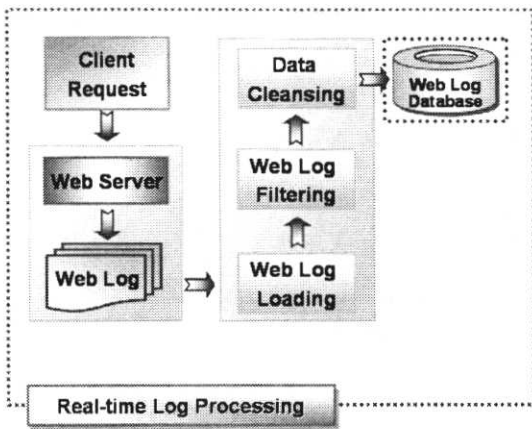
<표 1> 조사 연구 사례

	HTML2WML	TransWAP	ArgoGroup's WAP Tool
Link	같은directory아래있는것만 링크 가능. 다른서버나, 가상directory로 링크된경우 안됨	안됨	안됨
Table	안됨	안됨	안됨
Image	[IMG]로 표시됨	안됨	IMAGE: 'alt' 속성값으로 표시됨
Image map	Image map: http://주소로 표시됨	안됨	안됨
Form	안됨	안됨	가능
Frame	안됨	안됨	안됨
Java Script	안됨	안됨 (source code 그대로 보여줌)	안됨
기 타 사 항	하나의 HTML file을 한번에 다 conversion하지않기 때문에 다음 page(deck)으로 넘어갈 때 Reload버튼을 눌러 conversion해줘야 함	하 나 의 HTML file을 한번에 다 conversion해주는 기능이 있음	Online에 있는 page는, conversion할 수 없음[hard disk에 저장된 HTML file만 가능]

3. 상품추천시스템 설계 및 구현

본 장에서는 웹서버에 의해 수집된 데이터를 정제/변환하여 웹로그 데이터베이스에 기록하는 로그데이터는 고객 파악에 있어 어느 정도 한계를 갖게 된다. 이는 웹서버 측면에서 본 기록이므로 콘텐츠적인 면에서 보았을 때 필요한 기록은 고려되지 않았기 때문이다. 이러한 로그기록의 한계를 극복하기 위해 고객의 행위패턴을 실시간으로 수집할 수 있는 방법에 대해 설계 구현하였다.

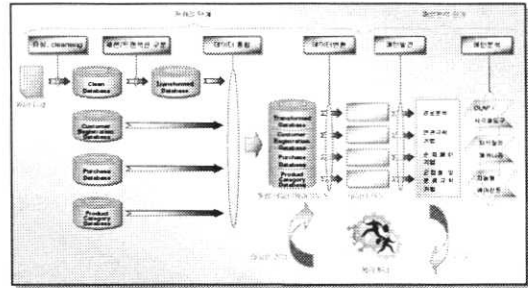
[그림 1]은 웹서버에 수집된 Web Log의 내용을 추출하는 방법을 도식화하여 나타낸 것이다.



[그림 1] Web Log 처리 시스템

웹사용마이닝은 웹 사용자의 사용 패턴을 분석하는 것이다. 이를 통해 웹 사용자의 행동을 접속 통계정보 이상으로 이해할 수 있고, 또한 웹 페이지의 이용 패턴을 알 수 있게 된다. 결국 이 정보는 사용자에게 더욱 친숙하게 페이지를 재구성하거나, 웹 서버 로드 밸런싱, 사용자별 맞춤형 웹 페이지 구성 등에 이용된다.

[그림 2]는 본 논문에서 구현한 상품추천시스템의 전체적인 시스템 구성도로서 전처리과정은 데이터 정제(data cleaning), 사용자 구분(session identification), 세션보정(path completion), 트랜잭션 구분(transaction identification) 등으로 구성되어 있다. 본 논문에서 구현한 상품추천시스템은 크게 세부분 구성되었다. 즉, [그림 1]에서 보여주고 있는 Web Log 처리 시스템과 이를 통해 얻어진 데이터와 기타 고객 관련 데이터를 통합하여 재구성해주는 시스템, 그리고 재구성된 데이터들을 통해 상품추천 DB를 구축하는 시스템으로 나눌 수 있다는 것이다. 여기서 세 번째에 해당하는 상품추천 DB를 구축하는 시스템에서 웹사용마이닝이 적용되었다. 또한, 상품추천 DB는 사용자의 패턴 등을 분석하여 상품들 간의 연관성을 찾아내고, 특정 부류의 고객에게 적합한 상품 리스트정보를 담고 있도록 구성된다.



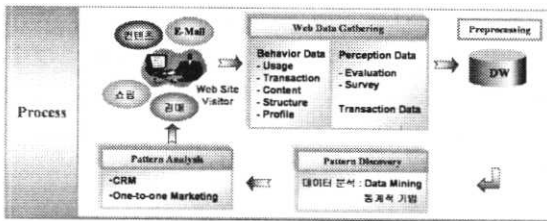
[그림 2] 전처리 단계와 패턴분석 단계

상품들 간의 연관성 측정의 기본은 얼마나 자주 구매되었는가하는 빈도다. 이 빈도를 기반으로 연관정도를 정량화하기 위해서는 전체 거래중 항목 X와 항목 Y를 동시에 포함하는 거래가 어느 정도 인가를 나타내주며 전체적 구매도에 대한 경향을 파악할 수 있는 지지도(Support), 항목 X를 포함하는 거래중에서 항목 Y가 포함될 확률은 어느 정도인가를 나타내주며 연관성의 정도를 파악할 수 있는 신뢰도(Confidence), 항목 X를 구매한 경우 그 거래가 항목 Y를 포함하는 경우와 항목 Y가 임의로 구매되는 경우의 비를 나타내는 리프트(Lift/Improvement)등 세 가지 기준을 고려하게 된다.

본 논문에서는 위에서 말한 지지도, 신뢰도, 리프트를 이용하여 상품들 간의 연관성을 분석하였고, 이를 통해 상품추천시스템을 개발하였다.

웹의 출현과 함께 인터넷 업체들은 훨씬 더 낮은 원가를 투입하여 고객이 어떠한 분야를 더욱 선호하는지에 관한 정보를 얻을 수 있게 되었으며, 이러한 정보를 바탕으로 실시간으로 일대일(one-to-one)원리에 따라 정보를 추천할 수 있게 되었다. 넓은 의미에서 이와 같은 추천 기술들은 개별 고객에게 가장 적합한 웹 페이지의 제공을 위해 협조하기 때문에 웹사이트 개인화의 한 부분이라 할 수 있다. 보편적으로 이러한 시스템을 추천시스템(recommender system)이라 하고 Amazon.com, CDNOW, Drugstore.com, eBay, MovieFinder.com, Reel.com등과 같은 인터넷업체들은 그들의 웹사이트에 하나 또는 그 이상의 다양한 추천시스템 기술을 사용하고 있다.

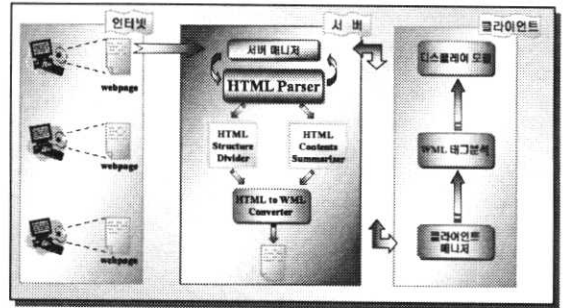
[그림 3]은 상품추천 엔진을 통한 상품추천이 이루어지는 과정을 도식화한 그림이다. 일반적으로 클라이언트에 해당하는 E-Customer가 상품추천물에 들어와서 쇼핑을 하고 상품을 구매하는 등의 일련의 과정을 거치게 된다. 또한 상품추천물은 각각의 고객에 대한 구분을 위해 세션을 활성화하고 고객행위 등을 기록하게 된다. 즉, 세션이 활성화되게 되면 상품추천엔진의 고객의 행위패턴 수집기가 고객의 행위 등을 기록하게 된다. 또한 세션이 비활성화 되면 수집은 끝나게 된다. 상품추천엔진의 주된 기능으로 이렇게 수집된 기록 등을 토대로 위에서 말한 전처리과정을 통해 상품추천 DB를 구축하게 되는 것이다. 이와같이 상품추천 DB는 실시간 추천시스템에 적용되게 된다. 본 논문에서는 [그림 3]의 전체적인 상품추천흐름도와 전처리과정, 개인화된 정보추천 방법 등을 이용하여 상품추천시스템을 구현하였다.



[그림 3] 상품추천 흐름도

다음으로는 이러한 상품추천시스템을 무선단말기에서 서비스하기 위하여 변환시스템을 구현하였다.

웹페이지 변환시스템은 [그림 4]와 같이 크게 서버모듈과 클라이언트 모듈로 구성된다. 서버모듈은 웹상의 HTML로 구성된 페이지를 무선단말기의 서비스를 위한 WML로 변환하여 클라이언트에게 전달한다. 클라이언트는 서버로부터 전달받은 무선단말기용 페이지를 WML뷰어를 통해서 적절하게 디스플레이 한다.



[그림 4] 시스템 구조 (서버모듈과 클라이언트 모듈)

서버모듈은 웹상에 존재하는 HTML 페이지를 무선단말기를 위한 WML페이지로 변환하여 클라이언트에게 전송한다. 또한, 서버모듈 내에는 HTML 파서가 존재하고, HTML 파싱을 통한 내용요약자(Contents Summarize), 구조구분자 (Structure Divider)는 요청받은 HTML 페이지를 무선단말기로의 서비스를 위한 WML 문서변환이 가능하도록 HTML 페이지를 재구성한다. 내용요약자(Contents Summarizer)는 요청된 HTML 문서가 HTML 파서를 통해 무선단말기로 서비스를 위한 변환용 HTML문서로 재구성될 경우, 웹페이지에 들어있는 내용을 기록할 수 있는 부분과 그렇지 않은 부분으로 나누어질 수 있게 되며, 이 둘을 기반으로 HTML2XML 변환기는 무선단말기용 WML문서를 생성한다. 구조구분자 (Structure Divider)는 요청된 HTML 문서는 HTML 파서를 통해 트리구조로 표현된다. 즉, 본래의 웹 내용을 그대로 유지함과 동시에 트리구조의 각 태그에 따른 깊이를 가지게 된다. 트리구조로 표현된 HTML 문서는 구조구분자 (Structure Divider)에서 구조적인 태그만으로 이루어지는 재구성된 HTML문서를 만들게 되고, 내용요약자((Contents Summarizer)와 함께 HTML2XML 변환기에서 무선단말기용 WML문서를 생성하게 된다.

클라이언트 모듈은 WML태그분석기, 디스플레이 모듈로 구성된다. 서버로부터 전송된 무선단말기용 WML문서는 WML태그분석기를 통해 태그 정보를 저장한다. 디스플레이 모듈은 저장된 WML문서의 정보를 사용자에게 효과적으로 보여준다. WML 태그분석기는 서버의 응답으로 전송된 WML 문서 중에는, HTML 문서에 포함되어

질 수 있는 다양한 콘텐츠와 미디어들로 인해 문서변환이 이루어지지 않는 경우가 있다. 따라서 변환이 가능하지 않은 무선단말기용 WML문서를 요청받게 되면 WML태그분석기에서는 이를 적당한 WML문서로 재구성하여 무선단말기에 디스플레이 해야 한다. 클라이언트의 흐름은 서버의 응답으로 전송된 WML문서를 WML 태그분석기를 통하여 WML태그 분석을 한 후, 태그의 정보를 데이터구조에 저장한다. WML문서로부터 생성된 데이터는 디스플레이 모듈에 의해 무선단말기를 통하여 사용자에게 효과적으로 보여진다.

본 논문에서는 HTMLtoWML 변환모듈을 개발하여, 기존의 Converter에 비해 다음과 같은 확장기능을 제공한다. 먼저 다양한 Web Image를 Wbmp Image File로 변환 가능하도록 하였고, Link되어진 다른 HTML페이지도 자동적으로 변환될 수 있게 했다. 또한 작은 무선단말기 화면에 적절하게 디스플레이 될 수 있도록 WML상의 card로 나누어 서비스 되도록 구현하였다[10][11].

본 논문에서 설계하고 구현한 방법은 기존의 상품 추천 시스템에서 단편적으로 사용해왔던 로그 분석 방법이나 데이터 마이닝의 연관 규칙에서 로그 분석을 할뿐만 아니라 데이터 마이닝의 연관 규칙을 적용하기 위하여 행위자의 패턴 분석을 훨씬 효율적으로 수행하였으며, 모바일 기기에서 서비스를 받기위하여 행해져왔던 WML 문서로 작성하는 것 대신에 기존에 HTML로 작성된 내용을 모바일 기기에서 서비스 받을 수 있도록 하였다는 것이 차별화된 내용이다.

#### 4. 구현 예

서론에서 언급한 바와 같이 웹기술은 많은 발전과 성장을 거듭해 왔고, 그 대표적인 기술로는 JAVA 기반의 기술과 닷넷 기반의 기술을 말할 수 있다. 본 논문에서는 이 두 가지 기반 기술중 JAVA 기반 기술을 이용하여 전체 시스템을 구현하였다.

본 논문에서는 웹 모듈에 해당하는 부분 즉, 상품 추천물은 JSP 기술을 사용하였고, 실시간

상품추천에 해당하는 부분은 EJB 기술을 적용하였다. 또한 비즈니스 로직에 관련된 부분은 EJB 컨테이너에서 처리하도록 구성하여 웹서버 성능향상을 고려하여 구현 하였다.

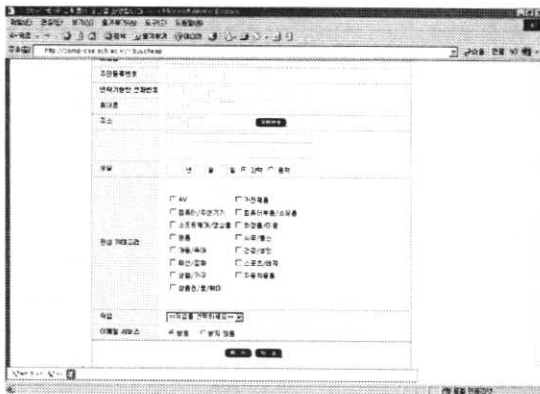
또한 JAVA의 MVC 패턴기법을 적용하여 구현함으로써 웹페이지와 프로그램을 분리할 수 있었다. 즉, 프리젠테이션 부분과 프로그래밍 부분을 완전히 분리했다는 것이다. 이는 차후 업데이트 및 추가적인 개발에 있어 효율적인 방안을 마련해 놓았다고 볼 수 있다.

[그림 5]는 상품추천물의 메인화면을 보여주고 있다. 상품카테고리는 페이지의 좌측에 구성하였다. 상품카테고리 구성은 기존 쇼핑물에서 주로 사용하고 있는 대분류, 중분류, 소분류 체계를 따르지 않고, 각각의 카테고리 아이템들을 같은 형식으로 정의하였으며, 전체 상품카테고리 구성은 부모카테고리와 자식카테고리 형식으로 연결고리를 만들어 주었다. 물론 최상의 카테고리는 ROOT 카테고리로 선언하였다. 이와같이 카테고리를 트리구조 형태로 구현하였다.



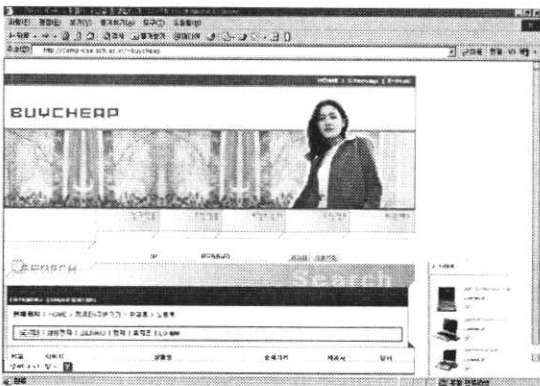
[그림 5] 상품 추천물의 메인화면

[그림 6]은 상품 추천물의 회원가입 부분을 보여주고 있다. 회원가입 부분에서는 고객의 관심 카테고리를 설정해주는 부분이 있는데, 최초 고객 분류는 바로 이 설정부분에 의해 이루어진다. 즉, 회원가입을 하게 되는 고객에 대한 데이터가 전혀 없기 때문이다.



[그림 6] 상품 추천물의 회원가입

[그림 7]은 상품 추천물의 상품진열을 나타내고 있는 화면이다. 본 논문에서 구현한 상품 추천물의 상품진열은 product\_list.jsp에 의해 상품진열이 이루어지고 있다. product\_list.jsp는 현재 카테고리에 상품아이템이 없을 경우 자식 카테고리들에서 상품추천 리스트를 얻게 된다. [그림 7]은 현재 카테고리에 상품아이템이 없으므로 자식카테고리에서 얻어낸 추천 상품을 우측에 보여주고 있다.

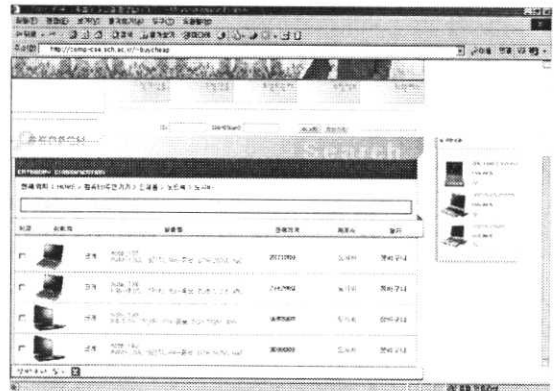


[그림 7] 상품 추천물의 상품진열 1

[그림 8]은 [그림 7]과 같이 product\_list.jsp에 의해 보여진다. 하지만 [그림 7]과는 달리 여러 개의 상품들이 진열되어 있는 것을 확인할 수 있다. 이와같이 현재 카테고리에 상품들이 존재하게 되면 현재 카테고리 내에 있는 상품들 중에서 고객에게 알맞은 상품을 추천해 준다.

또한 고객이 상품을 구매하기 위해 장바구니에 상품을 담게 되면 장바구니에 담겨져 있는 상품

리스트들을 대상으로 상품추천이 이루어진다. 즉, 고객이 장바구니에 상품을 담지 않고 있을 경우에는 현재 카테고리에 대한 관심도만으로 상품을 추천하고, 상품을 장바구니에 담게 되면 장바구니에 담겨져있는 상품을 대상으로 상품추천이 이루어진다.



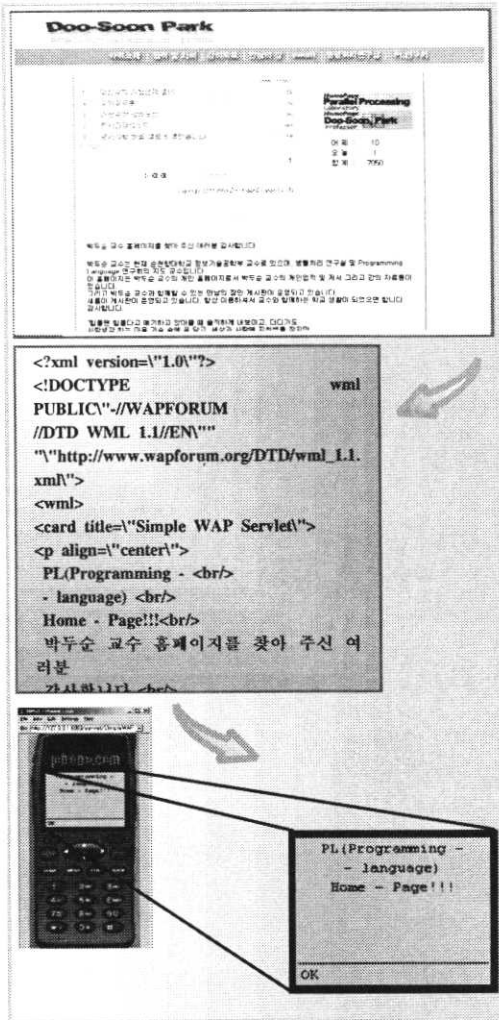
[그림 8] 상품 추천물의 상품진열 2

마지막으로 웹상에서 서비스가 이루어지고 있는 상품추천서비스를 M-Commerce상으로의 서비스를 위해 웹컨텐츠 변환 모듈을 개발하였다.

[그림 9]는 설계된 각 모듈을 구현하고 예제를 통하여 실행 예를 보인 것이다. 서버모듈은 웹상에 존재하는 요청된 HTML 페이지를 HTML 파서에 의하여 내용요약자 (Contents Summarizer)와 구조구분자(Structure Divider)로 넘겨주고 HTML2WML 변환기에 의한 무선단말기용 WML문서를 클라이언트에게 전달한다.

[그림 9]는 실제 웹페이지가 서버로부터 전송 받은 무선폰을 위해 재구성된 WML문서의 일부 분이다.





[그림 9] 시스템 구조

## 5. 결 론

본 논문에서는 상품추천엔진을 설계 및 구현하기 위해 데이터마이닝의 연관규칙기술을 이용한 상품추천 알고리즘을 개발하였고, 이를 웹로그 추출 모듈설계 및 실시간 E-Customer 행위 추출 모듈을 개발하였다. 이는 상품추천엔진을 구성하는데 필수적인 요소들로 이러한 요소들을 통해 상품추천엔진개발과 이를 상품추천물에 적용하여 상품추천서비스를 가능하게 하였다.

끝으로 본 논문에서는 이러한 상품추천서비스를 M-Commerce상에서의 서비스가 가능하게 하기 위한 웹정보의 표현기술인 HTML을 무선 단말기로의 표현기술(WML)로 변환하는 변환기를

개발하였다.

이는 학습자에게 개별화된 학습 객체의 추천 시스템으로 활용이 가능 할 것이다.

## 참 고 문 헌

[1] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-Commerce recommendation applications", *Data Mining and Knowledge Discovery*, 5, pp 115-153, 2001

[2] P. Giudici and R. Castelo, "Association models for web mining", *Data Mining and Knowledge Discovery*, 5, pp 183-196, 2001

[3] B. A. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and evaluation of aggregate usage profiles for web personalization", *Data Mining and Knowledge Discovery*, 6, pp 61-82, 2002

[4] W. Lim, S. Alvarez, and C. Ruiz, "Efficient adaptive-support association rule mining for recommender systems", *Data Mining and Knowledge Discovery*, 6, pp 83-105, 2002

[5] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining : Discovery and applications of usage patterns from web data", *SIGKDD Explorations*, ACM SIGKDD, Vol 1, Issue 2, pp 12-22, Jan. 2000

[6] Web Analyzer-Web Analytics Ver 3.0, ECMIner

[7] <http://www.pionsoft.com/>, PIONSOFT

[8] 마이크로소프트웨어 월간지, 5, 2000

[9] 장근녕, 박유진, 인터넷 쇼핑물을 위한 개인화된 정보추천기법, *Telecommunications Review* 제 12권 1호, 1~2월, 2002

[10] <http://www.gnu.org/directory/>

[11] <http://argogroup.com/Waptool>

## 이 경 호



2002 순천향대학교 전산학과  
(공학사)  
2002~현재 순천향대학교  
일반대학원 전산학과  
석사과정

관심분야: 데이터마이닝, eCRM  
E-Mail: khlee24@lycos.co.kr

## 윤 창 현



2002 순천향대학교 전산학과  
(공학사)  
2002~현재 순천향대학교  
일반대학원 전산학과  
석사과정

관심분야: 모바일, 데이터마이닝  
E-Mail: ych54321@hanmail.net

## 박 두 순



1981 고려대학교 수학과  
(이학사)  
1983 충남대학교 전산학과  
(전산학 이학석사)  
1988 고려대학교 대학원

전산학과 (전산학 이학박사)  
1992~1993 미국 U. of Illinois at Urbana-  
Champaign CSRD 객원교수  
1985~현재 순천향대학교 정보기술공학부 교수  
2002~현재 순천향대학교 공과대학 학장  
관심분야: 멀티미디어, 병렬처리, 가상교육,  
인터넷 컴퓨팅, 데이터마이닝  
E-Mail: parkds@sch.ac.kr