# 웹 서버의 로그파일 분석이 첫걸음

인터넷 라이프가 시작될 21세기를 앞두고 웹사이트는 각 기업, 개인들에게 천숙한 인터넷 가상공간이 되고 있다. 홈페이지를 통해 기업들은 홍보를 하고, 제품판매를 하며, 개인들인 친구와 모이고 관심거리를 이야기 하고 있다. 명확한 목적을 가지고 홈페이지를 구축했을 경우, 지금 현재의 웹사이트가 얼마만큼 왔으며 어디로 가야하는 가를 알고 싶을 것이다. 이러한 웹사이트에 대한 통계분석방법은 웹사이트에 대한 객관적인 분석을 통해 웹사이트를 평가하고, 문제점을 찾아주며, 나아가야 할 곳을 일러준다.

■ 황태욱
한국전산원(twhwang@nca.or.kr)

## 웹 로그 파일 분석

포탈 사이트라는 용어가 어느덧 신문을 장식하면서 "야후 1,000만 페이지 뷰 달성" 등과 같은 기사가 신문을 장식하고 있다. 이러한 통계 결과는 단순히 사용자 숫자, 접속 수 만을 의미하지는 않는다. 인터넷 비즈니스의 모델은 대부분 이러한 통계수치를 근거로 한 인터넷 광고를 중심으로 이루어지기에 웹사이트의 이용현황을 분석하는 통계방법은 상당히 중요한 의미를 가진다. 또한 직접적으로 광고수익 등의 목적이 아닌 홍보성 기업 웹사이트의 경우에도 통계결과를 통해 투자대비 효과 등의 객관적인 생산성을 측정할 수 있다.

통계학의 유용성에 힘입어 웹사이트 통계분석방법은 다양한 접근방법이 있다. 웹사이트에서 사용자 만족도를 이벤트를 통해 조사할 수도 있으며, 온라인이 아닌 다른 매체를 통해서도 리서치 작업을 할 수 있다. 하지만 가장 기본적이면서도 객관적인 통계분석방법은 웹 서버의 로그파일 분석방법이다.

로그파일 분석방법은 웹 서버에 남겨진 사용자 접속 기록을 분석하는 방법이다. 웹 서버의 로그파일에는 사용자가 언제 어느 곳에서 어떤 작업을 수행하였는지가 자세히 기록된다.

물론 웹 프로토콜인 HTTP의 방식으로 인해(Connect less 방식) 지속적인 사용자 추적이 어려운 것은 사실이지만 그래도 상당히 많은 정보를 웹 서버의 로그파일을 통해 얻을 수 있다.

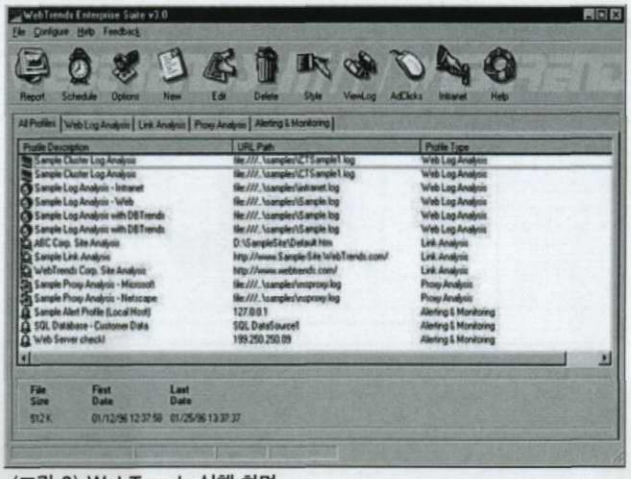
웹 서버의 로그파일 형태는 각각의 웹 서버마다 다르고 구성역시 조금씩 차이가 있을 수 있다. 하지만 일반적인 웹 서버의 로그파일 형태는 일반적으로 텍스트 파일 형태며 다음에 그 예가 있다.

```
210.105.55.66 - - [16/May/1999:22:23:30 +0900] "GET /content/y2000/faq5-3.gif HTTP/1.1" 200 14683
test-scooter.av.pa-x.dec.com - - [16/May/1999:22:44:05 +0900] "GET /law/law_4.html HTTP/1.0" 200 26346
203.227.34.23 - - [16/May/1999:22:47:53 +0900] "GET /content/y2000/snu/tsld026.htm HTTP/1.1" 200 1001
```

간단한 구조를 살펴보면 먼저 사용자의 IP 주소 또는 도메인 주소가 나오며, 다음으로는 접속시간, 수행한 행동이 기록된다. 이러한 로그파일 기록은 사용자가 웹 브라우져를 작동시키면 HTTP 프로토콜을 통해 웹 서버에 요청되고 웹 서버는 이러한 요청에 응답함과 동시에 로그파일에 이러한 내용을 기록하게 된다.

그런데 실제로 웹 브라우저가 웹 서버


〈그림 1〉 WebTrends의 인터넷 사이트




〈그림 2〉 WebTrends 실행 화면

를 통해 웹사이트를 살펴보는 경우 다양한 경로를 통하게 된다. 경우에 따라 방화벽을 지나가기도 하며, 프락시 서버를 통해 접근하기도 한다. 또한 캐쉬 서버나 자신의 PC의 로컬 캐쉬에 자료가 있을 경우에는 웹 서버에 요구하지 않기도 한다. 이는 웹 브라우져와 웹 서버 사이의 동작 메커니즘의 다양성에 의하여 사용자의 모든 행위(웹사이트를 대상으로 한)가 기록되지는 않음을 의미한다. 물론 이러한 한계에도 불구하고 웹 로그파일은 상당히 정확한 기록을 제공하고 있다.

웹 로그파일은 상당히 많은 자료를 보관하는 커다란 파일이다. 실제로 하루에 1,000명 정도 접근하는 홈페이지의 경우 히트 수로 계산하면 거의 30,000번 정도의 기록이 남는다. 즉 한달 정도면 약 900,000줄 정도의 기록이 남는다고 예상할 수 있다. 웹 서버 관리자에게는 로그파일의 적절한 백업과 분배역시 중요한 작업일 것이다. 이런 대용량의 로그파일을 보다 정확하고 세밀하게 분석하기 위해 여러 웹 로그파일 분석 도구들이 있다.

## 웹트랜드, 로그파일 분석 도구

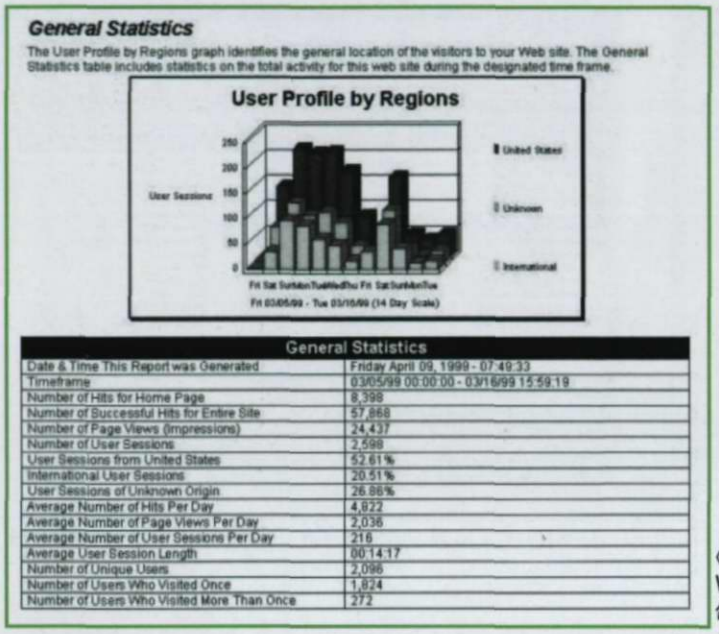
최근에는 웹 서버 자체에 웹 로그파일 분석 도구가 결합되어 있기도 하지만 아직은 전문적인 로그 분석을 위해서는 별도의 로그파일 분석 제품을 구매하는 것이 타당하다. 이러한 로그파일 분석 도구 중 가장 널리 사용되는 WebTrends(웹트랜드)의 기능과 사용법을 간단히 설명하고자 한다. WebTrends는 미국의 여러 인터넷 잡지 등에서 가장 뛰어난 웹 로그 분석 도구로 평가받고 있으며 국내에서도 가장 많이 사용되고 있는 로그 분석 도구이다. 보다 자세한 정보는 인터넷 사이트에서 얻기 바란다( http://www.webtrends.com ).

웹트랜드를 통해 로그 분석과 함께 웹 서버의 HTML 오류 검사 등의 기능도 있지만 로그 분석 기능을 중심으로 살펴보자. 웹트랜드에는 로그파일을 필터링을 통해 시간별로, 특정 도메인별로 구분해서 분석할 수 있다. 또한 이러한 분석을 미리 일정을 잡아서 분석할 수도 있다. 웹트랜드에서 뛰어난 기능은 결과에 대한 커스터마이징 기능이다. 각 표와 항목의 개수를 지정할 수 있고 여러 출력 파일 형태를 지원한다.

이미 한글화가 되어 출력 결과 등이 한글로 출력 가능하다. 단지 아래한글을 지원하지 않고 있으나 곧 이를 지원할 예정이라고 한다. 웹트랜드를 사용할 때 분석할 로그파일이 커지면 이를 분석하는데 걸리는 시간도 만만치 않은데 관리하고 있는 사이트의 약 3G byte 정도의 자료를 분석하는데 266MHz 팬티엄 II PC에서 5~7시간 정도 소요되었다.

웹트랜드에서 로그파일을 통해 분석하는 항목은 접속 통계 요약, 최고 빈도 요구 페이지, 최저 빈도 요구 페이지, 진입 페이지, 최종 페이지, 광고 클릭률, 광고 노출도, 국가별 접속수, 도메인별 접속수, 일간 주간 접속 수, 시간별 접속 평균 등 많은 항목에 대한 통계 항목을 분석한다. 이외에도 서버 오류 수치, 사용자 브라우져 종류 등에 대한 정보도 제공한다. 이러한 세부 항목은 로그분석 프로그램마다 차이가 있으며 경우에 따라서는 여러 분석 도구의 분석 결과를 같이 사용할 필요도 있다. 접속 통계 요약에 대한 리포트 예제를 살펴보자. 이 예제에서 사용자에게 보여주는 것은 웹 서버 로그 분석 결과의 생성시간, 분석한 로그의 기간(Timeframe), 첫페이지에 접속한 횟수, 전체 사이트에 접속한 횟수(Hit수), HTML 접속 수(Page View), 사용자 접속 수(user session)에 대한 통계 수치를 제공한다.

히트수와 페이지 뷰, 사용자 접속 수에



**General Statistics**

The User Profile by Regions graph identifies the general location of the visitors to your Web site. The General Statistics table includes statistics on the total activity for this web site during the designated time frame.



| General Statistics | |
|---|---|
| Date & Time This Report was Generated | Friday April 09, 1999 - 07:49:33 |
| Timeframe | 03/05/99 00:00:00 - 03/16/99 15:59:19 |
| Number of Hits for Home Page | 8,398 |
| Number of Successful Hits for Entire Site | 57,868 |
| Number of Page Views (Impressions) | 24,437 |
| Number of User Sessions | 2,599 |
| User Sessions from United States | 52.61% |
| International User Sessions | 20.51% |
| User Sessions of Unknown Origin | 26.86% |
| Average Number of Hits Per Day | 4,822 |
| Average Number of Page Views Per Day | 2,036 |
| Average Number of User Sessions Per Day | 216 |
| Average User Session Length | 00:14:17 |
| Number of Unique Users | 2,096 |
| Number of Users Who Visited Once | 1,824 |
| Number of Users Who Visited More Than Once | 272 |

〈그림 3〉 WebTrends 실행 결과치

대해 좀더 설명하면 히트수는 웹 서버를 통해 요청된 모든 객체들의 총 요구량이다. 예를 들어 첫 홈페이지가 3개의 그림 파일이 포함된 HTML로 구성되어 있다면 사용자가 첫 페이지에 접근하는 것만으로 4 hit 수치를 기록한다. 페이지 뷰는 순수하게 HTML 파일에 대한 접속 숫자만을 의미하게 되고, 사용자 접속수치는 웹 서버 로그파일에서 하나의 세션으로 취급하는 접속에 대한 수치 통계이다. 보통 동일한 IP 주소에 대해 30분 간격 이내의 요구 사항들은 하나의 세션으로 취급하여 측정된다. 물론 이러한 세션이 정확한 사용자 수치를 반영하지는 않지만 그래도 어느 정도 객관성이 있는 하나의 참고 수치이기는 하다.

그 밖에도 일일 평균 히트수, 일일 평균 페이지 뷰, 일일평균 사용자 수치와 평균적인 사용자들의 이용시간(User Seesion 시간의 평균 수치)에 대한 통계를 제공한다. 개별적 유저(Unique User)에 대한 통계는 개별적인 IP에 대한 접근 통계를 의미하며 단 한번 사용한 유저의 수치와 여러 번 홈페이지에 접근한 사용자 수치를 보여준다. 평균적인 사용자 이용시간 수치는 일반적으로 사용자들의 홈페이지 충성도와 연관이 높은 수치이며, 한번이상 접근한 사용자 수치는 단골 고객에 대한 수치와 관련이 높다.

만약 필터링을 통해 *.co.kr에서 접근한 사용자에 대해서만 분석한다면 앞의 수치는 국내 사용자의 통계 수치를 보여주게 될 것이다. 이러한 로그분석 결과는 객관적이고 정량적인 웹 서버 통계분석이다. 이러한 통계분석 수치를 바탕으로 정성적인 평가와 분석 예측 등을 통해 더욱 의미 있는 자료를 얻을 수 있다.

예를 들어 주말 이용자가 많고 근무시간 이용자가 적다면 이용자는 사무실이 아닌 가정에서 접근하는 경우가 많다고 가정할 수 있다. 도메인별 접속자 수치에서 *.ac.kr 수치가 많다면 대학생이나 대학 관계자들의 접근이 많고 이는 웹사이트의 콘텐트가 그들에게 유용하다고 할 수 있다. 이와 같이 웹 서버 로그 분석결과에 어떠한 해석과 분석을 통해 더욱 전략적인 자료를 얻을 것인가 하는 것이 웹 서버 로그파일의 결과를 가지고 작업해야 할 일이다.

이러한 분석을 위해서는 몇 가지 기초 자료와 함께 분석하는 것이 좋다. 웹사이트를 구축할 때 설정했던 목표, 주 사용자 고객, 주 사용자 고객에 대한 성향 등을 기록한 내용으로 로그파일 결과를 예측하고 오차가 있을 경우 어떠한 상황인가를 분석하여 사이트의 방향성을 잡을 수 있다.

## 사용자 DB 분석 필요

그럼 이러한 로그분석은 언제 얼마만큼 하는 것이 좋은가? 이는 웹사이트의 성격에 따라 다르다고 할 수 있다. 보통의 경우 월간, 분기별로 하는 정기적인 분석과 웹사이트 개편, 디자인 변경 등과 같은 이벤트에 맞추어 분석하는 비정기적인 로그분석을 겸해서 수행하여야 한다. 통상적으로 한달에서 분기별 1회가 바람직하나 특별히 접속수가 높거나 중요도가 큰 사이트의 경우 보다 빨리 분석하여 정량적 데이터로 축적하고 이를 모아서 다시 정성적인 분석을 하는 것도 한 방법이다.

로그파일 분석방법은 통계기법을 통한 웹사이트의 분석방법으로 가장 보편적이면서, 가장 중요한 분석방법으로 사용될 것이다. 향후 좀더 자세한 사용자 분석을 위해서는 이러한 로그파일 분석기법과 함께 사용자 DB를 별도로 구축하여 각 개인 사용자별 특색과 중요한 행위-쇼핑몰 사이트에서의 구매 행위-를 통해 세밀한 분석을 제공하는 형태로 발전할 것이다. 이것은 인터넷 비즈니스에서 추구하는 1:1 마케팅을 위해서는 필수적인 형태의 분석 기법이 될 것이다.

지금 웹사이트를 평가하고, 새롭게 방향을 잡고 싶은가? 그렇다면 맨 먼저 해야 할 일은 웹 서버의 로그파일을 분석하는 일이다.