

Log Analysis System Design using RTMA

Hee Chang Park¹, Ho Min Myung²

Abstract

Every web server comprises a repository of all actions and events that occur on the server. Server logs can be used to quantify user traffic. Intelligent analysis of this data provides a statistical baseline that can be used to determine server load, failed requests and other events that throw light on site usage patterns. This information provides valuable leads on marketing and site management activities. In this paper, we propose a method of design for log analysis system using RTMA(realtime monitoring and analysis) technique.

Keywords : log analysis, pushing, real time monitoring

1. 서론

WWW(world wide web) 환경에서 하나의 웹 사이트는 하이퍼 링크(hyperlink)로 연결 되어 있는 수많은 HTML(hypertext markup language) 문서들의 집합으로 이루어져 있다. 초기의 웹 사이트는 각 HTML 문서들이 지닌 의미와 문서간의 상호 관계 등을 고려해 최상의 웹 사이트를 구현하고자 하는 웹 마스터의 의도가 반영된 것이다. 하지만 동적으로 변화해가는 사용자들의 요구를 반영하여 보유정보를 효과적으로 제공할 수 있도록 지속적으로 웹 사이트(web site)의 설계를 바꾸어 가는 일은 매우 어려운 일이다. 하나의 웹 사이트에 대한 사용자들의 접근 로그 데이터는 이 사이트에서 일반적으로 사용자들이 보여주는 정보접근 패턴을 알아낼 수 있는 중요한 자료를 제공한다. 그리고 이러한 사용자 접근패턴을 바탕으로 보다 효과적으로 보유정보를 제공할 수 있도록 기존 웹 사이트의 구조와 표현방식을 개선시킬 수 있을 것이다. 본 논문에서는 이와 같이 사용자 접근 패턴을 학습해서 웹 사이트의 구조나 외형을 자동적으로 개선시켜 나가는 적응형 웹 사이트를 구축하는 한 가지 방안을 제시하고자 한다.

인터넷이라는 새로운 사이버스페이스는 지금까지와는 다른 형태의 고객 접근 환경을 제공하였다. 이를 기반으로 신규사업 모델에 의한 인터넷 기업들이 생겨났으며, 전통적인 기업들도 홈페이지

¹Professor, Department of Statistics, Changwon National University, Changwon, Kyungnam, 641-773, Korea. E-mail : hcpark@sarim.changwon.ac.kr

²Graduate Student Department of Statistics, Changwon National University, Changwon, Kyungnam, 641-773, Korea

지를 구축하여 사이버 공간에서의 고객 접근을 유지하고 있다. 가상공간에 자리 잡은 이러한 모든 형태의 홈페이지는 결국 실세계의 잠재 고객이나 일반 대중에게 열려있는 일종의 상품 전시관 혹은 상품 판매점으로 볼 수 있다. 그리고 로그분석 등으로 방문한 고객의 수 또는 방문 빈도를 분석할 수 있지만 이것은 오프라인(off Line)이라는 단점이 있다. 관리자 또는 담당자가 원하는 시점에서 웹(web)상에 접속한 사용자(user)들의 정보는 알 수가 없는 것이다. 본 논문에서 제안하는 RTMA(realtime monitoring analysis) 기법은 이러한 문제점을 해결하여 마케팅상의 약점을 극복할 수 있는 방안이다. 즉, RTMA의 가장 큰 목적은 실시간으로 웹 사이트에 접속한 사용자들의 현황을 파악하여 마케팅의 효율을 극대화 시키는데 있다.

RTMA 기법은 기존의 로그분석에 사용되었던 IP(internet protocol)운용 방식(요즘은 IP공유기의 등장으로 중복된 IP가 발생)과 이의 단점을 보완한 mac address(media access control address)를 사용함으로써 사용자들의 정보를 축적함으로써 사용자들의 패턴을 예측하여 이제까지 익명일 수밖에 없었던 방문자의 성향을 파악하고 고객 유형에 따른 맞춤형 웹 서비스(web service) 혹은 인터넷 마케팅수행의 능력을 향상시키고자 한다. 한편, 웹 서버에 축적되는 대용량의 로그 데이터로부터 데이터 마이닝(data mining) 기술을 도입하여 유용한 정보를 발견하고 새로운 패턴을 찾으려는 시도가 이루어지고 있다. 이를 웹 마이닝(web mining)이라 하는데, 웹 마이닝은 최근 학계에서의 활발한 연구 활동뿐만 아니라 실제 비즈니스에 많은 응용이 이루어지고 있다.

이에 본 논문에서는 RTMA 기법을 이용하여 웹상에서의 고객 행위 흔적이라고 할 수 있는 웹 로그 데이터를 분석하는 시스템의 설계방안을 제시하고자 한다. 2장에서 RTMA의 기능에 대해 전반적으로 살펴보고, 3장에서는 RTMA의 설계 및 활용방안에 대하여 기술하며, 4장에서는 결론을 맺고자 한다.

2. RTMA의 기능

온라인 마케터(marketer)는 끊임없이 자사의 웹 사이트를 방문하는 사람들을 표적화 하는 방법을 찾고 있고, 그 방법으로 웹 로그 분석을 이용 하고 있다. 그리고 많은 관광관련 웹 사이트도 자사의 홈페이지 방문객과 그들의 방문 경험에서 귀중한 판매 정보를 얻기 위해 웹 로그 분석을 하고 있다(Sweeney Susan, 2000). 이러한 이유로 웹 로그 분석에 대한 연구들이 이루어지고 있으며, 로그 파일 분석 도구들도 많이 개발되어 사용되고 있다. 최근에는 데이터의 증가와 더불어 웹 로그 파일을 토대로 한 e-CRM(e-customer relationship management) 으로의 확장 및 웹 데이터마이닝 기법까지도 대두되고 있다. 웹 로그 분석과 관련되는 연구들의 대부분은 단순하게 빈도분석에 그치는 논문들이 대부분이었으나 Buchner와 Mulvenna(1998)는 OLAP(OnLine Analytic Processing)를 이용해서 그 결과를 인터넷 마케팅에 활용하기 위해 고객을 유인하고, 유지하는 방법 등을 제시하였다. Borges와 Levene(1999), 그리고 Spiliopoulou 등(1999)은 웹 사이트를 분석하고, 사이트의 구조와 링크관계를 분석함으로써 접속자들이 편리하게 탐색할 수 있는 웹 사이트의 디자인 방안을 제시하였

고, 이화영(1999)은 로그파일 분석의 결과를 통해서 웹 서버 관리 및 웹 사이트 디자인 등의 전략을 수립하는 방안을 제시하였다. 기노채(2000)는 시장 세분화 방법의 새로운 도구로서 로그 파일을 활용하는 방법론을 제시하였으며, 정강용과 박나연(2000)은 웹 서버의 로그파일을 추출하여 분석하고 웹 사이트의 재구성을 위한 방향을 제시하였다. 이성백(2001)은 e-CRM에 전략적으로 활용할 수 있도록 확장된 개념의 로그분석에 대하여 기술하면서, 웹으로부터의 정보 추출이 e-CRM을 위한 첫 단계로 설명을 하고 있다. 사용자 접근 패턴을 알아내기 위해 데이터에 적용되는 알고리즘으로 조건부 확률 알고리즘과 APRIORI 알고리즘(Agrawal과 Srikant(1994))이 있다. Perkowiz와 Etzioni(1997)가 제안한 조건부 확률 알고리즘에서는 문서P1과 P2가 있다고 가정했을 때 P1을 방문한 사용자가 P2를 방문할 확률과 P2를 방문한 사용자가 P1을 방문할 확률 중 최소치를 선택해 동시 발생(co-occurrence)빈도수로 정하고 이를 행렬과 그래프로 도식화해서 클래스(class)를 생성한다. 반면에 연관 규칙(association rule) 탐사에서 적용되는 APRIORI 알고리즘에서는 최소 지지도(support degree)와 신뢰도(confidence degree) 이상을 만족하는 문서들의 집합으로 클래스를 생성한다.

본 논문에서는 실시간 모니터링 분석 방법인 RTMA를 이용한 웹 로그 분석 시스템 설계에 대하여 기술하고자 한다. RTMA를 이용한 웹 로그 분석 시스템은 현행의 로그 분석 툴(tool)보다는 좀 더 발전된 형태라고 할 수 있다. 본 논문에서 제안하고자 하는 시스템의 장점은 다음과 같다.

1) 실시간 시스템(realtime system) 기능

실시간 시스템은 사용자들의 문의(request)에 실시간으로 반응(response)하는 시스템을 말한다. 실시간 기능을 통한 실시간 고객의 다양한 정보를 수집 및 분석할 수 있으며, 실시간으로 이루어지는 고객 상담을 통하여 고객의 신뢰와 만족도를 높여줄 수 있다. 실시간으로 이루어지는 다양한 고객 지원 시스템 기능, 사용자들의 요구에 실시간으로 유연한 대응을 할 수 있고, 실시간 고객들의 이동경로 관찰 및 모니터링 기능이 가능하며 실시간 고객 데이터를 통한 분석 기능 및 실시간 고객 정보 제공 기능을 할 수 있다.

2) 모니터링 기능

모니터링이란 웹 사이트의 CCTV(closed circuit television)를 의미한다. 관리자는 웹 사이트 모니터링 페이지(web site monitoring page)만 클릭하면 사용자들의 흐름을 실시간으로 한 눈에 볼 수 있다. 웹 사이트에는 무수히 많은 콘텐츠(contents)가 존재한다. 하지만 많은 콘텐츠들을 관리한다는 것은 힘겨운 일이다. 웹 사이트 내에서는 많은 사용자들이 다양한 콘텐츠를 보기 위하여 지속적으로 이동한다. 웹 마케터들에게 필요한 것은 이렇게 지속적으로 이동하여 다니는 사용자들의 흐름을 실시간으로 모니터링하는 것이다. 예를 들어, 웹 사이트 콘텐츠를 모니터링하여 사용자들의 연령, 학력, 소득, 지역 등의 고객 데이터베이스와의 연계를 통하여 고객의 특성에 맞는 타겟 마케팅이 가능하며 사용자들의 흐름을 파악하여 콘텐츠를 좀더 고객 친화적으로 재구성 할 수 있다.

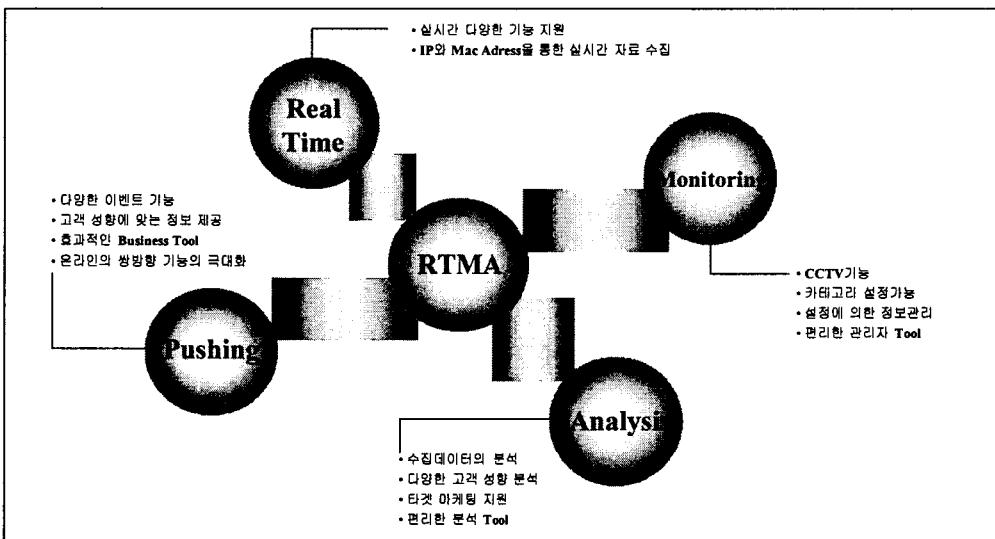
3) 푸싱(pushing) 기능

모니터링에서 사용자들의 흐름 및 성향을 파악하면 사용자 중심의 마케팅을 실시하여야 한다. 푸싱기법은 타겟 마케팅의 한 기법중의 하나로 실시간 정보 제공 및 이벤트(event)를 제공하는 기법이다. 일반적으로 배너(banner)광고는 이제 더 이상 사용자들의 관심을 끌지 못할 뿐 아니라 귀찮다는 생각까지도 들게 한다. 하지만 푸싱기법은 수동 또는 자동적인 기법을 병행하여 사용자들에게 이벤트 및 원하는 화면을 팝업(popup)시켜 주는 기법으로 사이트 내를 옮겨 다니지 않을지라도 이벤트행사 등을 화면에 팝업 시킬 수 있다. 수동으로서의 푸싱은 관리자가 원하는 그룹, 개인, 특정 콘텐츠(contents)에 위치한 사용자 등 다양한 계층으로 팝업 시킬 수 있으며, 자동으로서의 푸싱은 분석과정에서 생성된 결과를 캠페인 서버(campaign server)를 통하여 사용자들의 성향에 맞게 푸싱 해주는 기법이다. 푸싱기법의 사용으로 사용자들의 성향에 맞는 정보를 제공할 수 있어 고객 맞춤 서비스를 할 수 있다.

4) 실제적 분석 기능

본 시스템의 분석은 기존의 로그분석과 같이 회원, 비회원의 구분이 없이 일괄적 분석 기법을 탈피하여 실제 회원과 비회원간의 관계를 명확히 설정하여 분석을 실시한다. 다시 말해서, 회원의 정보와 모니터링 결과를 종합하여 분석하는 기능을 말한다. 분석의 결과는 고객의 정확한 분석을 통하여 고객 개개인의 속성을 파악하여 고객 맞춤 마케팅 방향을 설정하는데 사용된다. 또한, 다양한 분석 기법을 통한 분석자료 도출 기능 및 모니터링에서 축적한 자료의 활용성이 높다.

<그림 1>은 시스템 기능 및 장점에 대한 그림이다.



<그림 1> 시스템 기능 및 장점

3. RTMA를 이용한 로그분석시스템의 설계

3.1. 시스템 개발 환경 및 시스템 흐름

구현된 시스템의 개발 환경에서 개발 언어는 Visual C++ 6.0, php, html등이며, 운영 체제는 Linux를 사용하며, 데이터베이스는 MySQL-Ver 4.0.17을 이용하고자 한다. 본 논문에서 제안한 설계를 바탕으로 시스템의 구현을 위해 사용할 로그분석 파일 항목은 다음과 같다.

항목	설명
IP	◎ 접속한 사용자의 IP 주소 또는 도메인
REMOTE_IDENT	◎ 서버가 RFC 931 을 지원하는 경우 이 환경 변수에 클라이언트 시스템에서 CGI프로그램을 실행시킨 사용자 이름이 저장. 현재 RFC931 부분을 사용하고 있는 웹 서버는 없으며 항상 “-”로 표시
사용자 이름	◎ .htaccess .htpasswd 에 정의된 사용자 id. 웹 서버에서 특정 디렉토리를 지정하여 사용자 이름과 암호를 설정해 놓은 경우, 해당 디렉토리는 등록된 사용자에게만 자료검색을 허락. 현재 사용자가 등록 사용자가 아닌 경우 “-”로 표시
시간	◎ 클라이언트(사용자 브라우저)의 접속시간정보
클라이언트 요청종류	◎ 전송방식(get, post)
요청한 자료 위치	◎ 클라이언트가 요청한 홈페이지 URL 주소
프로토콜 버전	◎ 전송 프로토콜은 현재 대부분이 HTTP를 사용하며, 버전번호와 함께 기록된다.
상태코드	◎ 접속 상태와 데이터 이동상태의 현황을 나타낸다. 100 카테고리는 시스템, 200 카테고리는 성공, 300 카테고리는 무시, 400 카테고리는 에러, 500 카테고리는 서버에러.
전송데이터 크기	◎ 데이터량을 기록한 것 단위는 byte. 상태코드 일부 304는 “-”로 표시
Service	◎ 서비스 제공자의 이름. 웹 서버의 경우는 W3SVC, FTP서버는 MSFTPSVC, Gopher, GopherSvc라고 기록한다.
ServerName	◎ name 서버의 이름
ServerIP	◎ 서버의 IP 주소.
DownLoadTime	◎ 서비스를 제공하는데, 즉 파일이 전송하는데 걸린 시간으로 단위는 msec, 1/1000초.
Receipt	◎ 웹 브라우저로부터 수신한 바이트의 수량으로 그 내용은 서비스 요청이다.
time-taken	◎ 데이터 이동에 걸린 시간
s-port	◎ 연결된 서버의 포트번호
Cookie	◎ 쿠키내용
레퍼럴(Referrer)	◎ 검색 엔진을 사용한 경우 참조 검색엔진, 검색 엔진을 사용한 경우 가장 많이 사용된 키워드 값, URL경로, 이동 페이지 접속 시간
에이전트 로그	◎ 방문자의 웹 브라우저 타입, 버전, OS의 종류, 화면 해상도, 애플리케이션 프로그램 종류
에러 로그	◎ 에러 시간, 에러 종류, 에러 IP(요청 사용자 IP)

본 시스템의 구현을 위해 설계한 테이블은 다음과 같다.

◎ RTM_raw : 로그분석 기본 테이블

필드 이름	필드 설명	타입
Idx	고유번호	int(10) unsigned
Ip	사용자 ip 주소	mediumint unsigned
Day	사용자 접속 시간	datetime
Method	클라이언트 요청종류	tinyint unsigned
Url	요청 자료, 자료 위치	mediumint unsigned
Code	상태코드	tinyint unsigned
Size	전송데이터 크기	mediumint(8) unsigned
Time_taken	데이터이동에 걸린시간	smallint(5) unsigned
Cookie	쿠키정보	varchar(255)
Refer_url	참조 url 정보	mediumint unsigned
Mac_address	맥어드레스 정보	mediumint unsigned

◎ RTM_contents : Html, asp, jsp등의 콘텐츠(또는 페이지)를 저장 하는 테이블

필드 이름	필드 설명	타입
Idx	콘텐츠 고유 번호	int(10) unsigned
Contents	해당 콘텐츠 url	varchar(255)

◎ RTM_group : 콘텐츠를 그룹(group)화 하는 테이블

필드 이름	필드 설명	타입
Idx	그룹고유번호	int(10) unsigned
Group	그룹이름	varchar(255)
Contents	해당 콘텐츠 고유 번호	varchar(255)

◎ RTM_agent : RTM_group에서 구성한 콘텐츠 그룹을 관리하는 사용자 구성 테이블

필드 이름	필드 설명	타입
Idx	사용자 고유 번호	int(10) unsigned
Agent	각각의 그룹번호	varchar(255)

◎ RTM_error : 에러 코드가 발생하였을 경우 에러 페이지와 시간을 기록

필드 이름	필드 설명	타입
Idx	고유번호	int(10) unsigned
Error	error 페이지 정보	varchar(255)
Error_time	error 발생 시간 정보	datetime
Error_ip	error ip 정보	mediumint unsigned

◎ RTM_server : 하나의 서버에 다수의 호스팅이 존재할 경우 이루 구분하는 테이블

필드 이름	필드 설명	타입
Idx	고유 번호	int(10) unsigned
Server	서버 도메인 정보	varchar(255)

◎ RTM_ip : 지역별 정보를 분석하기 위하여 ip별로 테이블을 생성

필드 이름	필드 설명	타입
Idx	고유 번호	int(10) unsigned
Ip	Ip 번호 정보	varchar(15)

⊙ RTM_day : 일일 로그 집계 테이블

필드 이름	필드 설명	타입
Idx	고유번호	int(10) unsigned
Data	시간 정보	daytime
Contents	해당 컨텐츠 url	varchar(255)
Count	hit 수	mediumint unsigned

⊙ RTM_month : 일별 로그 집계 테이블

필드 이름	필드 설명	타입
Idx	고유번호	int(10) unsigned
Data	일자 정보	daytime
Contents	해당 컨텐츠 url	varchar(255)
Count	hit 수	mediumint unsigned

⊙ RTM_mac : 회원 연동을 위한 연결고리 테이블

필드 이름	필드 설명	타입
Idx	고유 번호	int(10) unsigned
Macaddress	맥어드레스 정보	varchar(17)

⊙ RTM_code : 해당 코드 기록 테이블

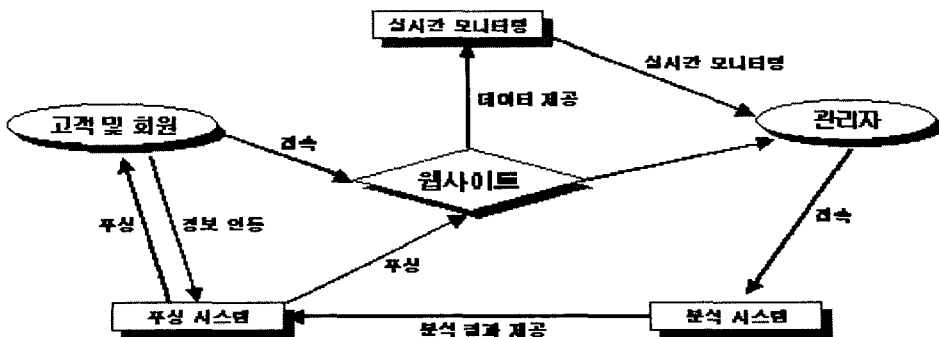
필드 이름	필드 설명	타입
Idx	고유 번호	int(10) unsigned
Code	코드 번호	varchar(17)

⊙ RTM_method : get, post 등 전달 방식을 기록

필드 이름	필드 설명	타입
Idx	고유 번호	int(10) unsigned
Method	메소드 종류	varchar(4)

3.2 시스템의 구성

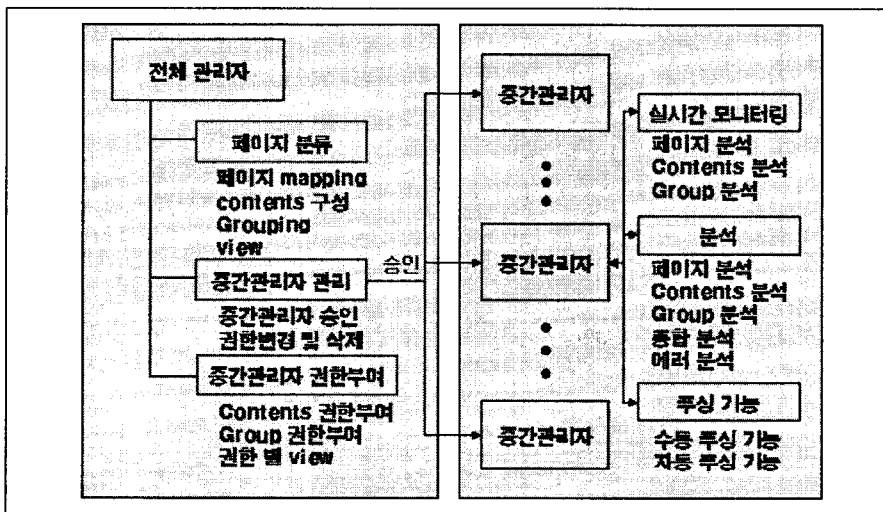
<그림 2>는 RTMA를 이용한 로그분석시스템 흐름도이다.



<그림 2> 시스템 흐름도

시스템 흐름도를 살펴보면 웹 사이트를 방문하는 사용자들에 대하여 관리자는 사용자들의 웹 사이트 이동경로 및 접속 패턴들에 대한 데이터를 실시간 모니터링 하여 얻을 수 있고, 이 데이터를 바탕으로 한 웹 로그 분석을 실시한다. 웹 로그 분석의 결과를 바탕으로 웹 사이트의 페이지나 콘텐츠 등에 정보를 푸싱 할 수 있고 사용자 데이터베이스의 속성들과 연동하여 사용자 특성에 맞는 정보를 푸싱 할 수 있다.

본 시스템의 구성은 전체 관리자(admin)와 중간 관리자(agent)로 기능이 나누어지며, 시스템 구조도는 <그림 3>과 같다.



<그림 3> 시스템 구조도

전체 관리자는 페이지 분류, 중간관리자 관리, 중간관리자 권한부여의 기능을 제공한다. 페이지 분류 기능에는 페이지 mapping, 콘텐츠 구성 및 변경, 그룹 구성 및 변경 기능이 있다.

- 페이지 mapping : 필요한 주소(url)을 얻기 위해서 사용한다. 주소 뒤에 오는 값(parameter)는 대부분 get방식의 변수명과 값일 수도 있지만, 주소경로를 의미하는 경우도 있다. 이를 구분기 위해서 주소의 적정 부분을 mapping함으로써 관리자에게 편리함을 주는 기능이다.
- 콘텐츠 구성 및 변경 : 여러 개의 페이지들을 묶어서 하나의 콘텐츠를 구성하는 기능이다. 각 페이지들 보다 이것들을 묶은 콘텐츠가 분석이 용이할 수 있다. 각 페이지는 변경 가능하다.
- 그룹 구성 및 변경 : 여러 개의 콘텐츠를 묶어서 하나의 그룹을 구성하는 기능이다. 각 콘텐츠들은 변경이 가능하다.
- 콘텐츠, 그룹 구성보기 : 각 콘텐츠나 그룹에 속한 페이지들을 볼 수 있다.

중간관리자 관리 기능으로는 중간 관리자 승인, 중간관리자 권한변경, 탈퇴 기능이 있다.

- 중간관리자 승인 : 가입을 원하는 중간관리자를 승인, 승인거부 기능.

- 중간관리자 권한변경, 탈퇴 : 중간관리자들에 대한 분석 권한을 변경하거나 중간관리자를 삭제할 수 있는 기능이다.

중간관리자 권한부여 기능으로는 콘텐츠 권한부여 및 변경, 그룹 권한부여 및 변경, 콘텐츠 및 그룹별 중간관리자 보기 기능이 있다.

- 콘텐츠 권한부여, 변경 : 중간관리자에게 콘텐츠를 실시간으로 분석할 수 있는 권한을 부여한다. 권한 변경 가능하다.
- 그룹 권한부여, 변경 : 중간관리자에게 그룹에 대한 실시간으로 분석할 수 있는 권한을 부여한다. 권한 변경 가능하다.
- 콘텐츠 및 그룹별 중간관리자 보기 : 각각의 콘텐츠, 그룹들에 대한 분석 권한을 가지고 있는 중간관리자들을 볼 수 있다.

중간 관리자는 실시간 분석, 비실시간 분석, 푸싱 기능을 제공한다. 실시간 분석에는 페이지분석, 콘텐츠 분석, 그룹 분석의 기능을 지원하고 그래프 형식 및 텍스트 형식으로 지원되는 모니터링은 실시간 들어오는 접속자들을 바로 볼 수도 있고 누적해서 볼 수도 있다.

- 페이지 분석 : 텍스트 형식으로 결과를 볼 수 있다.
- 콘텐츠 분석 : 각 콘텐츠별로 그래프 형식과 텍스트 형식으로 실시간 분석이 가능하다.
- 그룹 분석 : 각 그룹별로 그래프 형식과 텍스트 형식으로 실시간 분석이 가능하다.

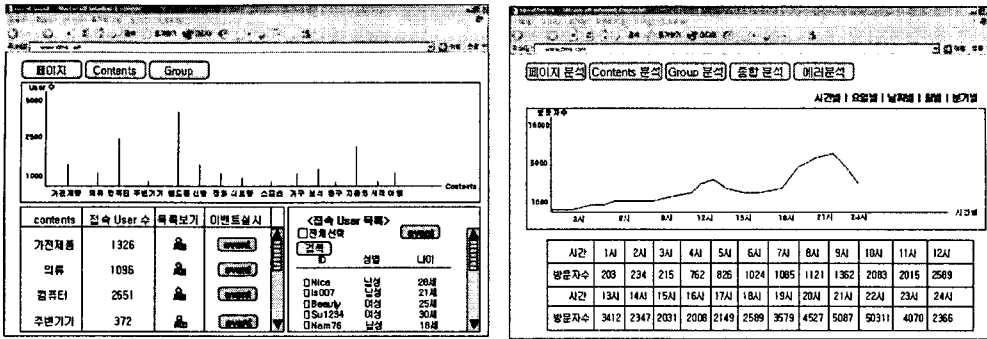
비실시간 분석에는 페이지분석, 콘텐츠 분석, 그룹분석, 종합분석, 에러분석 기능이 있다.

- 페이지 분석 : 페이지별 가장 count(page view)가 높은 것부터 나열. 페이지 count가 0인 것은 나타내지 않음. 시간별, 요일별, 날짜별, 분기별 분석은 3차원 데이터 및 그래프 지원.
- 콘텐츠 분석 : 각 콘텐츠별로 count(page view)를 보여줌. 시간별, 요일별, 날짜별, 분기별 분석은 3차원 데이터 및 그래프 지원.
- 그룹 분석 : 각 그룹별로 count(page view)를 보여줌. 시간별, 요일별, 날짜별, 분기별 분석은 3차원 데이터 및 그래프 지원.
- 종합분석 : 사이트의 총 count(page view) 및 하루평균, 주일평균, 월 평균, 분기 평균 및 시간별, 요일별, 날짜별, 분기별로 보여준다.
- 에러분석 : 에러의 종류별로 총 에러수, 시간별, 요일별, 날짜별, 분기별로 보여준다.

푸싱기능에는 수동 푸싱 기능과 자동 푸싱 기능이 있다.

- 수동 푸싱 기능 : 일반 사용자들에게 이벤트, 메시지 등을 직접 푸싱하는 기능.
- 자동 푸싱 기능 : 분석 결과에 따른 캠페인 서버들 통한 사용자의 성향에 맞는 푸싱 기능.

<그림 4>와 <그림 5>는 각각 실시간 모니터링 화면과 로그분석 화면을 나타낸 것이다.



<그림 4> 실시간 모니터링 화면 <그림 5> 로그분석 화면

본 논문에서 설계한 시스템 활용 방안은 다음과 같다.

1) 고객 및 회원에 대한 타겟 마케팅

: RTMA를 통하여 고객 및 회원의 다양한 정보를 수집하여 개개인에 맞게 분석하여 기호와 취향에 맞춘 맞춤형 타겟 마케팅이 가능하다. 고객의 사이트 방문 시간은 길지 않으며, 고객의 사이트 방문 시에 구매까지의 경로는 다양하고 쉽지 않다. 고객에게 맞는 정보를 제공하는 것은 고객의 만족도 증가는 물론 매출의 증가로 이어지게 된다.

2) 사이트 내 다양한 이벤트 실시(실시간)

: 오프라인에서 이루어지는 반짝 세일과 시간별 마케팅을 온라인에 도입하여 다른 사이트와의 차별화와 흥미를 유발할 수 있다.

3) 고객 데이터베이스 정제(cleaning)

: 사람들의 신상정보는 평균 2-3년에 한번씩 또는 빠르게 바뀌고 있다. 이런 고객의 유동 정보를 실시간 접속을 통하여 다양한 정보를 수집, update되어야 하며, 이를 통하여 정확한 마케팅이 수행될 수 있다. update된 정보를 자동으로 DB Cleaning Logic에 의하여 갱신되어 정확한 고객 데이터베이스가 구축된다.

4) 타 시스템과 연계

: 고객관계관리, 전사적 자원관리(enterprise resource planning: ERP)등 타 시스템과의 연계를 통하여 더욱 정확한 고객 정보의 수집과 마케팅 활동을 지원할 수 있다. mac address와 IP를 통하여 얻어진 고객의 우수한 정보와 실시간 모니터링을 통한 성향의 분석을 통하여 개개별의 고객은 물론 사이트 전체 고객에 대한 정보의 수집과 마케팅 기능을 향상 시킬 수 있다.

5. 결론 및 향후과제

본 논문에서는 RTMA 시스템을 구현하여 웹 로그를 분석하는 단계에서 이전과는 다른 방식을 사용 하였다. 분석의 효율성을 좀 더 극대화시키기 위한 방법 이었다. 온라인 쇼핑몰(online shoppingmall)에서의 1:1 타겟 마케팅 지원 시스템, eCRM 지원 시스템, 푸싱 시스템을 통한 쌍방향

커뮤니케이션 및 실시간 이벤트 기능, 그리고 홈페이지 사이트 관리 부분에서 홈페이지 사용에 대한 모니터링 및 활용도 측정 및 콘텐츠 활용 측정 분야 및 다양한 회원 분석과, 기업 관리·보안 시스템에서 기업 내부 직원 관리를 위한 모니터링 시스템 및 기업 인트라넷 활용, 모니터링을 통한 기업 정보 보안 시스템에 사용 할 수 있도록 구현 하였다. 향후 발전 계획으로서는 데이터마이닝기술을 응용하여 좀 더 정확한 정보를 전달 할 수 있는 기술이 요구된다.

참고문헌

1. 김형택, 민옥길(2001). 효과적인 인터넷 마케팅을 위한 웹 로그 분석, 비비컴, 2001.
2. 이성백(2001). "Web 로그분석과 e-CRM", 한국전자거래(CALS/EC)학회, e - BIZ World Conference 2001(CALS Korea 2001): 129-147.
3. 이화영(1999). "표준 로그파일을 이용한 웹마이닝에 관한 연구", 한국과학기술원 석사학위논문.
4. 정강용·박나연(2000). "웹서버의 로그파일 분석에 의한 웹 서비스 활용에 관한 연구", 한국OA학회, 논문집 13.
5. Agrawal, R, and Srikant, R(1994). Fast Algorithms for Mining Association Rules, In Proc. of the 20th VLDB Conference, 487-499.
6. Borges, J. and Levene, M.(1999). "Data mining of user navigation patterns", WebKDD'99.
7. Buchner, A.G. and Mulvenna, M.D.(1998), "Discovering internet marketing intelligence through online analytical web usage mining", SIGMOD Record, 27(4) : 54-61.
8. Chen, M.S., Park, J.S., and Yu, P.S.(1996). "Data Mining for Path Traversal Patterns in a Web Environment", In Proc. 16th Int. Conf. Distributed Computing Systems, 385-392.
9. Cheung, D.W., Kao, B., and Lee, J.(1997). "Discovering User Access Patterns on the World-Wide Web", In Proc. of PAKDD-97.
10. Cooley, R., Mobasher, B., and Srivastava, J.(1997). "Web Mining: Information and Pattern Discovery on the World Wide Web", In Proc. of TAI-97.
11. Fernandez M. Florescu D., Kang J., Levy A., and Suci D.(1997). "System Demonstration -Strudel: A Web-site Management System", In ACM SIGMOD Conference on Management of Data.
12. Fink J., Kobsa A., and Nill A., "User-oriented Adaptivity and Adaptivity in the AVANTI Project, In Designing for the Web: Empirical Studies"
13. Perkwowitz M. and Etzioni O.(1997). "Adaptive Web Sites: an AI Challenge", In Proc. 15th Int. Joint Conf. AI, 1997.
14. Piroll, P., Pitkow, J., and Rao, R.(1996). "Silk from a Sow's Ear: Extracting Usable Structures from the Web", In Proc. of CHI-96, 118-125.
15. Spiliopoulou, M., Pohle, C., and Faulstich, L. C.(1999), "Improving the effectiveness of a web site with web usage mining", WebKDD'99.

16. Susan, S.(2000), *Internet Marketing for your Tourism Business*, Maximum Press, Canada.
17. Zaiane, O.R., Xin, M., and Han, J. (1998). "*Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs*", In Proc. of ADL-98, 1998.