

WWW 이용자 추적을 통한 개별적 특성·취향 분석 시스템

이충석*, 정연홍**, 박규석***

경남대학교 컴퓨터공학과

e-mail:cslee@hawk.com.kyungnam.ac.kr

An Analysis System of Individual Characteristic & Taste through Pursuit of WWW User

Chung-Seok Lee*, Youn-Hong Jung**, Kyoo-Seok Park***

Dept. of Computer Science, Kyungnam University

요약

최근 기업에서의 인터넷 응용이 보편화되고 있으며, 인터넷 기반의 전자상거래(CALS/EC)는 선택의 문제가 아니라 생존의 수단으로 받아 들여 지고 있다. CALS/EC는 비즈니스 관련 정보를 전자적으로 교환함으로써 기존의 비즈니스 방식을 개선, 전환, 재정의하고, 비용절감, 품질향상, 서비스의 품질과 속도개선을 통해 고객만족과 기업이윤을 증대시키는 새로운 기업의 전술로 정의되고 있다. 이러한 인터넷 기반 CALS/EC의 활성화로 인해 여러 분야에서의 기업 상품에 대한 마케팅 전략 등이 최우선적으로 고려되고 있다. 본 논문에서는 이러한 전자상거래를 포함, 인터넷 이용자의 특성·취향을 분석하기 위한 이용자 특성·취향 분석 시스템의 구성원리와 그 구성요소 중 하나의 모듈인 특성화된 이용자 로그정보로부터 추적 알고리즘을 이용한 시뮬레이션을 수행함으로써, 보다 정확한 이용자 정보를 수집하여 유연성있는 정보를 제공할 수 있는 시스템을 구축하고 구현하였다.

1. 서론

전자상거래란 지난 수년간 기업간 거래에 널리 활용되어온 EDI, 최근 PC 통신을 통해 수행되어 온 홈뱅킹, 온라인 쇼핑, 인터넷을 통해 이루어지는 기업 대 개인 간의 구매 행위 등 정보통신 매체를 통하여 이루어지는 일체의 상거래 활동을 통칭한다고 할 수 있다.

이와 같이 인터넷의 사용이 활발해지고 있는 현재, 인터넷은 정보 전달이라는 단순한 매개체의 개념에서 벗어나 생산자와 소비자 사이에 대중매체를 통한 단방향 정보 전달 체제에서 양방향 통신의 정보 전달 통로로서의 역할을 담당하게 되었다.

이로 인해 상업화를 목적으로 인터넷을 이용하려는

시도가 사회 전반에 걸쳐 고조되고 있으며, 21세기에는 전자상거래를 이용한 실물거래가 주류를 이룰 것으로 예상된다. 특히 현대 사회에서 대중 매체를 통한 광고는 자신을 알리거나 상품을 소개하는데 있어 주요한 역할을 했다면, 인터넷상에서의 광고는 기존의 대중을 향한 광고와는 달리 개별화된 특성에 맞는 소비자와 공급자간의 대화를 통한 새로운 방향으로 전개되고 있다.

지금까지의 개인 신상 정보 자료는 등록 시 지면을 통해 이루어지고, 기타 정보는 설문을 통해 자료 수집이 이루어 졌으며, 자료수집은 이를 작성하는 작성자의 불성실한 응답으로 인해서 정확한 데이터를 수집할 수 없어 이를 이용한 분석은 큰 효과를 기대하기 어려웠다.

현재 인터넷에서 이용되고 있는 전자상거래를 위한 솔루션이나 광고를 이용한 기타 상거래 웹사이트의 경우 이용자에 대한 관리가 거의 일방적으로 이루어지고 있으며, 일부 전자상거래를 발전시키려는 연구자들에 의해 이용자의 취향을 분석하려는 시도가 있기는 하지만 그 성과가 미흡한 실정이다.

본 논문에서는 웹사이트를 이용하는 개개인의 특성·취향을 수집하고 이를 바탕으로 한 유용한 정보를 제공하고자 하며, 구성은 다음과 같다. 2장에서는 본 시스템과 관련한 연구를 살펴보고, 3장에서는 특성·취향 분석 시스템, 4장에서는 시뮬레이션 및 추적 알고리즘을 이용하여 특정 모듈의 시뮬레이션 모델을 구현하며, 5장에서는 결론을 내린다.

2. 관련연구

이용자의 특성 및 취향에 관한 정보를 얻기 위하여 회원등록에 대한 자동화 및 이용자 로그인 정보, 설문조사를 이용한 정보의 수집과정부터 로봇 스케줄러를 이용한 검색 Agent System의 전체 단계에서 수동으로 이루어지던 분류 과정을 실시간으로 자동화시킴으로써 국가나 지역, 시간에 상관없이 인터넷 이용자 관리가 용이해지고 이에 따라 많은 사람의 노력을 감소시키게 된다.

데이터를 암호화함으로써 인터넷상에서의 문서 암호화에 가까이 접근할 수 있으며 Agent Mail Service 기능을 통해 이용자의 특성 및 취향별로 특정 광고나 정보를 예약된 시간에 보낼 수 있으므로 인터넷상에 이루어지는 정보교환을 활성화시키며, 검색엔진의 경우는 웹에서 사용하는 에이전트 로봇이 정보를 수집할 때 관심 있는 요구 정보(사용자 특성 및 취향)만을 수집할 수 있는 필터로 사용할 수 있다. 이 경우 정보를 수집할 때 필요한 정보의 옳고 그름을 판별할 수 있으므로 문서를 수집한 후에 필터링 하는 경우에 비해서 네트워크 트래픽(traffic)을 줄일 수 있으며, 불필요한 정보를 저장하여 저장 공간을 낭비하는 경우가 없게 된다. 또한 이러한 정보에 대한 분류작업이 이루어지게 되면, 정보를 가져오는 시점에 분류항목이 정해지므로 정보를 새로 추가 할 때마다 재분류할 필요 없이 새로운 정보에 대한 분류항목의 서비스가 이루어진다.

표1은 본 연구와 관련한 항목을 나타내고 있다.

[표 1] 이용자 특성·취향 분석 기술 항목

기능분류	연구·개발
이용자등록	정보에 대한 암호화 데이터 자동 변환기 등
이용자 분석	원시데이터분석 및 가용한 정보 변환 등
Mailing System	실시간 Mailing Service 이용자 취향별, 그룹별 Mail Service Mail 예약 및 CGI를 이용한 취향분석 등
Agent 검색엔진	이용자 취향 정보 분석 색인어, 질의어생성 및 검색 스케줄러 구성 데이터베이스 스키마 생성기 등
실시간 광고	취향별 실시간 배너 및 텍스트 광고 통계처리 및 문서 생성 등
이용자 정보처리	실시간 그래픽 통계처리 실시간 문서 생성 등

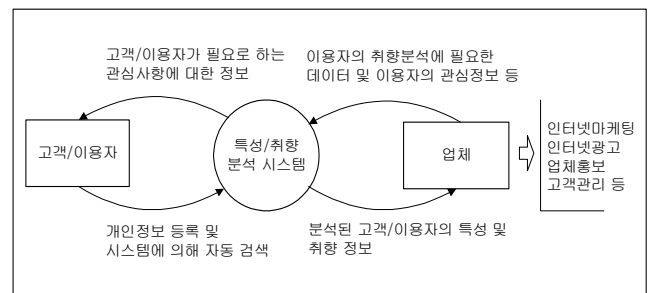
3. 특성·취향 분석 시스템

본 논문의 연구·개발은 다중 에이전트 기반구조에 기초한 자바 언어(Java Language)와 CGI(Common Gateway Interface) 프로그램으로 구성되어 있으며, Multi Thread 방식을 이용하여 네트워크(Network) 및 서버(Server)의 부하를 최소화하여 처리량을 향상시키고 효율적인 실시간 정보제공이 가능하게 하였다.

인터넷 이용자의 특성·취향을 분석하기 위해 특정 사이트 방문자의 프로필과 본 논문에서 제안한 특성·취향 분석 시스템의 각 모듈에서 수집한 정보를 이용자 추적 알고리즘을 이용해서 Rule을 생성하고 이에 따른 Dynamic Matching을 통한 정보를 제공하며, 개인 정보를 바탕으로 이용자가 구매 행위나 원하는 정보를 찾고자 할 때 전혀 불편을 느끼지 않는 트랜잭션을 제공한다.

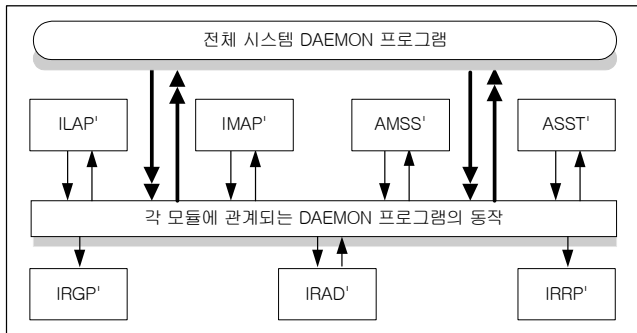
3.1 시스템 구성

고객은 인터넷을 통해 원하는 정보를 얻고, 기업체는 자사의 상품이나 정보를 제공하는 것으로 쌍방향의 목적을 충족시켜주기 위한 본 특성·취향 분석 시스템은 그림1과 같이 나타낼 수 있다.



[그림1] 이용자 특성·취향 분석시스템

본 시스템은 여러 방향에서 이용자 정보를 수집, 분석하여 데이터 필터링(Data filtering)을 통해 보다 정확한 이용자 정보를 생성하며, 특히 전체모듈을 감시, 관리하는 Daemon과 각각의 모듈들은 시스템 콜(Call)이 아닌 TCP/IP Message 전송 체계를 이루며 이러한 메시지 전송은 미리 정해진 코드체계에 의한 각 모듈 관리를 담당하는 Daemon들로 구성되어진다.



[그림2] 각 모듈간의 통신관계

그림2에서 나타내고 있는 각 모듈들을 설명하면 다음과 같다.

· ILAP'(Internet Log Analysis Processor)

서버에 접근하는 이용자에 대해 실시간으로 특성화된 로그를 분석하여 이용자의 특성·취향과 행위에 대한 정보를 수집한다.

· IMAP(Internet Members Automation Processor)

인터넷상에서 자동화 회원등록 프로그램을 구현한 것으로 회원 개인에 대한 정보는 공개키 암호화 알고리즘(RSA)을 사용하여 처리하며, 이용자의 정적인 특성·취향 정보를 수집한다.

· AMSS(Agent Mailing Service System)

실시간 메일링 서비스를 지역별, 시간대별, 이용자의 특성·취향별 등 예약, 실시간으로 개별적, 그룹별 처리하며, CGI(Common Gateway Interface) 프로그램을 이용한 특성화된 메시지로부터 이용자의 특성·취향 정보를 수집한다.

· ASST(Agent Search System)

특성화된 로봇 제어정책을 통해 Internet Network Traffic을 최적의 상태로 유지하면서 실시간으로 등록된 이용자 서버의 내용을 검색하여 특성·취향 정보를 수집하며, 서버의 구성 인식과 오류 수정을 한다.

· IRGP'(Internet Real-time Graphic Processor)

각 모듈에서 구성된 데이터베이스로부터 분산객체

인 CORBA와 Java를 사용하여 관리자, 이용자가 원하는 정보를 다양한 방법으로 실시간 그래픽처리로 나타낸다.

· IRAD'(Internet Real-time Advertisement)

인터넷 광고에 특성을 부가하여 실시간 처리하며, 이용자의 특성·취향 정보를 수집하여 전자상거래를 비롯하여 인터넷에서 이루어지는 모든 거래에 대한 광고효과를 창출한다.

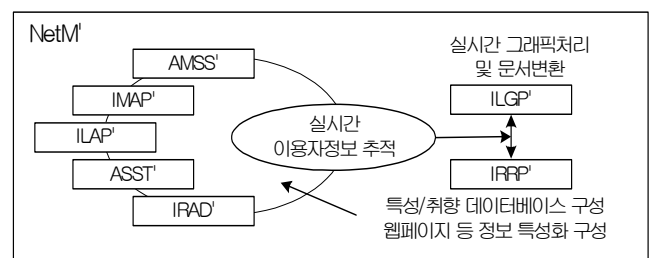
· IRRP'(Internet Real-time Report Processor)

각 모듈에서 구성된 데이터베이스로부터 필요한 정보를 다양한 방법을 통해 문서로 기록해서 나타낸다.

3.2 이용자추적 시스템

통합시스템의 각 모듈로부터 수집된 정적, 동적 이용자 정보를 실시간으로 분석하여 수치 좌표와 문서로 나타내어준다.

이것은 기존 시스템의 Web-Server를 분석할 때 단순한 접속횟수 혹은 데이터량과 기계적이고 정적인 수치만을 제공해 주는 것으로부터 얼마간의 이용자에 대한 정보를 취하는 것과는 달리 실시간 이용자 추적을 통해 실시간으로 변화하는 이용자의 특성·취향 정보를 수집, 분석하게 된다. 또한 이용자의 이동경로를 분석하여 현재 이용자의 위치나 이동경로를 실시간으로 확인할 수 있으며, 그에 따른 정보는 통계 수치로 나타내어 실시간 그래픽 변화로 처리한다.



[그림3] 각 모듈간의 관계와 이용자 정보 추적

그림3은 각 모듈로부터 실시간으로 수집되는 이용자 정보를 재구성하여 이용자 정보 추적을 통해 실시간으로 그래픽 및 문서 변환함을 보여주고 있다.

4. 구현 결과 및 분석

먼저 각 모듈들의 행위에 대해 관리하는 통합관리 모듈(NetM)은 각 모듈들의 처리를 위한 우선 순위 스케줄링과 각 모듈간의 데이터베이스의 일관성을 관리한다. 또한 TCP/IP Message를 통한 Multi Thread 방식으로 구성되며, 우선 순위 설정 및 데이터베이스 일관성 유지를 위한 프로그램 부분은 그림 4 및 5와 같다.

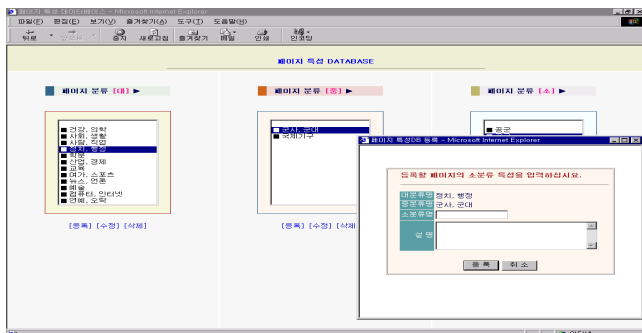
```
.....
System.out.println("SOCKET [PORT]: "+svSocket.getLocalPort()+" Waiting for connection...");
LOG.writeLog("SOCKET [PORT]: "+svSocket.getLocalPort()+" Waiting for connection...");
//-----
for(;chk;){
// accept()함수로 클라이언트의 접속을 기다림
try{
new SubThread(svSocket.accept(),JobID_M_DataPro,LOG,MCODE).start();
}
catch(IOException e){System.err.println(e.toString());}
}
svSocket=null;
.....
Client와 각 모듈간의 메시지전송을 담당하며 다중 Thread를 생성
```

[그림4] 각 모듈간의 메시지전송 및 다중 Thread 생성

```
.....
//ID가 작업중임을 알리기 위한 설정
JobID_M_SetJobID($Target_ID,this.getName());
//우선순위의 설정
if($Job_Code=500&&$Job_Code=6){this.setPriority(5-$Job_Code);}
yield();
//동일ID가 작업중인가 판별
for(;JobID_M_ChkJobID($Target_ID,this.getName());){
System.out.println("W'+this.getName()+" -> Waitting...");
OUT.println("wait");
//접속한 서버모듈에 대기신호 전송
continue;
}
OUT.println("ok"); // 접속된 서버모듈에 작업 허용 신호 전송
switch($Job_Code){
.....
각 모듈간의 우선순위를 설정하며, 데이터베이스 일관성 유지
```

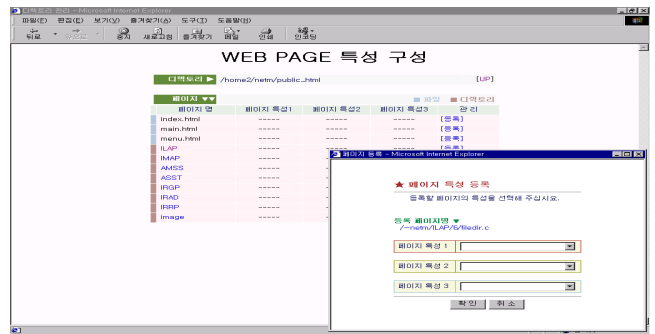
[그림5] 우선순위설정 및 데이터베이스 일관성 유지

그림6은 웹 상에서 특성·취향 데이터베이스를 생성하는 목록으로 대·중·소로 구분한다.



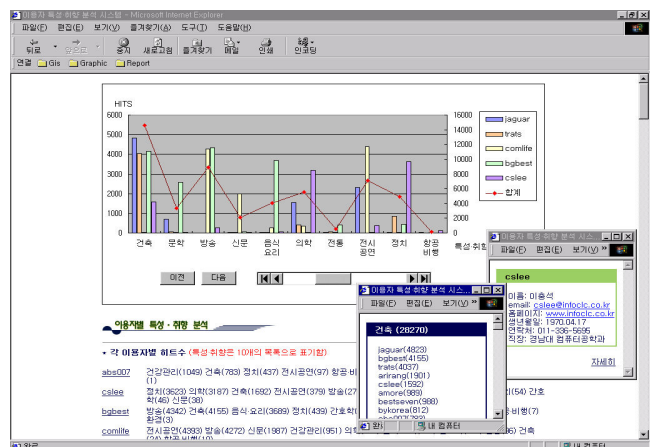
[그림6] 특성·취향 목록 작성

그림7은 특성·취향 데이터베이스로부터 웹 페이지를 특성화하고 있음을 나타내고 있다.



[그림7] 웹 페이지 특성·취향 등록

그림8은 사용자 정보를 실시간으로 분석한 그래프 및 문서를 나타내고 있다.



[그림8] 실시간 그래프 및 문서 변환

5. 결론

본 논문에서는 기존의 정적인 사용자 정보 수집과는 달리 실시간으로 이용자의 행위나 상태를 분석하고 이들의 특성 및 취향 정보를 필터링 함으로써 보다 정확한 사용자 정보를 얻을 수 있는 사용자 특성·취향 분석 시스템의 각 모듈별 성능을 제시하였으며, 각 모듈로부터 수집된 사용자 정보와 특성화된 로그정보로부터 사용자 추적을 통해 실시간 이용자들에 대한 포트폴리오 구성이 용이해 질 수 있었다.

이에 따라 인터넷상에서의 정보교환이 더욱 용이해지고, 의사 반영 경로는 직접적으로 이루어지며 그 응답이 즉각적으로 이루어지며, 기존의 단방향이었던 문제점들을 다소 해결하였으며, 앞으로 이용자의 특성·취향을 분석에 대한 추가적인 방안에 대한 계속적인 연구가 요구된다.

5. 참고 문헌

- [1] Edith Cohen, Balachander Krishnamurthy, Jennifer Rexford, "Efficient Algorithms for Predicting Requests to Web Servers," Proceedings of the IEEE INFOCOM'99 - Volume 1 ,V.1 ,1999 -03 -21.
- [2] Garofalakis J, Kappos P, Mourloukos D, "Web site optimization using page popularity," IEEE Internet Computing ,V.3 N.4 ,1999 -07 -01.
- [3] Michael D.Cooper, "Design Considerations in Instrumenting and Monitoring Web-Based Information Retrieval Systems," Journal of the American Society for Information Science ,V.49 N.10 ,1998 -08 -01.
- [4] Olfa Nasraoui,Raghu Krishnapuram,Anupam Joshi, "Relational Clustering Based on a New Robust Estimator with Application to Web Mining," Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society , ,1999 -06 -10.
- [5] Nabe M, Murata M, Miyahara H, "Analysis and modeling of World Wide Web traffic for capacity dimensioning of Internet access lines," Performance Evaluation ,V.34 N.4 ,1998 -12 -01.
- [6] Dan Backman with Jeffrey Rubin, "Web Log Analysis: Finding a Recipe for Success," Network Computing ,V.8 N.11 ,1997 -06 -15.
- [7] Noritaka Osawa, Toshitsugu Yuba, Katsuya Hakozaki, "An Evaluation of Generation Replacement Schemes Based on WWW Caching Proxy Server Logs," Transactions of Information Processing Society of Japan ,V.39 N.3 ,1998 -03 -01.
- [8] S. Comai, E. Damiani, L. Tanca, "Semantics-aware querying in the WWW: the WG-Log Web Query System," Proceedings of the IEEE Multimedia Systems '99 - Volume 2 ,V.2 ,1999 -06 -07.
- [9] Arun K.Iyengar, Edward A.MacNair, Mark S.Squillante, Li Zhang, "A General Methodology for Characterizing Access Patterns and Analyzing Web Server Performance," Proceedings of the 6th International Symposium on Modeling,Analysis

and Simulation of Computer and Telecommunication Systems , ,1998 -07 -19.

- [10] Rajiv Shukla, Jon McCann, "TONICS for Operation Support Systems System Management using the World Wide Web and intelligent software agents," Proceedings of the 1998 IEEE Network Operations and Management Symposium - Volume 1 ,V.1 ,1998 -02 -15.