

WebTracker by Matthai, v. 1.1

Copyright (C) 2002 Matej Kovačič

About the program

WebTracker is an advanced counter of website visits and visitors.

It tries to identify users which are periodically coming back to the website using cookies. Unique cookie ID's are stored in the database. It also saves information about all opened sessions. For each session it saves IP address, hostname, browser type, etc. In the third database is stored information about visited location and in the last database information about robots visiting your website.

Version 1.1 is able to handle data from WebTracker utilities. See a list of utilities at WebTracker website.

All these data could be exported to the tab delimited text files using export modul. More about the databases see in the chapter "*Tables description*" at the end of this document. Finally those data could be imported to the SPSS using SPSS syntax file.

Upgrade from previous version

If you want to upgrade this software with a newer version just replace all the files in the package with the new ones.

In version 1.1 only *wlexport.php* and *webtracker.sps* were changed, so you can replace only this two files.

This package includes

WebTracker package is available in ZIP format. ZIP file includes:

- *webtracker.php* - a main script;
- *webtracker.sql* - script to create SQL tables;
- *webtracker.sps* - script to import data to SPSS;
- *webtracker.pdf* - program manual in PDF format;
- *install.txt* - program manual in TXT format;
- *index.php* - example file, which shows how to include this program into your website;
- *export/wlexport.php* - data export modul;
- *export/cookies.dat* - empty file where data will be exported;
- *export/sessions.dat* - empty file where data will be exported;
- *export/moving.dat* - empty file where data will be exported;
- *export/robots.dat* - empty file where data will be exported.

License

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE

PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

Support

Since this program is a free software I do not offer any technical support.

Latest version

The latest version of the software could be downloaded from my website:
<http://www.ljudmila.org/matej/webtracker>

About the author

See my website at <http://www.ljudmila.org/matej> if you want to know more about me.

Requirements

This program is intended to be installed to the web site. WebTracker has been designed to use PHP with MySQL. Whatever webserver you use (Apache, IIS, etc.) PHP and MySQL must be installed. It is always recommended to use the latest stable version of all the software.

Security issues

It is up to you to properly secure your database, web server and host settings. If you are looking for information in securing your system some good starting points are:

- <http://www.php.net>
- <http://www.mysql.com>
- <http://www.apache.org>
- <http://www.devshed.com>
- <http://www.securityfocus.com>

Special thanks

I want to send out thanks to users of Slo-Tech forum (<http://www.slo-tech.com>), who helped me with many useful suggestions.

Installation

1. Open webtracker.php with some text editor, for instance with TextPad (<http://www.textpad.com>) and edit configuration.
First you can set variable `$forbidden_browsers` - browsers which are not logged in the database. You can put there webbots and spiders - in that case you will get only real users. Then you should set your MySQL database settings - database name, username and host.
Finally you can set how long does cookie lives - this is actually expiration

time of a cookie. By default it is one year. If your cookie lives one year, you will be able to recognize a user who returns to your website in that amount of time. If user visits your website within one year, cookie life will be automatically prolonged.

2. Copy webtracker.php script to your server.
3. Create MySQL database. Fill it with three tables using webcount.sql. You can do this using PHPMyAdmin web script or in command line:
 - `mysqladmin create yourdatabase`
 - `mysql yourdatabase < webtracker.sql`
4. Include this line to each of your PHP scripts:
`include_once("webtracker.php");`
5. Open wlexport.php with some text editor and edit configuration. You should set settings to connect to the MySQL database. It is also highly recommended to change administrator's password. Default is *test*. If you intend to change *\$robotsfile* variable, it *should not be changed to robots.txt*, since this is in general a special file which contains information for web bots. Web bots could be misled if they find file with this filename in your system containing other data than expected.
6. Copy wlexport.php script to your server (you can also copy it in some subdirectory).
7. Copy cookies.dat, moving.dat, sessions.dat and robotvisits.dat to the same subdirectory you placed wlexport script. *Make them writable*. In Unix environment you can use command 'chmod 666 *.dat'.
8. Please note: after exporting data it is highly recommended to initialize files with exported data. Use appropriate option in the wlexport script.
9. After downloading files with exported data use webtracker.sps script to import data to SPSS.

Tables description

Data are written in four MySQL tables.

Table **wt_cookieident** contains the information about all cookies sent to your web site users, and has the following variables:

- *cookieid*: unique cookie ID. Actually it is session ID at the first visit of specified user;
- *num_visits*: number of visits. When user with the cookie in the database comes back, this counter is increased;
- *firstlogindatetime*: date and time of user's first visit;
- *nocookie*: if user do not accept the cookies, this is set to 1. This could be identified only if user do not block session cookies. When user visit the site, WebTracker establishes session and tries to send him a cookie. If user blocks the cookie, and tries to reload a page or clicks to a file which also has WebTracker included, WebTracker finds out that user blocked the cookie in the previous step;
- *cookieerror*: if user edit cookie ID on his computer (that means s/he spoils the cookie), this variable is set to 1. This variable will also be set to 1 if you delete all values from the table. In that case WebTracker will found that

user already has a cookie, but he will not be able to find the cookie in the database.

Table **wt_sessionident** contains the information about all sessions and has the following variables:

- *sessid*: unique session ID. Unfortunately it seems that PHP sometimes has problems with session variables and it is possible that this session ID could be written more times in the database. WebTracker is aware of this problem and is solving it;
- *cookieid*: cookie ID, the same as in table wt_cookieident;
- *IP*: user's IP address;
- *host*: host name of the user. Some web servers do not offer this information because system administrators are trying to reduce the load on the server. Your web server administrator should check if NO_DNS_HOSTNAMES was defined in config.h before compilation;
- *referer*: HTTP referer, i. e. where user come to your site (from which link);
- *browser*: browser's type. That information can reveal you also which operating system is using the visitor of your web site.
- *acclang*: contains the contents of any "Accept-Language:" headers supplied by the visitor's client;
- *sessionstartdatetime*: date and time of session start;
- *sessappear*: number of session appears. Usually it should be 1, but as mentioned,

PHP has some problems with session variables. So WebTracker checks if session ID is already in the database and if it is, it just increase this counter.

Table **wt_moving** contains the information about visited subpages and has the following variables:

- *sessid*: session ID, the same as in table wt_sessionident;
- *location*: visited location (i. e. subpage of your website);
- *accessdatetime*: date and time of the access.

Table **wt_robots** contains the information about robot's visits (robots (also known as spiders or bots) are set in the \$forbidden_browsers variable in webtracker.php) and has the following variables:

- *robotname*: name of the robot, truncated to 80 characters max. Contains the contents of the "User-Agent:" header supplied by the client;
- *firstIP*: IP of the robot. It could be possible that the same robot is runned from several hosts, but only the IP of the first visit is recorded;
- *firsthost*: host name of the robot's server;
- *robotcounter*: number of robot's visits.

How to use a program

After correct installation your only care will be how to export and analyse collected data.

Exporting data

You can export data using wtexport modul. Simply open wtexport.php script in your web browser (<http://www.yourdomain.org/export/wtexport.php>) and type your password. After that wtexport modul will export all four tables in tab-delimited format. You can download them by clicking on them. After

downloading, it is highly recommended to click to the "initialize link". Initializing means that files with exported data are deleted. That is good for security reasons and also to save space on your webserver.

Importing data to SPSS

When you have data in tab-delimited text files, you can import them to SPSS using *webtracker.sps* script. SPSS is a statistical package mostly used in social sciences. You can also use other statistical programs, even Excel, but importing your data to that programs is up to you.

Webtracker.sps script imports data, equip them with labels and creates SPSS data files. Cookies, sessions and movings are also merged to one big database. Analysis is, again, up to you. It depends of a level of your statistical knowledge.

Future features

I found WebTracker data useful for some other types of analyses, for instance network analysis. I am preparing a special modul to export data to Pajek format to make some cute network analysis on the data. I am also preparing a program for analysing web sites and this could be somehow connected with WebTracker data.

It is also possible to connect WebTracker with web survey system, for instance with a system used by RIS project (<http://www.ris.org>). Modul doing this also needs to be prepared.