

클릭스트림 분석을 위한 웹 서버 시스템의 설계 및 구현

강 미 정[†] · 정 옥 란[†] · 조 동 섭^{††}

요 약

인터넷을 통한 비즈니스에 수익 모델에 대한 관심이 높아지면서 방문자별로 개인화된 서비스를 제공하려는 노력이 커지고 있다. 개인화(Personalization)란 고객 한 명을 대상으로 하여 그 고객 한 사람을 위한 정보나 제품을 제공하는 작업을 말한다. 개인화 서비스를 위해서 전처리과정인 사용자 프로파일 생성과정이 필요하며, 적극적인 개인화 서비스를 제공하기 위해서는 충분한 고객 데이터가 필요하다. 본 논문에서는 웹사이트 상에서 사용자 행위 패턴을 파악할 수 있는 클릭스트림 정보를 모듈화 하였으며, 이를 이용하여 확장된 웹 로그 시스템을 구현하였다. 클릭스트림 정보를 웹 로그정보에 포함시켜 사용자의 행위 패턴을 파악할 수 있도록 웹 서버 시스템을 설계하고 구현하였다. 그리고 이 웹 서버는 웹사이트로부터 얻은 클릭스트림 정보를 분류하고 저장하여 관리자가 쉽게 분석할 수 있다. 이때 데이터베이스 저장 기술로 OLE DB Provider 상에서 수행되는 ADO(ActiveX Data Object) 기술을 사용함으로써 확장된 웹 로그 처리 시스템을 설계하였다. 확장된 웹 로그 DB를 패턴분석, 군집분석 등의 마이닝(Mining) 기법을 통하여 맞춤형서비스에 대한 사용자 프로파일을 구축할 수 있다.

Design and Implementation of Web Server for Analyzing Clickstream

Mi-jung Kang[†] · Ok-Ran Jeong[†] · Dong-sub Cho^{††}

ABSTRACT

Clickstream is the information which demonstrate users' path through web sites. Analysis of clickstream shows how web sites are navigated and used by users. Clickstream of online web sites contains effective information of web marketing and to offers usefully personalized services to users, and helps us understand how users find web sites, what products they see, and what products they purchase. In this paper, we present an extended web log system that add to module of collection of clickstream to understand users' behavior patterns in web sites. This system offers the users' clickstream information to database which can then analyze it with ease. Using ADO technology in store of database constructs extended web log server system. The process of making clickstreaming into database can facilitate analysis of various user patterns and generates aggregate profiles to offer personalized web service. In particular, our results indicate that by using the users' clickstream. We can achieve effective personalization of web sites.

키워드 : 개인화(personalization), 웹로그(web log), 클릭스트림(clickstream), ADO

1. 서 론

웹을 통한 전자상거래는 개별고객의 거래 내역을 추가적인 노력을 기울이지 않고 손쉽게 추적하여 획득할 수 있다. 웹사이트에서 사용자들의 성향을 직접적으로 분석할 수 있는 클릭스트림 분석에 대한 중요성이 대두되었다. 일반적으로 개인화를 위해 사용되는 기술은 로그파일을 분석하는 전처리과정을 거친다. 로그분석이란 사용자들이 웹 사이트를 이용하면 기록이 로그라는 형태로 흔적을 남기게 되는

데, 이 데이터를 기반으로 다양한 정보를 추출해 내는 것이다. 또한, 로그 데이터를 이용하여 웹사이트의 페이지뷰, 사용자별 페이지뷰, 접속 장소 및 방식, 시간별 페이지뷰, 방문자수 등에 대한 현황 및 추세를 분석하는 것이다[16]. 이 방법은 사용자가 웹 사이트에서 무엇을 원하는 지를 구체적으로 상세하게 파악하는 것은 불가능하다. 그러므로 사용자가 웹 페이지 내에서의 행위패턴을 발견하기 위해서는 클릭스트림 데이터를 수집하여 기본 웹 로그 정보를 확장할 필요가 있다. 본 논문은 단순한 고객 데이터 분석과 구매 자료 수집에 의존한 프로파일링 보다 진보된 기술인 클릭스트림 데이터를 수집, 저장, 분류하는 확장된 웹 로그 서버 시스템을 제안한다. 사용자들의 클릭 내용을 추적하는

* 이 논문은 2001년도 두뇌한국21사업에 의하여 지원되었음.

† 준 회 원 : 이화여자대학교 과학기술대학원 컴퓨터학과

†† 종 신 회 원 : 이화여자대학교 컴퓨터학과 교수

논문접수 : 2002년 3월 27일, 심사완료 : 2002년 7월 2일

클릭스트림 데이터를 데이터베이스화였다. 필요한 경우 데이터마이닝 기법을 통해 사용자들의 여러 패턴들을 파악해 각각의 개별 사용자들에 대한 정보를 이용해 보다 정확한 개인화된 서비스를 제공할 수 있다.

클릭스트림(clickstream)이란 웹 사이트 상에서의 사용자들의 행위 패턴을 파악할 수 있는 정보이다. 웹 브라우저에서 사용자가 취하는 동작의 복합체라 할 수 있다. 클릭스트림에는 실제 클릭과 그 요청에 대한 서버의 반응 모두가 포함된다[4]. 웹 사이트상의 사용자들이 컴퓨터 화면상의 어떤 객체를 선택하거나 원하는 항목을 실행하도록 하기 위해 마우스를 이용하여 선택하는 일련된 흐름의 정보들이다. 이러한 클릭스트림을 분석함으로써 사용자들의 성향을 알아낼 수 있다. 맞춤서비스를 제공하기 위해서는 사용자들의 성향을 올바르게 파악하고 사이트의 매출에 직접적으로 기여할 수 있는 행위 패턴을 찾아야 한다.

이에 본 논문에서는 현재 웹 사이트상의 개인화와 개인화된 웹 사이트에 사용되는 기존의 기술을 바탕으로 사용자들의 행위 패턴을 수집하기 위해 클릭스트림을 정의하고 수집하는 방법을 제안하였다. 그리고 기존 웹 서버 시스템을 확장하여 클릭스트림에 의해 추가된 웹 로그 서버 시스템을 제안하고 구현하였다. 본 논문의 구성은 다음과 같다.

제 2장에서 이와 관련한 연구내용들을 소개하고, 제 3장에서 본 논문이 제안하는 클릭스트림을 위한 웹 서버 시스템을 설계를 보인다. 제 4장에서는 이 시스템을 구현하고, 데이터베이스화된 클릭스트림 정보를 이용한 구체적인 적용 사례를 들며, 제 5장에서 결론과 향후 연구를 기술한다.

2. 관련 연구

2.1 개인화된 웹 사이트

개인화(Personalization)란 인터넷 비즈니스 상의 모든 상품과 광고, 메뉴 등의 콘텐츠에 대해 고객 개개인의 성향과 취향에 맞도록 제공하여, 획일적인 서비스가 아닌 고객과 친근한 일대일 관계를 맺기 위한 전략이다[1]. 이러한 솔루션은 웹 사이트에 들어오는 고객의 프로파일을 결정하여, 고객 개개인의 요구에 맞는 상품이나 콘텐츠를 추천하고, 정교하게 타겟팅된 마케팅 캠페인을 실행하도록 지원한다. 한 명의 고객에 대한 특성과 구매패턴에 따라 적절한 정보와 상품을 제공하는 것은 인터넷의 특성을 매우 잘 활용하는 마케팅 기법이라 할 수 있을 것이다. 그러므로 웹 고객은 웹 사이트를 향해서 자신의 취향에 맞는 개인화된 서비스를 경험할 수 있다.

2.1.1 개인화된 웹 사이트의 현황 및 문제점

미국에서는 1997년 이후, 개인화란 개념이 구체적으로 웹 사이트에 구현되기 시작하면서 1998년 이후에는 범처럼 확산되어 우수한 웹 사이트들은 대부분 개인화된 솔루션을

갖추어 왔고, 이미 미국 인터넷 사용자의 29%가 웹 사이트 중 개인화된 페이지를 이용하고 있으며, 88%가 개인화된 웹 페이지를 보다 선호한다고 한다[1].

또한 이러한 개인화된 서비스를 제대로 제공하기 위해서는 충분한 프로파일 구축과 맞춤서비스를 위한 전처리 과정이 중요하다. 현재의 솔루션들은 불충분한 고객 데이터에 의존하기 때문에 정확한 고객 프로파일을 구축할 수 없다. 완벽한 프로파일 구축을 위해서는 모든 채널에서 발생하는 데이터를 통합하는 것이 필요하다. 웹 로그를 비롯한 Cache Log, Proxy Log, WAP Log, BBS Log, Chat Log, E-mail Log 등의 웹 데이터는 물론, 고객의 인구통계 정보와 거래 데이터 등 여러 유형의 데이터를 이용해야 한다.

먼저 사용자의 인터넷 사용 환경에서, 보다 편리하고, 보다 다양한 서비스를 제공받을 수 있는 체계가 제공되어야 하며, 그 기업의 특성에 맞추어 개인화 정도를 고려해야만, 개인화는 진정한 가치를 가질 수 있다. 다음은 웹 사이트의 개인화에 있어서 다룰 수 있는 주요 문제점을 정리한 것이다[12].

- 불충분한 고객 데이터의 활용
- 추천 알고리즘의 단순한 적용
- 데이터의 통합적 분석 및 피드백 부족의 구조적 한계

2.2 개인화된 웹 사이트에 사용되는 기술

2.2.1 로그파일분석

사용자가 웹 사이트를 이용하면 이에 대한 기록이 로그라는 형태로 흔적이 남는다.

로그분석이란 이 데이터를 기반으로 다양한 정보를 추출해 내는 것이다. 사용자가 웹 사이트에 접속한 후 모든 작업들은 웹 서버를 통해 이루어진다. 사용자가 웹 서버에 접속을 하게 되면 그 이후의 모든 작업들은 웹 서버에 미리 정해 놓은 위치에 데이터로 남게 된다. 특정 웹 페이지를 보기 위한 사용자의 요구로, 웹 서버는 해당 웹 페이지와 관련된 여러 파일 등에 접근하게 된다. 따라서 사용자가 요청하는 특정 웹 페이지뿐만 아니라 해당 웹 페이지와 관련된 이미지 파일, 이미지 데이터, Include 파일 등에 대한 정보가 로그 파일에 저장된다.

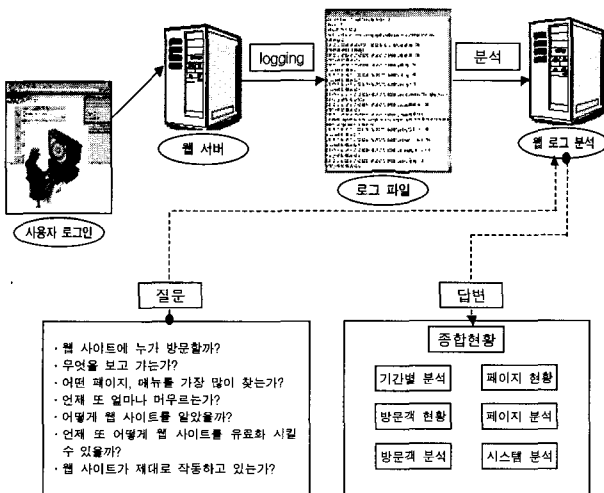
웹 서버 로그의 표준은 CLF(Common Log file Format)이다. 이러한 로그파일형식에는 Host, Identification, AuthUser, Time, Request, Status, Bytes등의 7가지필드가 기록된다. ECLF(Extended Common Log Format Standard)에는 두 개의 요소가 추가되었다. 대부분의 서버들이 NCSA의 표준로그파일형식(CLF)을 지원하고 있기 때문에 표준로그파일은 보통 Transfer 또는 Access Log File이라 불리며 파일이름은 access_log와 같이 정하고 있다. 다음 <표 2-1>은 대부분의 웹 서버에서 얻을 수 있는 전형적인 데이터 요소를 나열한 것이다[4].

<표 2-1> 웹 서버 로그 데이터 요소

데이터 요소	CLF*	ECLF*	설 명
Host	√	√	클라이언트의 정식 도메인 이름이나 클라이언트의 IP주소
Ident	√	√	클라이언트가 활성화된 것으로 파악된 경우 클라이언트가 제공한 ID정보
Authuser	√	√	비밀 번호로 보호된 문서에 대한 요청에 대해서는 요청에 사용된 사용자 ID가 된다.
Time	√	√	요청이 CLF형식으로 서버에 도달한 시간 (dd/Mmm/yyyy : hh : mm : ss영역)
Request	√	√	클라이언트에서 온 첫 번째 요청 라인
Status	√	√	클라이언트로 반환된 세 자리 상태 코드
Bytes	√	√	클라이언트로 반환된 바이트 수에서 HTTP 헤더를 제외한 것
Referrer		√	참조하는 서버의 URL
User-agent		√	클라이언트(브라우저)의 이름과 버전
Filename			파일 이름
Time-to-seve			요청을 처리한 시간(초)
IP-address			원격 호스트의 IP주소
Server-port			요청을 처리하는 서버의 정식 포트
Process-ID			요청을 처리한 자식 프로세스 ID
Formatted-time			지정된 strftime(3)형식으로 된 시간
URL-requested			요청된 URL경로
Server-name			요청을 처리한 서버의 정식 이름
Cookie			클라이언트의 쿠키 파일에서 불러온 쿠키의 값

* CLF(Common Log Format)
* ECLF(Extended Common Log Format Standatd)

로그 분석을 통해 웹 사이트는 사용자에 대해 보다 정확히 파악할 수 있게 된다. 하지만 사용자에 대한 분석만으로는 의미가 없다. 이러한 로그 분석을 통해 웹 사이트는 보다 개인화된 접근이 가능해지고, 마케팅 활동의 효과를 높일 수 있게 된다. 웹 사이트의 페이지뷰, 사용자별 페이지뷰, 접속장소 및 방식, 시간별 페이지뷰, 방문자수 등에 대한 현황 및 추세를 분석할 수 있다. 다음 (그림 2-1)은 웹 로그 프로세스를 나타낸 것이다.



(그림 2-1) 웹 로그 프로세스

2.2.2 쿠키파일분석

쿠키(Cookie)는 인터넷 웹 사이트와 인터넷 사용자 자신의 컴퓨터 사이에서 통신을 매개해주는 데이터 파일의 이름으

로 웹 브라우저에서 인터넷 사용자가 어떤 웹 사이트에 접속했는지 또 그 사이트 내에서 어떤 정보를 읽어 들이고 어떤 정보를 남겼는지 기록하는 것이다. 예를 들면, 어떤 사용자가 특정 웹 사이트에 접속한 후 그 사이트 내에서 어떤 정보를 보았는지 등에 관련된 기록을 남겨 놓았다가 다음에 접속하였을 때 그것을 읽어 이전의 상태로 유지하면서 검색할 수 있게 하는 역할을 한다.

방문자가 접속을 했을 경우, 방문자의 헤더(Header)부분에 관련 정보를 첨가해서 해당 서버로 접속을 시도했을 경우 쿠키를 헤더부분에 싣고 전송함으로써 쿠키를 설정할 수 있다. 일반적으로 배너 광고를 회전시키기 위해 사용되기도 하지만, 사용자가 쓰고 있는 브라우저의 형식 또는 그 웹 사이트에 이미 제공했던 다른 정보에 기초를 두어 서버에서 보낼 웹 페이지들을 사용자에게 맞추는 데에도 사용된다.

쿠키는 성과평가정보를 추출하기 위한 기초적인 정보를 제공할 수 있다. 즉, 쿠키를 통하여 방문자가 특정한 제품을 선택한 후, 구입을 위하여 쇼핑 카트에 제품목록을 등록한 사항 등을 파악할 수 있다. 쿠키파일 분석은 전체적으로 웹 사이트 사이를 이동하는 고객들의 욕구를 파악할 수 있게 하며, 고객의 반응을 검토함으로써 효과적인 마케팅을 위한 가치 있는 정보를 제공한다. 그러나 쿠키의 정보가 항상 정확하다고 볼 수 없으며, 다음과 같은 요인들이 쿠키의 정보를 부정확하게 만들 수 있다.

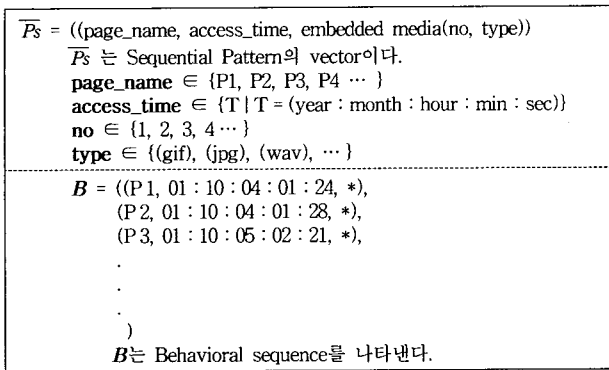
- 1인당 1PC 사용이 아닐 수 있음
- 오래된 쿠키 데이터는 지워질 수 있음
- 프락시 서버(Proxy Server)에 따라 쿠키가 달라질 수 있음
- 쿠키 사용을 허용하지 않는 클라이언트들이 있음

2.3 클릭스트림

클릭스트림(clickstream)이란 웹 사이트 상에서의 사용자의 행위 패턴을 파악할 수 있는 정보이다. 웹 사이트상의 사용자들이 컴퓨터 화면상의 어떤 객체를 선택하거나 원하는 항목을 실행하도록 하기 위해 마우스를 이용하여 선택하는 일련된 흐름의 정보들이다. 잠재 고객들이 경쟁자의 웹 사이트를 클릭 한번으로 손쉽게 넘나들기 시작하면서 클릭스트림 분석이 필요하다. 자사의 웹 사이트로부터 고객이 떠나지 않도록 하기 위하여 기업은 고객을 웹 사이트에 묶어두는 요소들을 끊임없이 첨가하여야한다. 이러한 요소들은 고객의 행동과 선호도, 구매 성향 등에 대한 지식을 기반으로 결정되며 고객의 최근 행위를 반영하는 고객 개개인에 맞춰진 역동적인 콘텐츠 제공으로 이어진다.

현재 기본적인 로그 분석들은 클라이언트로부터 요청 받은 개개의 파일에 대한 서버의 응답 단위인 히트수만을 이용해 사용자들의 행위 패턴을 분석하려고 한다. 히트(hit)란 웹 서버에 저장된 파일에 대해 이용자가 열람을 요청한 횟수, 즉 사용자가 인터넷 브라우저를 통해 특정 페이지를 보고자 클

릭하였을 때 이 요청한 내용 자체가 히트가 된다. 히트는 웹 서버에 있는 로그 파일에 기록된다. 히트수는 서버가 얼마나 많은 트래픽을 처리할 수 있는냐에 대한 의미 있는 척도가 되지만, 때로는 얼마나 많은 페이지가 보여졌느냐에 대해서는 잘못된 지표가 될 수도 있다. 따라서 웹 페이지내의 사용자들의 행위 패턴을 수집하기 위해 사용자들이 원하는 상품을 클릭 하였을 경우 즉, 클릭 이벤트가 발생할 경우 상품에 대한 정보를 수집하였다. 다음 (그림 2-2)는 사용자들의 웹 상에서의 행위 패턴을 수식으로 나타낸 것이다.

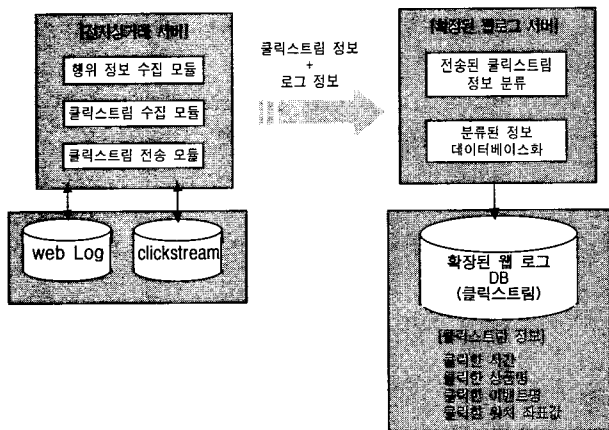


(그림 2-2) 사용자들의 행동 패턴의 정의

3. 클릭스트림을 위한 확장된 웹 서버 시스템의 설계

3.1 클릭스트림을 위한 확장된 웹 로그 시스템의 구성

본 절에서는 사용자가 전자상거래 사이트를 방문하여 관심 있는 상품을 클릭하면, 클릭스트림 정보를 웹 로그 서버로 전송하도록 모델링하였다. 이러한 클릭스트림 정보들은 쿠키에 저장하지 않고 클릭스트림이 발생할 때마다 실시간으로 전달되도록 설계하였다. 그리고 이렇게 전송된 데이터들은 사용자들의 행위패턴을 분석하기 쉽도록 하기 위해 분류하여 데이터베이스에 저장하였다. 이와 같은 내용은 (그림 3-1)과 같이 통합되어 사용자들의 웹 사이트내의 행위 패턴을 분석을 위한 웹 로그 시스템을 구성한다.



(그림 3-1) 클릭스트림 분석을 위한 전체적인 시스템 구성도

3.2 클릭스트림 수집과 전송을 위한 전자상거래 서버 설계

3.2.1 클릭스트림 수집을 위한 전자상거래 서버 설계

웹 사이트에 방문한 사용자가 관심 있는 상품을 클릭하는 경우 기본적인 웹 로그 정보와 함께 클릭한 이벤트 명, 상품에 대한 정보인 클릭한 시간, 상품명, 클릭한 위치정보를 수집할 수 있도록 설계하였다. 여기서 클릭스트림 정보를 전송하기 위해 쿠키를 사용하지 않고, 클릭 이벤트가 발생할 때마다 사용자의 클릭정보가 웹로그 서버로 전송되도록 설계하였다.

사용자가 클릭 이벤트를 발생시켰을 때 정보를 수집하는 방법은 다음 5 단계로 처리된다.

- 단계 1: 먼저, 웹 사용자는 URL을 이용해 전자상거래 서버에 문서를 요청한다. 두 대의 컴퓨터가 네트워크를 하기 위해서는 IP주소와 포트(PORT)를 사용하는데, 웹서버는 기본적으로 80 포트를 사용하지만, 본 논문에서는 90 포트를 사용한다.
- 단계 2: 전자상거래 서버는 웹 사용자로부터 90 포트를 통해 들어온 요청을 해석해 URL에 해당하는 문서를 찾는다.
- 단계 3: 웹 사용자가 요청한 문서가 정상적으로 존재할 때, 웹 서버는 해당 문서를 웹 사용자에게 전달한다.
- 단계 4: 웹 사용자는 웹 서버가 전달해 준 문서를 해석해 사용자의 화면에 보여준다. 이때 사용자는 자신이 관심 있는 이미지를 클릭한다.
- 단계 5: 해당 문서내에서 자바스크립트 함수를 이용해 클릭한 이미지에 대한 이벤트 명, 클릭한 시간, 클릭한 위치 좌표, 클릭한 이미지 정보를 수집하기 위해 웹 로그 서버로 전송한다.

3.2.2 클릭스트림 정보 전송을 위한 확장된 웹 로그 설계

본 절에서는 클릭스트림 정보를 수집하기 위해 기본적인 웹 로그 정보와 클릭스트림 정보를 추가하여 사용자가 원하는 상품을 클릭 하였을 경우, 기본적인 웹 로그 정보와 함께 클릭스트림 정보가 수집되도록 설계한다. 본 논문에서는 모든 로그의 기본이 되는 CLF(Common Log file Format)에 사용자의 클릭스트림 정보를 포함한 부분으로 확장하였다. 다음 <표 3-1>은 클릭스트림에 의해 확장된 웹 로그의 종류와 내용을 설명하였다.

전자상거래상의 사용자가 원하는 상품을 클릭한 경우 클릭스트림 정보가 기본적인 웹 로그 정보에 추가되어 수집된다. <표 3-1>에서 보는 바와 같이 CLF형식에 사용자의 클릭스트림 정보인 'EventName', 'ImgName', 'ClickTime', 'ClientX', 'ClientY' 등의 정보를 로그에 포함하도록 설계하였다.

<표 3-1> 클릭스트림에 의해 확장된 웹 로그의 종류와 내용

포맷 구분		처리 내용	
Extended Web Log	CLF (Common Log file Format)	Host	클라이언트의 정적 도메인 이름이나 클라이언트의 IP주소
		Ident	클라이언트가 활성화된 것으로 파악된 경우 클라이언트가 제공한 ID정보
		Authuser	비밀번호로 보호된 문서에 대한 요청에 대해서는 요청에 사용된 사용자가 ID가 됨
		Time	요청이 CLF형식으로 서버에 도달한 시간
		Request	클라이언트에서 온 첫 번째 요청 라인
		Status	클라이언트로 반환된 세 자리 상태 코드
		Bytes	클라이언트로 반환된 바이트 수에서 HTTP 헤더를 제외한 것
Clickstream		EventName	사용자가 발생시킨 이벤트명을 설명함.
		ImgName	사용자가 클릭한 아이템명을 설명함.
		ClickTime	사용자가 아이템을 클릭한 시간을 설명함.
		ClientX	사용자가 아이템을 클릭한 위치 좌표값
		ClientY	사용자가 아이템을 클릭한 위치 좌표값

3.3 확장된 웹 로그 서버 시스템의 설계

3.3.1 확장된 웹 로그 서버의 클릭스트림 처리

확장된 웹 로그 서버 시스템은 전자상거래 서버로부터 전송된 클릭스트림 데이터를 분류하고 저장하는 기능을 한다. 또한 클릭스트림 데이터를 데이터베이스화함으로써 관리자가 분석을 효율적으로 할 수 있도록 설계하였다.

기본적인 웹 서버는 사용자가 요청한 웹 문서를 웹 브라우저에 전송하는 역할을 담당한다. 즉, 웹 서버는 인터넷 서버의 메모리에 상주하고 있으며, 사용자의 요청이 발생하면 이를 효율적으로 해결하기 위하여 자식 프로세스(Child Process)를 생성하여 요청된 사안을 처리하도록 하고 웹 서버는 다시 사용자의 요청을 받기 위해 준비한다.

확장된 웹 로그 서버는 기본적인 웹서버이며 클릭스트림으로 확장된 웹 로그 데이터를 받아들이도록 설계하였다. 이 서버는 WWW의 기본 프로토콜인 HTTP 요청에 응답한다. 사용자가 브라우저를 통해서 전자상거래 서버에서 제공해 주는 페이지를 보고 클릭이벤트를 발생시켰을 경우 이 클릭 정보가 웹로그 서버로 전송된다. 전송된 클릭스트림 정보를 받은 웹 로그 서버는 이러한 데이터들을 분류하여 데이터베이스화한다.

GET과 POST는 웹 브라우저에서 입력한 값이 서버로 넘어가는 방식을 지정하는 것이다. GET 방식의 경우에는 QUERY_STRING 환경변수를 통해 전달되고, POST 방식의 경우에는 웹 서버와 CGI 프로그램간에 연결된 파이프라인을 통해 전달된다. 주로 LOGIN 과정에서 입력한 암호 데이터와 같이 보안상 중요한 데이터를 전달해야 할 경우, 파일 업로드와 같이 많은 양의 데이터를 전달해야 할 경우, 방명록 작성과 같이 입력 폼의 데이터가 많은 경우 POST 방식을 이용해서 매개변수를 웹 서버로 전달한다. FORM

태그의 METHOD 속성을 'POST'로 설정한 후 서비스를 요청하여 클릭스트림 정보를 확장된 웹 로그 서버로 전송한다. 그 수행 순서를 살펴보면, 다음과 같다.

단계 1 : 이 과정은 HTTP GET 방식의 요청과 달리 약간의 차이가 있다. 웹 브라우저는 POST 방식에 해당하는 HTTP 요청 헤더를 생성한 후, 폼에 입력된 데이터를 URL 인코딩한다. HTTP 요청 메시지의 엔티티 바디로 추가해 HTTP POST 방식의 요청을 수행한다. URL인코딩은 모든 매개변수를 name = value 형태로 만들고, 두 개 이상의 매개변수를 전달해야 할 경우에는 각 매개변수를 구분하기 위해 '&'를 사용한다.

단계 2 : POST 방식으로 받은 클릭스트림 정보들을 타이틀과 내용을 구분하여 본 연구에서는 OLE DB상의 ADO 기술을 이용하였다.

3.3.2 클릭스트림 데이터베이스 스키마 설계

본 논문에서는 사용자들의 웹 사이트내에서의 행위패턴을 분석하기 위해서, 즉 클릭스트림을 수집하기 위해 기본적인 웹 로그 정보에 클릭스트림 정보를 추가하여 웹 로그 정보를 확장하여 저장하도록 설계하였다. 또한 로그 데이터를 text 파일이 아닌 데이터베이스에 저장함으로써 검색이 용이하도록 설계하였고, 로그를 담당하는 서버에서 처리함으로써 전자상거래 서버 자체의 부하가 줄어들게 되며, Microsoft Access를 사용하여 클릭스트림 정보를 저장하였다. Microsoft Access는 사용이 간단하며 실시간으로 데이터를 조회할 수 있는 데이터베이스를 만들 수 있다. 그래서 검색이 용이하고 각종 분석들에 편리하게 이용할 수 있도록 설계하였다. 저장요소로는 <표 3-2>에서 보는 바와 같이 'SessionID', 'EventName', 'Imgname', 'Clicktime', 'ClientX', 'ClientY'을 들 수 있다. 여기서 'SessionID'는 사용자를 구분해주며, 'EventName'은 사용자가 발생한 이벤트에 대한 구분을 해주며, 'ImgName'은 사용자가 클릭이벤트를 발생시킨 상품명이다. 'ClientTime'은 사용자가 클릭이벤트를 발생시킨 시간이며, 'ClientX', 'ClientY'는 사용자가 클릭이벤트를 발생시킨 화면상의 좌표값을 나타낸다.

<표 3-2> 클릭이벤트에 의해 확장된 웹 로그 구성요소

구분	구성 요소	설 명
Who	SessionID	사용자 구분
Why	EventName	클릭한 이벤트 명(단순한 클릭, 장바구니 클릭 등)
What	ImgName	클릭한 상품명
When	ClickTime	클릭한 시간
WhereX	ClientX	클릭한 위치정보(X좌표)
WhereY	ClientY	클릭한 위치정보(Y좌표)

다음 표는 클릭스트림 정보를 저장하는 데이터베이스 스키마를 나타낸 표이다.

〈표 3-3〉 웹 로그를 저장하는 데이터베이스 스키마

필드 이름	형식	크기	설 명
sessionid	텍스트	50	사용자의 세션아이디
eventname	텍스트	255	클릭한 이벤트명
imgname	텍스트	50	클릭한 상품명
clicktime	텍스트	50	클릭한 시간
clientx	정수	4	클릭한 위치
clienty	정수	4	클릭한 위치

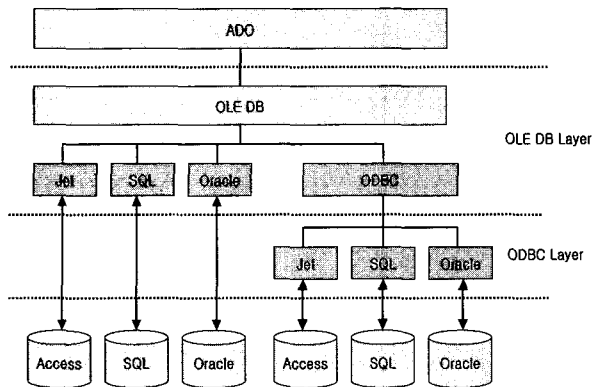
전송 받은 클릭스트림 정보를 버퍼에 저장하여 라인별로 다른 버퍼에 저장하여 타이틀과 내용을 분리하여 데이터베이스에 저장하였다. 이때 OLE DB를 활용하여 여러 가지 기존의 데이터베이스는 물론 엑셀등 스프레드시트파일, 텍스트 파일등 확장된 범위의 데이터를 활용하도록 고안된 객체인 ADO를 이용하여 액세스 DB로 저장하도록 설계하였다.

4. 클릭스트림에 의해 확장된 웹 로그 시스템의 구현 및 평가

본 장에서는 제안된 시스템을 구성하는 각 요소들의 구현 내용을 설명한다. 전자상거래 서버에서 사용자가 클릭이벤트를 발생시켰을 경우 클릭스트림 데이터를 수집하는 프로그램, 이 프로그램에 의해 생성된 클릭스트림 데이터를 웹 로그 서버로 전송하는 방법을 설명한다. 그리고 전송된 클릭스트림 데이터를 전송 받아서 저장, 분류하는 프로그램을 설명한다.

이러한 구현을 위하여 운영체제로 마이크로소프트사의 Windows 2000 Advanced Server, 전자상거래 서버로 ASP 엔진이 설치된 IIS 5.0을 사용하였으며, 전자상거래 서버 프로그램 개발과 클릭스트림 정보를 수집하는 프로그램은 JavaScript 언어를 사용하였다. 그리고 클릭스트림 정보를 분류하고 저장하는 웹 로그 서버는 마이크로소프트 Visual C++ 6.0언어를 사용하여 구현하였다.

(그림 4-1)은 OLE DB 프로바이더를 사용한 웹 로그 서버의 작업 환경이다. OLE DB 프로바이더가 ODBC를 포함하고 있어서 ODBC 드라이버를 사용한 시스템의 추가 확장이 가능하다.



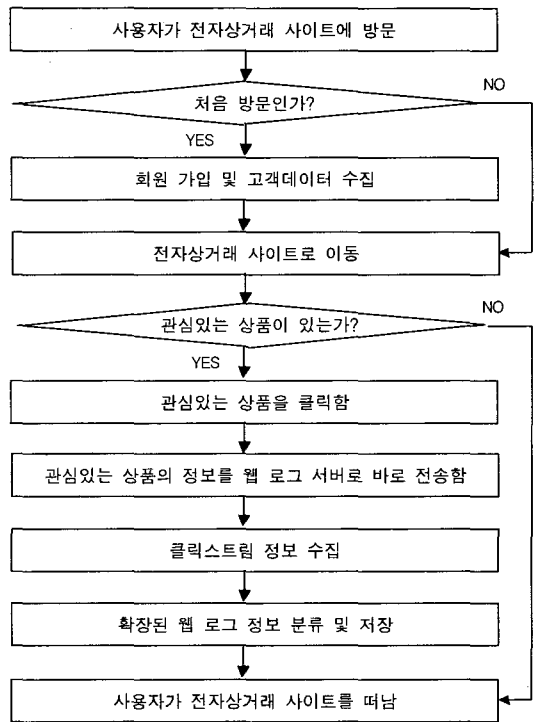
(그림 4-1) OLE DB 프로바이더를 사용한 웹 로그 서버의 작업 환경

OLE DB프로바이더가 ODBC를 포함함으로써 OLE DB가 기존의 ODBC 데이터 소스들을 액세스 할 수 있도록 하는 것이다.

4.1 전자상거래 서버상의 클릭스트림 정보 생성 모듈

전자상거래 서버는 고객관련 데이터와 고객 행동 데이터를 수집할 수 있도록 설계하였다. 사용자의 클릭스트림 정보를 저장하기 위한 모듈은 ASP 문서에서 JavaScript로 구현하였다. 웹 상에서의 회원 등록 과정을 거친 사용자가 관심 있는 상품을 클릭한 경우 기본적인 웹 로그 정보에 클릭스트림 정보를 추가하여 상품을 클릭한 즉시 쿠키에 저장하지 않고 바로 웹 로그 서버로 전송하도록 구현하였다.

그 수행과정을 살펴보면 (그림 4-2)와 같다.



(그림 4-2) 클릭스트림 정보의 수집 과정

(그림 4-2)의 각 단계는 사용자가 전자상거래 사이트에 방문하여 관심 있는 상품에 대해 클릭이벤트를 발생시킴으로써 사용자들의 행위정보인 클릭스트림 정보를 수집할 수 있도록 구현하였다.

4.1.1 클릭스트림 정보의 수집 방법

사용자가 관심 있는 상품을 클릭하였을 경우, 즉 클릭 이벤트를 발생시킨 경우 웹 페이지내의 JavaScript함수에 의해 클릭한 상품에 대한 정보 즉, 클릭한 시간, 클릭한 상품명, 클릭한 이벤트 명, 위치정보를 읽어 들이는 부분이다. 다음 (그림 4-3)은 사용자가 클릭이벤트를 발생하였을 경우, 클릭한 상품명과 클릭한 시간을 수집하는 부분이다.

```
function ackData (img_name) {
    var thisTime = new Date ();
    CTime = (thisTime.getYear () + "-" + (thisTime.getMonth () + 1) + "-" +
        thisTime.getDate () + "-" + thisTime.getHours () + "-" +
        thisTime.getMinutes () + "-" + thisTime.getSeconds ());

    document.form1.imgname.value = img_name ;
    document.form1.clickTime.value = CTime ;
    document.form1.submit ();
    ...
}
< a href = "javascript : ackData ('<% = rs ("book_isbn") %> ')"
... >
< imgsrc = "... / < % = rs ("book_s_image") % > " >
< / a >
...
```

(그림 4-3) 클릭한 상품명, 시간을 수집하는 방법

다음은 사용자가 클릭한 위치 좌표값(X,Y)을 수집하는 부분이다. 사용자가 관심 있는 상품을 클릭했을 경우 클릭한 위치 좌표값을 얻기 위한 부분이다.

```
function clientCoords () {
    var clientInfo = "";
    clientInfo = "xcoordinate : " + window.event.clientX + "\r";
    clientInfo += "ycoordinate : " + window.event.clientY + "\r";
    document.form1.clientx.value = window.event.clientX ;
    document.form1.clienty.value = window.event.clientY ;
}
< a ... onclick = "clientCoords()" >
< imgsrc = "... / < % = rs ("book_s_image") % > " > < / a >
...
```

(그림 4-4) 클릭된 위치 좌표값을 수집하는 방법

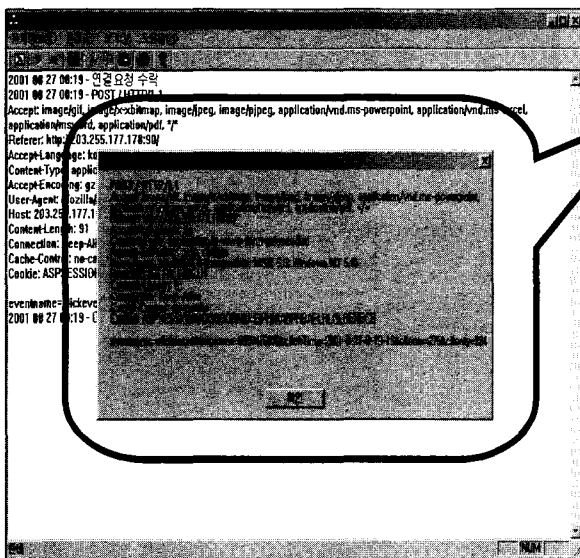
4.1.2 클릭스트림 정보의 전송 방법

생성된 클릭스트림 정보를 웹 로그 서버로 전송하기 위해 웹 브라우저를 통한 서버로의 데이터 전송방법 중 가장 대표적인 것인 HTML의 <FORM>과 </FORM>태그를 이용하였다. <FORM> 태그에서의 전송 방식에는 POST 방식과 GET 방식이 있는데, POST 방식을 이용하여 전송하였다. <FORM>은 입력이 전달될 프로그램의 이름과 전달방법을 표시하는데 사용된다. ACTION과 METHOD영역을 포함하고 있다. ACTION 영역의 값은 구동시킬 CGI 프로그램의 경로와 이름을 나타내며, METHOD는 GET이나 POST중 한 방법을 선택하게 된다. <FORM>에서 METHOD = "GET"을 쓸 경우 모든 입력이 QUERY_STRING이라는 환경변수(Environment Variable)로 전달된다. 반면 METHOD = "POST"를 사용하는 경우에는 표준 입력(Standard Input)으로 입력된다.

4.2 확장된 웹 로그 서버의 클릭스트림 정보 분류 및 저장 모듈

확장된 웹 로그 서버의 작업 순서는 사용자로부터 얻은 클릭스트림 정보를 받고 분류하여 데이터베이스에 저장하게 된다. 프로그램은 마이크로소프트 Visual C++ 6.0 언어를 사용하였다. 다음 (그림 4-5)는 전자상거래 서버에서 구현한 웹 로그 서버로 보낸 클릭스트림 데이터를 로그 메시지 형태로 전송된 결과를 보인 것이다.

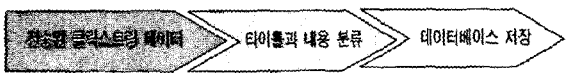
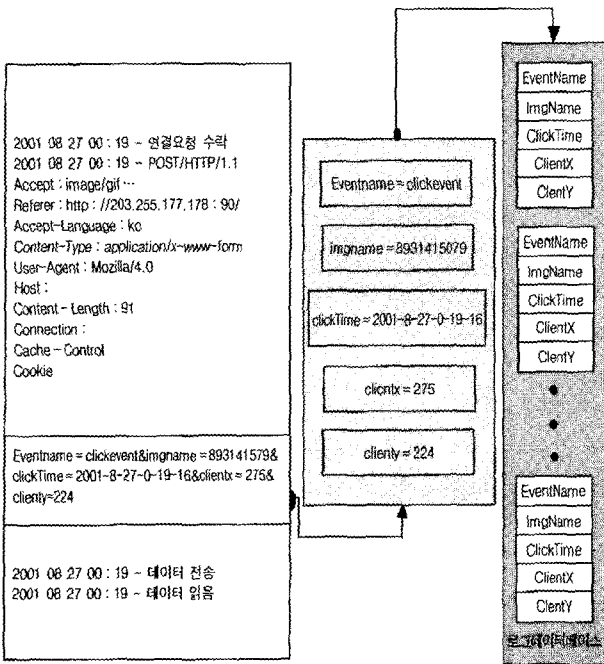
클릭스트림 데이터를 데이터베이스에 저장할 때 OLE DB Provider 의 ADO기술을 사용하였다. 다음 (그림 4-6)은 로그 메시지를 분류하고 데이터베이스에 저장되는 것을 보여준다.



```
.....
2001 08 27 00:19 - 연결요청 수락
2001 08 27 00:19 - POST / HTTP/1.1
Accept: image/gif, .....
Referer: http://203.255.177.178:90/
Accept-Language: ko
Content-Type: application/x-www-form-urlencoded
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0
Host: 203.255.177.178:9111
Content-Length: 91
Connection: Keep-Alive
Cache-Control: no-cache
Cookie: ASPSESSI.....

eventname=clicktime=2001-08-27-0-19-16&clientx=175&
clienty=224
2001 08 27 00:19 - 데이터 전송
2001 08 27 00:20 - 데이터 읽음
.....
```

(그림 4-5) 웹 로그 서버로 전송된 클릭스트림 데이터의 예



(그림 4-6) 클릭스트림 데이터를 분류하고 저장하는 과정

전자상거래용 서버에서 전송된 클릭스트림 데이터는 메시지 형태로 웹 로그 서버로 전송하게 된다. 전송된 메시지는 임시로 버퍼에 저장된 후 의미 단위로 분석하기 위해 타이틀명과 내용으로 분리되어 저장된다. 다음 단계는 로그 메시지가 데이터베이스로 저장되는 알고리즘을 나타낸 것이다.

- 단계 1 : 전송된 메시지를 버퍼에 복사한다.
- 단계 2 : 버퍼의 내용을 한 라인씩 읽어서 버퍼에 저장한다.
- 단계 3 : 저장한 버퍼의 캐릭터를 하나씩 읽어 타이틀과 같은 내용부분을 각각의 저장소에 저장한다.
- 단계 4 : 전체 버퍼의 내용이 끝나면 버퍼를 비운다.
- 단계 5 : 데이터베이스에 연결하기 위해 ADO object포인터를 선언한다.
- 단계 6 : 데이터베이스 인터페이스 포인터를 생성함.
- 단계 7 : 포인터를 통해 데이터베이스와 연동 한다.
- 단계 8 : 데이터를 각 레코드 셋(record set) 변수에 저장한다.
- 단계 9 : 테이블에 데이터를 저장한다.
- 단계 10 : 데이터베이스와의 연결을 닫는다.

다음은 OLE DB Provider 상의 ADO 기술을 사용하여 분류된 데이터들이 데이터베이스로 저장되는 것을 보여주는 부분이다. 먼저 AOD object pointers 선언하고, 선언된 pointers를 초기화 한 후, 레코드 셋 포인터(pRstWeblog)와 데이터베이스에 연결하기 위한 포인터(pConnection)를 선언해 준다. 데이터가 저장될 위치를 설정한다.

```

HRSULT hr = S_OK ;
if(FAILED( :: CoInitialize(NULL)))
    return ;
if (SUCCEEDED(hr))
{
    _RecordsetPtr pRstWeblog = NULL ;
    _ConnectionPtr pConnection = NULL ;
    IADORecordBinding *picRs = NULL ;
    CWebLogRs m_pLogRs ;
    HRSULT hr = S_OK ;

    _bstr_tstrConn("Provider = Microsoft.JET.OLEDB.4.0 ;
        DataSource = D : \\...\\logaccess.mdb") ;
    
```

(그림 4-7) 클릭스트림 데이터를 데이터베이스에 저장하는 방법

4.3 확장된 웹 로그 데이터베이스의 사용 예

4.3.1 클릭스트림 데이터의 이용 사례 I

본 논문에서 설계, 구현한 전자상거래용 서버에서 수집된 클릭스트림 데이터를 사용자 중심으로 상품의 선호도를 평가할 수 있도록 클릭 이벤트 중심으로 피벗테이블을 이용하여 분석하였다. 다음 표는 실제 분석에 사용된 구성요소이다.

<표 4-2> 사용자들의 선호도를 분석하기 위해 사용된 구성요소

필드 이름	형식	크기	설명
sessionid	텍스트	50	사용자의 세션아이디
eventname	텍스트	255	클릭한 이벤트명
imgname	텍스트	50	클릭한 상품명

다음 실험을 통해 사용자별로 어떤 상품을 단순히 클릭만 하였는지, 구매를 하기 위해 장바구니에 담았는지, 어떤 상품을 선호하는지 쉽게 분석할 수 있다. (그림 4-8)에서 ①은 상품을 클릭한 모든 사용자를 나타내는 것이며, ②는 사용자들이 관심있는 상품을 클릭한 상품명이며, ③은 단순히 상품을 클릭한 횟수를 나타내며, ④는 관심있는 상품을 장바구니에 담기 위해 클릭한 횟수를 나타낸다.

세션 ID	이벤트명	상품명	클릭 횟수	장바구니 담기 횟수	총 합계
8988751019	clickevent	precart	11	3	14
8988751248			1	3	4
8988751310			17	2	19
8931409311			2		2
8931414978			1		1
8931415079			21	1	22
8985228959			6	1	7
8988751019			9	3	12
8988751027			1		1
8988751108			2	5	7
8988751167			7	1	8
8988751248			5		5
8988751310			3		3
총 합계			86	19	105

(그림 4-8) 사용자별 클릭스트림 통계량

다음 (그림 4-9)는 개인 사용자별 통계를 분석한 그림이다. 이런 분석을 통해서 사용자별로 선호된 상품을 추천할 수 있다. (그림 4-8)에서와 마찬가지로 ①은 상품을 선택한 개별 사용자를 ②는 사용자들이 관심 있는 상품을 클릭한 상품명이며, ③은 단순히 상품을 클릭한 횟수를 나타내며, ④는 관심 있는 상품을 장바구니에 담기 위해 클릭한 횟수를 나타낸다.

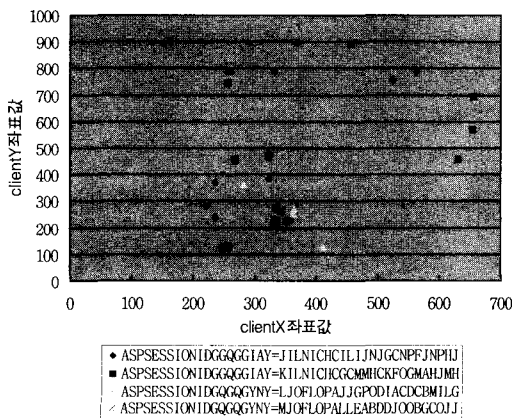
sessionid		ASPSSESSIONIDGGGGIAY=JLNCHCLUNUGXFFNPHU ①		
개수	imgname	eventname	precart	총합계
	imgname ②	clickevent ③		
	8988751248		1	1
	8988751310		2	2
	8931415079		1	1
	8985226959		1	1
	8988751027		1	1
	8988751018		1	2
	8988751167		1	1
총합계			6	9

(그림 4-9) 사용자별 클릭스트림 통계량

이러한 분석으로 어떤 사용자가 어떤 상품의 관심을 가지고 있는지 클릭한 횟수를 통해서 알 수 있으며 몇번의 클릭 후에 제품을 구매하는지도 알 수 있다. 그러므로 한번도 클릭하지 않은 상품이나, 액세스 된 적이 없는 페이지를 찾아낼 수도 있다. 또한 방문자를 실제 구매자로 끌어들이는 전환율을 높이기 위한 전략을 세우는데도 이용될 수 있다.

4.3.2 클릭스트림 데이터의 이용 사례 II

사용자들이 클릭한 메뉴, 광고, 클릭한 위치는 사용자들의 행동 성향을 분석하기 위해 중요한 자료가 된다. 그러므로 사용자들이 자주 클릭 되는 위치 정보를 찾아내어 광고나 추천 상품을 배치함으로써 개인화된 서비스를 제공할 수 있으며, 역동적인 콘텐츠 제공으로 이어진다. 다음 (그림 4-10)은 사용자들의 클릭한 위치정보를 나타낸 그림이다. X, Y좌표값은 사용자들이 클릭한 Pixel 좌표값을 나타냄으로써 사용자들이 주로 클릭하는 위치를 정보를 한눈에 파악할 수 있다.



(그림 4-10) 사용자별 웹 사이트상의 클릭 이벤트가 발생한 위치 값

5. 결 론

본 논문에서는 개인화를 위한 전처리과정인 클릭스트림 분석 결과로 얻어진 확장된 웹 로그 정보를 통해서 사용자들의 성향을 분석할 수 있는 확장된 웹 로그 서버 시스템을 제안하였다. 단순히 클릭 한번으로 사용자가 관심 있는 사이트로 이동하므로 클릭스트림 데이터에 대한 분석은 중요하며, 사용자들의 관심을 빠르게 파악할 수 있게 되었다. 이러한 클릭스트림을 분석해 사이트의 디자인을 최적화 시켜주는 기능을 할 수 있으며, 사용자들이 가장 많이 클릭하는 화면의 메뉴를 가장 클릭하기에 좋은 위치에 배치하고, 사용자들이 최종적으로 원하는 정보나 상품화면으로 찾아 들어가기 위한 과정에 있어서의 편의성을 제공할 수 있다.

또한, 단순한 웹 로그 정보가 아닌 클릭스트림에 대한 정보를 수집하고 전송하는 전자상거래용 서버와 전송된 클릭스트림 데이터를 분류하고 데이터베이스에 저장하는 웹 로그 시스템을 제안하였다. 클릭스트림 데이터를 데이터베이스화함으로써 관리자가 정보를 효율적으로 관리, 기존의 데이터를 삭제, 변경시키는 작업을 저장된 데이터가 일관성을 유지할 수 있도록 하였다.

본 연구를 통하여 사용자의 성향을 가장 빠르게 파악할 수 있는 클릭스트림 데이터를 수집, 분류, 저장함으로써 맞춤형 서비스에 대한 개선된 사용자의 프로파일을 구축 할 수 있었다. 보다 효율적인 웹 로그 분석을 통한 사용자 프로파일링을 위해서는 사용자의 등록정보인 고객 데이터와 웹 로그를 통한 사용자 행위 데이터를 통합한 분석이 필요하며, 웹 로그 분석에 적절한 패턴분석, 군집분석 등 마이닝(Mining) 기법을 적용하여 웹사이트 성격에 맞는 사용자 프로파일을 생성에 대한 연구가 이루어져야 할 것이다.

참 고 문 헌

- [1] 강성진, 박주석, 정태명, "전문가를 위한 전자상거래", 매일경제신문사, pp.147-197, 2000.
- [2] 김기수, 이상우, "인터넷 전자상거래 쇼핑몰 구축에서 운영까지", 비엔씨, pp.17-56, 1993.
- [3] 박용찬, "e-비즈니스 파워", pp.27-58, 2000.
- [4] Ralph Kimball, Richard Merz, "The Data Webhouse Toolkit," WILEY, pp.39-80, 2000.
- [5] 광미라, 조동섭, "패킷 정보를 이용한 웹 사용 분석에 관한 연구", 대한전기학회/대한전자공학회 시스템 및 제어분야 합동 추계학술대회논문집, pp.453-455, 2001.
- [6] Buchner A. G, Anand S. S., Mulvenna M. D. and Hughes J. G., "Discovering Internet Marketing Intelligence through Web Log Mining," Proc. Unicomm99 Data Mining & Data-

warehousing : Realising the full Value of Business Data, pp.127-138, 1999.

- [7] Chung Wei-Lun, Yuan Soe-Tsyr, "A Synthesized Learning Approach for Web-Based CRM," WEBKDD'2000, pp.43-47, 2000.
- [8] Cooley Robert, Tan Pang-Ning, Srivastava Jaideep, "Discovery of Interesting Usage Patterns from Web Data," LNCS/LNAI Series. Springer-Verlag, 2000.
- [9] Mobasher Bamshad, Dai Honghua, Luo Tao, Nakagawa Miki, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization," WEBKDD-Special Issue for the Data Mining and Knowledge Discovery journal, 2000.
- [10] Moe Wendy W., Fader Peter S., "Capturing Evolving Visit Behavior in Clickstream Data," 2001.
- [11] Pretschner Alexander, Gauch Susan, "Personalization on the Web," 1999.
- [12] Tsvi Kuflik, Peretz Shoval. "User Profile Generation for Intelligent Information Agents-Research in Progress," 1999.
- [13] Osmar R. Zaiane, Man Xin, Jiawei Han, "Discovering Web Access Patterns and Trends by Applying OLAP and Data mining Technology on Web Logs," 1998.
- [14] Welcome To OpenSymphony, "Clickstream Overview," <http://www.opensymphony.com/clickstream/>
- [15] Owen Graduate School of Management Vanderbilt University "Modeling the Clickstream : Implications for Web-Based Advertising Efforts," <http://www2000.ogsm.vanderbilt.edu/papers/clickstream/clickstream.html>
- [16] personalization, <http://www.personalization.co.kr/>



강 미 정

e-mail : woodrock@ewha.ac.kr
 2000년 경성대학교 컴퓨터과학과 졸업
 (이학사)
 2002년~이화여자대학교 과학기술 대학원
 컴퓨터학과 졸업(공학석사)
 현재 쌍용정보통신 SI 기술연구소 연구원

관심분야 : 지식공학, 웹 마이닝, E-Commerce



정 옥 란

e-mail : orchung@ewha.ac.kr
 1993년 전북대학교 전자계산학과 졸업
 (이학사)
 1995년 전북대학교 대학원 정보과학과
 졸업(이학석사)
 1999년~현재 이화여자대학교 과학기술
 대학원 컴퓨터학과 박사과정

관심분야 : 웹 마이닝, E-Commerce, Multimedia



조 동 섭

e-mail : dscho@ewha.ac.kr
 1979년 서울대학교 전기공학과 졸업(공학사)
 1981년 서울대학교 대학원 전기공학과
 졸업(공학석사)
 1986년 서울대학교 대학원 컴퓨터공학과
 졸업(공학박사)

1985년~현재 이화여자대학교 컴퓨터학과 교수
 1996년 University of California, Irvine Visiting Scholar
 관심분야 : 컴퓨터구조 및 인터넷 공학, 컴퓨터비전, 컴퓨터그래
 피스, 가상교육