

WHAT DO WE (NOT) KNOW ABOUT OUR USERS

Jadranka Stojanovski

Ruđer Bošković Institute Library

jadranka@nippur.irb.hr

Web space

- most people, companies and organizations put information on the web because they want it to be seen by the world
- their goal is to have visitors come to the site, feel comfortable and stay awhile
- in the case of online businesses, more web site traffic increases the potential for sales

Library web space

- in the beginning focus was on the **content**
- rich content diversity among library web sites
 - access to the collections
 - interlibrary loan services
 - virtual reference collections
 - “subject” services
 - information retrieval services
 - free email etc.
- last few years a lot of attention is given on site's **users**

IRB Library web site (1994-)
<http://library.irb.hr>

RUDJER BOSKOVIC INSTITUTE LIBRARY

Promjena jezika

Zagreb, Croatia.

- GENERAL INFORMATION
- VIRTUAL REFERENCE LIBRARY
- SCIENCE ON THE INTERNET
- EDUCATION, INSTRUCTION, SUPPORT



- SEARCH
- ONLINE CATALOGUES
- ELECTRONIC RESOURCES
- NEWS



Questions

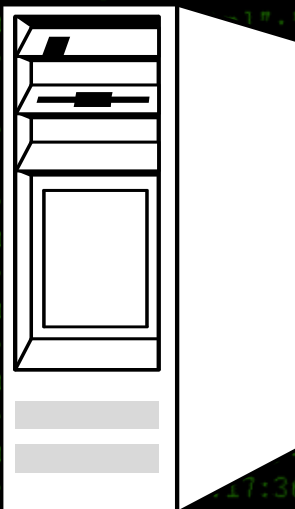
- who are users/visitors of my site?
- where are they come from?
- what are they doing there?
- how many of them is visiting my site?
- what are they clicking on?
- what are the most popular pages?
- for how long are they staying there?
- where and why do they leave my site?
- where are they going?

What's in the Log Files?



GET /library/catalogue.html

404 File Not Found



- view of client-server transactions – client request, with the server response
- multiple transactions can be required for a web page

Log files

- access.log, error.log, referer.log
- .log, .web, .clf, .txt extensions, usually in "logs" or "stats" directory
- text files in CLF (Common Log Format), ELF (Extended Log Format) or DLF (Combined Log Format)
- each line represents one request

Sample log file

```
fk-07-vusst.hr - - [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr - - [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1623
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr - - [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
194.186.163.241 - - [13/Feb/2002:15:00:01 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 0
"http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
odin.irb.hr - - [13/Feb/2002:15:00:39 +0100] "GET /hrv/index.html HTTP/1.1" 200 19411 "http://nippur.irb.hr/"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
odin.irb.hr - - [13/Feb/2002:15:00:40 +0100] "GET /images/ball_off.gif HTTP/1.1" 304 -
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
odin.irb.hr - - [13/Feb/2002:15:00:40 +0100] "GET /images/library-anim.gif HTTP/1.1" 304 -
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
odin.irb.hr - - [13/Feb/2002:15:00:40 +0100] "GET /images/ovid.gif HTTP/1.1" 304 -
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
odin.irb.hr - - [13/Feb/2002:15:00:40 +0100] "GET /images/biblio.gif HTTP/1.1" 304 -
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
odin.irb.hr - - [13/Feb/2002:15:00:40 +0100] "GET /images/priodoslovlje.gif HTTP/1.1" 304 -
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.0 [en]"
194.186.163.241 - - [13/Feb/2002:15:00:43 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 0
"http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
194.186.163.241 - - [13/Feb/2002:15:00:46 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 9853
"http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
194.186.163.241 - - [13/Feb/2002:15:00:54 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 32768
"http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
ivica.irb.hr - - [13/Feb/2002:15:00:55 +0100] "GET /hrv/wostrial HTTP/1.1" 404 288 "-" "Mozilla/4.0 (compatible; MSIE
6.0; Windows NT 5.0)"
194.186.163.241 - - [13/Feb/2002:15:00:56 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 32768 "-" "Mozilla/4.76
[en]" "http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
194.186.163.241 - - [13/Feb/2002:15:00:58 +0100] "GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 19885
"http://www.google.com/search?num=50&hl=en&q=R-matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0
(compatible; MSIE 5.00; Windows 98)"
ivica.irb.hr - - [13/Feb/2002:15:01:01 +0100] "GET /hrv/wostrial.html HTTP/1.1" 200 13430 "-" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.0)"
ivica.irb.hr - - [13/Feb/2002:15:01:01 +0100] "GET /images/logoback.jpg HTTP/1.1" 200 606
"http://nippur.irb.hr/hrv/wostrial.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)"
ivica.irb.hr - - [13/Feb/2002:15:01:01 +0100] "GET /images/home.gif HTTP/1.1" 200 134
"http://nippur.irb.hr/hrv/wostrial.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)"
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
```



```
194.186.163.241 - - [13/Feb/2002:15:00:46 +0100] "GET /izvjestaj/1993/1993.txt
HTTP/1.0" 206 9853 "http://www.google.com/search?num=50&hl=en&q=R-
matrix+synthesis+organic&btnG=Google+Search" "Mozilla/4.0 (compatible; MSIE
5.00; Windows 98)"
```

- user's IP address (can be resolved to DNS)
- (user's login name)
- date and time of the hit (plus user's time zone code)
- method used in request (GET, POST, ...)
- file requested by client including path
- web server's response code (200 OK, 404 – NOT FOUND)
- bytes sent back to the client in the response
- referrer path (referring URL - where did the user find the link to this request)
- user's agent
- cookies sent by the user
- cookies sent by the web server
- bytes received by the server in the request
- time taken by the server to process the request

Terminology used (1)

- **hit** – each HTTP request made to server
 - 6 images on a web page would equal 7 hits every time the page is viewed
- **page view (impression)**
 - number of pages viewed, i.e., the number of files retrieved with extensions such as .htm, .html, or .asp.
 - does not include the graphic and supporting files
 - if a site has 1 web page with 4 graphics on it, it would be reported as 5 hits, but only 1 page view or impression
- **visit / user session**
 - number of unique users who visited a web site during a certain time period

Terminology used (2)

• total hits

- the total number of files that are requested from the server, including files for graphics, audio/video files, and other supporting files, in addition to the actual html page itself
- include all requests for files, regardless of whether or not they were successfully retrieved

• visitors

- number of unique IP address (or cookies) that visited the Web site during a specified time period
- when there is no activity for a specified period of time the visit ends
- different analysis programs have discrepancies as to how long the period of inactivity is, and set their default settings differently, which produces inconsistent results between the programs

Logging / Tracking / Analysis

1. remote tracking services – house all the traffic recording scripts and reports on their own servers;
2. logging programs – scripts that you install on the server, which then generate both log files and reports
3. log analysis programs – commercial and freeware programs that analyze server logs

```
fk-07-vusst.hr-- [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr-- [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/novi-korisnici-on.gif HTTP/1.1" 200 1623
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr-- [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr-- [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/info-on.gif HTTP/1.1" 200 1029
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr-- [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/o-centru-on.gif HTTP/1.1" 200 1492
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:30+0100] "GET / HTTP/1.1" 200 892
"http://www.hr/wwwhr/education/libraries/index.hr.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:33+0100] "GET /images/ball_on.gif HTTP/1.1" 200 104
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:33+0100] "GET /hrv/index.html HTTP/1.1" 200 19411 "-" "Mozilla/4.0
(compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:34+0100] "GET /images/ball_off.gif HTTP/1.1" 200 155
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:35+0100] "GET /images/ovid.gif HTTP/1.1" 200 610
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
gita.srce.hr-- [13/Feb/2002:17:36:40+0100] "GET /hrv/searchglobal.html HTTP/1.0" 200 32768
"http://knjiznica.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:38+0100] "GET /images/priodoslovlje.gif HTTP/1.1" 200 892
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr-- [13/Feb/2002:17:36:40+0100] "GET /images/biblio.gif HTTP/1.1" 200 957
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:40+0100] "GET /ovid/ HTTP/1.0" 304 -- "[unknown origin]" "Mozilla/4.76 [en]
(Win98; U)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:40+0100] "GET /ovid/style/main-ie.css HTTP/1.0" 200 469 "-" "Mozilla/4.76
[en] (Win98; U)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:42+0100] "GET /ovid/periodika/podloga.gif HTTP/1.0" 200 1049
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/novi-korisnici-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/hrv/index.htm" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/back-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr-- [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
```



Log files can tell us

- the actual load on a server
- what pages are requested the most
- the IP address of the computer used for the request
- the status code of the request - success or failure
- the URL of the page the visitor was on when they requested your page
- which words and phrases are using to find us?
- which countries visitors come from
- which broken links visitors tried to use
- which search engines/other websites are providing us with users?
- how long is each page viewed?
- top entry pages
- top exit pages
- least popular pages
- single access pages
- the most popular navigation routes through the site

Log files cannot tell us

- e-mail addresses of visitors
- names of visitors or their employers
- connection speeds of the visitor
- exact numbers of visitors to a site

Groups of statistical information

- traffic (activities)
- audience (visitors, users)
- resources accessed (top pages, top documents...)
- browsers/platforms
- errors (reliability of the web site)
- referrers and keywords

Most requested pages

- the pages that receive the most requests (hits or page views) are usually the home pages, but can also indicate the popularity of certain pages and tell you what people come to get from your site

Least requested pages

- less popular pages or lack of access to your site through links, but it can also indicate the need for determining what it is about the page that is unappealing

Top entry pages

- indicates that the page is linked from a search engine or another site.

Top exit pages

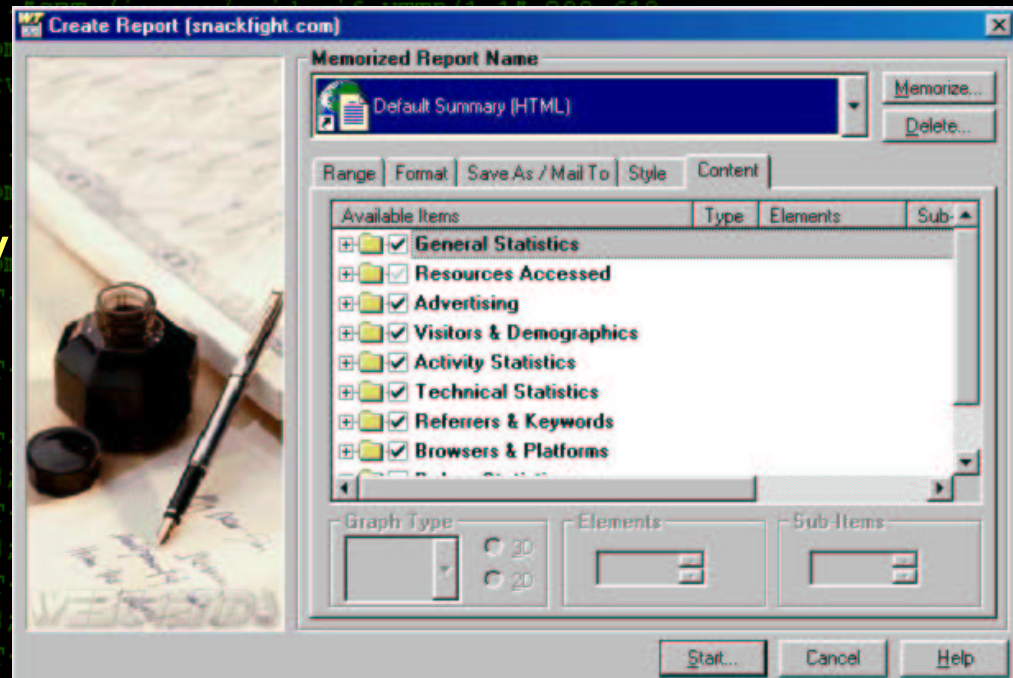
- indicates that the user is passing through quickly without using other features or links on the page

Single access pages

- the user accessed this web page and exited the site without viewing the other pages

WebTrends

- ease-of-use and ability to generate many different kinds of reports
- WebTrends software can interface with all of the big Microsoft Office products, which means the reports could be dumped into MS Excel or MS Word
- does attempt to identify the number of users to a site by using mathematical algorithms



Sowmill

- interface is entirely browser-based
- quick start option for the anxious webmaster

This view shows only the traffic described by the filters. The current filters are described in the filters bar above. You can change the filters using the Filters Editor. To see a different view of the same traffic (with the same filters), choose a different view from the Views menu.

For more information, see Statistics Pages. To hide these instructions, choose Instructions from the Hide menu below. To hide them permanently, use the Stats Views section of the configuration options.

Hour of day	Page views	Bandwidth	Visitors	Page views bar
1:00 AM - 2:00 AM	2	12.75k	1	█
3:00 AM - 4:00 AM	1	7.35k	2	█
6:00 AM - 7:00 AM	2	11.00k	1	█
7:00 AM - 8:00 AM	2	14.70k	2	█
9:00 AM - 10:00 AM	2	14.70k	2	█
10:00 AM - 11:00 AM	5	22.19k	3	█
11:00 AM - noon	12	11.88M	5	█
noon - 1:00 PM	7	4.74M	2	█
1:00 PM - 2:00 PM	3	7.35k	2	█
3:00 PM - 4:00 PM	1	7.35k	1	█
4:00 PM - 5:00 PM	2	4.09M	1	█
6:00 PM - 7:00 PM	7	42.07k	1	█
7:00 PM - 8:00 PM	11	42.14k	2	█
10:00 PM - 11:00 PM	1	1441	1	█
11:00 PM - midnight	2	11.00k	1	█
Average	4	1426.57k	1	
Total	68	20.90M	27	

show more rows show fewer rows show 10 rows show all rows

Options: Sort:

Sawmill sample (R. Polytechnic Institute, NY)

```
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1623
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
```

Filters: This page shows hits for all the log data

- Overview
- Single-page summary
- Graphs of traffic over time
- Top days
- Top years/months/days
- Top weekdays (C)
- Top weekdays, avg. (C)
- Top weekdays
- Top hours
- Top pages
- Top pages/directories
- Top file types
- Top visitor domains/hosts
- Top visitor domains
- Top referrers
- Top search terms (C)
- Top referring domains
- Top search engines
- Top search phrases
- Top web browser types
- Top operating systems
- Top spiders/robots
- Top server domains
- Top authenticated users
- Top server responses
- Sessions (summary)
- Entry pages

This view shows the traffic, broken down by domain. The domains shown are English descriptions of Internet domains; for instance .edu is Educational and .fr is France. A value of "IP Address" usually means that there was no domain information available because the log data contained only the IP address of the visitor. You can use this view to see who is visiting the site, and which domains they are browsing from.



Page Views

	Domain description	Page views	Visitors	Page views bar
1	Educational (edu)	783870	21672	<div style="width: 100%;"></div>
2	Commercial (com)	673594	39591	<div style="width: 95%;"></div>
3	IP Address	317575	54422	<div style="width: 40%;"></div>
4	Network (net)	198817	47939	<div style="width: 25%;"></div>
5	Canada (ca)	14981	5013	<div style="width: 10%;"></div>
6	Non-profit (org)	11211	1069	<div style="width: 8%;"></div>
7	United Kingdom (uk)	7515	1352	<div style="width: 6%;"></div>
8	United States (us)	6582	1423	<div style="width: 5%;"></div>
9	Germany (de)	5886	1476	<div style="width: 4.5%;"></div>
10	France (fr)	5423	1425	<div style="width: 4.2%;"></div>
11	Military (mil)	5026	418	<div style="width: 3.8%;"></div>
12	Netherlands (nl)	4594	1097	<div style="width: 3.5%;"></div>
13	Japan (jp)	4195	1357	<div style="width: 3.2%;"></div>
14	Australia (au)	3783	1232	<div style="width: 2.8%;"></div>
15	Government (gov)	3390	489	<div style="width: 2.5%;"></div>
16	Italy (it)	3366	776	<div style="width: 2.4%;"></div>
17	Spain (es)	3358	969	<div style="width: 2.4%;"></div>
18	Belgium (be)	2489	468	<div style="width: 1.8%;"></div>

```
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U) " "q
```

```
fk-07-vusst.hr [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916  
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
```

Filters: This page shows hits for all the log data

Overview This view shows the traffic on each page and file of the site.

- Single-page summary
- Graphs of traffic over time
- Top days
- Top years/months/days
- Top weekdays (C)
- Top weekdays, avg. (C)
- Top weekdays
- Top hours
- Top pages
- Top pages/directories
- Top file types
- Top visitor domains/hosts
- Top visitor domains
- Top referrers
- Top search terms (C)
- Top referring domains
- Top search engines
- Top search phrases
- Top web browser types
- Top operating systems
- Top spiders/robots
- Top server domains
- Top authenticated users
- Top server responses
- Sessions (summary)
- Entry pages

(parenthesized items omitted)

Page or directory	Page views	Visitors	Visitors (%)	Visitors bar
1 / (default page)	131214	34138	2.98%	
2 /rpinfo/ (default page)	221379	21688	1.89%	
3 /rpinfo	86917	8528	0.74%	
4 /About/site_index.html	12311	7606	0.66%	
5 /cgi-bin/hphf?(parameters)	28898	7428	0.65%	
6 /favicon.ico	7095	5506	0.48%	
7 /Contacts/directories.html	9454	4744	0.41%	
8 /Departments/Academic.html	5249	3714	0.32%	
9 /dept/athletics/ (default page)	5632	3617	0.32%	
10 /dept/athletics/menubottom.htm	5513	3542	0.31%	
11 /dept/athletics/main.htm	5586	3469	0.30%	
12 /AutoGen/univweb.html	3541	2595	0.23%	
13 /dept/llc/writecenter/web/resume.html	2925	2544	0.22%	
14 /dept/NewsComm/sub/gateway/students.html	3160	2509	0.22%	
15 /dept/athletics/varsity.htm	3445	2345	0.20%	
16 /application/ (default page)	3741	2284	0.20%	
17 /dept/catalog/calendar.html	2897	2263	0.20%	
18 /dept/catalog/ (default page)	3302	2261	0.20%	
19 /dept/llc/writecenter/web/handouts.html	2314	1938	0.17%	
20 /dept/athletics/m_hockey/ (default page)	2693	1827	0.16%	
21 /web/writingcenter/cover_letter.html	1877	1608	0.14%	
22 /dept/NewsComm/Calendar/ (default page)	1995	1529	0.13%	
23 /dept/student-life/www/ (default page)	1930	1512	0.13%	
24 /rpinfo/index-nn.html	5365	1440	0.13%	
25 /Contacts/contact.html	1612	1350	0.12%	

```
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"  
zgrabaric.pbf.hr [13/Feb/2002:17:36:44 +0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304 --  
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"  
zgrabaric.pbf.hr [13/Feb/2002:17:36:44 +0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247  
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
```



- Windows
- OS unknown
- Known robots
- Unix
- Macintosh

The wedges are plotted by the number of requests for pages.

Analog

(Statistical Laboratory, Univ Cambridge, UK)

```

home/korisnici-on.gif·HTTP/1.1"·200·1916·
ndows·98;·DigExt) "¶
home/novi-korisnici-on.gif·HTTP/1.1"·200·1623·
ndows·98;·DigExt) "¶
home/baze-on.gif·HTTP/1.1"·200·1633·
ndows·98;·DigExt) "¶
home/info-on.gif·HTTP/1.1"·200·1029·
ndows·98;·DigExt) "¶
home/o-centru-on.gif·HTTP/1.1"·200·1492·
ndows·98;·DigExt) "¶
1.1"·200·892·
0·(compatible;·MSIE·6.0;·Windows·98) "¶
/ball_on.gif·HTTP/1.1"·200·104·
E·6.0;·Windows·98) "¶
dex.html·HTTP/1.1"·200·19411·"-·"Mozilla/4.0·
/ball_off.gif·HTTP/1.1"·200·155·
E·6.0;·Windows·98) "¶
/ovid.gif·HTTP/1.1"·200·610·
E·6.0;·Windows·98) "¶
1.html·HTTP/1.0"·200·32768·
MSIE·6.0;·Windows·98) "¶
/prirodoslovlje.gif·HTTP/1.1"·200·892·
E·6.0;·Windows·98) "¶
/biblio.gif·HTTP/1.1"·200·957·
E·6.0;·Windows·98) "¶
1.0"·304·--·"[unknown·origin]"·"Mozilla/4.76·[en]·
main-ie.css·HTTP/1.0"·200·469·"-·"Mozilla/4.76·
lka/podloga.gif·HTTP/1.0"·200·1049·
a-home/novi-korisnici-off.gif·HTTP/1.0"·304·--·
a-home/back-off.gif·HTTP/1.0"·304·--·
a-home/baze-off.gif·HTTP/1.0"·304·--·
a-home/english-off.gif·HTTP/1.0"·200·1247·

```

Listing operating systems, sorted by the number of requests for pages.

no.:	reqs:	pages:	OS
1:	18,894,144:	5,856,874:	Windows
:	7,161,183:	2,099,191:	Windows 98
:	4,199,106:	1,285,393:	Windows 2000
:	3,702,948:	1,222,100:	Windows NT
:	2,179,552:	744,558:	Windows 95
:	945,320:	266,072:	Windows ME
:	616,882:	186,784:	Windows XP
:	80,164:	50,010:	Unknown Windows
:	6,182:	1,860:	Windows 3.1
:	650:	574:	Windows 32-bit
:	1,885:	262:	Windows CE
:	272:	70:	Windows 16-bit
2:	4,532,330:	3,475,443:	OS unknown
3:	1,238,518:	1,029,456:	Known robots

HTTP Analyze (Netstore)

Total unique URLs	2639	
Total unique sites	56347	
Total user sessions	72219	
Total unique agents	5991	
Total unique referrer URLs	16215	
	Maximum	Average
Hits per day	29022	21358
Files sent per day	23316	17585
Files cached per day	5035	3277
Pageviews per day	4756	3168
Sessions per day	2959	2329
KB sent per day	152594	98294
Logfile statistics		
Total logfile entries read	662160	
Total logfile entries processed	662108	
Total authenticated requests (skipped)	51	
Empty Request Method	1	
Other Response Codes		
No Content (Code 204)	28	
Partial Content (Code 206)	2076	
Moved Temporarily (Code 302)	5334	
Unauthorized (Code 401)	1237	
Forbidden (Code 403)	438	
Not Found (Code 404)	4096	
Internal Server Error (Code 500)	362	
Request Methods other than GET/POST		
HEAD	1775	
PUT	2	
OPTIONS	4	

Mozilla/4.76



Mozilla/4.76



Mozilla/4.76 [en]



"On the Internet, nobody knows you"



ad6-m100.net.hinet.h
"http://nippur.irb.h
gita.srce.hr... [13
"http://knjiznica.ir
ad6-m100.net.hinet.h
"http://nippur.irb.h
ad6-m100.net.hinet.h
"http://nippur.irb.h



"http://nippur.irb.hr/ovid/ Mozilla/4.76 [en] (Win98; *0) 1

"Users" may vary from ordinary library users...

- spiders and crawlers
- someone abusing the site
- someone stealing the content
- someone harvesting email addresses
- companies trying to sell us their products
- users searching for information



Pitfalls

- ambiguity of log file terminology
- multiple files downloading from one Web page
- multiple hits per person
- dynamic addressing (DHCP)
- firewalls
- cacheing
- proxy servers

Audience (visitors, users)

- the main goal is to identify how many "people" (unique users) visited your web site for a specified period

1. IP addresses

2. cookies (are connected with the machine, not with the user; also could be rejected)

3. login names (tracking heaven, but you'll lose a lot of users)

fk-07-vusst.hr... [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/novi-korisnici-on.gif HTTP/1.1" 200 1623
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/info-on.gif HTTP/1.1" 200 1029
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49+0100] "GET /ovid/grafika-home/o-centru-on.gif HTTP/1.1" 200 1492
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:30+0100] "GET /images/ball-on.gif HTTP/1.1" 200 104
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33+0100] "GET /images/ball-on.gif HTTP/1.1" 200 104
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33+0100] "GET /images/ball-on.gif HTTP/1.1" 200 104
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:35+0100] "GET /images/ovid.gif HTTP/1.1" 200 610
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
gita.srce.hr... [13/Feb/2002:17:36:35+0100] "GET /hrv/search/obraz.html HTTP/1.0" 200 32768
"http://knjiznica.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:35+0100] "GET /images/biblio.gif HTTP/1.1" 200 892
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:40+0100] "GET /images/biblio.gif HTTP/1.1" 200 957
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:42+0100] "GET /ovid/ HTTP/1.0" 304 -- "[unknown origin]" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:42+0100] "GET /ovid/style/main-ie.css HTTP/1.0" 200 469 -- "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:43+0100] "GET /ovid/periodika/podloga.gif HTTP/1.0" 200 1049
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:43+0100] "GET /ovid/grafika-home/novi-korisnici-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/back-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304 --
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44+0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"

• *"There is absolutely no way to count visitors reliably. Until Big Brother ties people to their computers and those computers scan their retinas or fingerprints to supply you with this information, you'll never be sure who's visiting your site."*

Bill Winett, 2002

What do we know about our users?

- what kind of OS and browser are they using
- from which domain, companies are they coming
- what are they searching for before they visited our site
- what pages are popular by users
- for how long they are staying on our site (per session)

Traffic (activities)

web traffic analysis is the process of measuring extent and character of activity of users on a web site, interpreting the measurements, and applying the conclusions

- what's traffic like to any given page?
- are there certain pages that stand out as high-traffic areas?
- what is the overall traffic on my site?
- is there any traffic jump after new content publishing?
- is there any difference between Fridays and Mondays?

Browsers/Platforms

- are user primarily Linux or Windows users?
- does pages look well only with IE?
- is the content "plug-in" required?
- does Linux users have their preferences on the site?

Referrers and keywords

- a page that links to your site (or result set of a search engine)
- indicates where a user was referred from: a link somewhere else on your site, a link on some one else's site or and advertisement

Log analysis and privacy

In general, two types of data are available about web visitors:

- information provided by the individual
- information collected by observing behavior (log files)
- the data we typically receive from visitors includes registration and subscription data, and delivery data (for product shipment)
- those sites using personalization can collect data about users' interests, preferences, and decision criteria
- One problem with information gathered directly from individuals is that the data can be inaccurate

```
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/novi-korisnici-on.gif HTTP/1.1" 200 1623
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/info-on.gif HTTP/1.1" 200 1029
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/o-centru-on.gif HTTP/1.1" 200 1492
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33 +0100] "GET /images/ball-on.gif HTTP/1.1" 200 104
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33 +0100] "GET /hrv/index.html HTTP/1.1" 200 19411
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33 +0100] "GET /hrv/index.html HTTP/1.1" 200 19411
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:33 +0100] "GET /hrv/index.html HTTP/1.1" 200 19411
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98) "
gita.srce.hr... [13/Feb/2002:17:36:36 +0100] "GET /hrv/searchglobal.html HTTP/1.0" 200 32768
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:38 +0100] "GET /images/prirodoslovlje.gif HTTP/1.1" 200 892
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt) "
ad6-m100.net.hinet.hr... [13/Feb/2002:17:36:40 +0100] "GET /hrv/index.html HTTP/1.1" 200 19411
"http://nippur.irb.hr/hrv/index.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98) "
zgrabaric.pbf.hr... [13/Feb/2002:17:36:42 +0100] "GET /ovid/ HTTP/1.0" 304 --
[unknown origin] "Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:43 +0100] "GET /ovid/style/main-ie.css HTTP/1.0" 200 469
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:43 +0100] "GET /ovid/periodika/podloga.gif HTTP/1.0" 200 1049
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:43 +0100] "GET /ovid/grafika-home/novi-korisnici-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/back-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247
"Mozilla/4.76 [en] (Win98; U) "
```

- the only way a web site can obtain a users' name, home address or email address is to request that the user register with the site, then the user will use a password to log onto the site in the future
- the user's profile will be kept in a database that can generate information about individual usage



```
eb/2002:17:36:43 +0100] "GET /ovid/style/main-ie.css HTTP/1.0" 200 469
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:43 +0100] "GET /ovid/periodika/podloga.gif HTTP/1.0" 200 1049
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:43 +0100] "GET /ovid/grafika-home/novi-korisnici-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/back-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304 --
"Mozilla/4.76 [en] (Win98; U) "
eb/2002:17:36:44 +0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247
"Mozilla/4.76 [en] (Win98; U) "
```

More advanced tracking users process

- installing so called (server, network) monitor on the web server
- network monitors perform packet sniffing
- client requests, server responses, cookies, and HTML files
- can track stop requests issued from the browser, making it possible to list the pages that are taking too long to generate
- can measure the Web server's response time to different requests
- certain network monitors can report on content-related HTML tags, such as <TITLE>, <META> or comment tags
- they can even capture "form data" transmitted via a POST request when the visitor hits a submit button
- MODELLING VISITOR BEHAVIOR

disadvantages:

- processor time consuming
- can't track encrypted information from secure Web servers

Deeper analysis

- web mining
- usability – is it easy to use?
- navigability – are people finding their way from one page to another?
- visibility – how are people finding your site?
- computer forensics

Conclusion

- web server log file analysis is an important part of your monitoring of your web servers
- log file analysis alone will not give you the complete picture of your web server, but you can't get the complete picture without it
- know what is useful in the log files, what limitations are inherent in them, and how to analyze them
- importance of long-term logging

Further reading:

• <http://www.ftc.gov/privacy/>

• [Server Log Analysis](http://www.libraryofthegoddess.com/ServerLogAnalysis)

• <http://usability.gov/serverlog/index.html>

• [Correlating Web-User Data](http://www.clickz.com/mkt/precis)

• <http://www.clickz.com/mkt/precis/mkt/article.php/841181>

• [Measuring User Motivation from Server Log Files](http://www.microsoft.com/usability/webconf/fuller/fuller.htm)

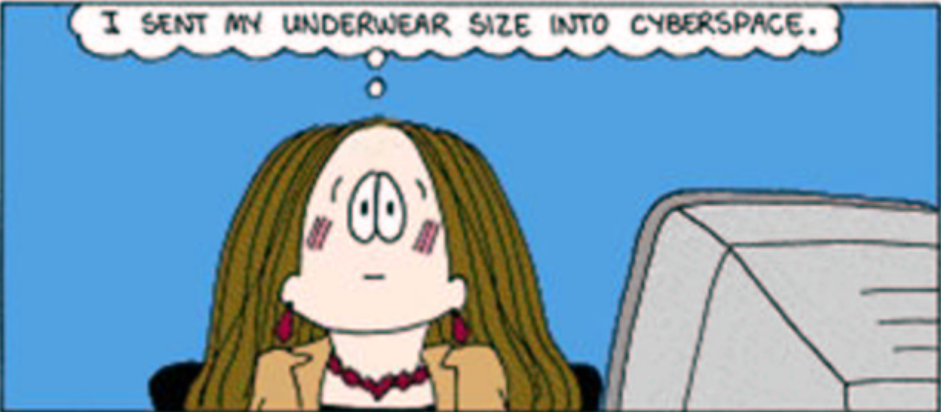
• [http://www.microsoft.com/usability/webconf/fuller/fuller.h](http://www.microsoft.com/usability/webconf/fuller/fuller.htm)

[tm](http://www.irsg.eu.org/irsg2000online/papers/he.htm)

• [Detecting session boundaries from Web user logs](http://www.irsg.eu.org/irsg2000online/papers/he.htm)

• <http://irsg.eu.org/irsg2000online/papers/he.htm>

```
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/korisnici-on.gif HTTP/1.1" 200 1916
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/novi-korisnici-on.gif HTTP/1.1" 200 1623
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/baze-on.gif HTTP/1.1" 200 1633
"http://nippur.irb.hr/ovid/" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
fk-07-vusst.hr... [13/Feb/2002:17:34:49 +0100] "GET /ovid/grafika-home/info-on.gif HTTP/1.1" 200 1000
```



```
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:43 +0100] "GET /ovid/grafika-home/novi-korisnici-off.gif HTTP/1.0" 304
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44 +0100] "GET /ovid/grafika-home/back-off.gif HTTP/1.0" 304
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44 +0100] "GET /ovid/grafika-home/baze-off.gif HTTP/1.0" 304
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
zgrabaric.pbf.hr... [13/Feb/2002:17:36:44 +0100] "GET /ovid/grafika-home/english-off.gif HTTP/1.0" 200 1247
"http://nippur.irb.hr/ovid/" "Mozilla/4.76 [en] (Win98; U)"
```