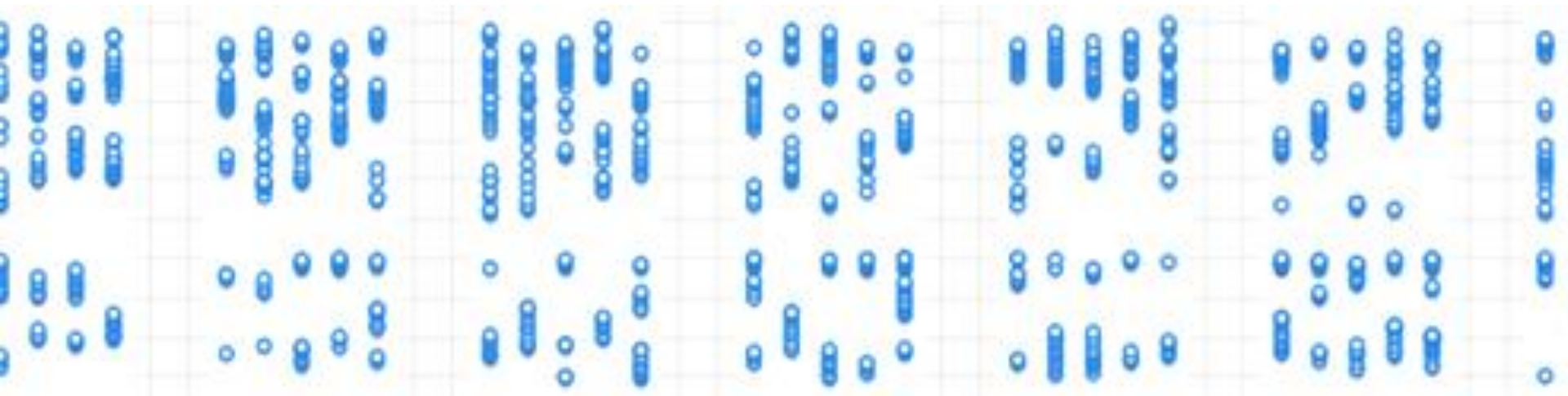


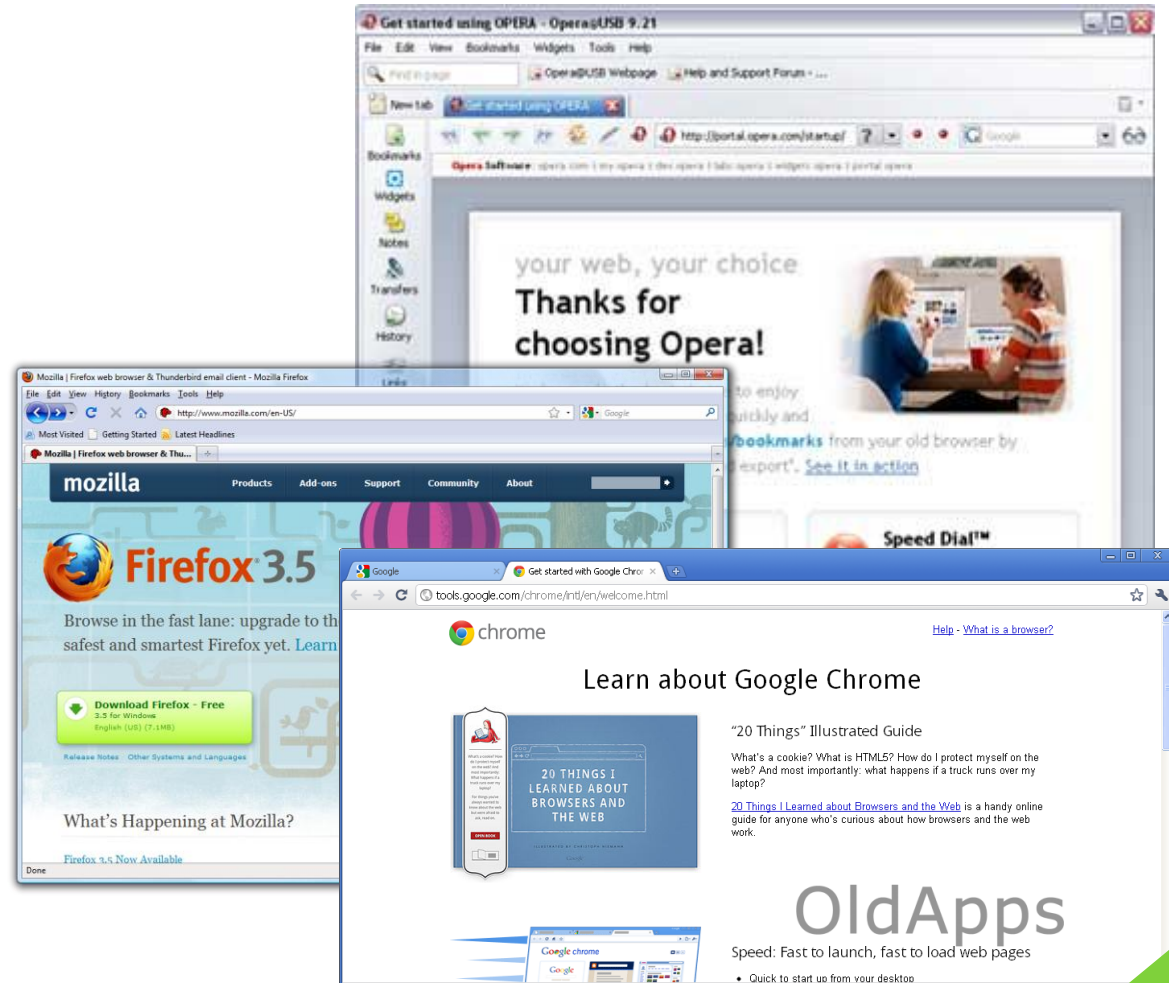
Web History Visualisation for Forensic Investigations

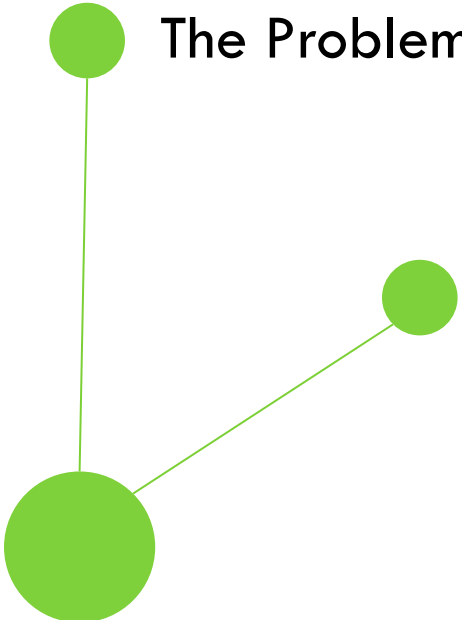
Sarah Lowman



Outline

- Background
- Visualisation
- Webscavator
- Testing
- Conclusions





The Problem + Proposed Solution

Related Work

Background

Web History Analysis

- Web history analysis is a common and important part of a digital investigation.
- Critical evidence can be found in the suspect's browser history, including websites visited, searches conducted and web-based e-mail.
- Web history may also provide an alibi, a general user profile or behavioural patterns.
- Neil Entwhistle was convicted of murdering his wife and baby daughter based on web evidence:



how to kill with a knife



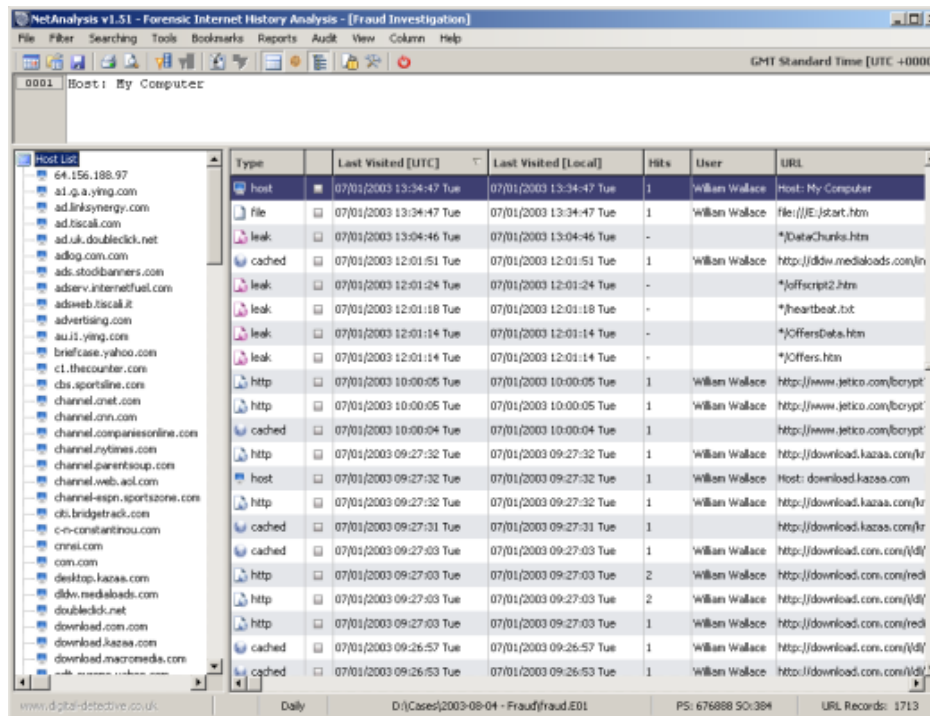
Search

About 51,300,000 results (0.27 seconds)

[Go to Google.com](#) [Advanced search](#)

Current Process

- Convert Internet history to spreadsheet format.
- Use an Internet history analysis tool such as Net Analysis or Web Historian.



The screenshot shows the NetAnalysis v1.51 interface. The main window displays a table of internet history records. A green arrow points to the table from the right side of the slide.

Type	Last Visited [UTC]	Last Visited [Local]	Hits	User	URL
host	07/01/2003 13:34:47 Tue	07/01/2003 13:34:47 Tue	1	William Wallace	Host: My Computer
file	07/01/2003 13:34:47 Tue	07/01/2003 13:34:47 Tue	1	William Wallace	file:///E:/start.htm
leak	07/01/2003 13:04:46 Tue	07/01/2003 13:04:46 Tue	-	-	*DataChunks.htm
cached	07/01/2003 12:01:51 Tue	07/01/2003 12:01:51 Tue	1	William Wallace	http://ddw.medialoads.com/in
leak	07/01/2003 12:01:24 Tue	07/01/2003 12:01:24 Tue	-	-	*joffscript2.htm
leak	07/01/2003 12:01:18 Tue	07/01/2003 12:01:18 Tue	-	-	*heartbeat.txt
leak	07/01/2003 12:01:14 Tue	07/01/2003 12:01:14 Tue	-	-	*OffersData.htm
leak	07/01/2003 12:01:14 Tue	07/01/2003 12:01:14 Tue	-	-	*Offers.htm
http	07/01/2003 10:00:05 Tue	07/01/2003 10:00:05 Tue	1	William Wallace	http://www.jetico.com/bcrypt
http	07/01/2003 10:00:05 Tue	07/01/2003 10:00:05 Tue	1	William Wallace	http://www.jetico.com/bcrypt
cached	07/01/2003 10:00:04 Tue	07/01/2003 10:00:04 Tue	1	-	http://www.jetico.com/bcrypt
http	07/01/2003 09:27:32 Tue	07/01/2003 09:27:32 Tue	1	William Wallace	http://download.kazaa.com/fr
host	07/01/2003 09:27:32 Tue	07/01/2003 09:27:32 Tue	1	William Wallace	Host: download.kazaa.com
http	07/01/2003 09:27:32 Tue	07/01/2003 09:27:32 Tue	1	William Wallace	http://download.kazaa.com/fr
cached	07/01/2003 09:27:31 Tue	07/01/2003 09:27:31 Tue	1	-	http://download.kazaa.com/fr
cached	07/01/2003 09:27:03 Tue	07/01/2003 09:27:03 Tue	1	William Wallace	http://download.com.com/vd/
http	07/01/2003 09:27:03 Tue	07/01/2003 09:27:03 Tue	2	William Wallace	http://download.com.com/red/
http	07/01/2003 09:27:03 Tue	07/01/2003 09:27:03 Tue	2	William Wallace	http://download.com.com/vd/
http	07/01/2003 09:27:03 Tue	07/01/2003 09:27:03 Tue	1	William Wallace	http://download.com.com/red/
cached	07/01/2003 09:26:57 Tue	07/01/2003 09:26:57 Tue	1	William Wallace	http://download.com.com/vd/
cached	07/01/2003 09:26:53 Tue	07/01/2003 09:26:53 Tue	1	William Wallace	http://download.com.com/vd/



Problems

1. The files are very big.
2. Each entry is a piece of text that needs to be read.
3. Correlations, patterns and anomalies are difficult to spot between lines of text.
4. Is it difficult to ask questions about the data. Columns can be sorted and searched only.
5. The data is difficult to present to those who are not technically minded.
6. Analysis requires high levels of concentration, patience and is error prone.

Proposed Solution

Build a tool that takes in web history and produces visualisations of the data





Hypothesis

Visualisations will improve:

1. The accuracy of investigators
2. The speed at which they can answer questions
3. The mood of the investigators

Related Work

- Forensic Timelines:
 - CyberForensic TimeLab (CFTL), a timeline-based forensic tool which finds and plots all forensic data based on timestamps [Olsson & Boldt, 2009].
- Visualisation for file attributes:
 - Coloured square blocks to represent files in a directory, with the intensity of the colour indicating an attribute such as file type or size [Teerlink & Erbacher, 2006].
 - Word clouds to focus on the words found in file contents [Stamps et al, 2009].



Visualisation Background

Examples

Visualisations

Visualisation

“One Picture is Worth Ten Thousand Words”
Chinese proverb



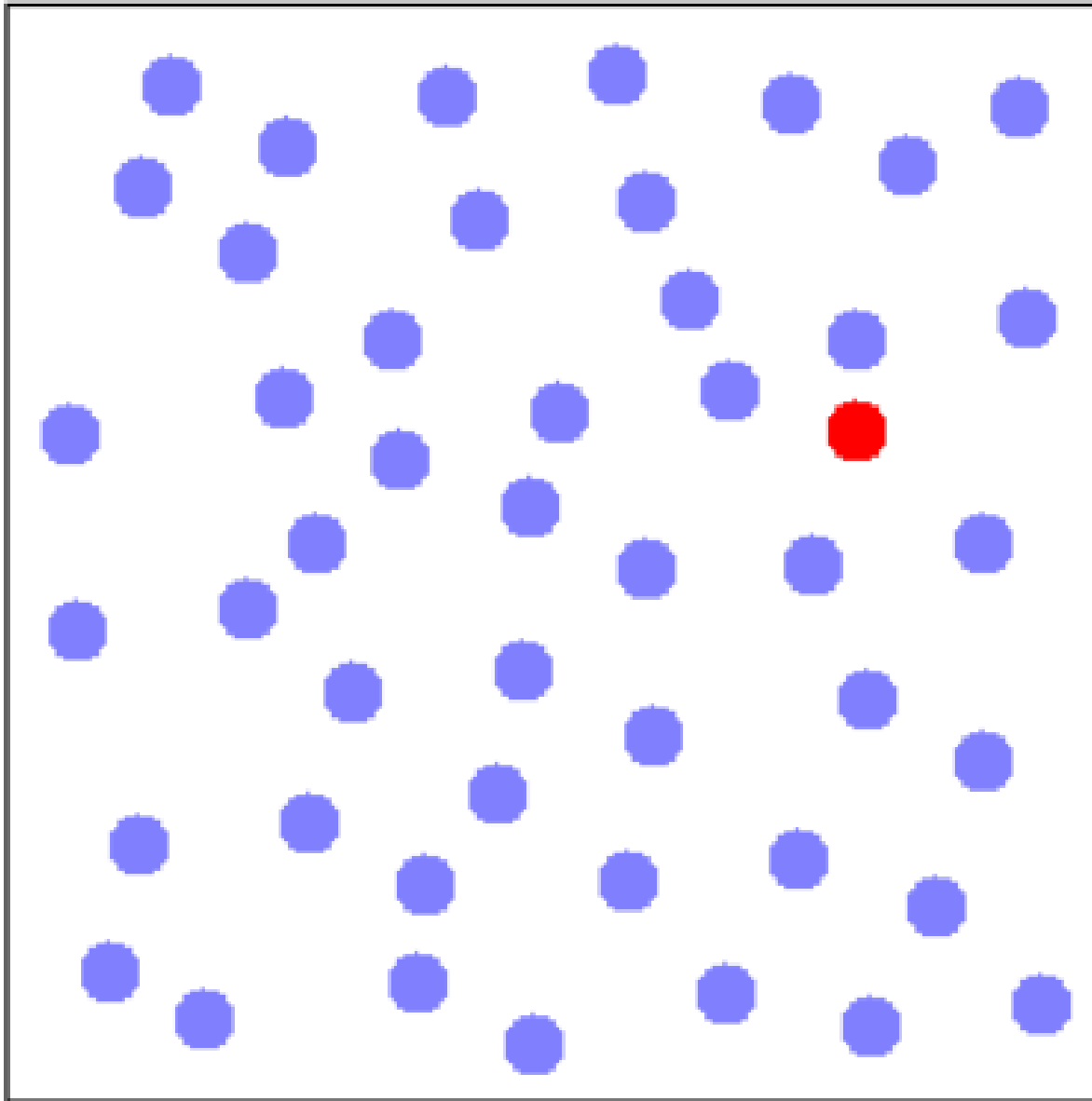
“Un bon croquis vaut mieux qu'un long discours”
 (“A good sketch is better than a long speech”)

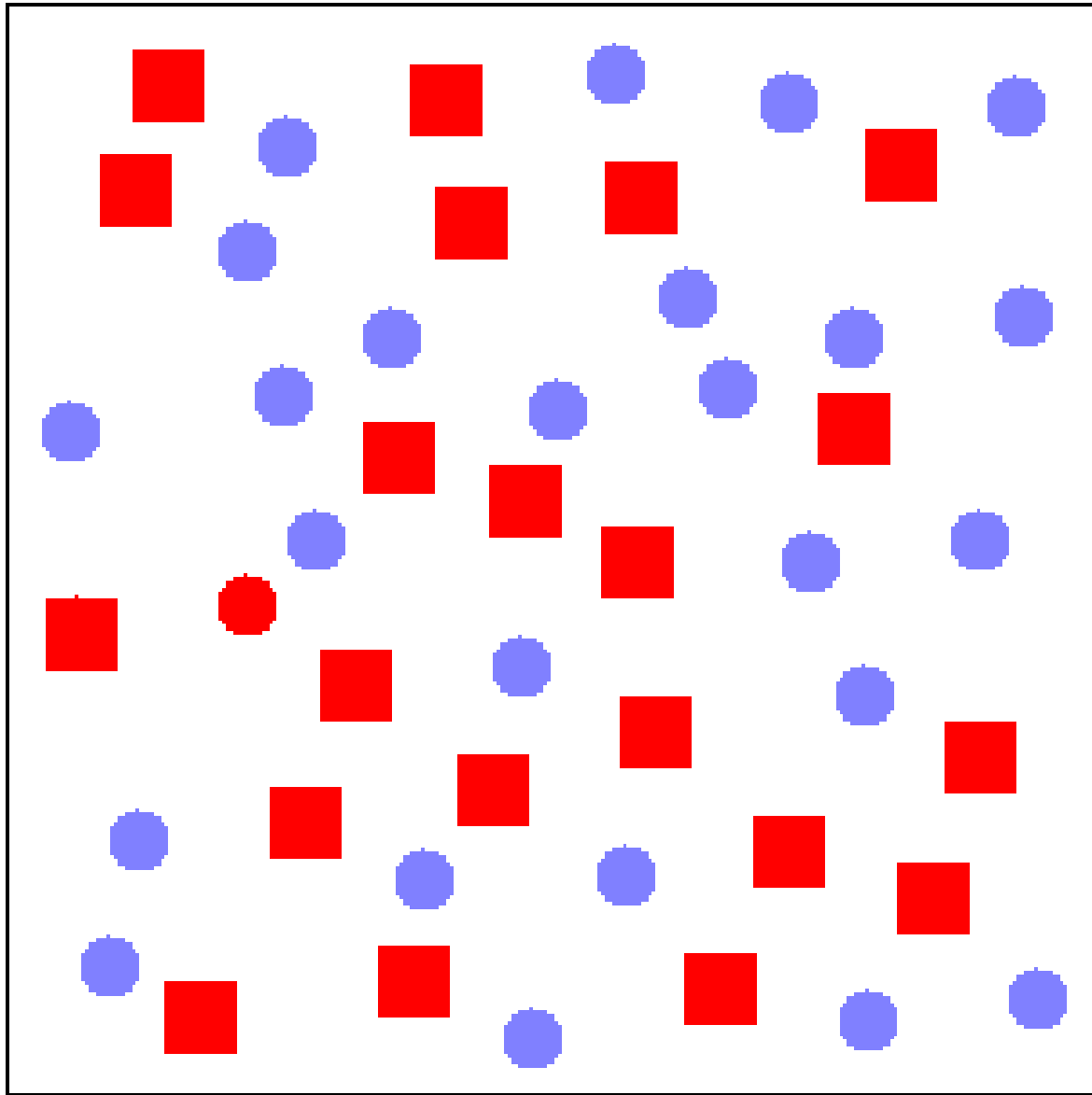
Napoleon Bonaparte

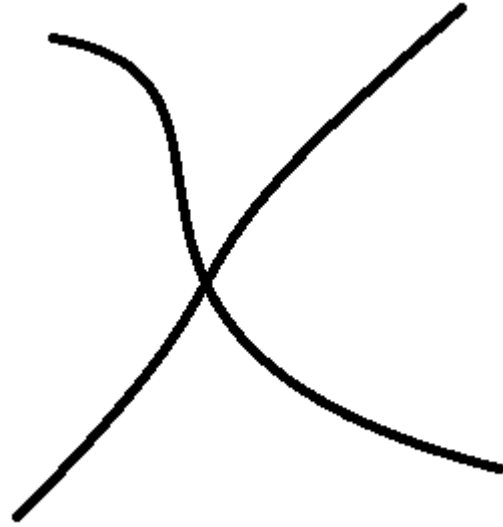


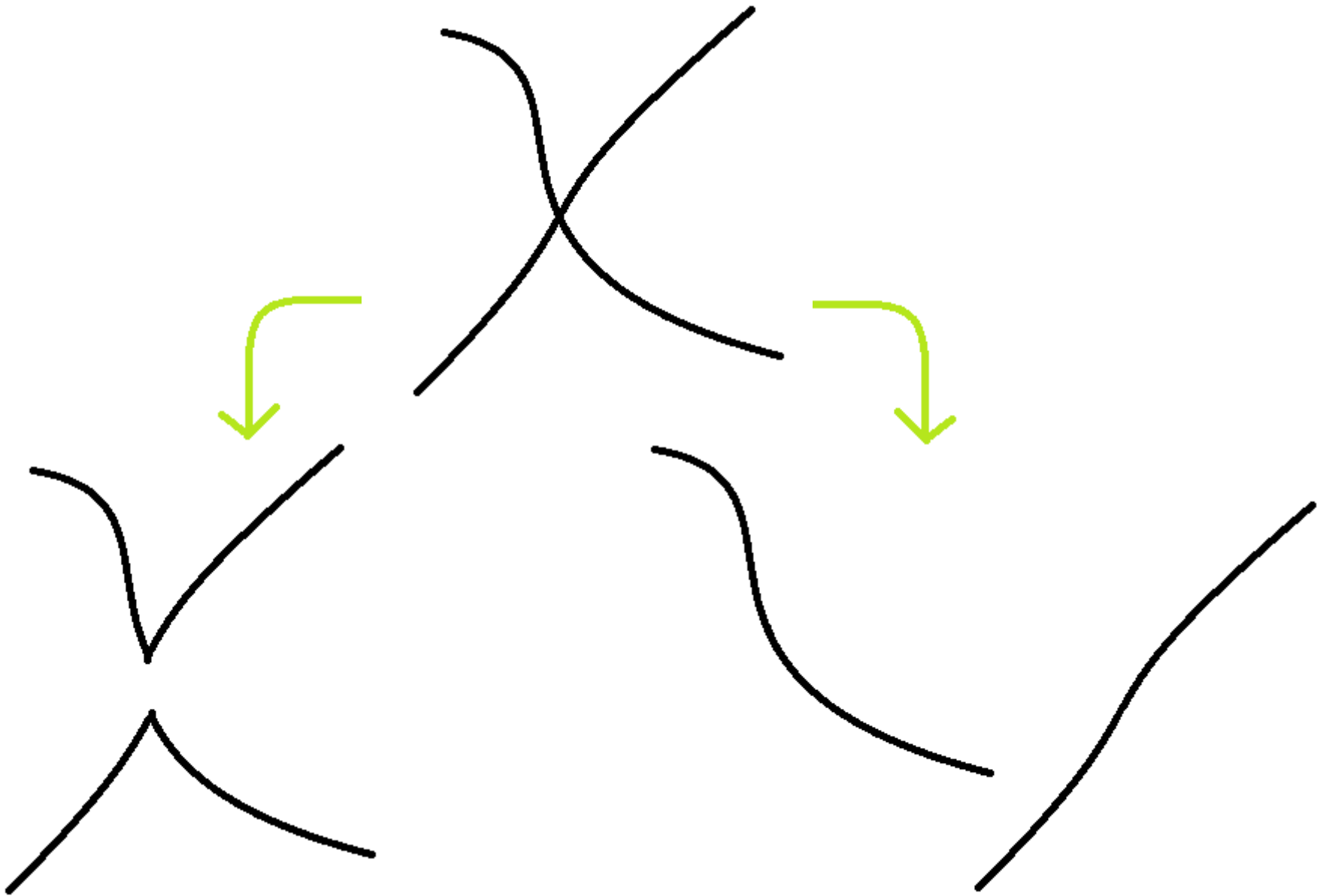
Visualisations

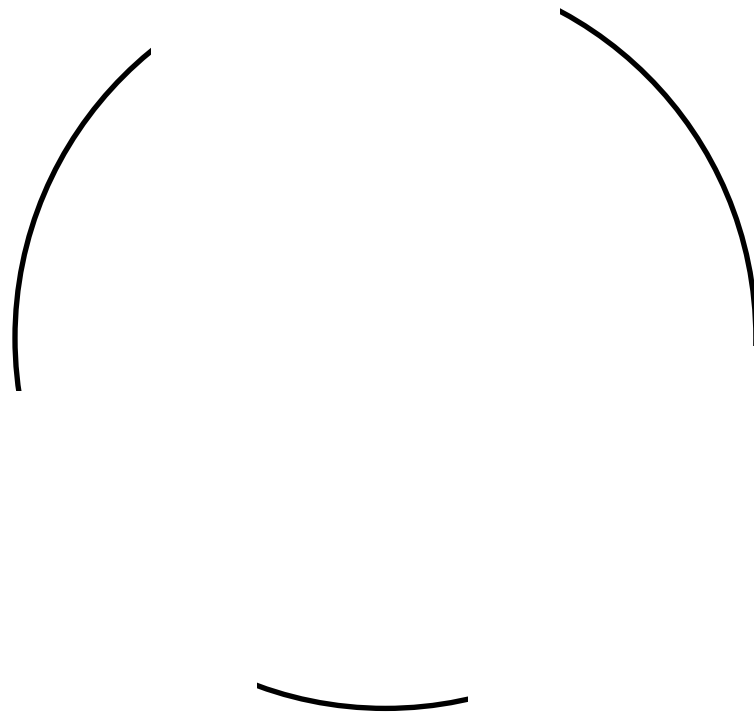
- Visualisations can act as a temporary storage area for human cognitive processes
 - Reduce the amount of information that must be held in working memory.
 - Allow the brain to hold and process more information simultaneously.
- The human visual system is able to perceive graphical information such as pictures, videos and charts in parallel, but text only sequentially [Hendee & Wells, 1997].
- Next couple of slides show examples of how the human visual system work
 - Preattentive Processing Theory
 - Gestalt Theory













Good Visualisations

- Information retrieval goals [Marchionini, 1997]:
 - To find a narrow subset of items that match a particular query
 - To develop an understanding of patterns within a set of data
- Visual information seeking mantra [Shneiderman, 1996]:
 1. Overview
 2. Zoom
 3. Filter
 4. Details-on-demand
 5. Relate
 6. History
 7. Extract



The Design

Screenshots

Webscavator

Webscavator Design

- Webscavator is a web application written in a mixture of Python and JavaScript.
 - Can make use of AJAX and JavaScript graphics libraries.
 - Autopsy (a web-based graphical interface for SleuthKit) is run in a similar way.
- By default, Webscavator will listen on a local port so that no other networked devices can gain access.
- For storing data, it uses SQLite, a lightweight open-source RDBMS.



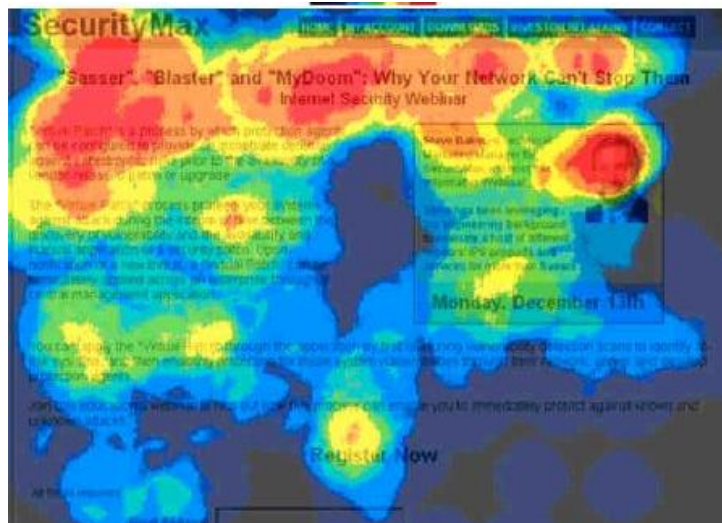
WEBSCAVATOR



Webscavator

- For input, Webscavator accepts CSV and XML files from programs such as Net Analysis and Web Historian
- Webscavator has five tabs, each showing a different visualisation.
- Filters can be applied to some of the tabs. Filters allow the user to highlight or remove web history entries from the visualisations by matching various attributes.
 - Narrow down results
 - Highlight particular features of interest
 - Reduce clutter due to irrelevant entries

Heat Maps



- In a heat-map, different values are represented as different colours – usually a range.
- Condense large amounts of information into a small space to bring out patterns.
- Webscavator's heat-map can highlight patterns in web browser usage times.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
00:00 - 00:59							
01:00 - 01:59							
02:00 - 02:59							
03:00 - 03:59							
04:00 - 04:59							
05:00 - 05:59							
06:00 - 06:59							
07:00 - 07:59							
08:00 - 08:59			6		1		
09:00 - 09:59	47	45	62	42	53		
10:00 - 10:59	12	39	13	31	29		
11:00 - 11:59	79	58	89	64	82		
12:00 - 12:59	6	2	2	4	4		
13:00 - 13:59	50	63	52	60	51		
14:00 - 14:59	99	75	61	84	69		
15:00 - 15:59	70	79	45	60	82		
16:00 - 16:59	89	67	83	84	78		
17:00 - 17:59	98	120	96	125	106		
18:00 - 18:59							
19:00 - 19:59							
20:00 - 20:59							
21:00 - 21:59							
22:00 - 22:59							
23:00 - 23:59							



Timeline

- Main visualisation
- The timeline has the day along the x-axis and the time along the y-axis. Each point on the graph is a web history entry.
- Hovering over a particular point will display a popup
- Clicking on a point will display more detailed information underneath the timeline.
- Double-clicking on part of the graph causes it to zoom centred on that part.
- The units of the y-axis are scaled automatically depending on zoom level. By clicking and dragging, the graph can be panned.

Filters

- Internet Explorer
- Firefox
- Chrome
- Google searches
- Local Files
- Work hours

[Reset Filters](#)

[Add New Filter](#)

Overview

Timeline

Websites visited

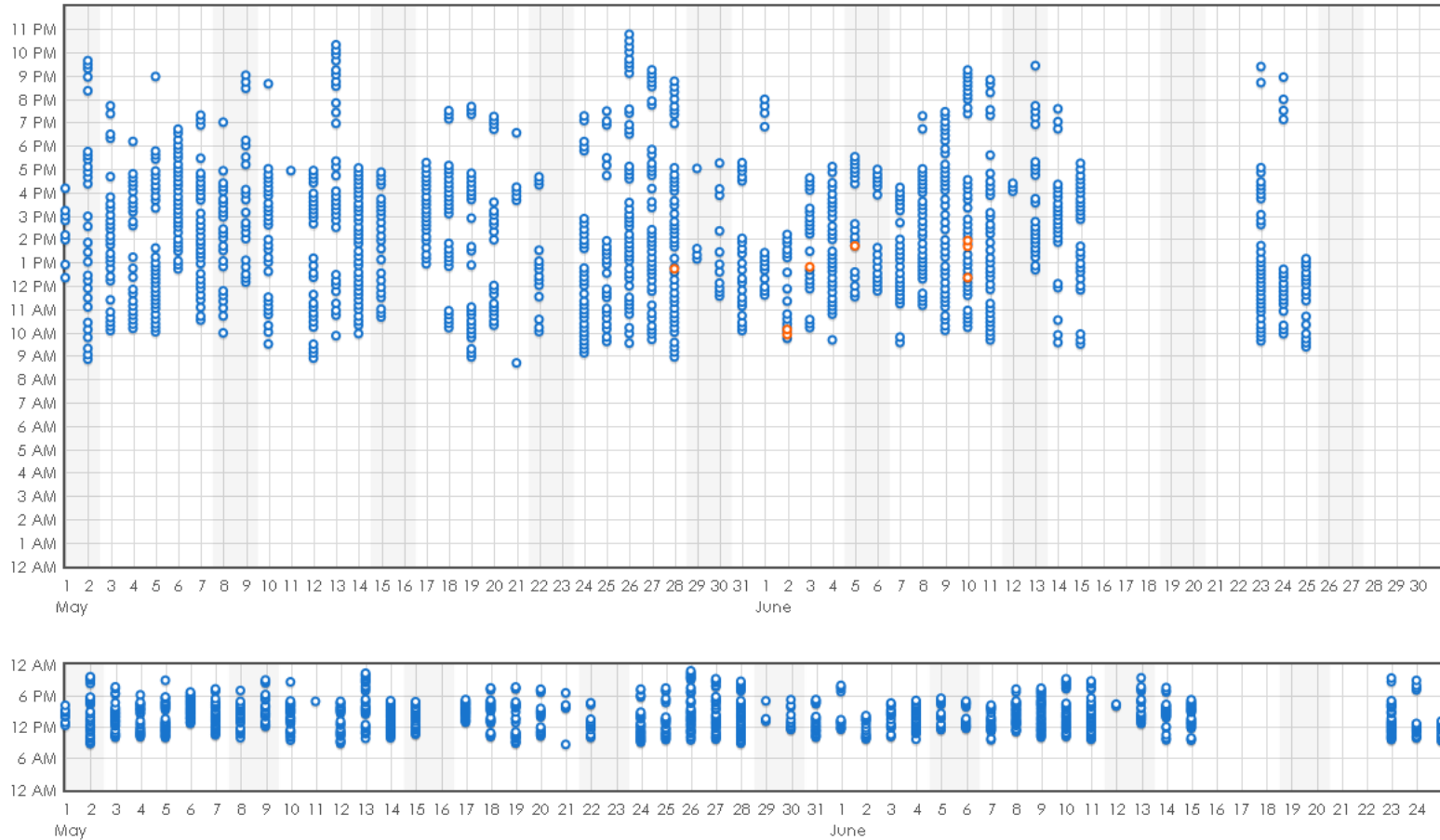
Online Searches

Files

Timeline for Test Data

[Reset Graph](#)

◀ [Previous month](#)



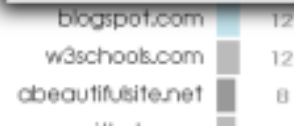
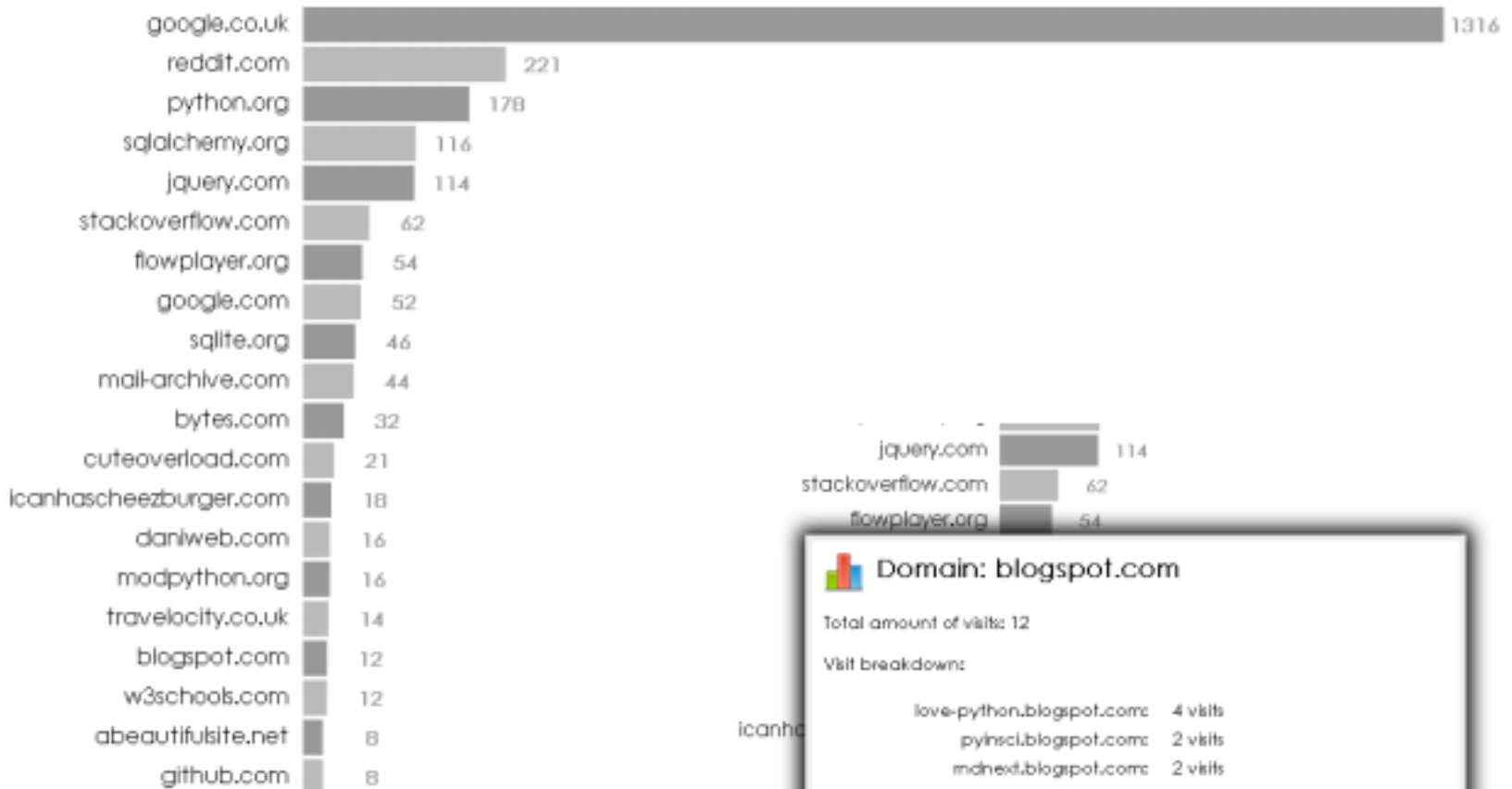


Bar Chart

- Shows the top domain names that appear in the web history.
- The chart can be configured to show the top 50, top 100 or all visited domain names.
- The total number of visits to a domain is appended to that domain's bar.
- A bar chart provides a quick visual indication of general browsing habits, and may help to determine useful filters to apply to other visualisations such as the timeline

Top 20

Domain Names Visited





Word Cloud

- A word cloud displays a set of words with the font size of each varying depending on how important the word is.
 - large words attract more attention than smaller words
- Webscavator's word cloud shows all the words entered into search engines in the web history.
 - The more a search words occurs, the larger it is displayed.
- When words in the word cloud are clicked, a pop-up is displayed showing a list of the full searches the word appeared in, which allows searches for the word to be viewed in context

Google

The **top 20** searched terms are below.

exception connect get list

jquery

datetime
sqlite
regular tools
database
one expression
string css
remove time


python

sqlalchemy javascript file


Tree Map + Pie Chart

- Tree maps show hierarchical relationships between objects
 - Tree views are a common way of displaying files in file manager applications
- Webscavator's tree map shows all the local files that have accesses recorded in the web history.
 - These only appear in Internet Explorer's index.dat files
- The indentation of folders gives context showing where files are in relation to each other
- Under each drive is a pie chart showing the breakdown of the different file types accessed.


In total, 205 files were accessed 406 times on 3 drives.

 **H:/ Drive [6 accesses]**

[show details](#)

 **C:/ Drive [385 accesses]**

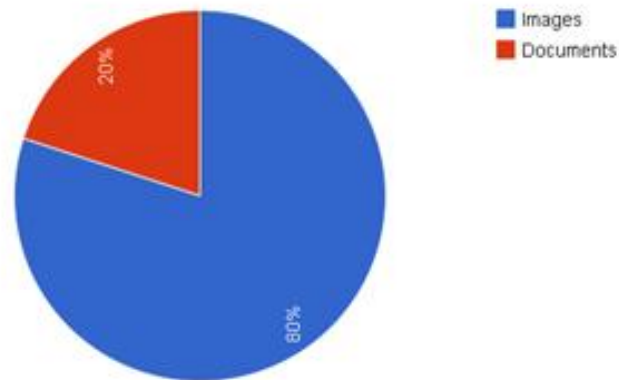
[show details](#)

 **D:/ Drive [15 accesses]**

[hide details](#)

Pie Chart of the different file types accessed

File Types for Drive D:/



Images [12 accesses]


[show details](#)


Documents [3 accesses]


[hide details](#)


 D:

 Images

 stuff

 names of security guards.docx [1 access]

 how to make the fake id card.docx [1 access]

 jewels in price order.docx [1 access]



The Participants and Test Setup

Results

User Testing

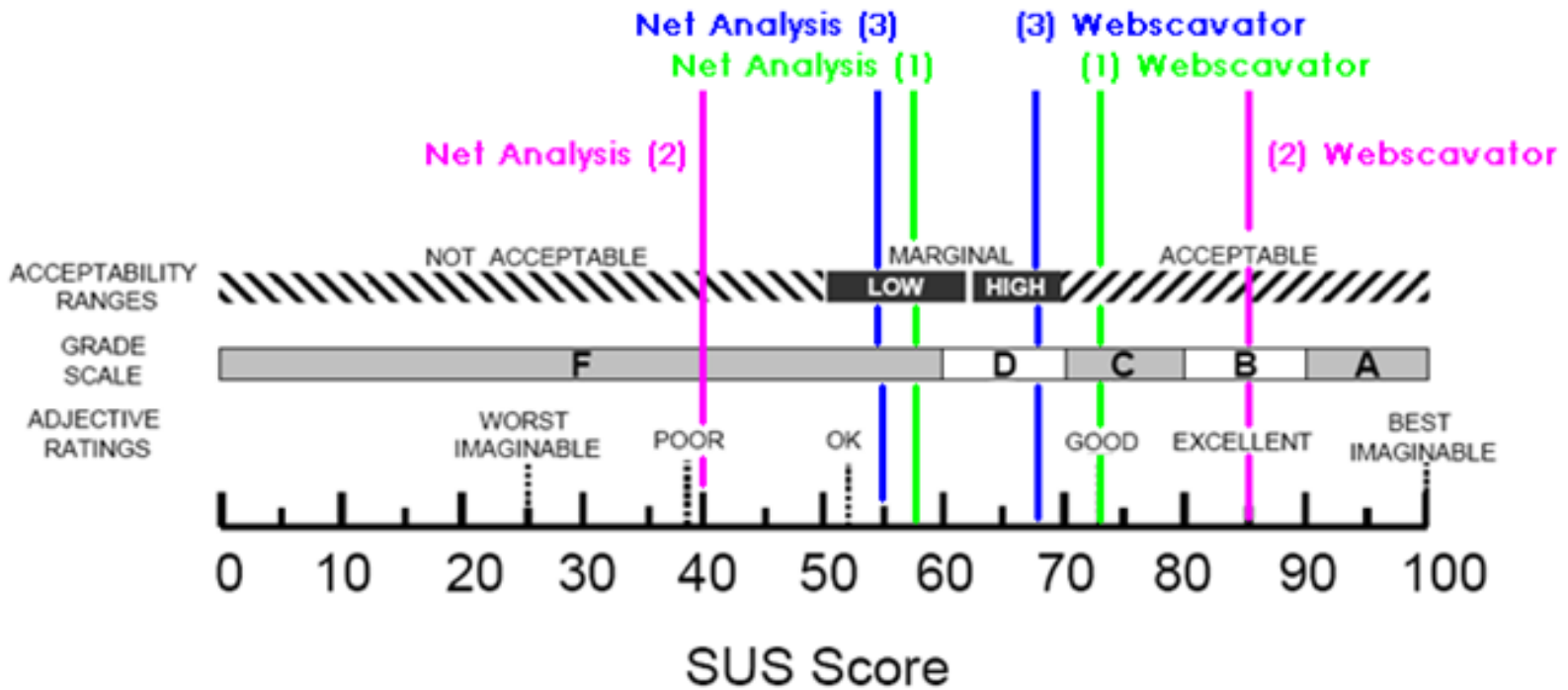
Testing Setup

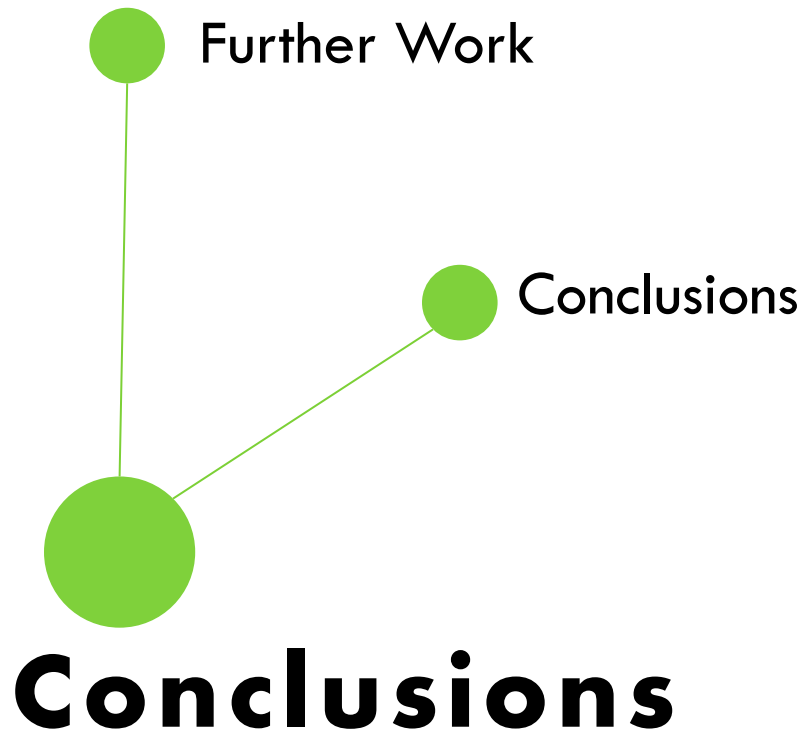
- Three participants – all forensics investigators
- Two scenarios
 - One based on industry forensics (employee surfing)
 - One for police forensics (jewellery store robbery)
- Each scenario had a set of questions to answer
- Each participant used Net Analysis and Webscavator
 - Used different datasets so the answers were not the same
 - Random whether Net Analysis/Webscavator was used first
- Rated on:
 - Number of correct answers
 - Speed at which the questions were answered
 - The participants “confidence” in their answer

Results

		Scenario A		Scenario B	
		Net Analysis	Webscavator	Net Analysis	Webscavator
User 1	Correct answer	67%	100%	33%	100%
	Confidence	37%	77%	64%	78%
	Time	18:54	17:01	17:57	09:50
User 2	Correct answer	67%	83%	78%	78%
	Confidence	73%	83%	82%	78%
	Time	19:35	10:13	17:12	08:17
User 3	Correct answer	83%	83%	78%	100%
	Confidence	57%	63%	78%	78%
	Time	14:16	8:42	15:25	07:47

Results





Further Work

- Allow filters to be viewed, edited and deleted.
- Add reporting.
 - Currently Webscavator is just a visualisation tool, however to be used in practice it needs to be able to generate reports and images.
- More integrated visualisations.
 - Following on from usability questionnaire results, more could be done to integrate the different visualisations.
- Fuzzy and periodic searching.
 - Match values that are homophones and spelt incorrectly
 - Matches values that are at regular intervals apart – remove tickers, RSS feeds.

Conclusions

- Investigators face several problems when using the non-visual tools that are currently used to analyse web history data:
 - Information overload
 - Difficulty spotting correlations and patterns
 - Difficulty in obtaining a summary overview of data
- Visualisation is particularly apt for solving these
- User testing results showed that Webscavator's visualisations perform well when compared to Net Analysis.
 - Average speed and confidence increased, errors decreased and investigators enjoyed using the Webscavator more
- These results show that visualisations should be taken seriously by the producers of forensic software.

References

- Olsson, J., & Boldt, M. (2009). Computer forensic timeline visualization tool. *Digital Investigation*, Volume 6, pp. S78-S87.
- Teerlink, S., & Erbacher, R. (2006). Improving the Computer Forensic Analysis Process through Visualization. *Communications of the ACM*, Volume 49. Issue 2, pp. 71-75.
- Stamps, A. S., Franck, J., Carver, J., Jankun-Kelly, T., Wilson, D., & Swan, J. E. (2009). A Visual Analytic Framework for Exploring Relationships in Textual Contents of Digital Forensics Evidence. *Proceedings of Workshop on Visualization for Cyber Security*, pp. 39.
- Hendee, W. R., & Wells, P. N. (1997). *The perception of visual information*. Springer.
- Marchionini, G. (1997). *Information Seeking In Electronic Environments*. Cambridge University Press: Issue 9 of Cambridge series on human-computer interaction.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *IEEE Symposium on Visual Languages*, pp. 336-343.
- Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*.



Questions

www.webscavator.org