



로그

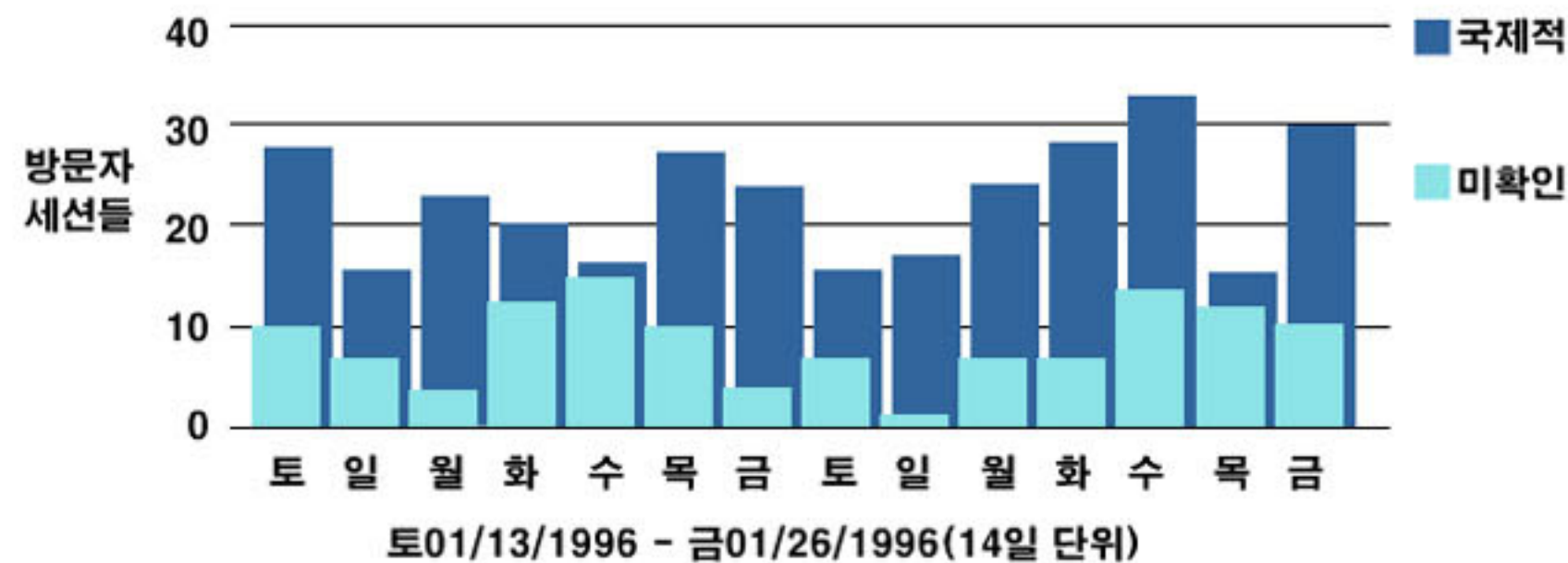
## 분석의 이해와 방법(1)

|          |                       |   |
|----------|-----------------------|---|
| <b>1</b> | 로그(Log) 이해            | 1 |
| <b>2</b> | 로그 분석이란?              | 1 |
| <b>3</b> | 통계 분석을 위한 데이터         | 2 |
| <b>4</b> | 용어 설명                 | 2 |
| <b>5</b> | 통계 분석 대상              | 2 |
| <b>6</b> | 웹 로그 파일(Web Log File) | 3 |
| <b>7</b> | 로그 분석 도구의 활용          | 6 |
| <b>8</b> | 요점 정리                 | 6 |



## 1. 로그(Log) 이해

로그(Log)는 웹 사이트에서 사용자들의 웹 사이트 이용에 대한 정보를 일컫는다.  
로그 정보를 통하여 트래픽 발생량, 페이지별 트래픽 발생량, 사용자의 위치, 평균적으로 머무른 시간 등 다양한 정보를 산출해 낼 수 있다.



## 2. 로그 분석이란?

### 1 일반적 의미의 로그 분석

일반적 의미의 로그 분석은 로그 데이터를 이용하여 트래픽을 파악하고, 이 트래픽이 지닌 의미를 분석해 나가는 것이라고 할 수 있다. 로그 데이터를 이용하여, 웹사이트의 페이지 뷰, 사용자별 페이지 뷰, 접속 장소 및 방식, 시간별 페이지 뷰, 방문자 수 등에 대한 현황 및 추세를 분석하는 것이다.

웹 사이트의 클릭 흐름을 분석하는 것 역시 이 범주에 들어간다. 사용자가 웹 사이트를 방문하는 경로와 서핑하는 경로에 대한 분석을 통하여 웹 사이트가 지닌 문제점을 찾고, 사용자가 웹 사이트에서 무엇을 원하는지를 보다 구체적으로 파악하는 것이다.

### 2 확장된 의미의 로그 분석

이에 반해, 확장된 의미의 로그 분석은 단지 로그 데이터뿐 아니라, 웹 사이트에서 보유하고 있는 고객 등록 정보, 구매 정보, 외부 환경 정보 등을 복합적으로 사용하는 분석을 말한다.

이러한 분석을 통하여 사용자 특성별로 웹 사이트의 이용, 구매에 대한 보다 폭넓은 분석이 가능해진다.



### 3. 통계 분석을 위한 데이터

#### 1 웹 로그 파일

- AccessLog(Transfer Log), RefererLog, ErrorLog, AgentLog

#### 2 쿠키 정보

- 쿠키는 브라우저가 사용자의 하드디스크에 text 파일 형태로 저장할 수 있는 정보  
- Cookie Central (www.cookiecentral.com)

#### 3 등록 정보

### 4. 용어 설명

- 히트(hit)=인터넷 사용자가 특정 사이트에 접속할 때마다 그 사이트를 관리하는 전산 시스템(웹 서버)에 저장돼 있는 구성 파일을 호출하는 횟수이다.
- 방문자 수(visitor)=특정 기간 동안 특정 사이트에 접속한 인터넷 사용자이다.
- 접속횟 수(login)=동일인 여부를 고려하지 않고 사용자가 사이트에 접속 한 숫자를 단순 합산한 숫자이다.
- 페이지 뷰(page view)=사용자가 특정 사이트에 들어가 본 홈페이지 수. 임프레션(impression)이라고도 한다. 히트 수나 방문자 수 등이 사이트의 영향력을 평가하는 일반적인 기준이라면 페이지 뷰는 사이트의 광고 단가를 결정하는 기준이다.

### 5. 통계 분석 대상

우리가 웹 서버에 접속을 하게되면 웹 서버의 로그 파일에 그 기록이 남게 된다. 이는 단순히 접속이라는 데이터로 기록되며, 한 페이지의 HTML 문서를 검색해도 접속은 해당 페이지에 포함된 다양한 형태의 데이터마다 기록을 남기게 된다. 이러한 하나의 데이터 이동을 접속(Hit)이라 한다면, 실제 사용자가 하나의 완성된 페이지를 보는 것을 검색(View)이라 한다. 그렇다면 우리는 특정의 사용자가 일정시간 내에 계속적으로 웹 서버를 검색하는 것을 방문(Visit)이라 할 수 있으며, 이러한 방문 데이터가 통계 분석의 대상이 되는 것이다.



## 6. 웹 로그 파일(Web Log File)

1) AccessLog(Transfer Log)

2) RefererLog

3) ErrorLog

4) AgentLog

### 1 AccessLog (TransferLog)

■ 개념: 클라이언트의 접속 내용을 기록

- Format : [host] [rfc931] [authuser] [date] [request] [status code] [bytes]

host : Client의 도메인이나 IP주소

rfc931 : information returned by identd for this person, otherwise "-"

authuser : if a userid sent for authentication, the user name, otherwise "-"

date : 시간과 날짜

request : 요청한 내용

status code : 상태 코드

bytes : 전송된 데이터의 크기

```
203.248.92.101 - - [26/Sep/1999:14:47:45 +0900] "GET /cwb-
data/icon/help.gif HTTP/1.1" 304 -
203.248.92.101 - - [26/Sep/1999:14:47:45 +0900] "GET /cwb-
data/icon/down.gif HTTP/1.1" 304 -
203.248.92.101 - - [26/Sep/1999:14:47:45 +0900] "GET /cwb-
data/icon/write.gif HTTP/1.1" 304 -
203.248.92.101 - - [26/Sep/1999:14:47:45 +0900] "GET /cwb-
data/icon/reload.gif HTTP/1.1" 304 -
203.248.92.101 - - [26/Sep/1999:14:47:57 +0900] "GET /cgi-
bin/board/CrazyWWWBoard.cgi?db=notice&mode=read&num=20&page=1&fty
pe=6&fval=&backdepth=1 HTTP/1.1" 200 6531
203.248.92.101 - - [26/Sep/1999:14:47:58 +0900] "GET /cwb-
data/icon/modify.gif HTTP/1.1" 200 386
203.248.92.101 - - [26/Sep/1999:14:47:58 +0900] "GET /cwb-
data/icon/delete.gif HTTP/1.1" 200 325
203.248.92.101 - - [26/Sep/1999:14:47:58 +0900] "GET /cwb-
data/icon/reply.gif HTTP/1.1" 200 347
203.248.92.101 - - [26/Sep/1999:14:47:58 +0900] "GET /cwb-data/icon/list.gif
HTTP/1.1" 304 -
203.248.92.101 - - [26/Sep/1999:14:49:54 +0900] "POST /cgi-
bin/board/CrazyWWWBoard.cgi HTTP/1.1" 200 213
```



```

211.37.134.119 ~ ~ [12/Dec/1999:04:37:02 +0900]
"GET /cgi-bin/way-board.cgi?db=flea03 HTTP/1.1" 200 11531
Host IP Status
211.37.134.119 ~ ~ [12/Dec/1999:04:37:06 +0900]
"GET /cgi-bin/way-board/img/bg.gif HTTP/1.1" 404 309
Request bytes
211.37.134.119 ~ ~ [12/Dec/1999:04:37:07 +0900]
"GET /img/icon-0.gif HTTP/1.1" 200 163
Date
211.37.134.119 ~ ~ [12/Dec/1999:04:37:07 +0900]
"GET /img/icon-1.gif HTTP/1.1" 200 163
203.239.41.196 ~ ~ [12/Dec/1999:04:37:46 +0900]
"GET /cgi-bin/way-board/way-board.cgi?db=data02 HTTP/1.1" 200 12289
  
```

```

#Software: Microsoft Internet Information Services 5.0
#Version: 1.0
#Date: 2000-03-14 00:00:01
#Fields: date time c-ip cs-uri-stem cs-uri-query cs(Cookie) cs(Referer)

2000-03-14 00:00:01 210.121.167.125 ~ /Directroy/viewpage.asp id=dir215131502
us Date 640 Time 2Ekr C~IP B4%AE% cs-uri-stem AE;+AS cs-uri-query
ht r.int

2000-03-14 00:00:01 210.121.167.125 ~ /Directory/viewpage.asp id=213172057
cs(Cookie)
userid=ha%40play%2Eco%2Ekr;+name=%B4%E8%C0%CE%C8%AF;+ASPSESSIONID.....
http://www.cnet.com

2000-03-14 00:00:01 16.194.143 ~ /Home/viewlist.asp ~
cs(Referer)
ASPSESSIONIDQGQGLIG=ODMININABMCODBOLBCNLGJKJ
http://www.imeca.com
  
```

## 2 RefererLog

- 개념: 클라이언트의 문서 요구에 대한 문서간의 연결 관계를 기록
- 사용자의 웹 서버 이용 패턴을 파악하는 정보 제공
- Format : [Referer] -> [Document]
  - Referer : 참조 위치
  - Document : 문서 위치

```

http://techpress.joins.com/bench/html/2000021/ ->
/benchmark/2000021/images/bench_chat_pic.gif
http://techpress.joins.com/bench/html/2000021/ ->
/benchmark/2000021/images/bench_chat_title.gif
http://www.webmania.co.kr/ -> /images/top_home_navi03.gif
http://www.webmania.co.kr/ -> /images/top_home_navi03a.gif
http://www.webmania.co.kr/ -> /images/top_home_navi04.gif
http://www.webmania.co.kr/ -> /images/top_home_navi04a.gif
http://techpress.joins.com/bench/html/2000021/chat_02.html ->
/benchmark/2000021/images/bench_chat_title.gif
http://www.webmania.co.kr/membership/ -> /cartoon/
http://news.myfolder.net/ -> /news/news_view.php
http://www.webmania.co.kr/ -> /images/home_webtool.gif
http://www.webmania.co.kr/ -> /images/home_mania.gif
http://www.webmania.co.kr/ -> /images/home_job.gif
http://www.webmania.co.kr/ -> /directory/images/icon_new.gif
http://www.webmania.co.kr/news/news_view.php?grp=all&number=363 ->
/include/style.css
http://www19.hanmail.net/Mail-
bin/view_mail?MSGID=I0000000000p3Dz&FOLDER=%B9%DE%
C0%BA%C6%ED%C1%F6%BA%B8%B0%FC%C7%D4 -> /news/news_view.php
http://www.webmania.co.kr/news/news_view.php?grp=all&number=362 ->
/include/style.css
http://www.webmania.co.kr/news/news_view.php?grp=all&number=359 ->
/images/top_navi03.gif
http://www.webmania.co.kr/news/news_view.php?grp=all&number=362 ->
/images/top_navi01.gif
  
```



### 3 ErrorLog

- 개념: 웹 서버의 에러 사항을 기록
- Format : [date] [level] [contents]
- 에러 메시지
  - 302 REDIRECT
  - 400 BAD\_REQUEST |
  - 401 AUTH\_REQUIRED
  - 403 FORBIDDEN
  - 404 NOT\_FOUND
  - 500 SERVER\_ERROR
  - 501 NOT\_IMPLEMENTED

```
[Fri Oct 1 17:22:45 1999] [error] [client 203.248.92.104] unable to include
"/include/channel_right_banner.inc" in parsed file
/usr/local/apache/htdocs/solution_channels/admin/domain/index.html
[Fri Oct 1 17:22:45 1999] [error] [client 203.248.92.104] File does not
exist:/usr/local/apache/htdocs/include/channel_left_navi.inc
[Fri Oct 1 17:22:45 1999] [error] [client 203.248.92.104] unable to include
"/include/channel_left_navi.inc" in parsed file
/usr/local/apache/htdocs/solution_channels/admin/domain/index.html
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] File does
notexist:/usr/local/apache/htdocs/include/channel_tit.inc
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] unable to include
"/include/channel_tit.inc" in parsed file
/usr/local/apache/htdocs/solution_channels/admin/network/index.html
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] File does not
exist:/usr/local/apache/htdocs/include/channel_right_banner.inc
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] unable to include
"/include/channel_right_banner.inc" in parsed file
/usr/local/apache/htdocs/solution_channels/admin/network/index.html
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] File does not
exist:/usr/local/apache/htdocs/include/channel_left_navi.inc
[Fri Oct 1 17:22:51 1999] [error] [client 203.248.92.104] unable to include
"/include/channel_left_navi.inc" in parsed file
/usr/local/apache/htdocs/solution_channels/admin/network/index
```

### 4 AgentLog

- 개념: 접속한 클라이언트의 에이전트(브라우저명)를 기록

Format : [User-Agent]

```
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt; dreamline)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt; dreamline)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt; dreamline)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt; dreamline)
Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt; dreamline)
```



## 7. 로그 분석 도구의 활용

- 접속 횟수에 따라 로그 내용이 누적되므로 거대한 로그 파일이 생성됨  
(주기적으로 로테이션 필요. 예: Logrotate )
- 로그 파일의 내용을 항목별로 분석하는 도구
- 일별, 주간별, 월별 로그 파일을 분석
- 분석된 로그 파일은 제거
- 내부 IP를 분리해서 관리한다.  
( 관리자, 운영자, 개발자로부터의 수치는 제외시킬 필요가 있을 때)

## 8. 요점 정리

- 웹 사이트를 분석하는 기초가 되는 웹 로그는 크게 AccessLog, RefererLog, ErrorLog, AgentLog로 나뉜다. 이러한 로그 정보로 어느정도 웹 사이트에 대한 통계는 낼 수 있으나 어디까지나 통계는 통계일 뿐 정확한 웹 사이트 분석을 하기 위해서는 다각도에서 연구해야 할 것이다.
- 추천 사이트  
아이비즈넷-웹로그 분석 가이드  
([www.i-biznet.com/logg/default.asp\\*http://www.webpro.co.kr](http://www.i-biznet.com/logg/default.asp*http://www.webpro.co.kr))  
웹매니아-솔루션 벤치마킹, 웹로그 분석  
([www.webmania.co.kr/reviews/solution\\_bench/2000061](http://www.webmania.co.kr/reviews/solution_bench/2000061))