

VISVIP: 3D Visualization of Paths through Web Sites

John Cugini and Jean Scholtz
Information Technology Laboratory
National Institute of Standards and Technology
{cuz, jean.scholtz}@nist.gov

Abstract

VISVIP allows web site developers and usability engineers to visualize the paths taken through the site by the subjects of usability experiments. They can dynamically customize and simplify the graphical layout of the web site, and select which subjects' paths to view. An animated representation of progress along the path through the web site is also available. The third dimension of the 3D display is used to represent the time spent on each page visit. The graph layout provided by VISVIP can be governed by either the web site topology or the intrinsic structure of the subject's path.

1. Background and Motivation

Given the growth of the Web, no one questions the importance of helping web site developers and usability engineers (UE) measure the effectiveness of their creations. The work described here is part of the NIST WebMetrics [5] project. The goal of the project is to develop experimental prototypes of remote, rapid, and automated tools for testing the usability of web pages and web sites.

In particular, the WebVIP tool allows the UE to instrument a web site so as to record the activity of a subject navigating the site. WebVIP captures only the transitions between pages, including use of the BACK button, and timing information. Recording activity *within* a page (e.g. filling out a form or clicking on checkboxes) is not yet supported, but this capability is currently in development for WebVIP and other systems [2]. The intended use of WebVIP is that subjects be given specific relevant tasks to perform, such as finding some information, or completing a transaction; the resulting trace is used to analyze the degree and source of any difficulty encountered. Of course, WebVIP could also record simple browsing, but the interpretation of activity that is not goal-driven becomes problematic.

The trace file generated by WebVIP, then, consists of a sequence of time-stamped jumps through the pages of a

Contribution of the National Institute of Standards and Technology. Not subject to copyright. Reference to specific commercial products or brands is for information purposes only; no endorsement or recommendation by the National Institute of Standards and Technology, explicit or implicit, is intended.

web site. We believe that this information represents a good opportunity to apply modern interactive visualization techniques for analysis and understanding.

2. Related Work

Although there are many tools to visualize web sites, we are aware of only one other effort, WebPath [3], in which the *path* of an individual subject within the web is depicted. The main purpose of WebPath, however, is not usability analysis but to provide the user with a more powerful history feature for general web navigation - it helps the user figure out where he or she has been; it does not help the UE analyze how well a specific web site supports some set of tasks.

Ed Chi et al [1] have presented some techniques for visualizing overall usage of pages within a web site, but do not try to depict paths (sequences of URLs). Their primary purpose is to show how a web site evolves over time.

The Footprints system [7] depicts typical paths through a web site as a navigation aid. These paths, however, are aggregates of several users. The system deliberately does not reveal individual paths.

3. Design and Features

Our goal was to produce a tool that enables the UE to explore the usage data compiled by WebVIP in a flexible way. We wanted the tool to exhibit reasonable default behavior so as to provide a good, first look at the data. We also identified many optional features that the UE might find desirable in order to control and customize the visualization. VISVIP uses familiar metaphors, highlights timing data, and, where useful, exploits 3D capabilities.

3.1 Web Site Topology

The web site is depicted as a directed graph, with pages as nodes, and links as edges. VISVIP automatically generates a 2D layout of the graph, using a force-directed algorithm. In our model, adjacent nodes exert a spring-like force on each other - they "try" to set themselves apart at a fixed distance. Non-adjacent nodes repel each other with a force inversely proportional to the distance between them. Initially we tried using the more familiar "inverse of distance squared" rule, but found that the graph did not spread out well enough because the repulsive force

effectively vanished over longer distances. Finally there is a third force that weakly attracts all nodes towards the origin; this serves to keep parts of the graph that become unconnected from flying off to infinity.

The UE can dynamically adjust the constants of the underlying force model so as to control the overall density and appearance of the graph. Also, for finer control, the UE can drag and drop individual nodes into precise locations as desired.

In our early implementations, we generated a 3D layout for the web site graph. While easy to produce, it was difficult to interpret, especially when the subject paths were overlaid. Therefore, we kept the web site layout as a 2D structure, and “saved” the 3rd dimension for timing information (see below).

3.2 Web Site Content

Because URLs tend to be long, a brief nickname is generated for each page, normally the ten characters immediately preceding the filetype. For example, an URL such as “http://www.wherever.com/animals/giraffe.gif” would be nicknamed “ls/giraffe”. Each node is shown as a small named box. The box is color-coded to reflect the page type (e.g. HTML, image, video, audio, mailto, directory, etc.). Sliding the mouse over a node causes it to be highlighted to indicate that it is currently selected, and the full URL and filetype of the page appear at the top of the window. The selected page can be displayed in Netscape.

3.3 Web Site Simplification

A web site may possess a simple structure that generates a comprehensible display. The site in Figure 1, for example, approximates a balanced tree (see [5] for full-size color figures). Others do not lend themselves so easily to good visualization. We found several common problems: first, some web sites have a lot of “noise” nodes, such as images and mailto nodes. These nodes usually are not significant in tracing subjects’ paths. Therefore VISVIP allows the user to request suppression of all nodes of a specified type, or of individual nodes. If such a node is traversed by a path, however, it will nonetheless be displayed as part of the graph.

A second problem was “over-connectivity”. In some web sites, all the pages have a link back to a home page. Also, there may be an image (such as a company logo) that appears on every page. The result is a graph with so many edges that the visualization is confusing (see Figure 2). While it may be reasonable to delete the node with a company logo, suppressing the home page is probably not a good solution. VISVIP offers two mechanisms to simplify the display by deleting edges: first, one can select a node and direct that all edges leading to that node be

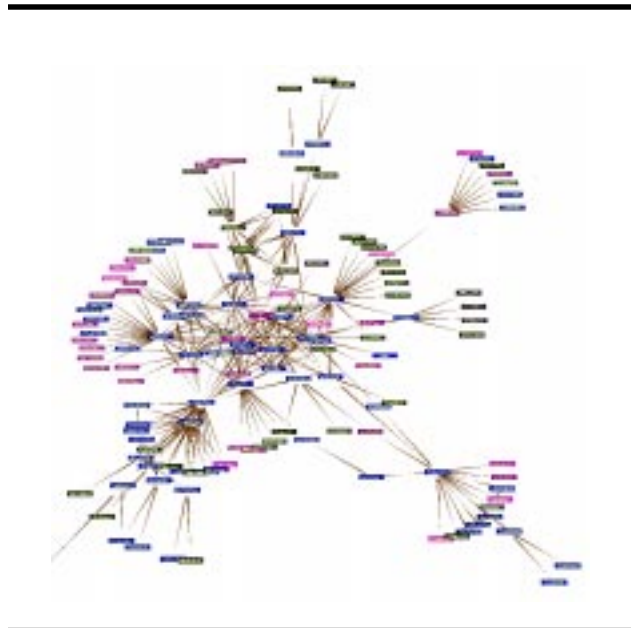


Figure 1. A typical web site layout

deleted. Second, and more drastic, you can select a node, such as the home page, and direct that the graph be simplified to a tree, using that node as the root.

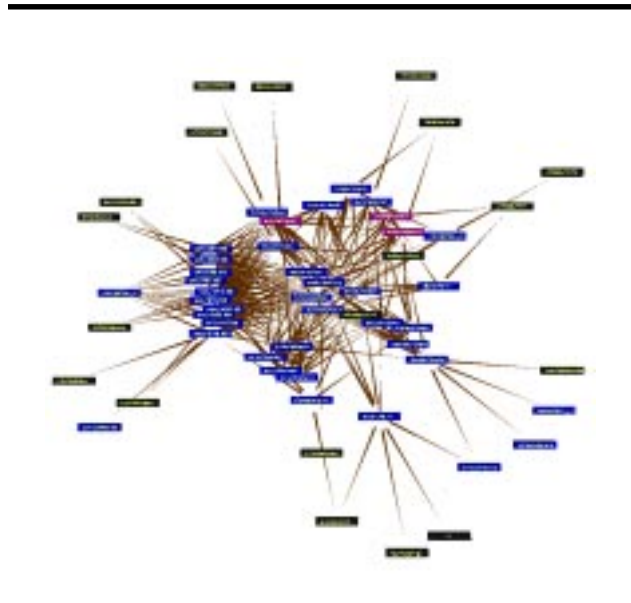


Figure 2. A Highly Connected Web site

Finally, if the web site is large, and a subject’s path is confined to a small section of it, it may be helpful to concentrate on only that part of the web site. Therefore VISVIP supports the suppression of nodes more than a given distance from any displayed path, e.g. all nodes more than two links away can be suppressed.

3.4 Path Sequence

Given this visual layout of the web site, our next step is to depict the subjects' paths through it. In early versions of VISVIP, we showed the paths simply as straight line segments connecting the successive nodes. Two problems became evident: first, it was hard to visually follow a single jagged path, and second, when several paths were shown at once or when a single path backtracked, it was difficult to distinguish among them.

We then adopted a spline representation: each subject path is depicted as a smooth curve overlaid on the directed graph. Each subject is coded with a distinct name and color. VISVIP also allows the UE to specify the colors manually so as to examine groups of subjects (e.g. novices vs. experts). The spline curves are decorated with arrowheads to indicate direction, and a special curvy arrow into and out of the plane of the graph highlights the starting and ending point of the path.

The UE may dynamically select any subset of paths to be viewed at any time. The appearance or suppression of nodes is automatically updated insofar as it is dependent on the paths. Often, we found it useful to start by viewing all the paths to get a general sense of which parts of the web site the subjects were visiting. One could then easily identify paths that deviated from the norm and examine these outliers individually.

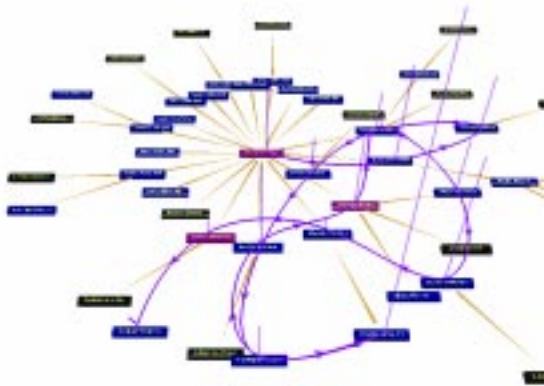


Figure 3. Path laid over a web site

3.5 Path Timing

Of course, an important aspect of the path data is the *time* the subject spent at each node. This is represented by a dotted vertical bar with its base on the node of the page where the time was spent, and its height (in the 3rd dimension, orthogonal to the plane of the graph) proportional to the amount of time. In order to avoid

overlap, VISVIP offsets the base location slightly in one dimension, depending on the subject, and in the second dimension, depending on total time spent so far (the latter to prevent overlap from a subject's revisiting the same page). Pages where subjects spent a long time or which were heavily visited are immediately evident (Figure 3).

In addition to the static display, VISVIP provides animation facilities for the visualization of path traversal. As the animation progresses, a horizontal "current time" indicator proceeds from the base of the vertical time bar to the top. When it hits the top (representing a jump from this page to the next), the next segment of the spline curve becomes visible, and the indicator starts moving up from the base of the next page. The UE can dynamically adjust the playback speed (10-30 times real-time is typical), and pause or re-start.

The UE can also directly control the animation with a slider to set the virtual time represented by the display. By moving the slider back and forth, the UE can effectively play the animation backward or forward, or pause at a time of special interest. The slider displays the number of seconds elapsed since the beginning of the paths.

Finally, the UE may select a node and generate a detailed timing table, displayed in Netscape, for the corresponding page. The table shows all visits by all subjects to that page, how long each visit took, and the total path time elapsed just prior to each visit (see Figure 4).

Time spent by users at node:
/~phsiao/culture/sitea/edweb/crashtst.html

	a1-amer-u05	a1-amer-u11	a1-amer-u15	a1-sand-u07
Visit #1	2:57 / 0:43	2:02 / 0:44	3:24 / 3:56	4:10 / 1:20
Visit #2	8:03 / 0:08		8:21 / 0:06	
Visit #3			12:42 / 0:05	
Visit #4			12:50 / 0:01	
Totals:	Total / 0:51	Total / 0:44	Total / 4:08	Total / 1:20

Format: [time of entry] / [time spent].
(All times expressed as minutes:seconds)

Figure 4. Detailed timing data

3.6 Path Topology

As Figure 3 illustrates, one cannot suppose that the structure of subjects' paths will neatly align with the

intrinsic static structure of the web site being traversed. Basing the organization of the graph on the inherent web site topology is good for seeing what parts of the web site are being heavily visited or avoided.

However, we found that the inherent topologies of the paths are also informative for discovering similarities and differences among subjects' navigation styles. Our force-directed graph layout algorithm depends only on whether pairs of nodes are adjacent (directly connected) or not. Therefore VISVIP lets the UE dynamically choose whether adjacency for a pair of nodes means "connected by a static link" as described so far, or "a displayed subject path jumps directly from one to the other". Paths that looked complicated when laid out directly over a web site often revealed a much simpler structure when laid out according to their own topology (see Figure 5). Indeed, our preliminary inspection seemed to show certain distinct patterns of activity by various subjects: some do very little backtracking; others seem to circle around; others keep returning to an anchor page, perhaps for orientation.

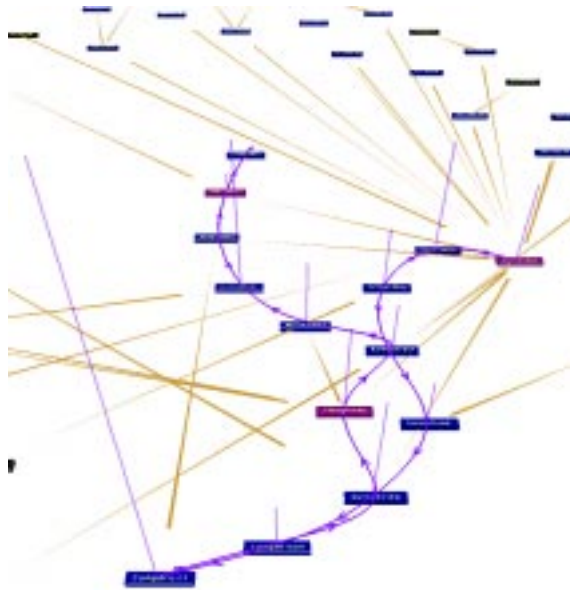


Figure 5. Natural topology of a path

4. Implementation

We used the free software Linklint [4] to determine the static structure of the web site. We then developed some PERL scripts to merge Linklint's analysis with the output of WebVIP (e.g. to resolve URLs), resulting in a clean set of files for VISVIP, two representing the web site *per se* (one for the nodes, the other for the edges), and one for each subject path to be depicted.

The visualization itself appears in a window managed by OpenGL. There is also a control menu, managed by Tcl/Tk, that presents the many non-spatial operations provided by VISVIP (see Figure 6). These two processes communicate via Xwindows [6]. The Tcl/Tk process sends X events to the OpenGL window. This allows the OpenGL process to be structured as a simple event-handling loop; those events may originate from Tcl/Tk or directly from a user action, such as a mouse-click.

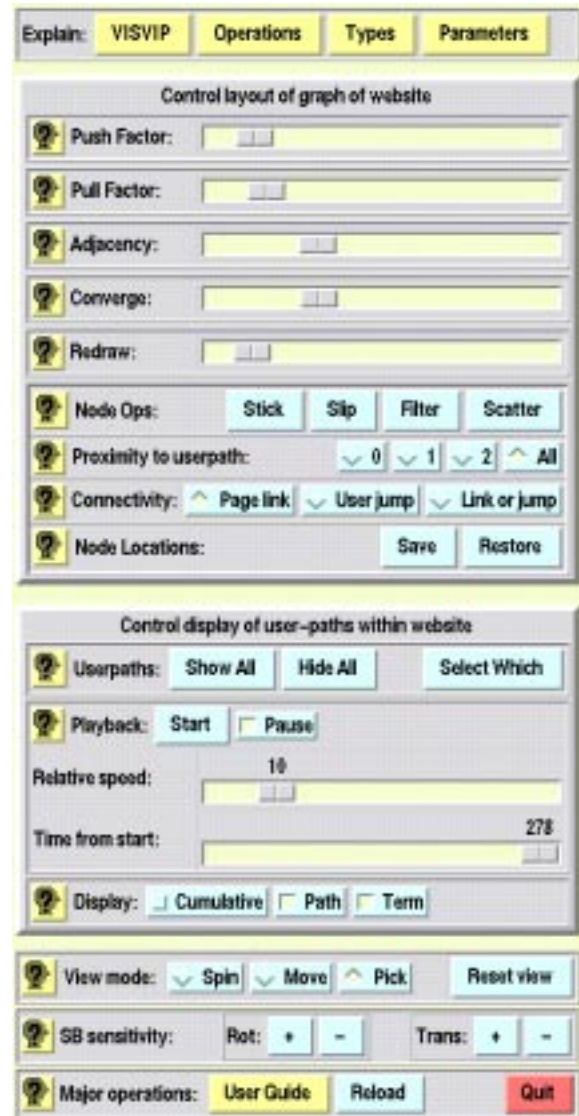


Figure 6. VISVIP Control menu

Although VISVIP can work with a mouse as the only input device, the preferred interaction method is to use a spaceball to control the position of the entire display (3D pan and zoom), and the mouse to pick out entities within

the display for various operations. If no spaceball is available, the mouse must be toggled between move mode and pick mode to achieve both of these functions.

VISVIP is currently running on SGI and HP platforms and should be easily portable to most UNIX systems as long as they support Xwindows. The current distribution is available at: <ftp://zing.ncsl.nist.gov/outgoing/visvip/>.

5. Conclusion

We believe that more formal usability engineering will play an increasingly important role in the design of sites as the use of the web becomes more pervasive. While there is an irreducibly creative aspect to good usability design, web site developers will come to rely on commonly accepted methods of testing and evaluation. These methods must be supported by appropriate tools to provide measurement and insight. In particular, visualization of user activity, such as that provided by VISVIP, should be one of the principal tools at the disposal of web site designers.

Acknowledgments

We thank Dr. Sharon Laskowski for her helpful comments and suggestions during the design of VISVIP, and Paul Hsiao for developing WebVIP.

6. References

- [1] E.H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, S.K. Card, "Visualizing the Evolution of Web Ecologies", *CHI 98 Conference Proceedings*, Los Angeles CA, April 1998.
- [2] M.P. Etgen, J. Cantor, "What does getting WET (Web Event-logging Tool) Mean for Web Usability?", *Proceedings of the 5th Conference on Human Factors and the Web*, Gaithersburg, MD, June 1999.
- [3] E. Frecon, G. Smith, "WebPath - A Three Dimensional Web History", *Proceedings: IEEE Symposium on Information Visualization*, Research Triangle Park, NC, October 1998.
- [4] Linklint: <http://www.goldwarp.com/bowlin/linklint/>
- [5] NIST WebMetrics: <http://zing.ncsl.nist.gov/~webmet/> and <http://zing.ncsl.nist.gov/~cugini/webmet/visvip/pix.html>
- [6] Nye, A., *Xlib Reference Manual*, O'Reilly & Associates, 1993.
- [7] A. Wexelblat, P. Maes, "Footprints: History-Rich Tools for Information Foraging", *CHI 99 Conference Proceedings*, Pittsburgh PA, May 1999.