

지능형 웹로그 서버를 이용한 전자상거래 사용자 패턴 수집 시스템

한지선*, 강미정, 조동섭
이화여자대학교 컴퓨터학과

Customer's Pattern Analysis System using Intelligent Weblog Server

Ji-Seon Han, Mi-Jung Kang, Dong-Sub Cho
Dept. of Computer Science and Engineering Ewha Womans University

Abstract - 전자상거래에서 쇼핑몰의 개인화된 서비스를 제공하기 위해서는 소비자의 구매 패턴을 분석하는 것이 필요하다. 이러한 패턴을 효과적으로 분석하기 위해 웹사이트 상에서 사용자 행동 패턴 정보를 수집해야 한다. 본 논문에서는 사용자 패턴 수집 시스템으로 쇼핑몰 서버에 기능을 추가하고 지능형 웹로그 서버를 정의하며 이를 설계, 구현하였다. 전자상거래 쇼핑몰 서버에는 사용자 행위 정보를 로그에 포함시켜 지능형 웹로그 서버에 전송하는 기능을 추가하였다. 그리고 지능형 웹로그 서버는 쇼핑몰 서버로부터 받은 로그 데이터를 분석하고 데이터베이스화하여 저장한다. 이때 데이터베이스 저장 기술로 OLE DB Provider상에서 수행되는 ADO기술을 사용한다. 그리고 저장된 데이터베이스를 레코드셋 단위로 원격에서 제어 가능하게 한다. 또 생성된 데이터베이스에서 필요한 데이터를 선별하여 XML DB로 저장한다. 이와 같은 사용자 패턴 수집 시스템은 데이터베이스 접근 속도가 빠르고, 관계형이나 비관계형 모두의 데이터베이스 접근이 가능하다는 장점을 가지며, 원격 데이터 베이스 접근 시 서버의 부하를 줄일 수 있다는 장점이 있다.

1. 서 론

인터넷을 매개로 한 전자상거래(Electronic Commerce)의 등장은 전통적인 무역거래의 관념과 패턴을 완전히 뒤바꾸어 놓았고 시간과 공간의 제약을 훌쩍 뛰어넘어 언제 어디서나 원하는 상품의 거래가 가능하도록 하였다[1]. 지식정보사회 구현을 가능하게 하는 핵심적인 요소로 전자상거래가 세계적인 관심의 초점이 되고 있고, 전자상거래 시스템을 도입하는 것은 이제 모든 기업이 경쟁력 향상을 위한 필수 조건이 되고 있다.

급변하는 인터넷 환경에서 이러한 전자상거래 시스템의 성공적인 운영은 방문자들의 행위 정보를 얼마나 잘 수집하고 분석하여 이해하느냐에 달려 있다[2]. 이 같은 정보는 웹로그 분석을 통해서 얻을 수 있으므로 이를 만족하기 위한 웹로그 분석 시스템들이 등장했다. 그 예로 국내의 Syswatcher나 국외의 Webalizer등이 있으며 이 외에도 많은 다른 웹로그 분석 시스템들이 존재한다. 그러나 이들 웹로그 분석 시스템들은 단순히 기존의 웹사이트의 로그파일 정보를 분석하고 가시화 하는 수준이다. 그러므로 본 논문에서는 사용자 행위 정보를 수집하고 이를 데이터베이스에 저장해서 원격으로 제어할 수 있는 사용자 패턴 수집 시스템을 설계, 구현하였다.

2. 관련 연구

2.1 로그의 포맷과 요소

로그 데이터란 웹 서버를 통해 이루어지는 모든 작업들에 대한 기록이라고 표현할 수 있다. 우리가 웹 서버에 접속을 하게 되면 그 이후의 모든 작업들은 웹 서버에 미리 정해 놓은 위치에 데이터로 남게 된다.

일반적으로 특정 웹 페이지를 보기 위한 사용자의 요구로, 웹서버는 해당 웹 페이지와 관련된 여러 파일 등을 접근하게 된다. 따라서 사용자가 요청하는 특정 웹 페이지뿐만 아니라 해당 웹 페이지와 관련된 이미지 파일, 이미지 데이터, Include 파일 등에 대한 정보가 로그 파일에 저장되는 것이다.

로그 파일에는 웹 서버가 수행한 작업들에 대한 풍부한 정보가 들어 있다. 이들은 단순히 특정 작업 요청과 성공 여부에 대한 것뿐만 아니라 실패했을 경우 그 해결책에 대한 정보도 들어있다.

2.1.1 로그의 포맷

모든 웹 서버는 클라이언트의 상호작용을 하나 이상의 로그 파일이나 데이터베이스로 로깅하거나 로그 정보를 다른 애플리케이션으로 실시간 파이핑할 수 있는 능력이 있다[3]. 이들 데이터 요소들은 웹 서버의 CGI(Common Gateway Interface)를 사용하여 실시간 애플리케이션으로 전송할 수 있다.

초기에 웹 서버에 저장하는 로그 데이터 항목들은 웹 서버마다 달라서 로그 분석에 어려움이 많았다. 항목들이 다르기 때문에 한 로그 분석 프로그램이 모든 로그 파일을 분석할 수 없었다. 그래서 등장한 것이 웹 서버 로그의 표준인 CLOG라고도 하는 CLF(Common Log Format)이다. 이 경우 공통적인 형식으로 로그 데이터를 저장하기 때문에 웹 서버가 다르다 하더라도 그 데이터 형식은 비슷하게 된다. 그리고 그 항목들은 대부분의 상황에 필요한 항목들을 제공한다. 하지만 CLF 이외에도 해당 웹 서버가 고유하게 제공하는 형식도 있다[4]. CLF 이외에도 몇 가지 요소가 더 추가된 ECLF(Extended Common Log Format Standard)가 있다.

2.1.2 로그의 요소들

CLF에는 보통 7가지 요소가 포함되어 있다[5]. ECLF에는 2가지 요소가 더 포함된다. 표1은 CLF와 ECLF의 요소들을 정리해 놓은 것이다.

표1 CLF와 ECLF의 요소들

데이터 요소	CLF	ECLF
Host	✓	✓
Ident	✓	✓
Authuser	✓	✓
Time	✓	✓
Request	✓	✓
Status	✓	✓
Bytes	✓	✓
Referrer		✓
User-agent		✓

2.2 기존 웹 서버 로그 파일

기존의 서버들 중 가장 많이 사용되고 있는 아파치와 IIS 서버들의 로그 정보를 살펴보면 사용자 행위 정보의 수집이 힘들다. 문제점은 다음과 같다.

- ▶ 로그를 텍스트 파일로 관리하여 검색이 힘듦
- ▶ 로그를 데이터베이스로 저장시 속도가 문제가 됨
- ▶ 로그에는 출입정보가 주로 기록됨

다음에서 각각의 서버에 대한 로그 파일의 특징을 설명하고 그 문제점들에 대해 자세히 지적해 보고자 한다.

2.2.1 아파치(Apache) 서버의 로그 파일

아파치는 로그를 기본적으로 CLF 형식으로 파일에 저장한다. 원하는 경우에 따라 LogFormat directive를 이용하여 다른 형식으로 저장할 수 있고 각 요소들은 스페이스로 구분되어 표 2와 같은 형식으로 기록된다.

표 2 아파치의 CLF 형식

host	ident	authuser	date	request	status	byte
------	-------	----------	------	---------	--------	------

그리고 어디서 무엇이 잘못되었는지를 알 수 있는 에러에 대한 정보를 Error 로그 파일에 기록한다. 또 방문자가 어디를 통해서 사이트에 방문했는지에 대한 정보를 가지는 Referer 로그를 파일에 기록한다. 이 Referer 로그는 내·외부적 네비게이션을 파악하는 자료로 사용할 수 있다. 이와 같이 아파치는 로그 자료들이 사용자의 이동이나 방문 등을 기록하고 있으며 로그 데이터를 텍스트 파일로 저장하여 검색이 힘든 상태이다. 로그 데이터가 방대해지는 경우 이를 감당할 수 없다.

2.2.2 IIS 서버의 로그 파일

IIS는 기본적으로 마이크로소프트 IIS 로그 파일 형식, NCSA 공용 로그 파일 형식, W3C 확장 로그 파일 형식, ODBC 로깅 등의 4가지 형식의 로그 형식을 지원한다.

로그 파일이 생성되는 시점은 매시간, 매일, 매주, 매월 등으로 필요에 따라 설정할 수 있다. 이 설정의 기준은 로그 데이터의 양으로 로그 데이터가 많을 경우는 로그 파일이 너무 커지게 되므로 관리하기가 불편해질 수 있다. ODBC 로깅을 제외하고 모두 ASCII 형태로 저장되어 다른 파일 형식으로 쉽게 변환해서 사용할 수 있다.

W3C는 IIS 로그 파일 형식 중 사용정도에 따라 가장 효율적으로 사용할 수 있는 것이다. 이 것은 사이트 성격에 따라 로그에 저장할 항목들을 선택할 수 있다는 것이 장점이다. 선택할 수 있는 항목들은 날짜, 시간, 클라이언트 IP 주소, 서비스 이름, 서버 이름, 서버 IP, 서버 포트, 메서드, URI 스템, URI 쿼리, HTTP 상태, Win32 상태, 보낸 바이트, 받은 바이트, 걸린 시간, 프로토콜 버전, 사용자 애이전트, 쿠키, 참조페이지 등이 있다.

IIS 역시 대부분의 로그들이 텍스트 파일로 저장되며 데이터베이스에 저장하는 ODBC 로깅은 웹 서버에서 이루어지는 모든 요청 때마다 데이터베이스에 저장하기 때문에 파일에 저장하는 것 보다 서버에 많은 부담을 주게 된다.

3. 전자상거래 사용자 정보 수집 시스템

2장에서 말한 기존 웹 서버들의 문제점을 해결하기 위해서 본 논문에서는 로그 데이터를 따로 처리하는 지능형 웹로그 서버를 두어 웹 서버에 부하를 줄이고 효율적으로 사용자 행위 정보를 수집하였다. 구현된 시스템은 크게 쇼핑몰 서버와 지능형 웹로그 서버로 나눌 수 있다. 그림 1은 전자상거래 사용자 정보 수집 시스템의

개요도이다.

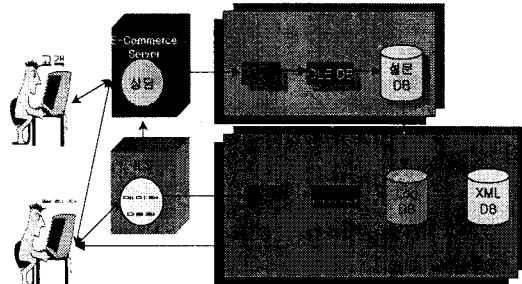


그림 1. 전자상거래 사용자 패턴 수집 시스템

3.1 전자상거래 쇼핑몰 서버의 구현

전자상거래 쇼핑몰 서버는 사용자가 쇼핑몰 상에서 구매나 사이트 이동 등의 행위를 할 때마다 그 정보를 쿠키에 추가하고 사용자가 쇼핑을 마치고 쇼핑몰을 빠져나갈 때 그 정보를 지능형 로그 서버인 ILS에 전송하여 주는 역할을 한다. 쇼핑몰의 환경은 다음과 같다.

- ▶ 운영체제 : Windows 2000 A/S
- ▶ 사용 언어 : ASP, VBScript
- ▶ 사용 서버 : MS IIS 5.0

전자상거래 쇼핑몰 서버의 처리 작업의 순서는 다음과 같다.

- step1 : 소비자의 사이트 방문, 이동 및 구매
- step2 : 소비자의 행위 정보를 쿠키에 저장
- step3 : 소비자가 사이트를 나감
- step4 : 쿠키에 저장된 정보를 ILS에 전송

다음 그림 2는 쇼핑몰에서 저장된 쿠키를 보여주는 화면이다.

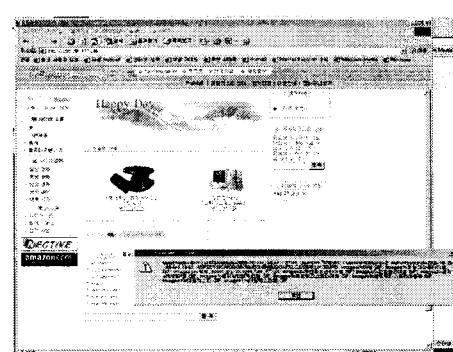


그림 2. 저장 쿠키 내용에 대한 메시지

지능형 웹로그 서버가 저장하는 데이터베이스에는 서버의 오투나 사용자의 실수로 분석 시 이용이 불가능한 데이터가 존재할 수 있다. 이러한 경우를 대비해서 관리자가 네트워크 환경에서 웹브라우저만 있으면 지능형 웹로그 서버에서 저장한 데이터베이스의 내용을 제어할 수 있는 기능을 추가하였다. 구현된 제어 기능으로 추가, 삭제, 개선, 정렬 등이 있다.

그리고 간략한 데이터의 검색 기능을 추가하여 원하는 데이터를 데이터베이스에서 찾지 않고 웹에서 바로 찾을 수 있도록 하였다.

다음 그림 3은 데이터베이스에 있는 로그 데이터의 내용을 웹 브라우저를 통해서 보는 화면이다.

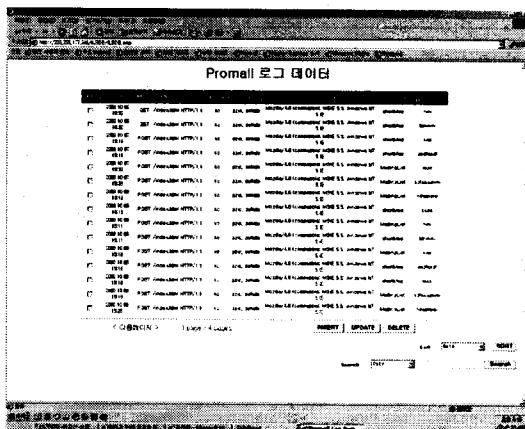


그림 3. 쇼핑몰의 로그 데이터 보기 화면

또 ASP 코드를 사용하여 데이터베이스를 XML 데이터베이스화 할 수 있다. 이는 사용자 패턴 수집 시스템에서 생성되는 mdb를 XML 기반의 환경에서 유용하게 사용할 수 있도록 XML 데이터베이스화 한다. 이 과정은 필요한 부분만을 다시 다른 데이터베이스로 재 저장하여 XML 데이터베이스화 할 수 있고 기존의 mdb 자체를 XML 데이터베이스화 할 수도 있다. 다음 그림 4는 mdb 로그 데이터가 XML 화된 화면이다.

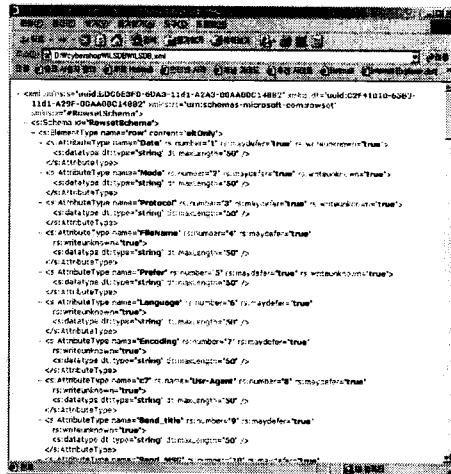


그림 4. mdb 로그 데이터의 XML화

3.2 지능형 웹로그 서버(II S)의 구현

지능형 웹로그 서버는 사용자의 데이터를 전송 받아서 이를 판별하여 각 요소별로 데이터베이스에 저장한다.
지능형 웹로그 서비스의 환경은 다음과 같다.

- ▶ 운영체제 : Windows 2000 A/S
 - ▶ 사용언어 : MS Visual C++ 6.0
 - ▶ 사용데이터베이스 : MS ACCESS의 MDB

다.
step1 : 쇼핑몰 서버로부터 로그 정보를 받음
step2 : 받은 로그 정보를 분류하여 DB에 저장
여기서 데이터베이스에 사용자의 행위 정보를 저장 시
데이터베이스 액세스 방법으로 OLEDB Provider상의

ADO기술을 사용하였다. 가장 많이 사용되는 ODBC는 관계형 데이터베이스만 액세스 가능한 반면 OLEDB Provider는 관계형이나 비관계형의 데이터베이스 액세스가 가능하다. 그리고 다른 데이터베이스 액세스 방법보다 ADO는 속도 면이나 기능 면에서도 상당히 빠르다. 다음 그림 5는 수집된 로그 데이터의 일부분이다.

그림 5. ADO 기술로 저장된 데이터베이스의 일부

4. 결 론

본 논문에서는 기존의 웹 서버와 달리 로그에 사용자 행위 정보를 추가하고 이를 데이터베이스에 저장하며 이를 웹 상에서 원격 제어할 수 있는 사용자 정보 수집 시스템을 구현하였다. 이 시스템에서는 많은 시간이 소요되는 데이터베이스 접근 시간을 ADO를 사용하여 단축하였고 [6], 웹을 통한 데이터베이스 제어로 관리자가 브라우저만 있으면 데이터베이스를 관리할 수 있는 환경을 구축하였다. 그리고 XML 기반의 환경에서 사용할 수 있도록 필요한 요소들을 부분적으로 XML DB화하였다.

앞으로는 본 시스템의 데이터베이스에서 사용자 프로파일 생성에 필요한 정보를 추출하는 기준을 정하고 이에 따라 데이터를 추출하는 알고리즘을 연구할 수 있다. 그리고 이를 이용하여 전자상거래 쇼핑몰에서 개인화된 서비스를 할 수 있을 것이다. 뿐만 아니라 분산된 쇼핑몰 서버에서의 정보 처리를 연구할 수도 있다.

[참 고 문 헌]

- [1] 안중호, 박철우, "인터넷과 전자상거래", 홍문사, 2000
 - [2] 이황규, 박규현, 이운령, "전자상거래의 개념에서 쇼핑몰 구축까지", 2000
 - [3] 김기수, 이상우, "인터넷 전자상거래 쇼핑몰 구축에서 운영까지", 비앤씨, 2000
 - [4] Osmar R. Zaiane, Man Xin, Jiawei Han, "Discovering Web Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs", 1998
 - [5] Ralph Kimball and Richard Merz, "The Data Webhouse Toolkit", 컴앤팩스, 2000
 - [6] Patrick Killelea, Web Performance Tuning, 권원상 역, O'REILLY, 2000
 - [7] <HTTP://www.i-biznet.com/logg/default.asp>
 - [8] <HTTP://www.cnk.or.kr/front/>
 - [9] <HTTP://www.microsoft.com/data/tch-vc21.htm>