**Webscavator User Guide**

# User Guide Part 1: Adding, loading and editing cases

## Getting the web history files

Firstly, you must have a program that exports web history files into a text file format. The programs currently supported by Webscavator are:

* Fox Analysis - exports to CSV files
* Net Analysis - exports to CSV files
* Web Historian - exports to XML files
* Pasco - exports to CSV files
* Chrome Cache Viewer - exports to CSV files

You can add more programs by following the guide [here](http://localhost:5000/help/addprograms/) (You'll have to do some Python programming though!).

### Using Fox Analysis with Webscavator

To obtain CSV data from Fox Analysis, go to File > Export To > CSV File. The will produce several files, please pick the one that ends with Website.csv.

### Using Net Analysis with Webscavator

To obtain CSV data from Net Analysis, click on File > Export History As > Tab Delimited Text. Rename the extension to .csv instead of .txt.

Sometimes Net Analysis produces files that do not conform exactly to Tab Delimted Text, and the converter will reject the file. If it is possible to locate the line(s) that cause the error, removing these will allow the file to be accepted.

### Using Web Historian with Webscavator

To obtain XML data from Web Historian DO NOT ask it to save as CSV. Web Historian currently uses a comma as a delimiter, and cannot be parsed correctly by any CSV parser when there are commas in the first field which is 'title' (very likely to occur). Instead save as an XML file.

### Using Pasco with Webscavator

To obtain CSV data from Pasco run:

> pasco index.dat > output.csv

### Using Chrome Cache Viewer with Webscavator

To obtain CSV data from Chrome Cache Viewer, select all the entries, and go to File > Save Selected Items. Save as a Tab-delimited Text File, and then change the extension from .txt to .csv. Make sure the file is in UTF-8 encoding, otherwise the CSV converter will not accept the file. You can do this by opening the CSV file in a program such as Notepad++, and choose Encoding > Encode in UTF-8.

## Starting Webscavator

To start Webscavator type the following command line in the Webscavator directory:

python launch.py runserver

In Windows, if you get the error:

'python' is not recognized as an internal or external command, operable program or batch file.

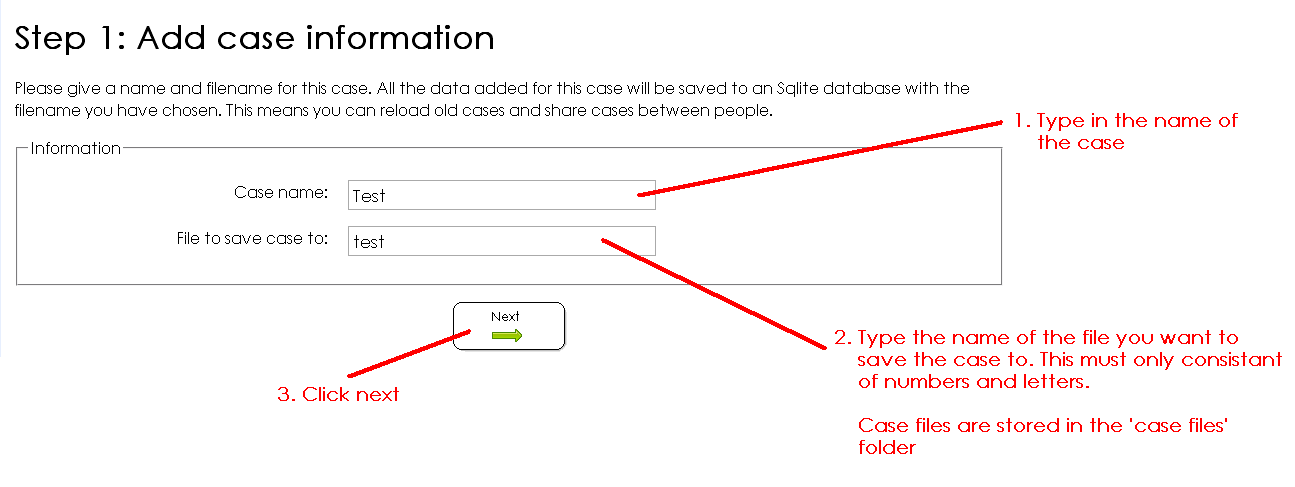
You will need to add python to your system environment %path% variable (this is easy to do, search online for *adding a path to system environment for [your operating system]*).

Finally, go to your favourite web browser (Firefox or Chrome is recommended) and go to <http://localhost:5000>.

## Adding a new case

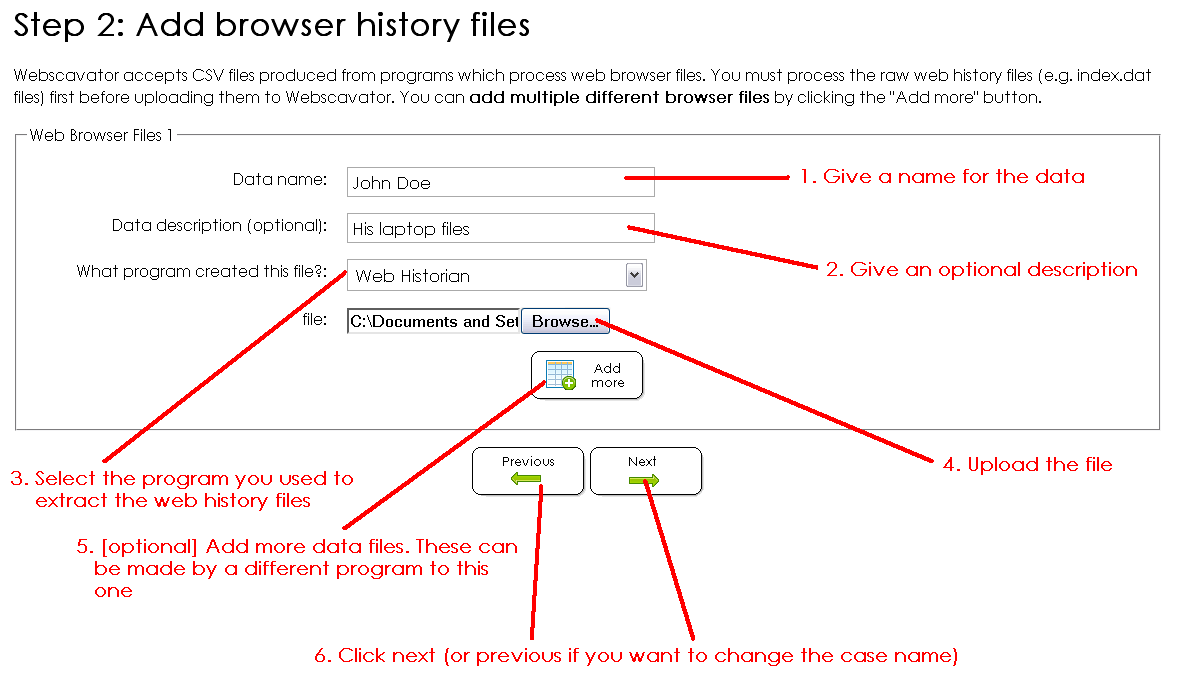
Once you have got your extracted web history files, and started Webscavator, you are ready to make a new case. To add a new case, follow the 'new case' wizard.

### Add case details

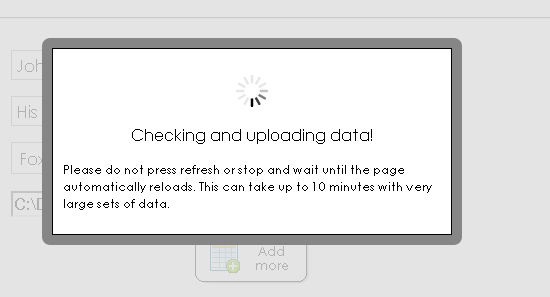


Each case is stored individually in its own Sqlite database in the 'case files' folder of Webscavator. You can store these somewhere else for safe keeping, but Webscavator will only load files from this folder. Having cases stored separately means you can easily share a particular case with a colleague without revealing any other cases.

### Add files

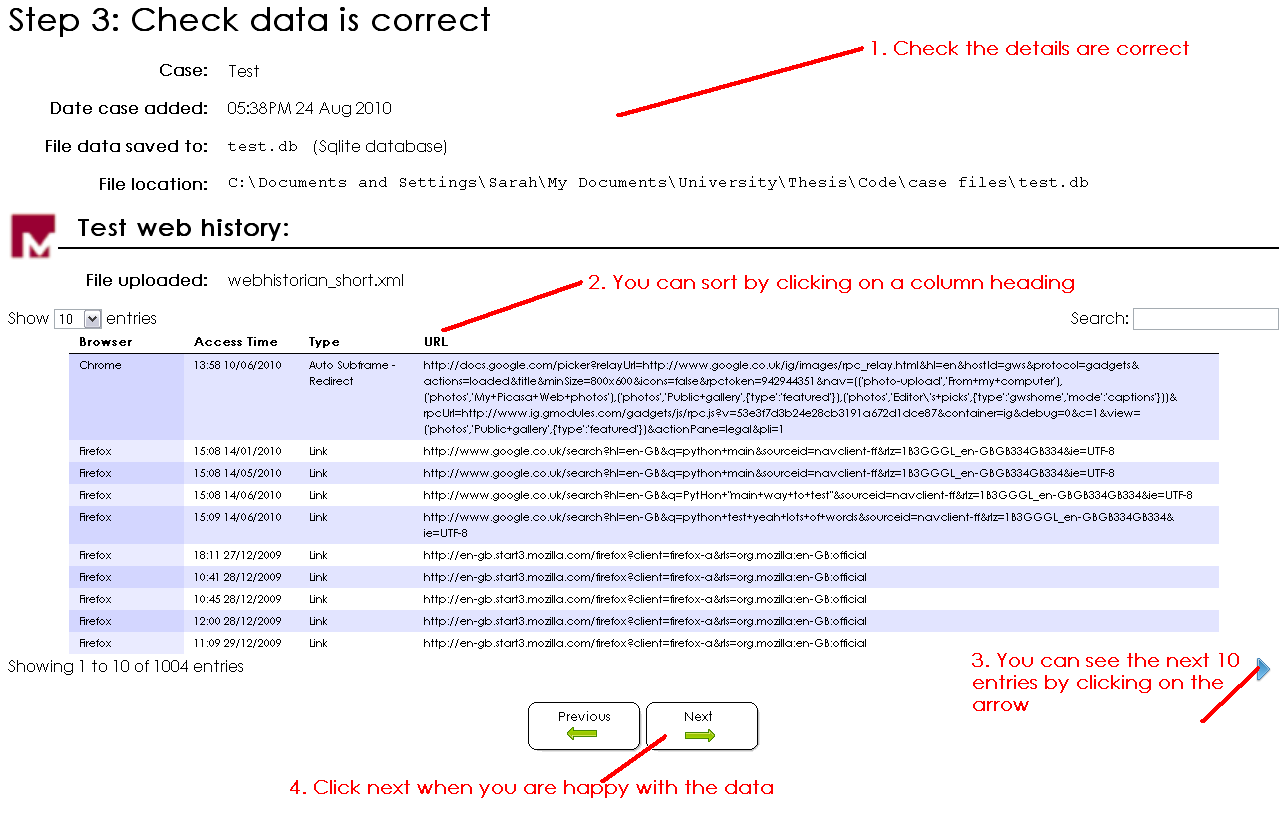


Make sure to select the correct program you used to make the file, otherwise the file will not be uploaded!

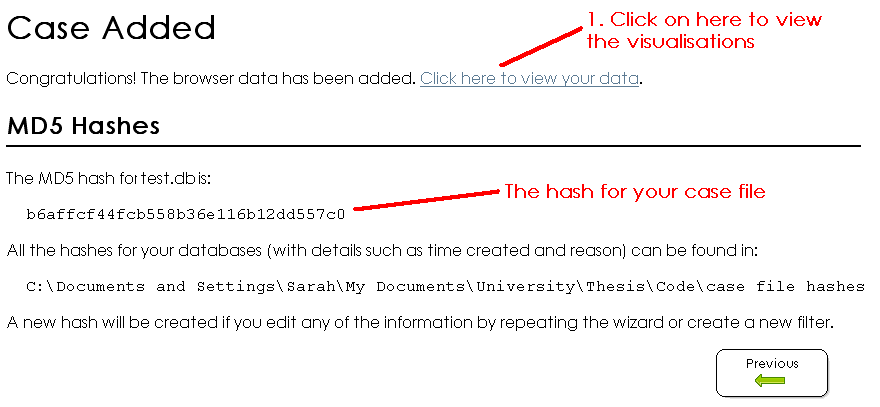


If the file is very large, it may take a long time to add all the data. You can add multiple files to a case if the person you are investigating has used multiple browsers or the program you have used has produced several files as output.

### Check details



### Finish wizard



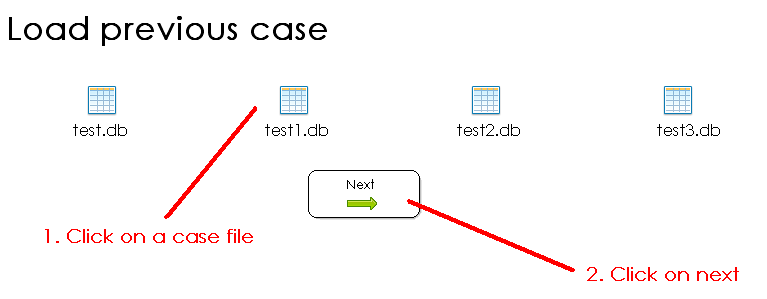
Congratulations, you have added a case. You can now see the visualisations.

### Security

Once you have added, edited or loaded any case an MD5 hash of the Sqlite database is created and stored in ‘case file hashes' folder of Webscavator, in a text file with the same name as the database file. For any subsequent edits, loads or filter additions, this file is updated with new hashes. The integrity of the database file can be checked by computing the MD5 hash of the database and comparing it to the latest entry in its MD5 hash file.

Webscavator has strong input validation. Because it only runs on a local port, the website is not available to anyone on the same network.

## Loading a case



If you have already added a case, you can load it into Webscavator. Make sure the case file is in the 'case files' folder of Webscavator, and then select it in the load case wizard.

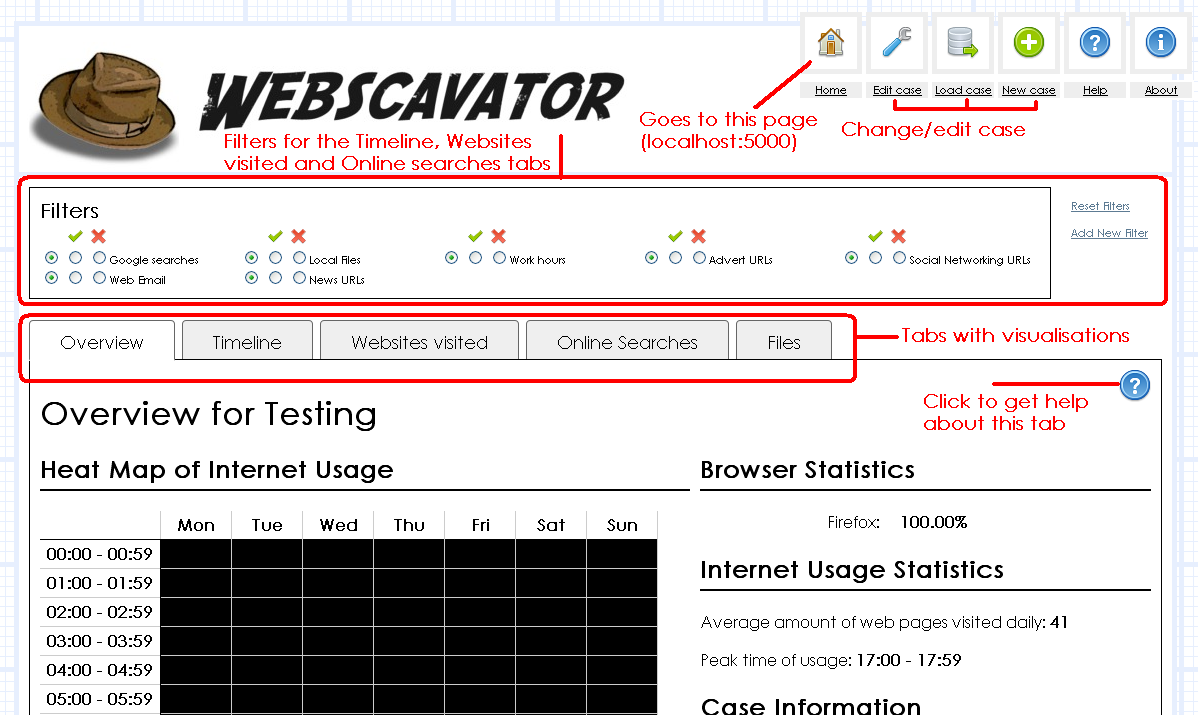
## Editing a case

Once you have added a case or loaded a case, you can edit it by following the same wizard you did to create the case. You can change any of the details, including changing the files. To change the name of the case database, just rename the file in the 'case files' folder (but not whilst Webscavator is using it!).

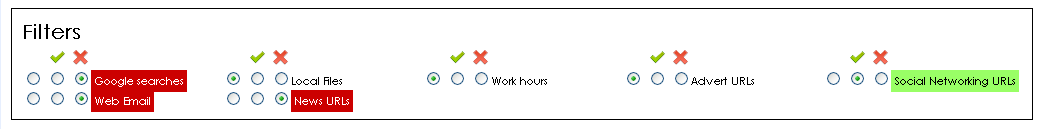
# User Guide Part 2: Using the visualisations

## Overview

The screenshot below shows the homepage of Webscavator once data has been loaded in. You can add, edit or load a case using the buttons in the top right corner. Below the logo is a filter box with a set of prebuilt filters. Below that are the visualisation tabs, with a different visualisation on each. At the top of each tab on the right is a question icon. Clicking this will bring up information about that tab.



## Filters



There are two kinds of filters available: those that remove/hide data from the visualisations and those that highlight/show data. Each filter has three radio buttons associated with it. Clicking on the button underneath the 'tick' will highlight data that corresponds to the filter, and the background colour of the filter will turn green. Clicking the button underneath the 'cross' will remove data that corresponds to the filter, and the background colour of the filter will turn red. The changes in background colour happen so that it is obvious that filters have been applied to the data, as it is not always easy at a glance to see which radio buttons are selected. All filters can be reset by clicking the 'Reset filters' link at the right hand side of the filters.

Removal filters always take preference over highlight filters. E.g. if a web history entry fits both a removal and highlight filter and they are both turned on, the entry will not be displayed as entries are removed first, then the remaining entries are highlighted.

### Adding a new filter

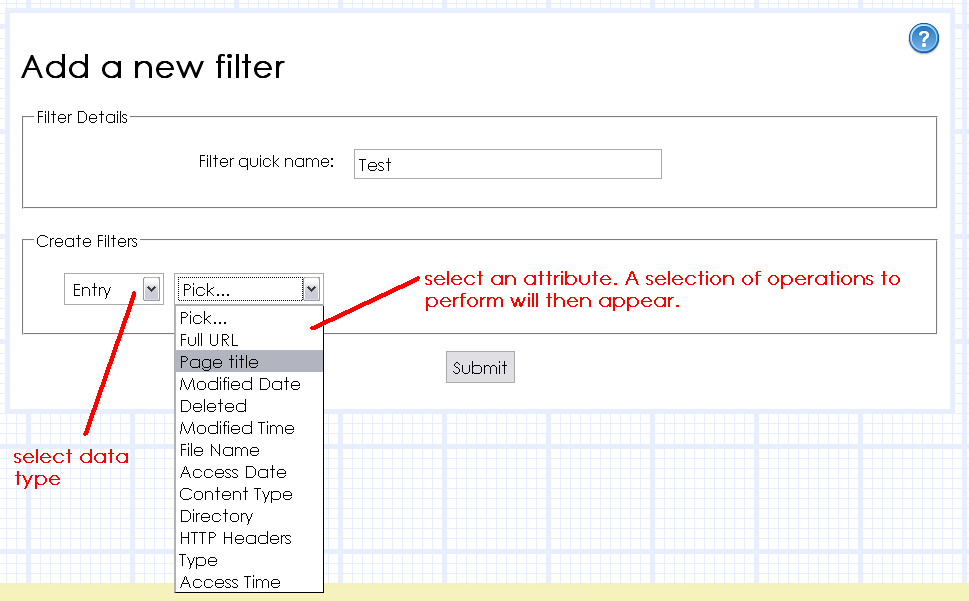
To add a new filter, click on the 'Add New Filter' link on the right of the filter box. This will bring up a pop up box. Enter a name for the filter and then you can choose what to filter on:

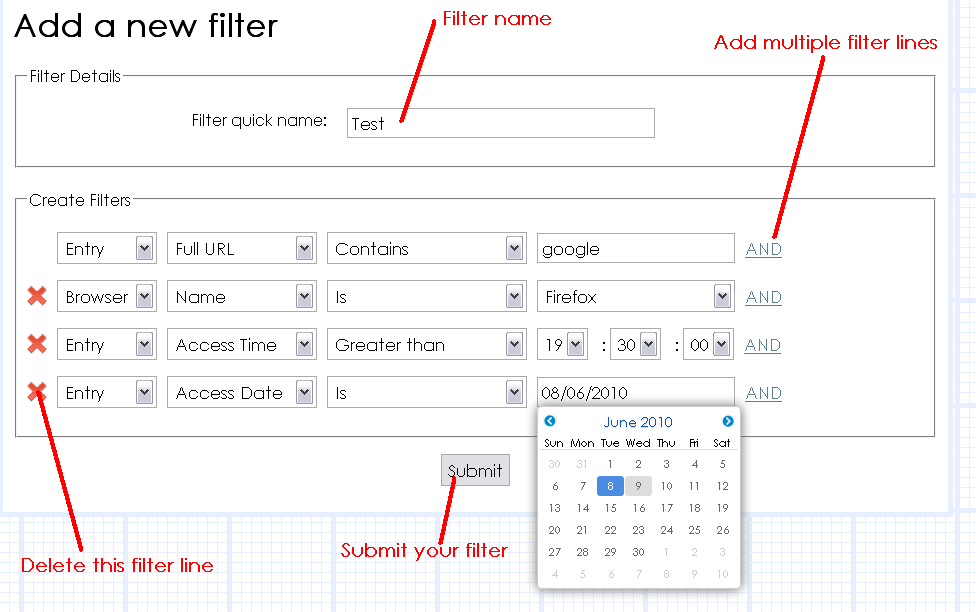
* Entry: can filter on the access date & time, the full URL, page title, HTTP headers, content type etc.
* Search Terms: can filter on the search engine used, search terms and term occurrence.
* URL parts: can filter on the different parts of the URL, e.g. the port, query strings, usernames, passwords etc.
* Web files: can filter on the program used to extract web history or different files uploaded. This is irrelevant if only one file was uploaded.
* Browser: can filter on browser name, version and profile used/IE location.

Some of the filters allow you to enter some text as the value, others you will select from a predefined list (e.g. 'browser name' will list all the browsers available to save you from typing), time values will ask you to select hours, minutes and seconds and date values have a date-picker (see screenshot on next page). The operation 'is in list' or 'not in list' will allow you to select text files located in the 'case lists' folder in Webscavator. These are simple files where each line is a value to compare to, e.g. a list of specific domain names you are interested in. You can add your own lists to this folder, and then select them in the select box. A few lists exist by default including a list of advert domains, social networking domains and online newspapers.

**Note:** If you are adding your own list, please make sure they are less than 450 lines as this is currently the maximum Webscavator allows.

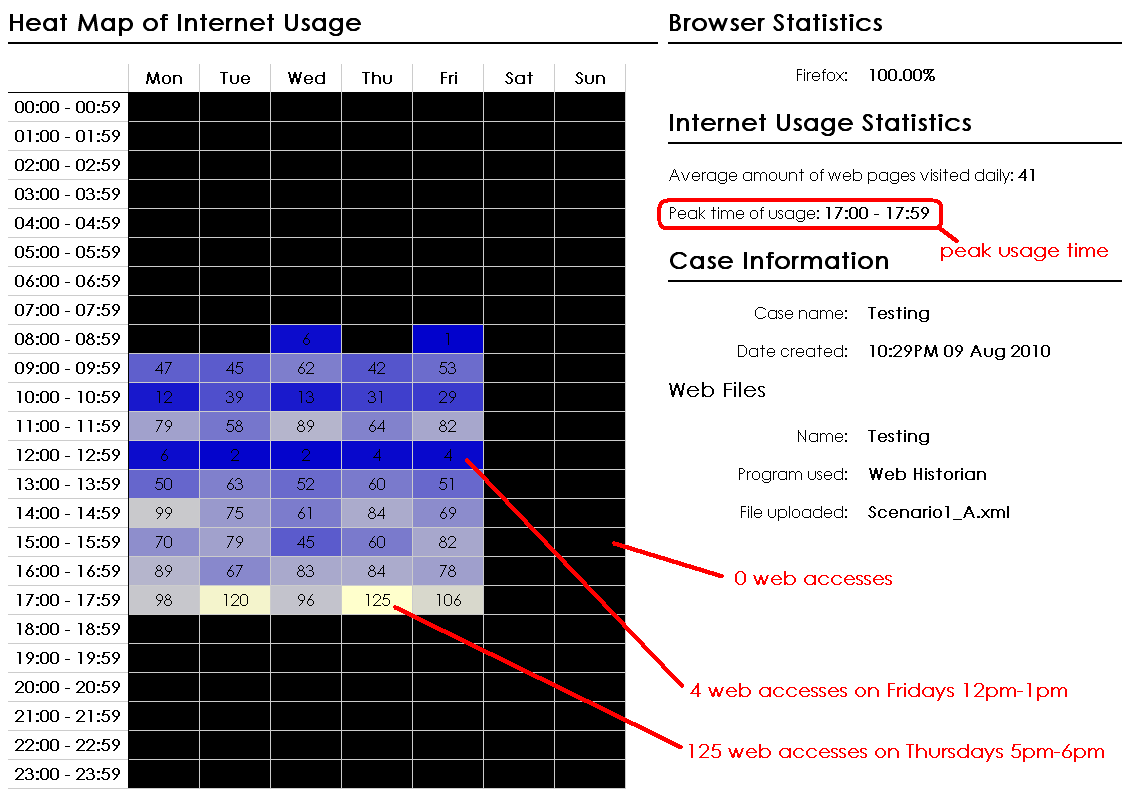
Currently Webscavator does not support the editing and deletion of filters.

[](http://localhost:5000/images/userguide/vis_filter_form_first.png)

[](http://localhost:5000/images/userguide/vis_filter_form.png)

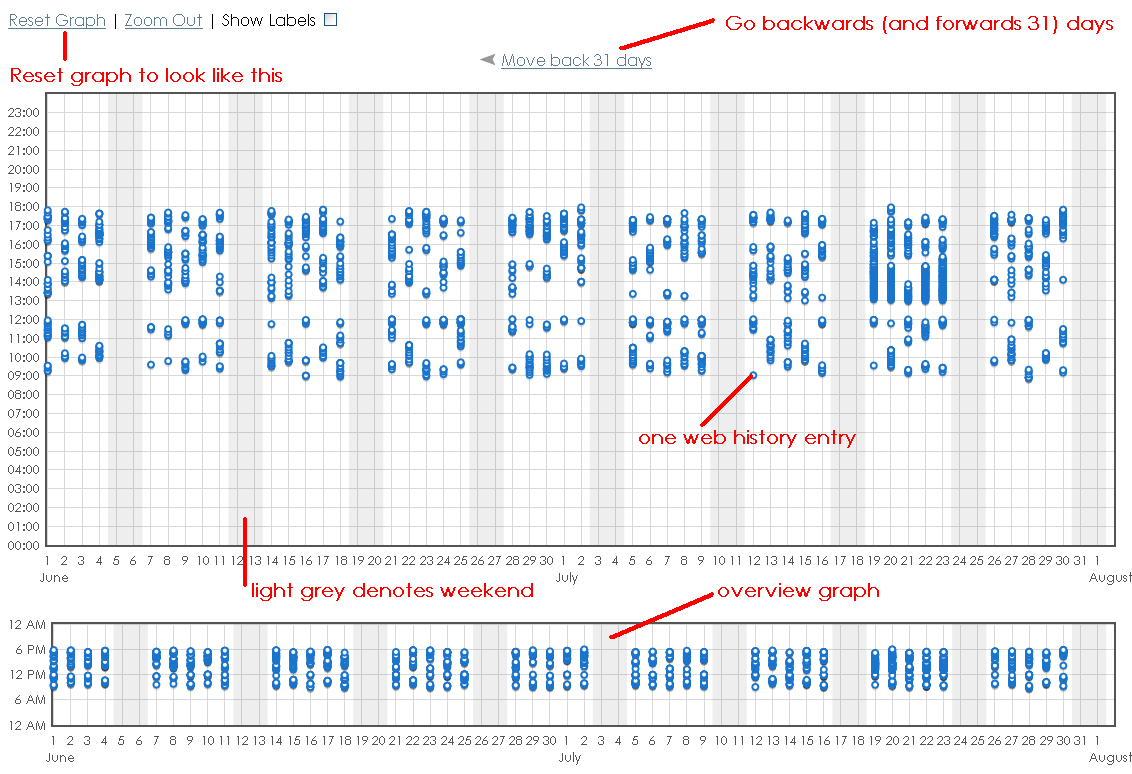
## Overview Tab

The overview tab has a heat-map showing the amalgamated number of entries for each day of the week for each hour. The more entries, the lighter the colour. Black means there are no entries. To the right are some overview statistics and information about the current case.

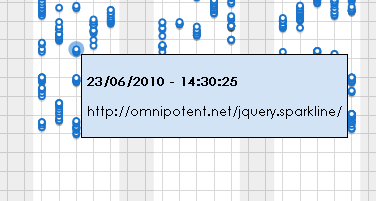
[](http://localhost:5000/images/userguide/vis_overview_tab.png)

## Timeline Tab

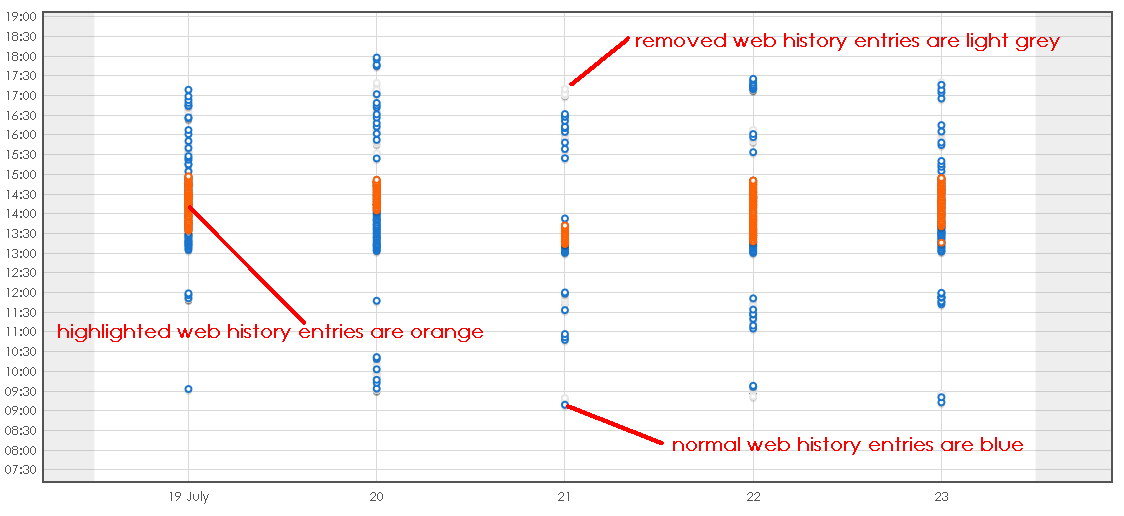
The timeline shows the web history in two graphs. Along the x-axis is the time, and the date along the y-axis. Each point is a web history entry. The smaller, bottom graph is an overview graph, and does not zoom in or change colours like the larger graph. This is to keep a sense of context. To move about, click and drag the top graph around, or click "move back 31 days".

[](http://localhost:5000/images/userguide/vis_timeline_overview.png)

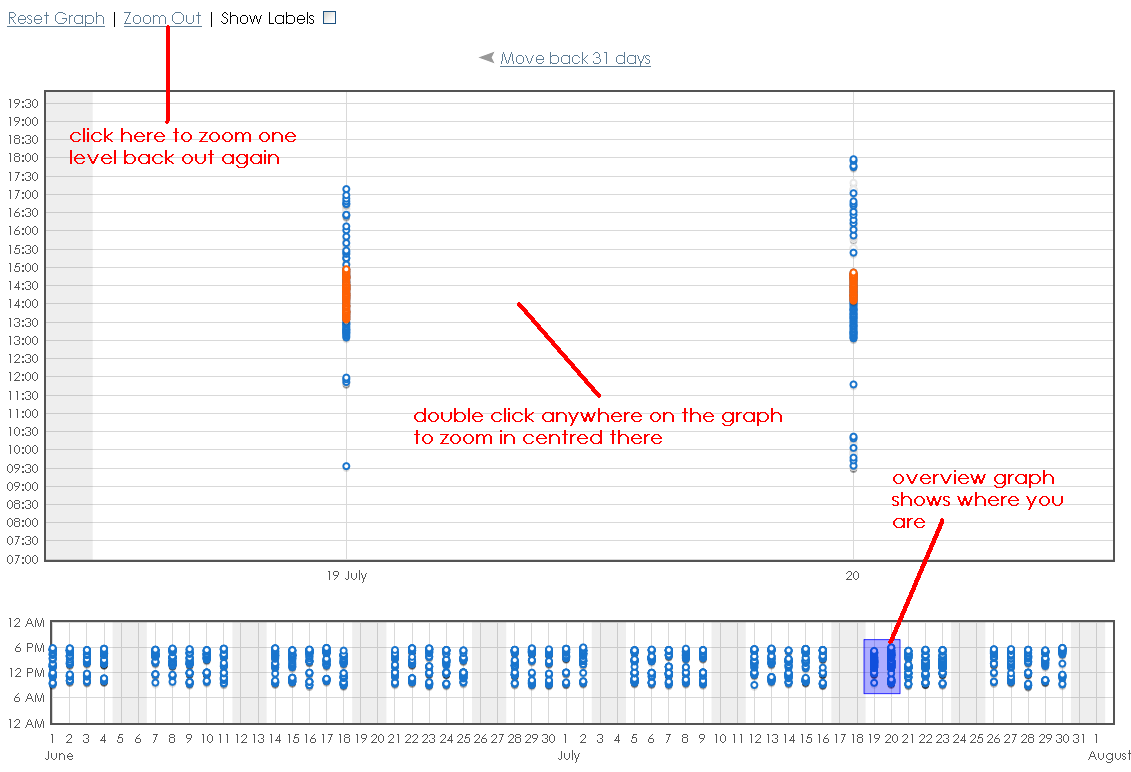
Hovering over a point gives the date, time and URL. Clicking on a point will produce more information below the graphs such as browser, page title etc.

[](http://localhost:5000/images/userguide/vis_timeline_hover.png)

Applying a removal filter will make any points that correspond to that filter turn a very light grey. Points are not completely removed from sight to keep a sense of context. Applying a highlight filter will make any points that correspond to that filter turn orange.

[](http://localhost:5000/images/userguide/vis_timeline_highlight.png)

There are two ways to zoom in. To zoom in slowly, you can double click on the top graph. This will zoom in centred where you clicked. Otherwise, you can drag a blue rectangle over the overview graph, and the top graph will zoom in there. The overview graph will always display a blue rectangle showing where you are zoomed into. You can still pan around by dragging your mouse. Clicking on 'Zoom out' will zoom out a small amount. Clicking on 'Reset graph' will zoom out completely.

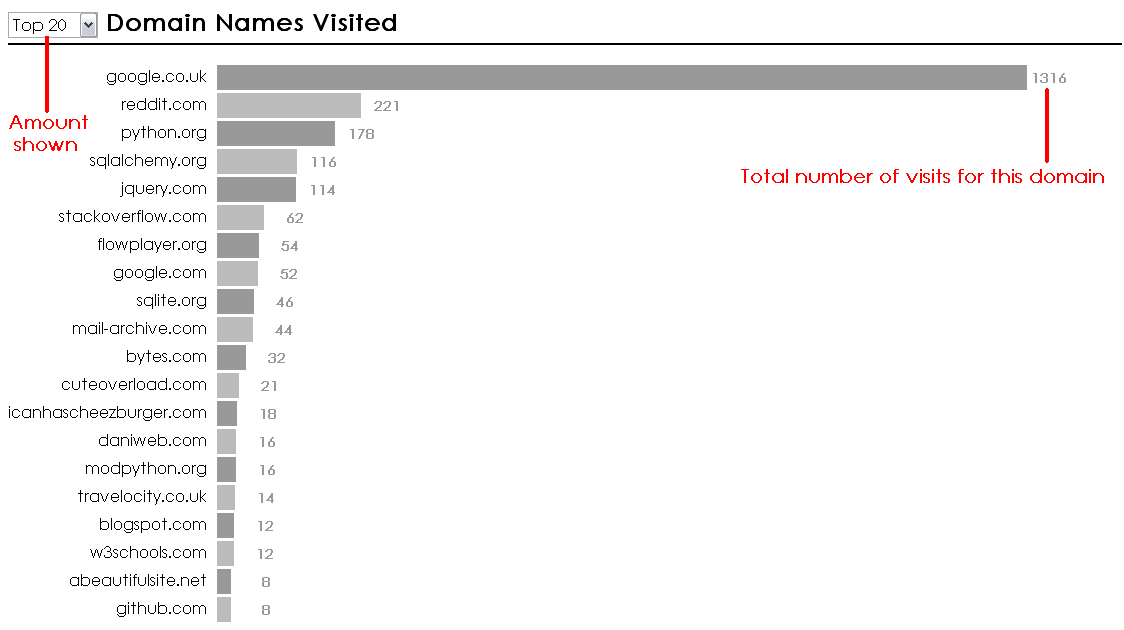
[](http://localhost:5000/images/userguide/vis_timeline_zooming.png)

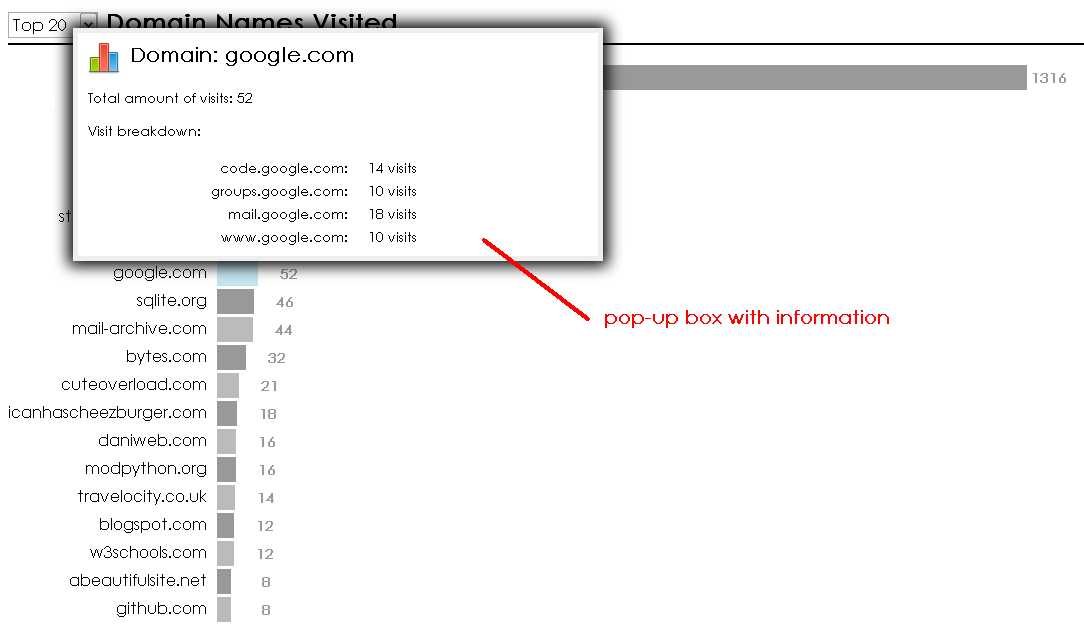
When you are zoomed in, you can check 'Show labels' which will show the URL beside the point.

[](http://localhost:5000/images/userguide/vis_timeline_labels.png)

## Websites visited Tab

This tab has a bar chart of the top domain names visited. You can change the amount viewed from Top 20 to Top 50, Top 100 or All. Clicking on a bar will bring up a pop-up with more information. The bar chart will not show data that corresponds to any removal filters that are active. If any highlight filters are active, only the data corresponding to those will be shown.

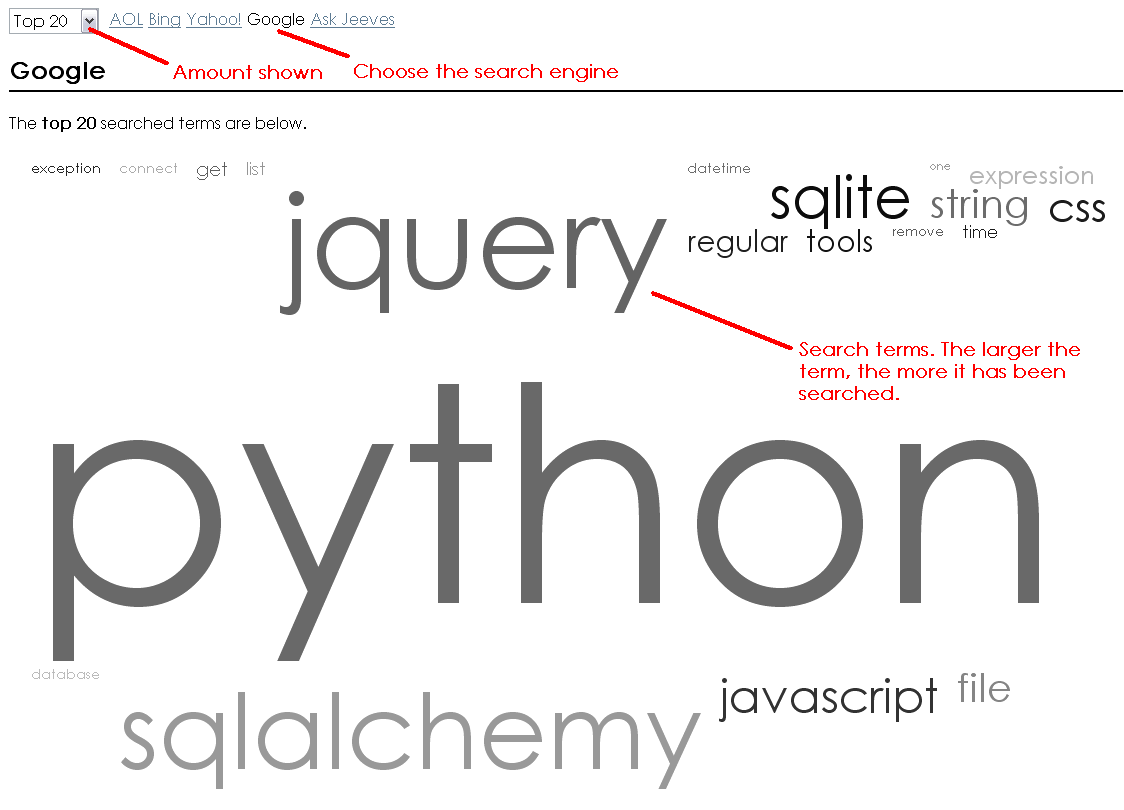
[](http://localhost:5000/images/userguide/vis_domains_tab.png)

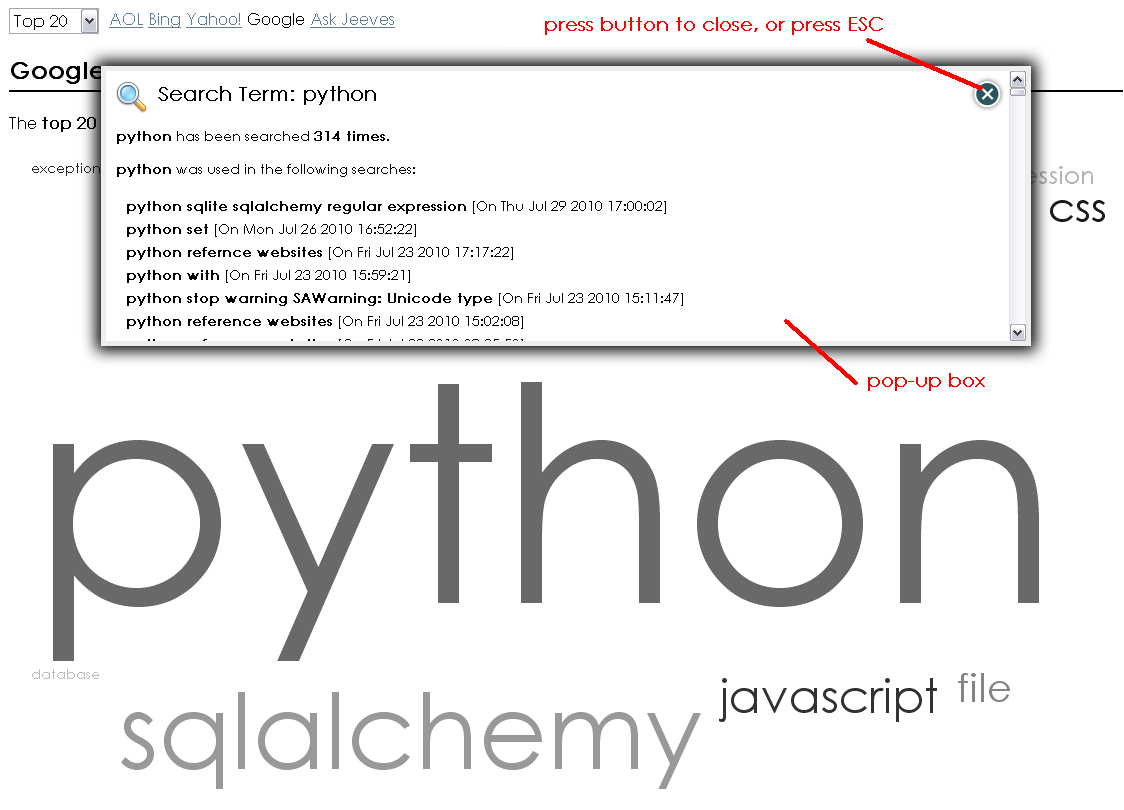
[](http://localhost:5000/images/userguide/vis_domains_tab_popup.png)

## Online Searches Tab

This tab shows the online search terms queried using search engines. Click on a search engine name to see the search terms queried. You can change the amount viewed from Top 20 to Top 50, Top 100 or All. The word cloud will not show data that corresponds to any removal filters that are active. If any highlight filters are active, only the data corresponding to those will be shown.

The larger the word, the more it has been searched for. Clicking on a word will show a pop-up with more details such as all the searches this word appeared in with the date and time.

[](http://localhost:5000/images/userguide/vis_wordcloud.png)

[](http://localhost:5000/images/userguide/vis_wordcloud_popup.png)

### Adding new search engines

It is possible to add more search engines by editing the config file located in the webscavator/config folder. Let's assume a typical search with the new search engine is http://search.foo.com/?terms=search+terms, and this search engine is called "Foo Search". All Foo Searches contain 'search.foo' in the URL, and 'terms' is always followed by the search terms. Add an entry underneath the [search\_engines] and [search] titles like so:

[search\_engines]

search.foo = terms

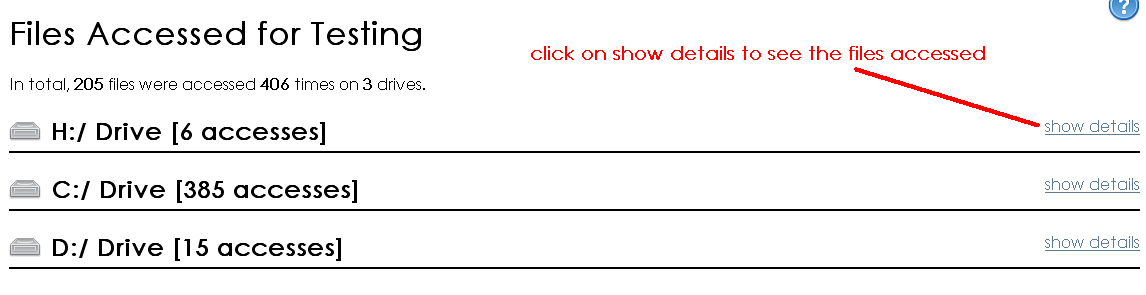
[search]

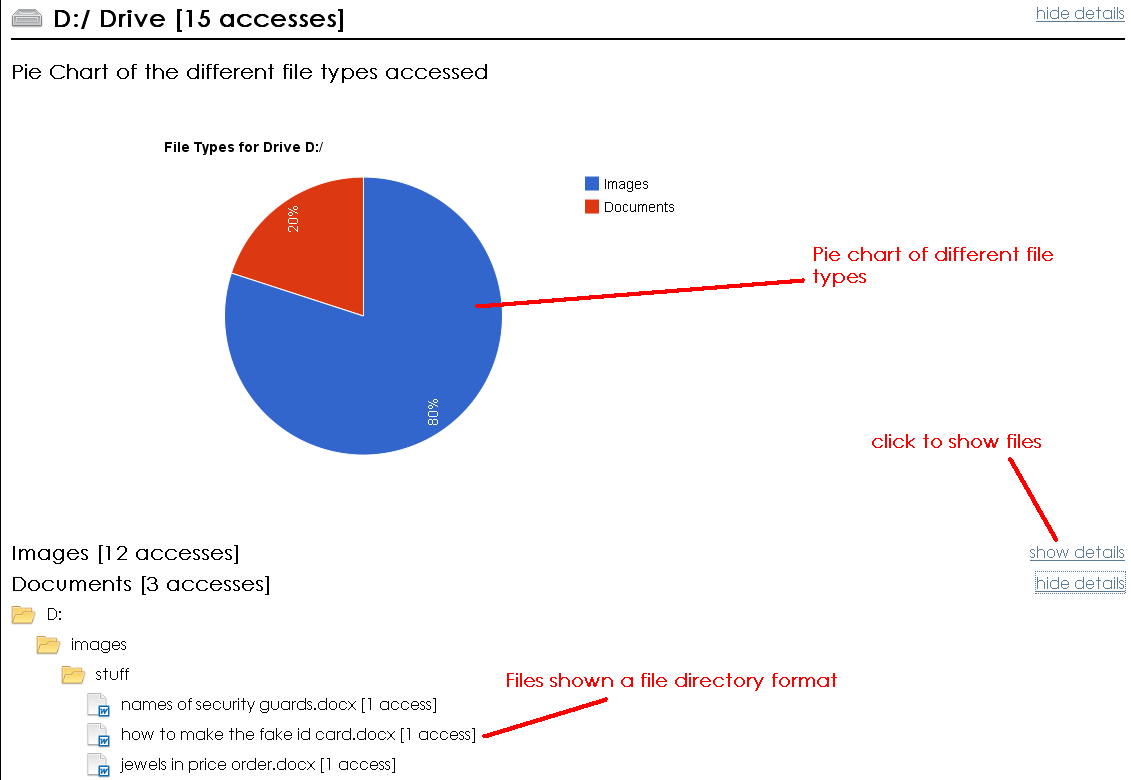
search.foo = Foo Search

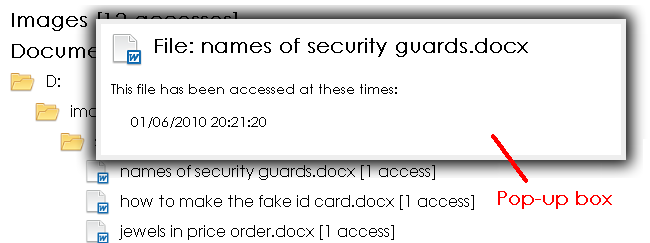
## Files Tab

This tab shows all the local file accesses divided up by drive. You can click 'show details' to show a pie chart of the file types on this drive, followed by a breakdown of the files accessed. Click on a file to pop-up details about when the file was accessed.

Currently only Windows drives are supported.

[](http://localhost:5000/images/userguide/vis_files.png)

[](http://localhost:5000/images/userguide/vis_files_drive.png)

[](http://localhost:5000/images/userguide/vis_files_drive_popup.png)