W3Perl

A free logfile analyzer

© Don Dixon 1999

Features

- Works on Unix / Windows / Mac
  - based on Perl scripts
- Web / FTP / Squid / Email servers
  - Others log format can be added easily
- Web Admin Interface
  - Allow to monitor your reports remotely
- PDF / Email / CSV /RSS reports
  - Get reports in your mailbox each morning
- Logfile parsing or page tagging
  - Read server log or log created from a javascript code
- Real time stats
  - View last entries
- Session tracking
  - Follow each user's path
- Detailed daily / weekly … reports
  - Full reports available on a different scale
- AJAX tools
  - Allow to sort table, to search for data …
- Can be run on a remote host
  - Log files can be retrieved, avoiding the need to compute on the server.

Monitoring technologies

Page tagging

- Cons :
  - Third party trust
  - No traffic data
  - Rely on javascript
- Pros :
  - No scripts to run or install
  - No caching problem
  - Can add extra data as screensize
  - Can log on javascript events
  - Doesn’t require logfile

Logfile analysis

- Cons :
  - CPU / Memory hungry
  - Need to install a package
- Pros :
  - No change in your website
  - Archive your logfile to retrieve data later
  - Spider activities
  - Fine tuning

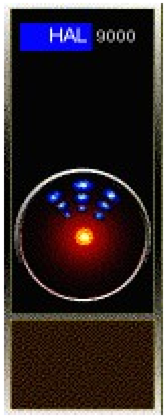
W3Perl can be used with both methods

Logfile

- Choose carefully your logfile format
  - Combined to get browsers/OS/referer stats
  - Reverse DNS to get country stats (w3perl can do it also)
- Hits vs Accesses

TOP SECRET

- Rotation / Compression
- Recommendation :
  - Apache : move to combined ECLF
  - IIS : use W3C or add extra fields to IIS format



- Create your own logfile if not available
  - Install the Page Tagging package first
- CLF logfile format :

%host %null %login %date %hourshift %method %page %protocol %status %requetesize

Logfile format

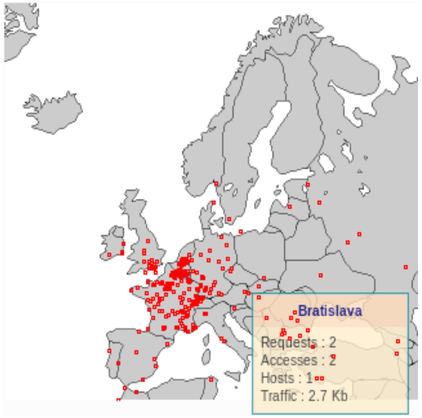
- Web : CLF, ECLF, W3C, IIS ...
- Squid : Native, CLF
- FTP : ProFTP, xferlog
- Mail : Exim / Postfix / Sendmail

Your can define your own format using predefined fields

=> %host, %date, %page, %protocol, %status, %requetesize ...

Data mining

- Main stats
  - Page, Hosts, Directories, File, Trafic, Country, City, Scripts ….



- Time stats
  - Real time, hours, days, weeks, months and years

- Additional stats
  - Referer, Browsers, OS, Errors, Sessions, RSS, URL mapping …

Using W3Perl

- Installation
  - Check requirements (Perl and Fly/Flydraw)
  - Choose Windows Binaries / Linux Packages or Tarball
- Configuration
  - Customize your stats : one configuration per report
  - Build your config from a Web administration interface
- Running the scripts
  - Master script : cron-w3perl.pl
    - -a to initialize
    - -e to update
    - -d <days> to compute the last days
  - Remotely via the Web administration interface or
  - Daily via a crontab
  - Load configuration via the –c flag
- Watch the results !

Installation

- Unix
  - RPM / Debian package
  - Tarball
    - Scripts -> /cgi-bin/w3perl/
    - Resources -> /htdocs/w3perl/
- Windows binaries (include predefined configuration files)
  - Apache
  - IIS
  - Abyss
  - No server
- Logfile access is not mandatory
  - A javascript tag to be inserted in your web pages can create logfiles for you
- Plug-in
  - PDF, Email, GeoIP ...

Avoid installing the package on a running server

Installation - Unix

- RPM / Debian package
  - Fly or Flydraw package needed
- Tarball
  - Install the fly or flydraw package
  - Extract w3perl package in your www document root
  - Edit the install.pl script
  - Change paths according to your system :
    - Perl path
    - CGI path
    - W3Perl package path
  - Run the install.pl
  - Use the web administration interface to build a config file
- Plug-in
  - Install optional third-party software : PDF, Email, GeoIP ...

Installation - Windows

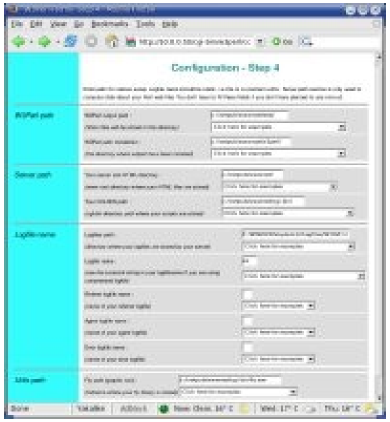
- ActivePerl
  - Install Perl first
- Binaries
  - IIS / Apache / Abyss server with default values
  - No server
- Configuration file
  - Apache / IIS / Abyss are ready to go
  - No server => give log files information
  - Customize values using a web interface
- Plug-in
  - Install optional third-party software : PDF, Email, GeoIP ...

Configuration

- Create a configuration file
  - Path, threshold, display, log format …
- If a web server is running :
  - Use the web administration interface
- If not :
  - You can use the online tool on w3perl website
  - Or edit manually the config.pl file
- Predefined config files :
  - Mandriva / Ubuntu / Debian
  - IIS / Apache / Abyss
  - FTP / Squid / Postfix / Exim
  - No logfile

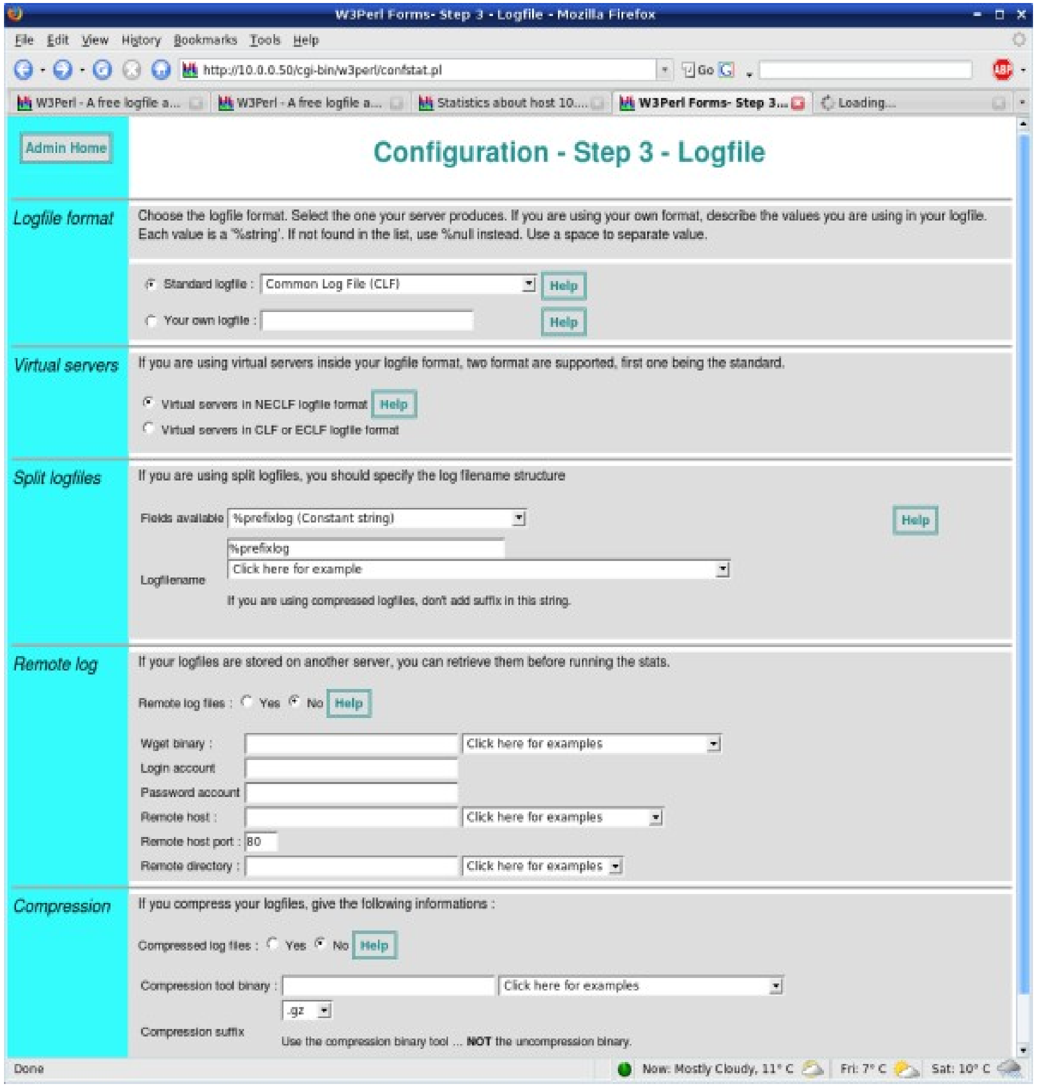
Web Administration Interface

- Manage configuration files
  - Create / Modify / Clone / Delete
- Run remotely your stats
  - Show status with a progress bar
- View stats reports
  - All your reports on one page
- Update package



Build a configuration file

- Customize display
  - Threshold, Graphs, URL / Hostname mapping, language …
- Logfiles
  - Paths, format, compression, split, filename
- Filtering rules
  - Hosts, URL, Countries, Directories, Robots to exclude
- Processing options
  - Reverse DNS, GeoIPCity, Domain stats, Scripts to run
- Automatic update
- Online tool available


W3Perl Forms- Step 3 - Logfile - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/cgi-bin/w3perl/confstat.pl
Go
W3Perl - A free logfile a...
W3Perl - A free logfile a...
Statistics about host 10....
W3Perl Forms- Step 3...
Loading...
Admin Home
Done
Now: Mostly Cloudy, 11° C
Fri: 7° C
Sat: 10° C

Configuration - Step 3 - Logfile

Logfile format

Choose the logfile format. Select the one your server produces. If you are using your own format, describe the values you are using in your logfile. Each value is a '%string'. If not found in the list, use %null instead. Use a space to separate value.

● Standard logfile : Common Log File (CLF) Help

○ Your own logfile : Help

Virtual servers

If you are using virtual servers inside your logfile format, two format are supported, first one being the standard.

● Virtual servers in NECLF logfile format Help

○ Virtual servers in CLF or ECLF logfile format

Split logfiles

If you are using split logfiles, you should specify the log filename structure

Fields available %prefixlog (Constant string) Help

Logfilename %prefixlog

Click here for example

If you are using compressed logfiles, dont add suffix in this string.

Remote log

If your logfiles are stored on another server, you can retrieve them before running the stats.

Remote log files : ○ Yes ● No Help

Wget binary : Click here for examples

Login account

Password account

Remote host : Click here for examples

Remote host port : 80

Remote directory : Click here for examples

Compression

If you compress your logfiles, give the following informations :

Compressed log files : ○ Yes ● No Help

Compression tool binary : Click here for examples

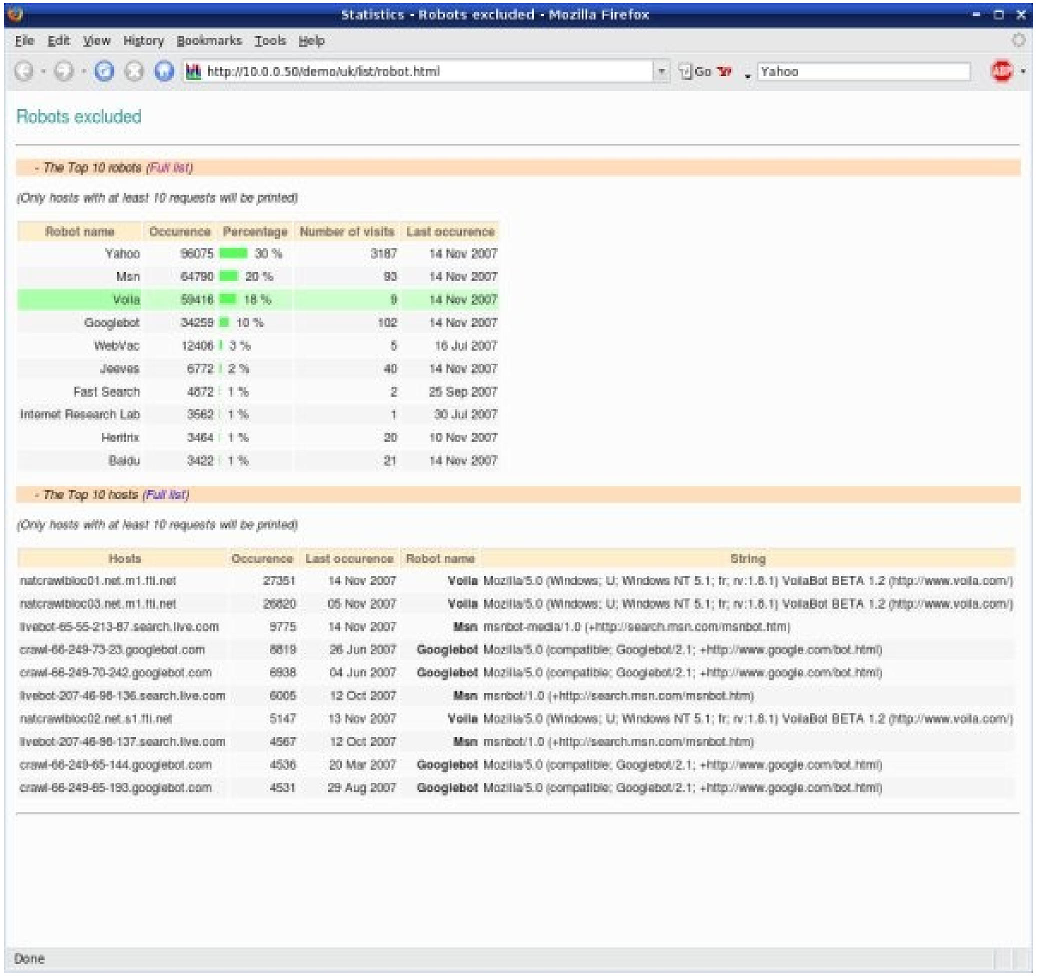
Compression suffix .gz

Use the compression binary tool ... **NOT** the uncompression binary.

Avoid to build configuration file manually

Filtering

- Robots detection
  - To reject spider visits
- Referer spam
- Countries
  - You can match your country
- URL
  - Selecting a specific web area
  - User’s report
- Directories
  - Avoid scanning private area
- Server’s status code
- Hosts
  - Exclude list of hosts


Statistics - Robots excluded - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/demo/uk/list/robot.html
Go
Yahoo
Done

Robots excluded

- The Top 10 robots (Full list)

(Only hosts with at least 10 requests will be printed)

Robot name	Occurence	Percentage	Number of visits	Last occurence
Yahoo	96075	30 %	3187	14 Nov 2007
Msn	64790	20 %	93	14 Nov 2007
Voila	59416	18 %	9	14 Nov 2007
Googlebot	34259	10 %	102	14 Nov 2007
WebVac	12406	3 %	5	16 Jul 2007
Jeeves	6772	2 %	40	14 Nov 2007
Fast Search	4872	1 %	2	25 Sep 2007
Internet Research Lab	3562	1 %	1	30 Jul 2007
Heritrix	3464	1 %	20	10 Nov 2007
Baidu	3422	1 %	21	14 Nov 2007

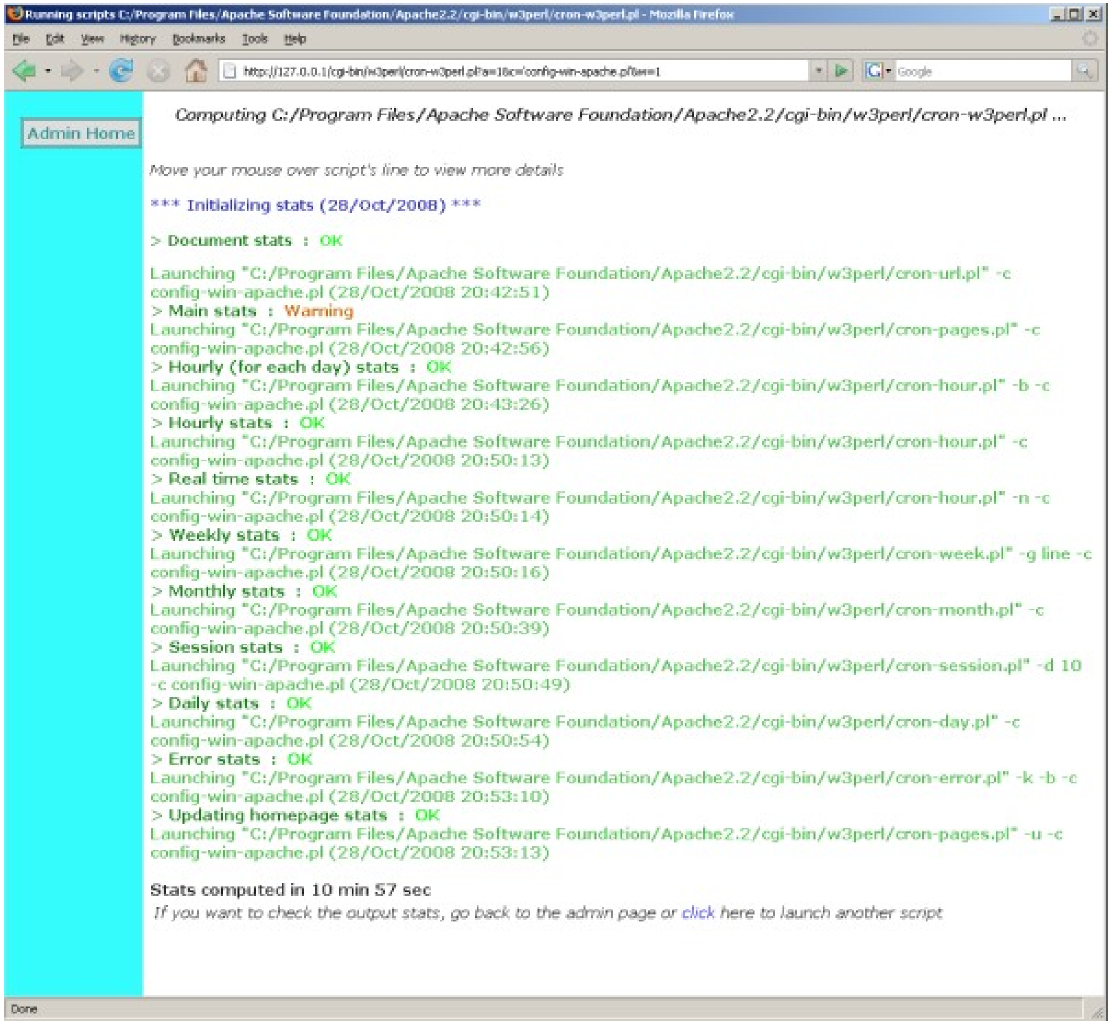
- The Top 10 hosts (Full list)

(Only hosts with at least 10 requests will be printed)

Hosts	Occurence	Last occurence	Robot name	String
natcrawlbloc01.net.m1.fti.net	27351	14 Nov 2007	**Voila**	Mozilla/5.0 (Windows; U; Windows NT 5.1; fr; rv:1.8.1) VoilaBot BETA 1.2 (http://www.voila.com/)
natcrawlbloc03.net.m1.fti.net	26820	05 Nov 2007	**Voila**	Mozilla/5.0 (Windows; U; Windows NT 5.1; fr; rv:1.8.1) VoilaBot BETA 1.2 (http://www.voila.com/)
livebot-65-55-213-87.search.live.com	9775	14 Nov 2007	**Msn**	msnbot-media/1.0 (+http://search.msn.com/msnbot.htm)
crawl-66-249-73-23.googlebot.com	8819	26 Jun 2007	**Googlebot**	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
crawl-66-249-70-242.googlebot.com	6938	04 Jun 2007	**Googlebot**	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
livebot-207-46-98-136.search.live.com	6005	12 Oct 2007	**Msn**	msnbot/1.0 (+http://search.msn.com/msnbot.htm)
natcrawlbloc02.net.s1.fti.net	5147	13 Nov 2007	**Voila**	Mozilla/5.0 (Windows; U; Windows NT 5.1; fr; rv:1.8.1) VoilaBot BETA 1.2 (http://www.voila.com/)
livebot-207-46-98-137.search.live.com	4567	12 Oct 2007	**Msn**	msnbot/1.0 (+http://search.msn.com/msnbot.htm)
crawl-66-249-65-144.googlebot.com	4536	20 Mar 2007	**Googlebot**	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
crawl-66-249-65-193.googlebot.com	4531	29 Aug 2007	**Googlebot**	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)

Running

- From a web admin
  - Remotely
  - Get report in one click
  - Click on ‘Init’ to start
- From command lines
  - Use -h for help
  - Master script : cron-w3perl
  - Flag –c to load a config
  - Flag –a to init
  - Flag –e to update
- Windows menu
  - Init / Update
  - Install / Uninstall
- Options
  - Date selection
  - Graph type
  - … and many more


Running scripts C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-w3perl.pl - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
http://127.0.0.1/cgi-bin/w3perl/cron-w3perl.pl?a=1&c=/config-win-apache.pl&w=1
Google
Admin Home
Done

Computing C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-w3perl.pl ...

Move your mouse over script's line to view more details

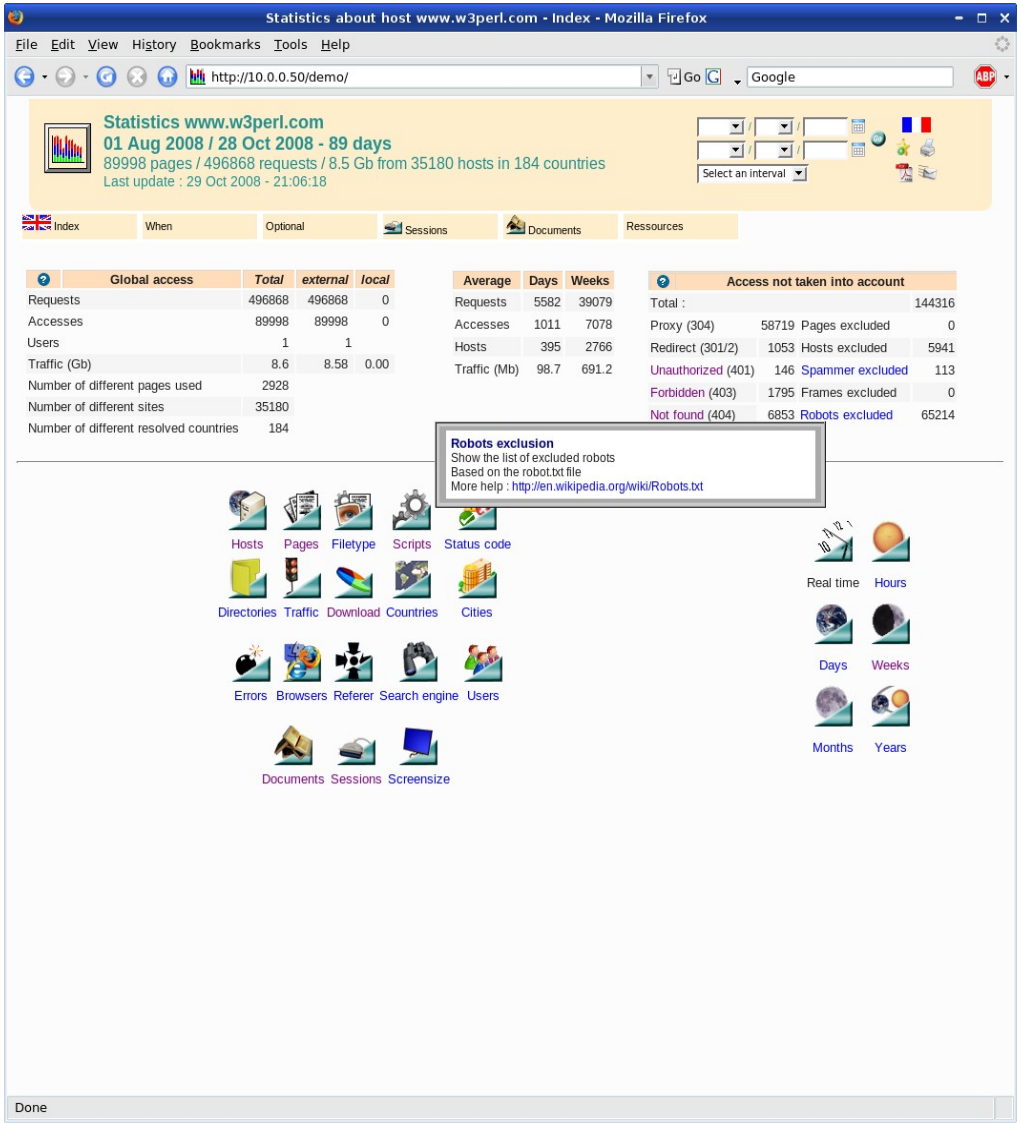
```
*** Initializing stats (28/Oct/2008) ***

> Document stats  :  OK

Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-url.pl" -c
config-win-apache.pl (28/Oct/2008 20:42:51)
> Main stats  :  Warning
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-pages.pl" -c
config-win-apache.pl (28/Oct/2008 20:42:56)
> Hourly (for each day) stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-hour.pl" -b -c
config-win-apache.pl (28/Oct/2008 20:43:26)
> Hourly stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-hour.pl" -c
config-win-apache.pl (28/Oct/2008 20:50:13)
> Real time stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-hour.pl" -n -c
config-win-apache.pl (28/Oct/2008 20:50:14)
> Weekly stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-week.pl" -g line -c
config-win-apache.pl (28/Oct/2008 20:50:16)
> Monthly stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-month.pl" -c
config-win-apache.pl (28/Oct/2008 20:50:39)
> Session stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-session.pl" -d 10
-c config-win-apache.pl (28/Oct/2008 20:50:49)
> Daily stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-day.pl" -c
config-win-apache.pl (28/Oct/2008 20:50:54)
> Error stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-error.pl" -k -b -c
config-win-apache.pl (28/Oct/2008 20:53:10)
> Updating homepage stats  :  OK
Launching "C:/Program Files/Apache Software Foundation/Apache2.2/cgi-bin/w3perl/cron-pages.pl" -u -c
config-win-apache.pl (28/Oct/2008 20:53:13)

Stats computed in 10 min 57 sec
```

If you want to check the output stats, go back to the admin page or click here to launch another script

Don’t mix web admin and command line run

Reports

- Top menu
  - Summary
  - Tools : date selection, pdf exportation, language …
- Pure CSS menu
  - Navigate through reports
  - Main/When/Optional/Sessions/ Documents/Resources
- Global summary
  - Hosts / Pages / Files / Traffic / Users / Exclude
- Icons field
  - Select reports
  - Updated when new reports are available



Statistics www.w3perl.com

01 Aug 2008 / 28 Oct 2008 - 89 days

89998 pages / 496868 requests / 8.5 Gb from 35180 hosts in 184 countries

Last update : 29 Oct 2008 - 21:06:18

? Global access	Total	external	local
Requests	496868	496868	0
Accesses	89998	89998	0
Users	1	1	
Traffic (Gb)	8.6	8.58	0.00
Number of different pages used	2928		
Number of different sites	35180		
Number of different resolved countries	184		

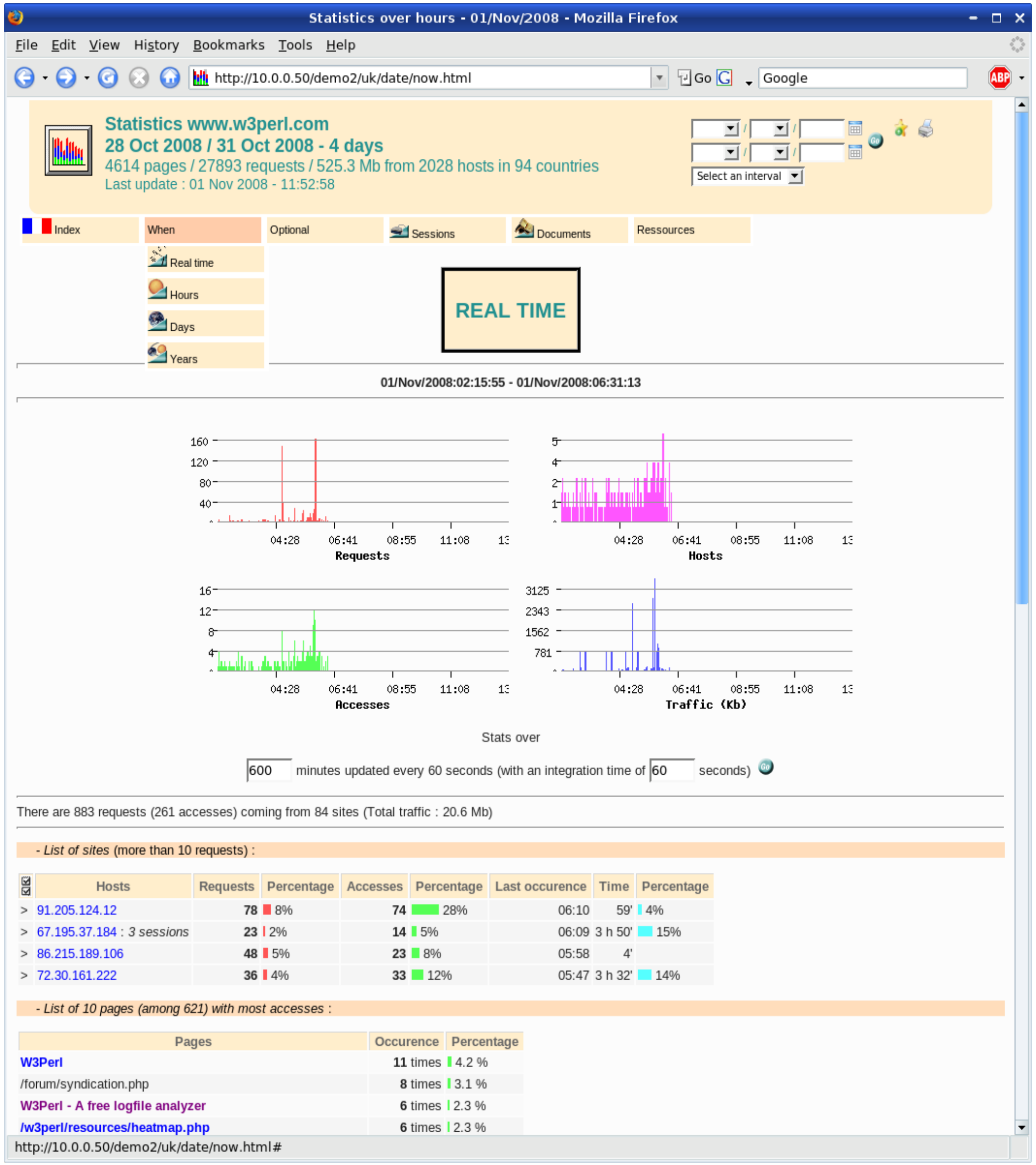
Average	Days	Weeks
Requests	5582	39079
Accesses	1011	7078
Hosts	395	2766
Traffic (Mb)	98.7	691.2

? Access not taken into account			
Total :			144316
Proxy (304)	58719	Pages excluded	0
Redirect (301/2)	1053	Hosts excluded	5941
Unauthorized (401)	146	Spammer excluded	113
Forbidden (403)	1795	Frames excluded	0
Not found (404)	6853	Robots excluded	65214

Robots exclusion
Show the list of excluded robots
Based on the robot.txt file
More help : http://en.wikipedia.org/wiki/Robots.txt

Real time

- Events monitoring
  - Updated every minute
  - Hits, Accesses, Traffic, Hosts…
- Graphics
  - Activity over the last hours
  - Length and integration can be changed in real time
- Show latest :
  - hosts / pages / scripts / files / traffic
- AJAX
  - Click on tables to sort them
  - Click on hosts to view page read



Statistics www.w3perl.com
28 Oct 2008 / 31 Oct 2008 - 4 days
4614 pages / 27893 requests / 525.3 Mb from 2028 hosts in 94 countries
Last update : 01 Nov 2008 - 11:52:58

01/Nov/2008:02:15:55 - 01/Nov/2008:06:31:13

Stats over

600 minutes updated every 60 seconds (with an integration time of 60 seconds)

There are 883 requests (261 accesses) coming from 84 sites (Total traffic : 20.6 Mb)

- *List of sites* (more than 10 requests) :

☑	Hosts	Requests	Percentage	Accesses	Percentage	Last occurence	Time	Percentage
>	91.205.124.12	**78**	8%	**74**	28%	06:10	59'	4%
>	67.195.37.184 : *3 sessions*	**23**	2%	**14**	5%	06:09	3 h 50'	15%
>	86.215.189.106	**48**	5%	**23**	8%	05:58	4'	
>	72.30.161.222	**36**	4%	**33**	12%	05:47	3 h 32'	14%

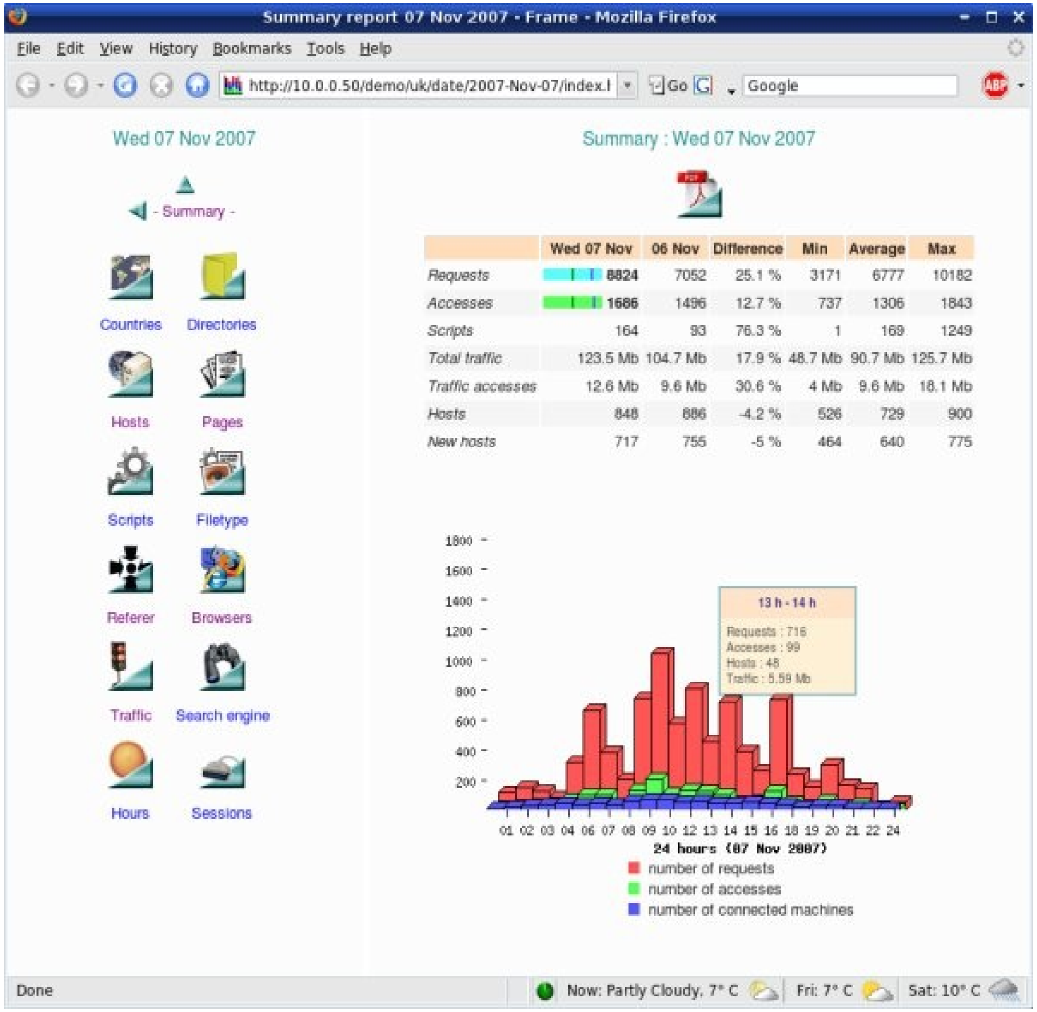
- *List of 10 pages (among 621) with most accesses* :

Pages	Occurence	Percentage
W3Perl	**11** times	4.2 %
/forum/syndication.php	**8** times	3.1 %
W3Perl - A free logfile analyzer	**6** times	2.3 %
/w3perl/resources/heatmap.php	**6** times	2.3 %

http://10.0.0.50/demo2/uk/date/now.html#

For each day

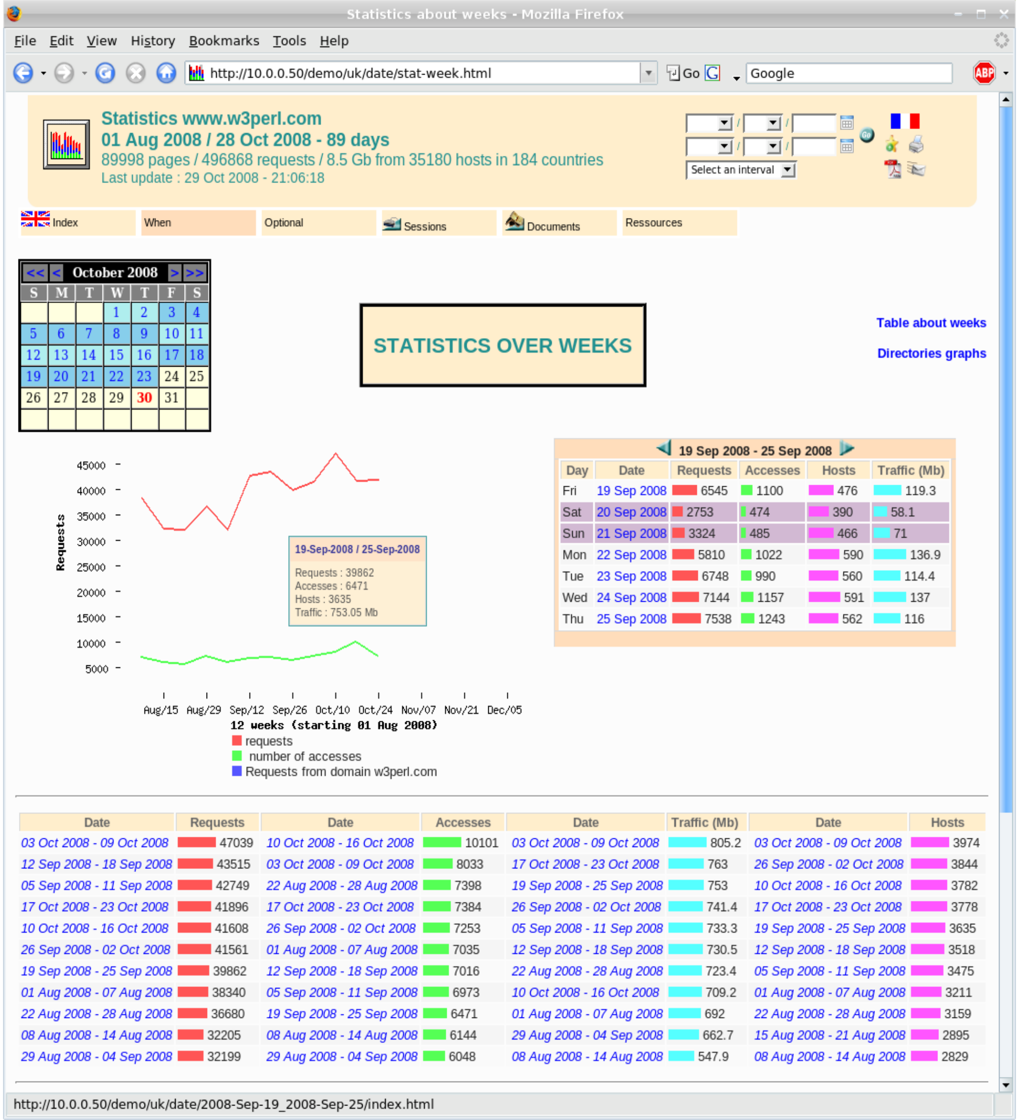
- Summary
  - Hits, Accesses, Traffic, Hosts…
- Graphic
  - Activity over 24 hours
  - Show peaks
- Complete reports
  - Countries / Directories / Hosts / Pages / Filetype / Traffic / Session / Referer / Browsers / Scripts / …
- Can be send via email
  - Html report
  - PDF as an attachment



	Wed 07 Nov	06 Nov	Difference	Min	Average	Max
Requests	**8824**	7052	25.1 %	3171	6777	10182
Accesses	**1686**	1496	12.7 %	737	1306	1843
Scripts	164	93	76.3 %	1	169	1249
Total traffic	123.5 Mb	104.7 Mb	17.9 %	48.7 Mb	90.7 Mb	125.7 Mb
Traffic accesses	12.6 Mb	9.6 Mb	30.6 %	4 Mb	9.6 Mb	18.1 Mb
Hosts	848	886	-4.2 %	526	729	900
New hosts	717	755	-5 %	464	640	775

Weekly / Monthly / Yearly

- Same template as daily reports
  - Hits, Accesses, Traffic, Hosts…
- Graphics
  - Hits / Accesses versus time
  - Hosts versus time
- Full list
  - Not restricted to the top ten
- External / Domain split
- Interactive graphics
  - Link to others time stats
  - Display popup
- Can be send via email
  - Html report
  - PDF as an attachment


Statistics about weeks - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/demo/uk/date/stat-week.html
Go
Google
Select an interval
Index
When
Optional
Sessions
Documents
Ressources

Statistics www.w3perl.com
01 Aug 2008 / 28 Oct 2008 - 89 days
89998 pages / 496868 requests / 8.5 Gb from 35180 hosts in 184 countries
Last update : 29 Oct 2008 - 21:06:18

<< <	October 2008				>	>>
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

STATISTICS OVER WEEKS

Table about weeks

Directories graphs

19-Sep-2008 / 25-Sep-2008

Requests : 39862
Accesses : 6471
Hosts : 3635
Traffic : 753.05 Mb

19 Sep 2008 - 25 Sep 2008

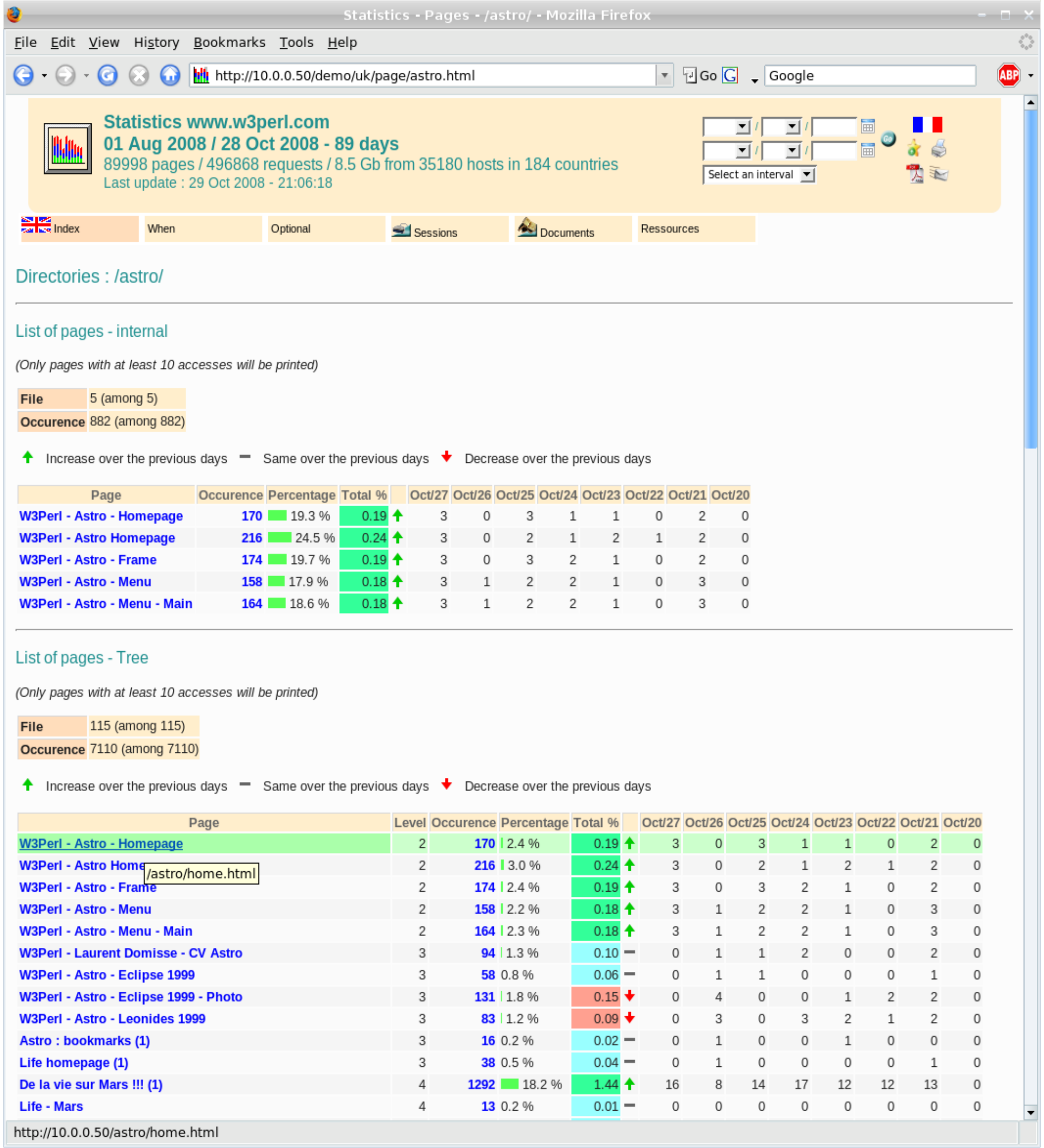
Day	Date	Requests	Accesses	Hosts	Traffic (Mb)
Fri	19 Sep 2008	6545	1100	476	119.3
Sat	20 Sep 2008	2753	474	390	58.1
Sun	21 Sep 2008	3324	485	466	71
Mon	22 Sep 2008	5810	1022	590	136.9
Tue	23 Sep 2008	6748	990	560	114.4
Wed	24 Sep 2008	7144	1157	591	137
Thu	25 Sep 2008	7538	1243	562	116

Date	Requests	Date	Accesses	Date	Traffic (Mb)	Date	Hosts
03 Oct 2008 - 09 Oct 2008	47039	10 Oct 2008 - 16 Oct 2008	10101	03 Oct 2008 - 09 Oct 2008	805.2	03 Oct 2008 - 09 Oct 2008	3974
12 Sep 2008 - 18 Sep 2008	43515	03 Oct 2008 - 09 Oct 2008	8033	17 Oct 2008 - 23 Oct 2008	763	26 Sep 2008 - 02 Oct 2008	3844
05 Sep 2008 - 11 Sep 2008	42749	22 Aug 2008 - 28 Aug 2008	7398	19 Sep 2008 - 25 Sep 2008	753	10 Oct 2008 - 16 Oct 2008	3782
17 Oct 2008 - 23 Oct 2008	41896	17 Oct 2008 - 23 Oct 2008	7384	26 Sep 2008 - 02 Oct 2008	741.4	17 Oct 2008 - 23 Oct 2008	3778
10 Oct 2008 - 16 Oct 2008	41608	26 Sep 2008 - 02 Oct 2008	7253	05 Sep 2008 - 11 Sep 2008	733.3	19 Sep 2008 - 25 Sep 2008	3635
26 Sep 2008 - 02 Oct 2008	41561	01 Aug 2008 - 07 Aug 2008	7035	12 Sep 2008 - 18 Sep 2008	730.5	12 Sep 2008 - 18 Sep 2008	3518
19 Sep 2008 - 25 Sep 2008	39862	12 Sep 2008 - 18 Sep 2008	7016	22 Aug 2008 - 28 Aug 2008	723.4	05 Sep 2008 - 11 Sep 2008	3475
01 Aug 2008 - 07 Aug 2008	38340	05 Sep 2008 - 11 Sep 2008	6973	10 Oct 2008 - 16 Oct 2008	709.2	01 Aug 2008 - 07 Aug 2008	3211
22 Aug 2008 - 28 Aug 2008	36680	19 Sep 2008 - 25 Sep 2008	6471	01 Aug 2008 - 07 Aug 2008	692	22 Aug 2008 - 28 Aug 2008	3159
08 Aug 2008 - 14 Aug 2008	32205	08 Aug 2008 - 14 Aug 2008	6144	29 Aug 2008 - 04 Sep 2008	662.7	15 Aug 2008 - 21 Aug 2008	2895
29 Aug 2008 - 04 Sep 2008	32199	29 Aug 2008 - 04 Sep 2008	6048	08 Aug 2008 - 14 Aug 2008	547.9	08 Aug 2008 - 14 Aug 2008	2829

http://10.0.0.50/demo/uk/date/2008-Sep-19_2008-Sep-25/index.html

Pages stats

- Based on file extension
  - Html, php …
- Most popular pages
  - Full list available
- Detailed pages
  - View hosts/date/occurrence
- Search
  - To find easily a target
  - To compare a group
  - Match URL or page’s title
- Link to hosts
  - View hosts for each page
- Over the last few days
  - Graphic activity
  - Increase / Decrease
- Sorted by directory


Statistics - Pages - /astro/ - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/demo/uk/page/astro.html
Go
Google
Select an interval
Index
When
Optional
Sessions
Documents
Ressources

Statistics www.w3perl.com
01 Aug 2008 / 28 Oct 2008 - 89 days
89998 pages / 496868 requests / 8.5 Gb from 35180 hosts in 184 countries
Last update : 29 Oct 2008 - 21:06:18

Directories : /astro/

List of pages - internal

(Only pages with at least 10 accesses will be printed)

File	5 (among 5)
Occurence	882 (among 882)

↑ Increase over the previous days — Same over the previous days ↓ Decrease over the previous days

Page	Occurence	Percentage	Total %		Oct/27	Oct/26	Oct/25	Oct/24	Oct/23	Oct/22	Oct/21	Oct/20
W3Perl - Astro - Homepage	170	19.3 %	0.19	↑	3	0	3	1	1	0	2	0
W3Perl - Astro Homepage	216	24.5 %	0.24	↑	3	0	2	1	2	1	2	0
W3Perl - Astro - Frame	174	19.7 %	0.19	↑	3	0	3	2	1	0	2	0
W3Perl - Astro - Menu	158	17.9 %	0.18	↑	3	1	2	2	1	0	3	0
W3Perl - Astro - Menu - Main	164	18.6 %	0.18	↑	3	1	2	2	1	0	3	0

List of pages - Tree

(Only pages with at least 10 accesses will be printed)

File	115 (among 115)
Occurence	7110 (among 7110)

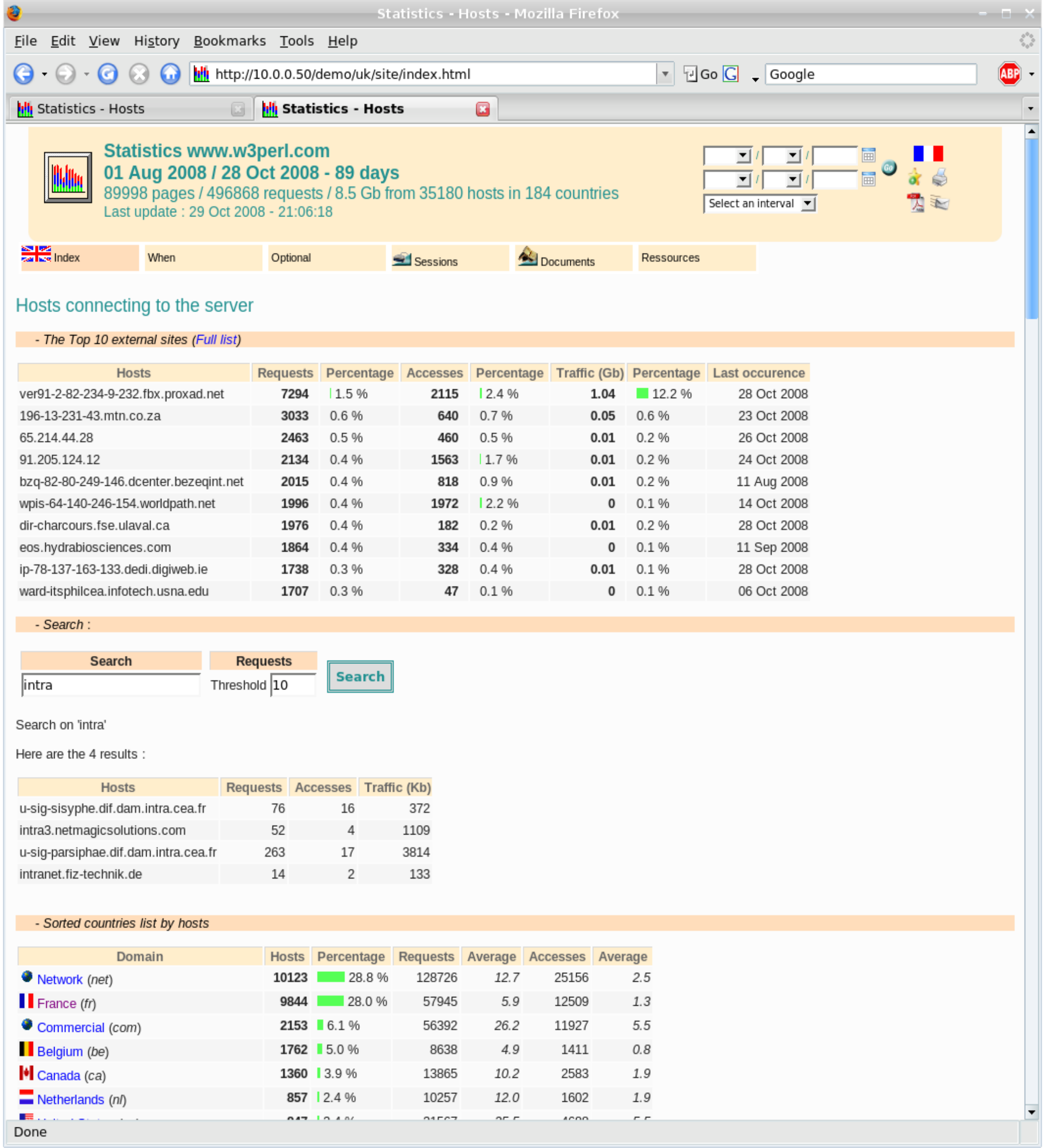
↑ Increase over the previous days — Same over the previous days ↓ Decrease over the previous days

Page	Level	Occurence	Percentage	Total %		Oct/27	Oct/26	Oct/25	Oct/24	Oct/23	Oct/22	Oct/21	Oct/20
W3Perl - Astro - Homepage	2	170	2.4 %	0.19	↑	3	0	3	1	1	0	2	0
W3Perl - Astro Home /astro/home.html	2	216	3.0 %	0.24	↑	3	0	2	1	2	1	2	0
W3Perl - Astro - Frame	2	174	2.4 %	0.19	↑	3	0	3	2	1	0	2	0
W3Perl - Astro - Menu	2	158	2.2 %	0.18	↑	3	1	2	2	1	0	3	0
W3Perl - Astro - Menu - Main	2	164	2.3 %	0.18	↑	3	1	2	2	1	0	3	0
W3Perl - Laurent Domisse - CV Astro	3	94	1.3 %	0.10	—	0	1	1	2	0	0	2	0
W3Perl - Astro - Eclipse 1999	3	58	0.8 %	0.06	—	0	1	1	0	0	0	1	0
W3Perl - Astro - Eclipse 1999 - Photo	3	131	1.8 %	0.15	↓	0	4	0	0	1	2	2	0
W3Perl - Astro - Leonides 1999	3	83	1.2 %	0.09	↓	0	3	0	3	2	1	2	0
Astro : bookmarks (1)	3	16	0.2 %	0.02	—	0	1	0	0	1	0	0	0
Life homepage (1)	3	38	0.5 %	0.04	—	0	1	0	0	0	0	1	0
De la vie sur Mars !!! (1)	4	1292	18.2 %	1.44	↑	16	8	14	17	12	12	13	0
Life - Mars	4	13	0.2 %	0.01	—	0	0	0	0	0	0	0	0

http://10.0.0.50/astro/home.html

Hosts stats

- Can exclude robots
  - Based on a file
- Most popular hosts
  - Full list available
- For each host
  - List of pages red and when
  - Hits / Accesses / Traffic
- Search
  - To find easily a target
  - To compare a group
- Explore by country domain
  - Country => Domain => Hosts => Host’s pages


Statistics - Hosts - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/demo/uk/site/index.html
Go
Google
Statistics - Hosts
Statistics - Hosts
Select an interval
Index
When
Optional
Sessions
Documents
Ressources
Done

Statistics www.w3perl.com
01 Aug 2008 / 28 Oct 2008 - 89 days
89998 pages / 496868 requests / 8.5 Gb from 35180 hosts in 184 countries
Last update : 29 Oct 2008 - 21:06:18

Hosts connecting to the server

- The Top 10 external sites (Full list)

Hosts	Requests	Percentage	Accesses	Percentage	Traffic (Gb)	Percentage	Last occurence
ver91-2-82-234-9-232.fbx.proxad.net	**7294**	1.5 %	**2115**	2.4 %	**1.04**	🟩 12.2 %	28 Oct 2008
196-13-231-43.mtn.co.za	**3033**	0.6 %	**640**	0.7 %	**0.05**	0.6 %	23 Oct 2008
65.214.44.28	**2463**	0.5 %	**460**	0.5 %	**0.01**	0.2 %	26 Oct 2008
91.205.124.12	**2134**	0.4 %	**1563**	1.7 %	**0.01**	0.2 %	24 Oct 2008
bzq-82-80-249-146.dcenter.bezeqint.net	**2015**	0.4 %	**818**	0.9 %	**0.01**	0.2 %	11 Aug 2008
wpis-64-140-246-154.worldpath.net	**1996**	0.4 %	**1972**	2.2 %	**0**	0.1 %	14 Oct 2008
dir-charcours.fse.ulaval.ca	**1976**	0.4 %	**182**	0.2 %	**0.01**	0.2 %	28 Oct 2008
eos.hydrabiosciences.com	**1864**	0.4 %	**334**	0.4 %	**0**	0.1 %	11 Sep 2008
ip-78-137-163-133.dedi.digiweb.ie	**1738**	0.3 %	**328**	0.4 %	**0.01**	0.1 %	28 Oct 2008
ward-itsphilcea.infotech.usna.edu	**1707**	0.3 %	**47**	0.1 %	**0**	0.1 %	06 Oct 2008

- Search :

Search	Requests
intra	Threshold 10

Search

Search on 'intra'

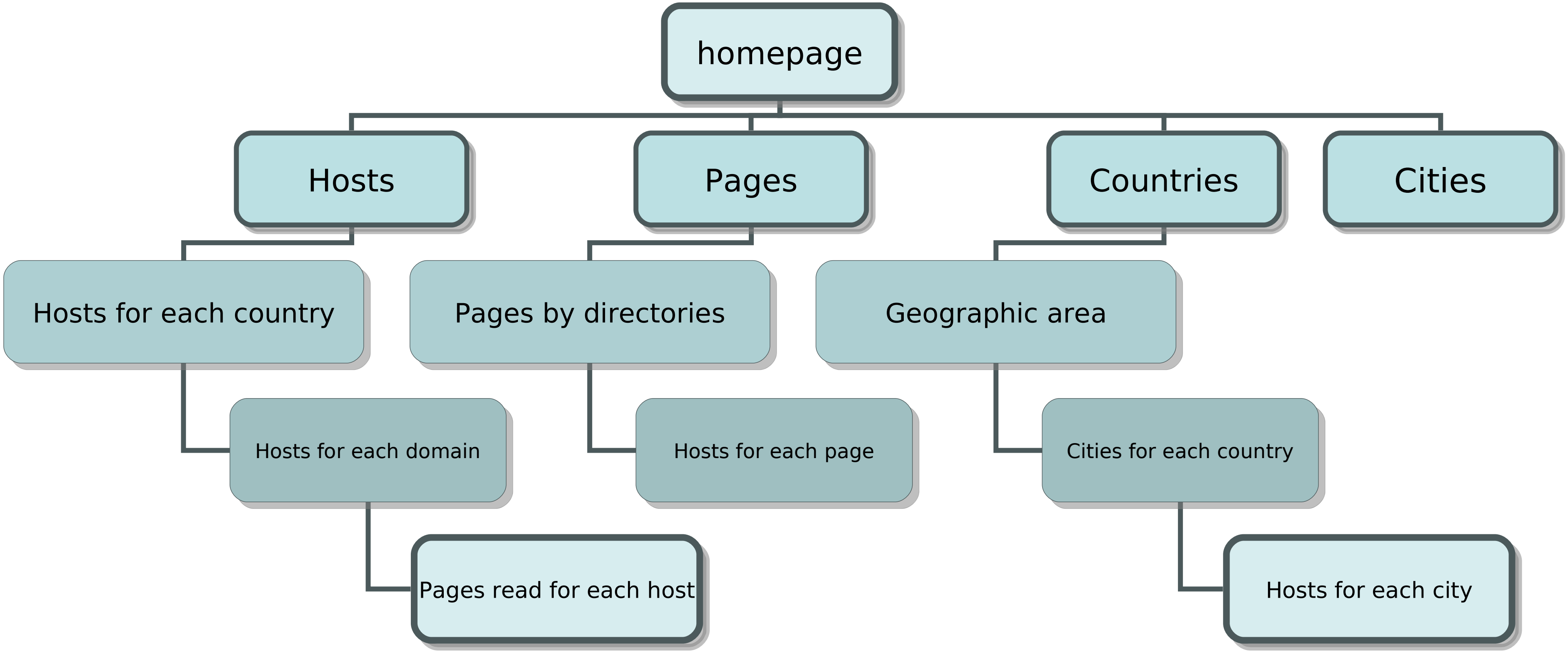
Here are the 4 results :

Hosts	Requests	Accesses	Traffic (Kb)
u-sig-sisyphe.dif.dam.intra.cea.fr	76	16	372
intra3.netmagicsolutions.com	52	4	1109
u-sig-parsiphae.dif.dam.intra.cea.fr	263	17	3814
intranet.fiz-technik.de	14	2	133

- Sorted countries list by hosts

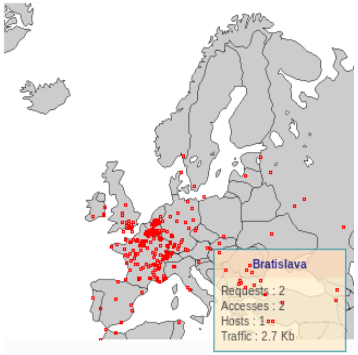
Domain	Hosts	Percentage	Requests	Average	Accesses	Average
Network *(net)*	**10123**	🟩 28.8 %	128726	*12.7*	25156	*2.5*
France *(fr)*	**9844**	🟩 28.0 %	57945	*5.9*	12509	*1.3*
Commercial *(com)*	**2153**	6.1 %	56392	*26.2*	11927	*5.5*
Belgium *(be)*	**1762**	5.0 %	8638	*4.9*	1411	*0.8*
Canada *(ca)*	**1360**	3.9 %	13865	*10.2*	2583	*1.9*
Netherlands *(nl)*	**857**	2.4 %	10257	*12.0*	1602	*1.9*

Navigation


homepage
Hosts
Pages
Countries
Cities
Hosts for each country
Pages by directories
Geographic area
Hosts for each domain
Hosts for each page
Cities for each country
Pages read for each host
Hosts for each city

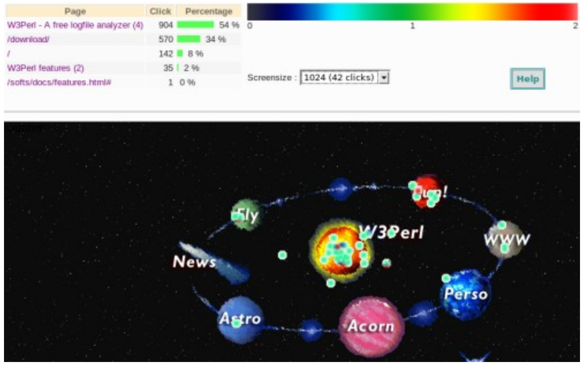
Others

- Using plug-in
  - GeoIP to get country stats from IP logfile
  - GeoLiteCity to get cities stats
  - When adding a javascript :
    - Heat map
    - Screen resolution / Color depth / Java support
- Directories
  - Hits / Accesses / Traffic for each path level
- Filetype
  - Check traffic by file extension (find easily unsolicited files)
- RSS
  - Who is reading your flux
- Status code
  - Check client / server errors
- Search engine
  - Which keyword has been used to reach your website ?
- Referer
  - Where does people come from ?



Bratislava

Requests : 2
Accesses : 2
Hosts : 1
Traffic : 2.7 Kb

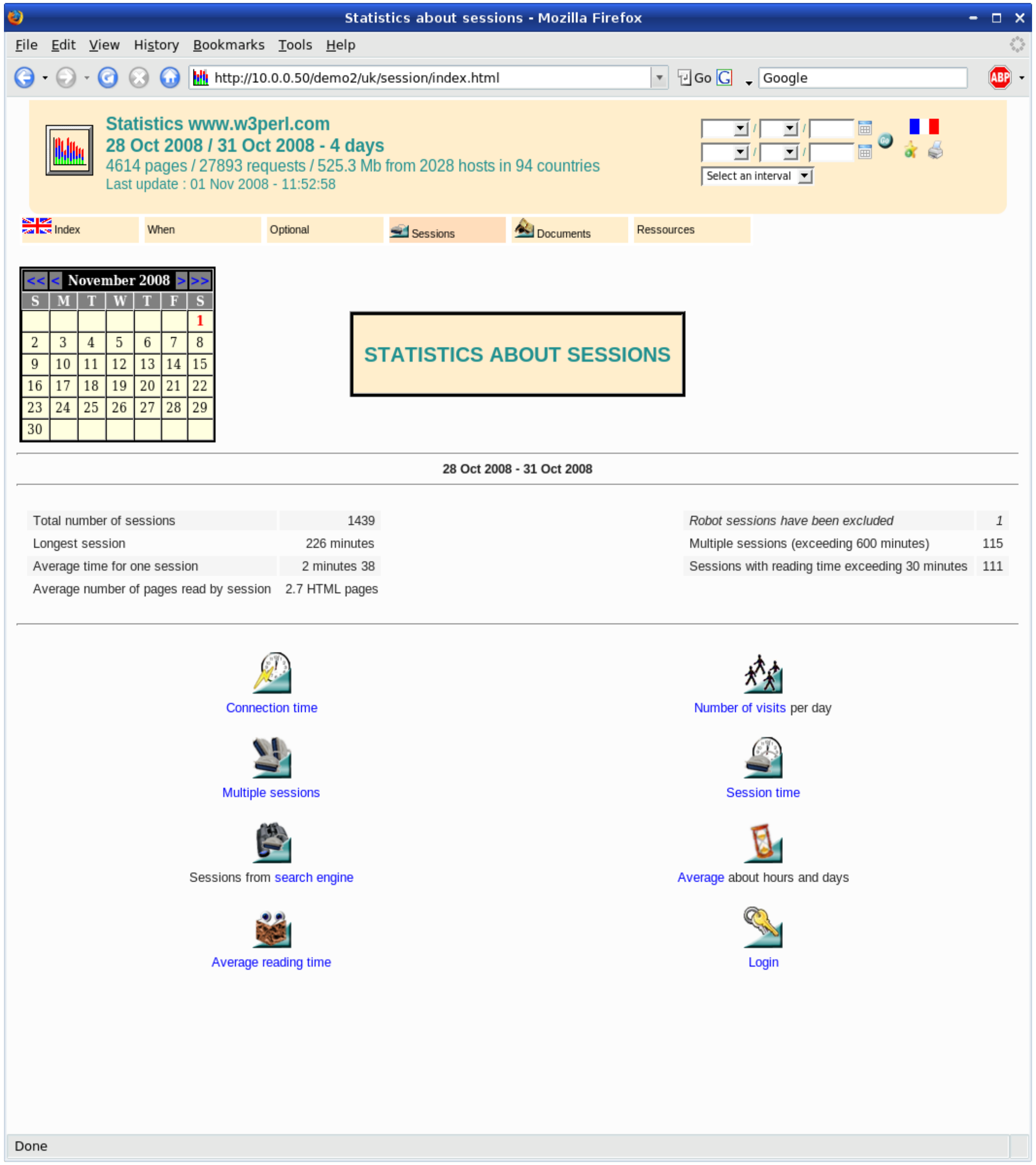


Page	Click	Percentage
W3Perl - A free logfile analyzer (4)	904	54 %
/download/	570	34 %
/	142	8 %
W3Perl features (2)	35	2 %
/softs/docs/features.html#	1	0 %

Screensize : 1024 (42 clicks)

Session problem

- Internet is an endless protocol
  - No delimiter
- Host = Visits ?
  - Same IP shared by different people
    - Proxy
    - NAT
    - Computer room
  - Same person with different IP
    - DHCP lease
    - Use different computer (home/work)
- Setting a Timeout
  - Usually 20 minutes without requests
- Session stats
  - Track user's path



http://10.0.0.50/demo2/uk/session/index.html

Statistics www.w3perl.com
28 Oct 2008 / 31 Oct 2008 - 4 days
4614 pages / 27893 requests / 525.3 Mb from 2028 hosts in 94 countries
Last update : 01 Nov 2008 - 11:52:58

<< < November 2008 > >>

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

STATISTICS ABOUT SESSIONS

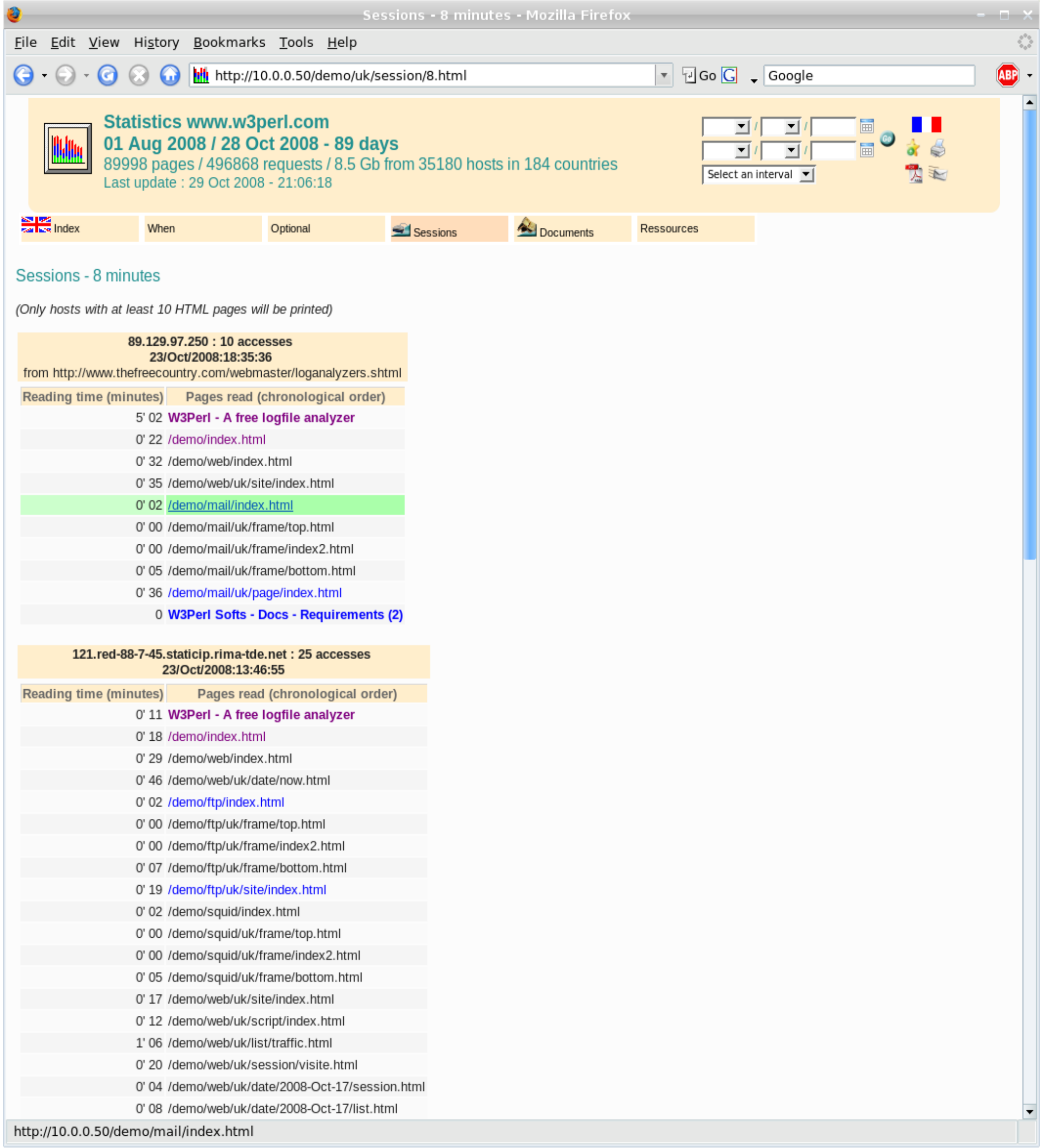
28 Oct 2008 - 31 Oct 2008

Total number of sessions	1439
Longest session	226 minutes
Average time for one session	2 minutes 38
Average number of pages read by session	2.7 HTML pages

Robot sessions have been excluded	*1*
Multiple sessions (exceeding 600 minutes)	115
Sessions with reading time exceeding 30 minutes	111

Follow user's path

- For each session
  - Host / Date / Referer / Accesses
  - View pages / Time spent
  - First / Ending page
- Fidelity
  - How many session for each host
  - How long they stay
- Session length
  - Histogram
- Page reading time
  - Most popular page (avoid page jump)
- Session hourly and daily histogram


Sessions - 8 minutes - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
http://10.0.0.50/demo/uk/session/8.html
Go
Google
Select an interval
Index
When
Optional
Sessions
Documents
Ressources
http://10.0.0.50/demo/mail/index.html

Statistics www.w3perl.com
01 Aug 2008 / 28 Oct 2008 - 89 days
89998 pages / 496868 requests / 8.5 Gb from 35180 hosts in 184 countries
Last update : 29 Oct 2008 - 21:06:18

Sessions - 8 minutes

(Only hosts with at least 10 HTML pages will be printed)

89.129.97.250 : 10 accesses
23/Oct/2008:18:35:36
from http://www.thefreecountry.com/webmaster/loganalyzers.shtml

Reading time (minutes)	Pages read (chronological order)
5' 02	**W3Perl - A free logfile analyzer**
0' 22	/demo/index.html
0' 32	/demo/web/index.html
0' 35	/demo/web/uk/site/index.html
0' 02	/demo/mail/index.html
0' 00	/demo/mail/uk/frame/top.html
0' 00	/demo/mail/uk/frame/index2.html
0' 05	/demo/mail/uk/frame/bottom.html
0' 36	/demo/mail/uk/page/index.html
0	**W3Perl Softs - Docs - Requirements (2)**

121.red-88-7-45.staticip.rima-tde.net : 25 accesses
23/Oct/2008:13:46:55

Reading time (minutes)	Pages read (chronological order)
0' 11	**W3Perl - A free logfile analyzer**
0' 18	/demo/index.html
0' 29	/demo/web/index.html
0' 46	/demo/web/uk/date/now.html
0' 02	/demo/ftp/index.html
0' 00	/demo/ftp/uk/frame/top.html
0' 00	/demo/ftp/uk/frame/index2.html
0' 07	/demo/ftp/uk/frame/bottom.html
0' 19	/demo/ftp/uk/site/index.html
0' 02	/demo/squid/index.html
0' 00	/demo/squid/uk/frame/top.html
0' 00	/demo/squid/uk/frame/index2.html
0' 05	/demo/squid/uk/frame/bottom.html
0' 17	/demo/web/uk/site/index.html
0' 12	/demo/web/uk/script/index.html
1' 06	/demo/web/uk/list/traffic.html
0' 20	/demo/web/uk/session/visite.html
0' 04	/demo/web/uk/date/2008-Oct-17/session.html
0' 08	/demo/web/uk/date/2008-Oct-17/list.html

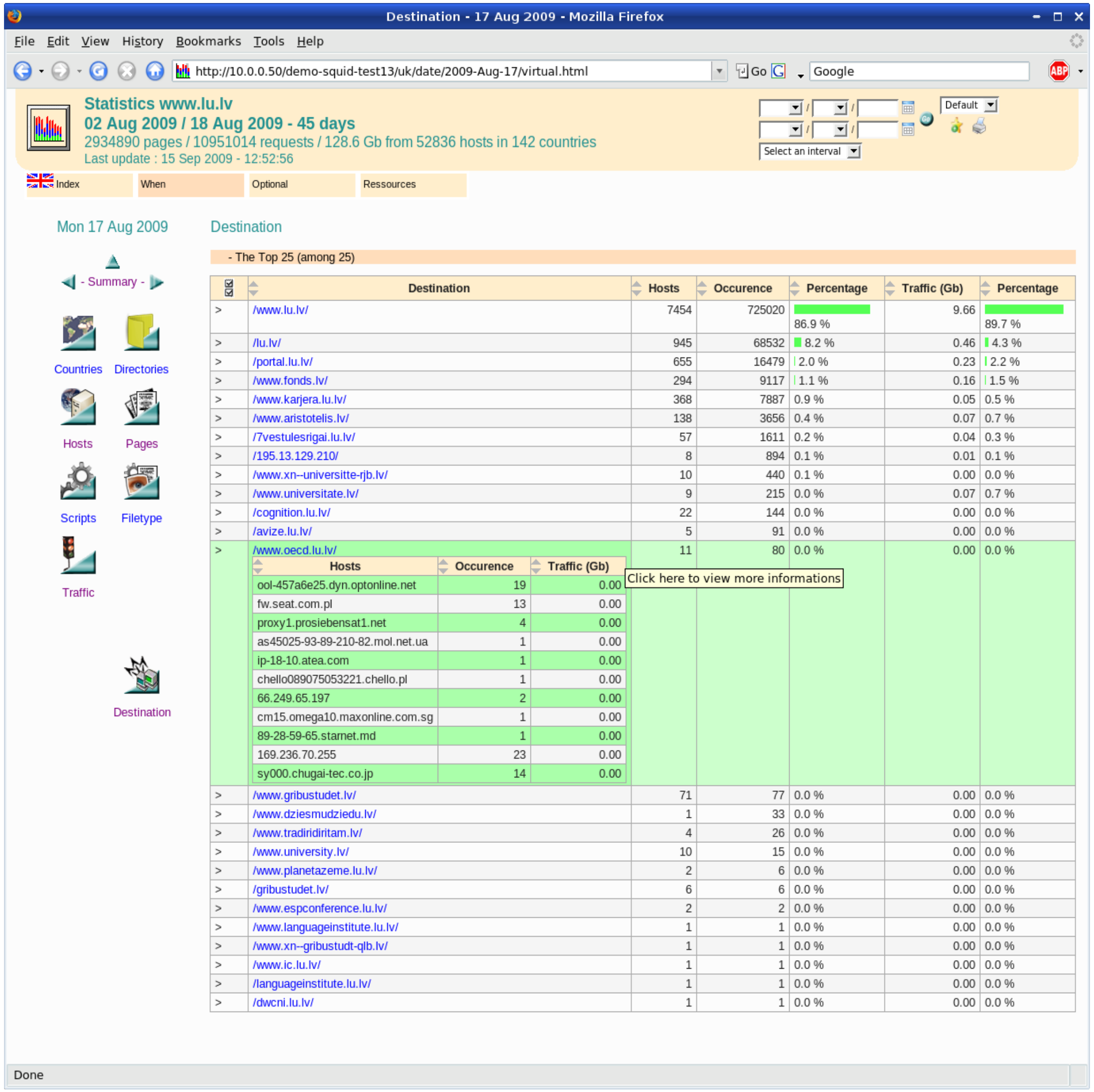
Squid / FTP / Mail

Add some extra reports

- Squid
  - Native and CLF format
  - Basic proxy stats (TCP status code, Traffic, elapsed time)
  - Destination sorted by countries
  - Users stats if authentified users found
- FTP
  - ProFTP, xferlog format
  - Users stats (files downloaded, traffic, from, date)
  - External / Local Top ten users
  - Incoming or Outgoing reports
- Mail
  - Can read Exim / Sendmail / Postfix
  - Emails list per users / domain / countries
  - Local / External server

Intranet

- Can map users to IP
  - Fill the resolv_users.csv
- Local domain
  - Regular expression to map domain IP
  - Parameter to exclude external hosts
- Works with secure server


Destination - 17 Aug 2009 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://10.0.0.50/demo-squid-test13/uk/date/2009-Aug-17/virtual.html
Go
Google
Default
Select an interval
Index
When
Optional
Ressources
Mon 17 Aug 2009
- Summary -
Countries
Directories
Hosts
Pages
Scripts
Filetype
Traffic
Destination
Done

Statistics www.lu.lv
02 Aug 2009 / 18 Aug 2009 - 45 days
2934890 pages / 10951014 requests / 128.6 Gb from 52836 hosts in 142 countries
Last update : 15 Sep 2009 - 12:52:56

Destination

- The Top 25 (among 25)

☑	Destination	Hosts	Occurence	Percentage	Traffic (Gb)	Percentage
>	/www.lu.lv/	7454	725020	86.9 %	9.66	89.7 %
>	/lu.lv/	945	68532	8.2 %	0.46	4.3 %
>	/portal.lu.lv/	655	16479	2.0 %	0.23	2.2 %
>	/www.fonds.lv/	294	9117	1.1 %	0.16	1.5 %
>	/www.karjera.lu.lv/	368	7887	0.9 %	0.05	0.5 %
>	/www.aristotelis.lv/	138	3656	0.4 %	0.07	0.7 %
>	/7vestulesrigai.lu.lv/	57	1611	0.2 %	0.04	0.3 %
>	/195.13.129.210/	8	894	0.1 %	0.01	0.1 %
>	/www.xn--universitte-rjb.lv/	10	440	0.1 %	0.00	0.0 %
>	/www.universitate.lv/	9	215	0.0 %	0.07	0.7 %
>	/cognition.lu.lv/	22	144	0.0 %	0.00	0.0 %
>	/avize.lu.lv/	5	91	0.0 %	0.00	0.0 %
>	/www.oecd.lu.lv/	11	80	0.0 %	0.00	0.0 %
>	/www.gribustudet.lv/	71	77	0.0 %	0.00	0.0 %
>	/www.dziesmudziedu.lv/	1	33	0.0 %	0.00	0.0 %
>	/www.tradiridiritam.lv/	4	26	0.0 %	0.00	0.0 %
>	/www.university.lv/	10	15	0.0 %	0.00	0.0 %
>	/www.planetazeme.lu.lv/	2	6	0.0 %	0.00	0.0 %
>	/gribustudet.lv/	6	6	0.0 %	0.00	0.0 %
>	/www.espconference.lu.lv/	2	2	0.0 %	0.00	0.0 %
>	/www.languageinstitute.lu.lv/	1	1	0.0 %	0.00	0.0 %
>	/www.xn--gribustudt-qlb.lv/	1	1	0.0 %	0.00	0.0 %
>	/www.ic.lu.lv/	1	1	0.0 %	0.00	0.0 %
>	/languageinstitute.lu.lv/	1	1	0.0 %	0.00	0.0 %
>	/dwcni.lu.lv/	1	1	0.0 %	0.00	0.0 %

Hosts	Occurence	Traffic (Gb)
ool-457a6e25.dyn.optonline.net	19	0.00
fw.seat.com.pl	13	0.00
proxy1.prosiebensat1.net	4	0.00
as45025-93-89-210-82.mol.net.ua	1	0.00
ip-18-10.atea.com	1	0.00
chello089075053221.chello.pl	1	0.00
66.249.65.197	2	0.00
cm15.omega10.maxonline.com.sg	1	0.00
89-28-59-65.starnet.md	1	0.00
169.236.70.255	23	0.00
sy000.chugai-tec.co.jp	14	0.00

Click here to view more informations

Squid report showing :

- Daily destination : which websites have been seen

- List of hosts for each destination : occurrence and traffic

Futur developements

- More logfile format support
  - PureFTP, Domino, Firewall …
- Web service
- More AJAX tools
  - Dynamic graphs
- More administration tools
  - Referer spammer management
- More tools for data-mining
- More mail reports
  - spam detection, bounced emails, delay
- Improve Windows installer
- Increase processing speed

Others packages

Logfile analysis

- AWStats
  - Widely used
  - No user’s path track
- Analog
  - Fastest in the world
  - Lack of features
- Webalizer
  - Only small updates from 2002
  - Fast
- AWFFull
  - Webalizer fork
- Visitors
  - Fast
  - Command line based

Page tagging

- Piwik
  - Require a database
  - Javascript code to include
- BBClone
  - PHP tag to include
- CrawTrack
  - Crawler stats
  - Require a database
- Google Analytics
  - No access to data
  - Popular

That’s all

- W3Perl homepage
  - http://www.w3perl.com
- New release every 3 months
- Demo : http://www.w3perl.com/demo/
  - Watch by yourself
- Feedbacks welcome
- Help available
  - FAQ
  - Forum
- Mailing list
  - Only new release update
- Comments ?
  - Send them to domisse@w3perl.com