

Example of an Analysis with the Scientific LogAnalyzer

by

Jonas Christen (†) and Ulf-Dietrich Reips

Psychology Department
University of Zurich
2005

updated by Ulf-Dietrich Reips
Universidad de Deusto
2011 and 2012

A complementary practical reading addition to:

Reips, U.-D., & Stieger, S. (2004). Scientific LogAnalyzer: A Web-based tool for analyses of server log files in psychological research. *Behavior Research Methods, Instruments, & Computers*, 36(2), 304-311.

Example of an Analysis with the Scientific LogAnalyzer

-> For more detailed information refer to the online help pages within the Scientific LogAnalyzer. Orange colored words are links and lead directly to the respective help page.

This text will guide you through an analysis with the Scientific LogAnalyzer. It includes all the details you need. You can find the Scientific LogAnalyzer at <http://www.sclog.eu>.

1. Your first step is to enter login information. If you do not have an account of your own, use "analyzer" as userID and as password. *Students in courses held by U. Reips will be provided with a specific account that offers access.*

After having conducted an experiment on the Internet a logfile with the participants' data has been created (see <http://wextor.org/wextor/en/faq.php> for an explanation of how to configure a Web server in a way that it produces an appropriate logfile). You have to link your logfile to the Scientific LogAnalyzer by either putting the URL in the first field or uploading the file from your local computer. In this example it has been uploaded from a local computer.



Preconditions for the use of Scientific LogAnalyzer:

- Javascript turned on
- Pop-up windows allowed
- a clear brain ;-)

Step: Please enter your ID and password.

If you are a guest user please enter "analyzer" for ID and password.

ID:

Password:

Step: Put in the URL to your logfile:

use this example if you are a guest user: <http://psych-wextor.unizh.ch/loganalyzer/Analyzer5/sdrexperiment.log>

Or upload the logfile from your local computer:

Idea und concept: *Ulf-Dietrich Reips*.

Programming: *Stefan Stieger*

2. In your logfile the data can be separated differently. Therefore you have to define how the columns are separated. In case of uncertainty you can use the "Separator Seeker" to find out which character has been used most frequently.

Once you know the separating character you select it from the list and continue.

- ! Students in courses held by U. Reips only need to check the default option (Webstar format) and can skip Steps 3 and 4.



Step: What separates the columns from each other?

- tabulator
 semicolon
 comma
 blank
 other character:

or is it one of the following predefined formats:

- Web Experimental Psychology Lab - Webstar (default)

go on

Step: If you are not sure, use our "Separator Seeker" (no guarantee)

seek

Idea und concept: *Ulf-Dietrich Reips.*

Programming: *Stefan Stieger*

3. In the next step you have to specify the positions of the relevant columns. The Scientific LogAnalyzer lists all the columns found, and your task is to attribute them to the
- Date (in format 12/03/05)
 - Time (e.g. 12:10:58)
 - IP address (e.g. 130.60.239.96)
 - URL (e.g. 35885:bbna:wilk7hbb.html)
 - Arguments (e.g. age=10&research=0&exper=0)
 - Browser (Firefox, Safari, Netscape, Internet Explorer etc.).
- You find a choice of the first 50 entries of your logfile's columns in the drop-down menus, which helps you to decide correctly. Note that there are several columns without any use in the analysis. An example is the first number below "1226".



Step: Please select the following columns.

In the column Nr. 1 of the logfile we found:

1226

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 2 of the logfile we found:

06/15/04

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 3 of the logfile we found:

22:42:48

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 4 of the logfile we found:

OK

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 5 of the logfile we found:

62.167.71.203

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 6 of the logfile we found:

.final_stud_exp:TheGlasses:scripts.js

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 7 of the logfile we found:

2204

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 8 of the logfile we found:

Mozilla/4.0 (compatible; MSIE 5.01; Windows NT 5.0)

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 9 of the logfile we found:

http://www.genpsylab.unizh.ch/final_stud_exp/TheGlasses/index.html

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 10 of the logfile we found:

1

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 11 of the logfile we found:

id=15161&vi_12=1&A2VL=0&var=0&i=10&rt_12=4497

What's its content: Date Time IP-adress URL Arguments Browser/System

In the column Nr. 12 of the logfile we found:

What's its content: Date Time IP-adress URL Arguments Browser/System

submit

Idea und concept: *Ulf-Dietrich Reips.*

Programmierung: *Stefan Stieger*

4. After having re-checked your choice you have to accept it in order to go on.
On this page you can also define the format of the date column.



Step: Your choice.

Date is in column 1

06/15/04

Time is in column 2

20:23:54

IP-Adress is in column 4

212.4.86.54

URL is in column 5

:final_stud_exp:TheGlasses:2ca15:Bild1VLa3.html

Arguments are in column 10

id=67867&vi_8=1&var=0&i=6&rt_8=5687

Browser/systeminfo are in column 7

Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)

Step: Format of column date:

(d = day; m = month; y = year)

mm / dd / yy submit

Idea und concept: *Ulf-Dietrich Reips.*

Programming: *Stefan Stieger*

- In case you have used pictures in your experiment and don't need their entries, the logfile can be cleaned from these unnecessary lines before you start analyzing. This improves the speed needed for the analysis. Use the button "clean the logfile" to choose this option. If there are several experimental data sets listed in your logfile you have to specify the name or the code number of the chosen experiment. Even more important is the number of experimental factors. In this example there is only one factor. To have a look at the optimized file you can use "preview".



Step: If you want to clean up the logfile from e.g. gif, jpg, ...

Step: If you want to preview the loaded logfile

Step: Experimental conditions

number of the experiment (if available):

number of factors:

Idea und concept: Ulf-Dietrich Reips.

Programming: Stefan Stieger

- For every factor you entered you have to define the number of levels. Furthermore you can optionally name the factors and specify a search string. More details are needed in the next step.



Step: Experimental conditions

Name of the factor	search string (no entry necessary)	Number of levels
<input type="text" value="Referenzglas"/>	<input type="text"/>	<input type="text" value="3"/>

Idea und concept: Ulf-Dietrich Reips.

Programming: Stefan Stieger

- The Scientific LogAnalyzer offers the possibility to mark certain IP or domain name (DN) addresses. For instance, you may later want to exclude participation from your own computer (testing your own materials) or any other undesired entries. You can also use the so-called "master list" with addresses that are updated by the Scientific LogAnalyzer administrator. You find this list on this same page further down.

To optimize your logfile you can exclude some unnecessary arguments from the file using "decline args". Remember to save your choice before closing the window.

- !** **Session management:** If you created your Web experiment with WEXTOR or used another way of identifying lines from the same user session by search argument, then click the second option: "Session separated through a search argument". This option will later save you much manual work of clarifying cases in the output file. WEXTOR-based Web

experiments use “?id” as the default search argument for session identification. Chose this option from the drop-down menu. If your logfile doesn't contain session IDs, then you will have to use the first option. Scientific LogAnalyzer will use the IP addresses and other information (operating system, Web browser, proximity in time) to identify lines that belong to the same user session. The session timeout defines the amount of time you allow for a break. You may change the standard setting of 15 minutes. Scientific LogAnalyzer will treat groups of lines coming from the same IP as separate sessions if the time between these groups of lines is longer than the time specified.

Speed-up: To increase the speed of your analysis you can check the box “don't make a dnslookup”, unless you need it for your experiment. In case of reloads or multiple submissions of the same page within your defined session time you have to decide whether you want to use the first or last value for arguments. In case of multiple submissions you will find a “1” for the participants concerned in the column “danger-args” in the output-file.

Search strings that define levels: To enable the analysis you have to provide the search-strings that define the levels of your experiment factors. Most important are the (first) distance and the search-string itself. The prefix and the suffix offer additional options for definitions, whereas the first distance indicates the number of characters between a possible prefix and the search-string. If there is no prefix you simply have to count the characters before the argument column's search-string. Therefore you need to open your logfile with some kind of text editor. TextWrangler (formerly BBEdit Lite) for Mac is recommended and is a freeware program that can be downloaded from the Internet (<http://www.barebones.com/products/textwrangler/>). UltraEdit is a software (\$60) for Windows with a 30-45 day free trial period available from <http://www.ultraedit.com/>

```

1.203 :final_stud_exp:TheGlasses:3f7d3:Antwor
?1.203 :final_stud_exp:TheGlasses:3f7d3:Intro
?1.203 :final_stud_exp:TheGlasses:3f7d3:Bid1
?1.203 :final_stud_exp:TheGlasses:3f7d3:Passov
?1.203 :final_stud_exp:TheGlasses:3f7d3:Bid2
?1.203 :final_stud_exp:TheGlasses:3f7d3:Antwor
?1.203 :final_stud_exp:TheGlasses:3f7d3:thank
67.71.203 :final_stud_exp:TheGlasses:index.h
67.71.203 :final_stud_exp:TheGlasses:style.cs
67.71.203 :final_stud_exp:TheGlasses:scripts
67.71.203 :final_stud_exp:TheGlasses:start.h
67.71.203 :final_stud_exp:TheGlasses:13d7a:de
67.71.203 :final_stud_exp:TheGlasses:13d7a:lr
67.71.203 :final_stud_exp:TheGlasses:13d7a:B
67.71.203 :final_stud_exp:TheGlasses:13d7a:Pe
67.71.203 :final_stud_exp:TheGlasses:13d7a:B
67.71.203 :final_stud_exp:TheGlasses:13d7a:A

```

Here you have a 27 characters distance between the beginning of the arguments and the search-string “1” (or “3” in the condition displayed in the top portion of the logfile).

The search-string identifies the level. You have to look it up in your logfile, too. In this example there are three levels with three different search-

strings ("1", "2" and "3"). These are the minimal search-strings, you could also use longer ones like "13d7a" as in the Figure below (and even leave out the distance, if the search-strings are unique).

If your experiment is more complicated there are some more sophisticated options concerning search-strings. For more details use the Scientific LogAnalyzer's online help pages.

For those who are interested in an analysis of dropout there is the button "show the paths". This gives you information for example about how many participants (1) chose which route through your study, (2) left your experiment at which point, or (3) the equality of your experimental conditions' random distribution. "Show the paths" is currently dysfunctional, until Stefan Stieger finishes his habilitation and finds some time to fix it.

- ! Once you are ready with your definitions you can start the analysis by clicking on the respective button. **IMPORTANT:** Do not interrupt and restart your analysis! If it happens – quit your Web browser, rename the logfile, and redo the analysis a few hours later.



Step: Ban IP or DN addresses:

Step: You can exclude search arguments from analyzing:

Step: Please choose your output format:

Output format:

You are using the testaccount. Therefore the number of lines is restricted to 2000

Step: Please choose the desired session management:

Session timeout after: seconds (Default 900s = 15 Minutes)

Session separated through a search argument (beta version):

Step: Some additional options:

suppress time-variable output

dont make a dnslookup

Lines in the logfile: 2118

With DN Lookup: approximately 01:13

Without DN Lookup: approximately 00:45

in case of reloads or multiple submissions within the above defined session time

use the first value use the last value

use our IP list use our DN list

Step: Provide the search strings:

Referenzglas ()

prefix	distance	search-string	distance	suffix
<input type="text"/>	<input type="text" value="27"/>	<input type="text" value="2ca15"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text" value="27"/>	<input type="text" value="3f7d3"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text" value="27"/>	<input type="text" value="13d7a"/>	<input type="text"/>	<input type="text"/>

Step: Start analyzing

Step: Show a tree of paths (beta-version):

Idea und concept: *Ulf-Dietrich Reips.*

Programming: *Stefan Stieger*

- You can chose between three different output files. All of them are explained below.
If you want to run another analysis you have to save your output first, and then delete your data on the Scientific LogAnalyzer server by clicking on "press to delete". Otherwise data from different experiments might get confused.

Experiment

[Tabdelimited File](#)

[Infopage](#)

[HTML-table of the output](#)

If you have finished your analysis and if you would like to delete your uploaded Serverlogfile, please press the button bellow.

CAUTION: If you do so you have to login again to analyze.

- Choosing the tabulator separated file you will find your output in a list format similar to the example below. Such a file can easily be imported into Excel, SPSS or any other statistical program. Remember to save this file on your local computer by right-clicking (Windows) or ctrl-clicking (Mac) on the link, and using "save link" (or similar).

factor_1	level_1	danger_1	banned-ip	ip	banned-dn	dn	session	start-date	start time	end-date	end ti	
1	3	0	0	62.2.188.67	0	not determined	1	06/19/04	00:33:43	06/19/04	00:34:41	58
1	1	0	0	217.162.170.111	0	not determined	1	06/17/04	22:24:36	06/17/04	22:26:28	112
1	2	0	0	80.219.208.149	0	not determined	1	06/16/04	14:11:42	06/16/04	14:14:54	192
1	2	0	0	62.202.227.28	0	not determined	1	06/17/04	22:42:09	06/17/04	22:45:04	175
1	3	0	0	146.136.2.187	0	not determined	1	06/17/04	10:18:11	06/17/04	10:19:46	95
1	2	0	0	62.202.31.66	0	not determined	1	06/19/04	09:05:43	06/19/04	09:07:09	86
1	1	0	0	81.62.71.151	0	not determined	1	06/17/04	22:30:28	06/17/04	22:31:40	72
1	1	0	0	146.136.2.97	0	not determined	1	06/17/04	11:05:46	06/17/04	11:08:51	185
1	2	0	0	146.136.2.97	0	not determined	2	06/17/04	13:01:41	06/17/04	13:02:58	77
1	2	0	0	131.152.29.120	0	not determined	1	06/17/04	08:26:27	06/17/04	08:28:50	143
1	3	0	0	217.254.23.223	0	not determined	1	06/18/04	19:00:10	06/18/04	19:01:01	51
1	1	0	0	195.226.8.34	0	not determined	1	06/18/04	16:03:52	06/18/04	16:04:59	67
1	1	0	0	146.136.2.215	0	not determined	1	06/17/04	18:04:43	06/17/04	18:06:17	94
1	1	0	0	213.3.231.127	0	not determined	1	06/17/04	16:03:15	06/17/04	16:06:02	167
1	3	0	0	213.3.254.14	0	not determined	1	06/17/04	07:15:02	06/17/04	07:15:52	50
1	1	0	0	62.202.224.184	0	not determined	1	06/18/04	23:02:16	06/18/04	23:05:59	223
1	3	0	0	146.136.2.203	0	not determined	1	06/17/04	10:21:13	06/17/04	10:22:27	74
1	2	0	0	213.3.186.96	0	not determined	1	06/17/04	10:16:49	06/17/04	10:18:31	102
1	3	0	0	212.152.19.40	0	not determined	1	06/17/04	02:19:37	06/17/04	02:22:06	149

- You should also save the Infopage – for further re-checking of your analysis. It is possible that you later will need the information about your search-strings or some other detail.
Note that test accounts analyze only up to 2000 lines. For larger files you have to apply for a proper account.



Log file analysis for experiment "

Path to the logfile: logfile.txt

Sessionlength: 900 Seconds
Our IP-List: not used
Our DN-List: not used
Suppressed time variable: not used
Dont make a DNS lookup: used
Lines analyzed: 2000
Duration of the analysis (hh:mm:ss): 00:00:13

	präfix	distance	search string	distance	postfix
Factor 1: Referenzglas ()					
	Level 1:	27	2ca15		
	Level 2:	27	3f7d3		
	Level 3:	27	13d7a		

Idea und concept: *Ulf-Dietrich Reips.*
 Programming: *Stefan Stieger*

- The last option presents the output in a browser window. This option is only suitable for small logfiles. For more detailed information and comments about the HTML-table of the output refer to "Table of results 2" in the Scientific LogAnalyzer's online help.

factor_1	level_1	danger_1	banned-ip	ip	banned-dn	dn	session	start-date	start time	end-date	end time	sessio length
1	3	0	0	62.2.188.67	0	not determined	1	06/19/04	00:33:43	06/19/04	00:34:41	58
1	1	0	0	217.162.170.111	0	not determined	1	06/17/04	22:24:36	06/17/04	22:26:28	112
1	2	0	0	80.219.208.149	0	not determined	1	06/16/04	14:11:42	06/16/04	14:14:54	192

You are done!