

TECHNICAL RESEARCH REPORT

Coordinating Overviews and Detail Views of WWW Log Data () -0.018723

by Harry Hochheiser, Ben Shneiderman

TR 2005-13



ISR develops, applies and teaches advanced methodologies of design and analysis to solve complex, hierarchical, heterogeneous and dynamic problems of engineering technology and systems for industry and government.

ISR is a permanent institute of the University of Maryland, within the Glenn L. Martin Institute of Technology/A. James Clark School of Engineering. It is a National Science Foundation Engineering Research Center.

Web site <http://www.isr.umd.edu>

Coordinating Overviews and Detail Views of WWW Log Data

Harry Hochheiser, Ben Shneiderman*

Human-Computer Interaction Lab, Department of Computer Science

*Institute for Systems Research and Institute for Advanced Computer Studies,
University of Maryland, College Park, MD 20742

{hsh,ben}@cs.umd.edu

ABSTRACT

Web server log analysis tools provide site operators with useful information regarding the visitors to their sites. Unfortunately, the utility of these tools is often limited by the use of aggregate summaries that hide the information associated with individual requests, and by the absence of contextual data that might help users interpret those summaries. Building upon earlier work in the use of starfield visualizations to display web site requests as individual data points [8], this paper describes the use of multiple-coordinated visualizations of web log data at varying granularities, and alongside additional related displays of appropriate contextual information.

Keywords

World Wide Web, Log File Analysis, Information Visualization, Snap-Together Visualization

1. INTRODUCTION

Analysis and visualization of WWW Log data is an area of active research and commercial development, with numerous products complementing a variety of research efforts [2,4,5]. Many of these systems use aggregation to handle the large data sets generated by Web servers. Reports that summarize the number hits per page, display period, or requesting domain (for example) provide useful feedback for site operators, at the expense of hiding the potentially useful information contained in the individual request points.

In earlier work [8], we have described the use of interactive starfield visualizations [1] for examination of WWW Log Data. Using the Spotfire visualization tool [13], we have developed visualizations of tens of thousands of individual web requests, with each request represented by a single point. Although this approach avoids the data loss present in tools that use aggregation, it suffers from the opposite problem: the lack of appropriate summary data for understanding of higher-level trends.

A further shortcoming of existing log analysis tools is the display of data without context: reports of web site usage are presented without any contextual information such as a site map or even the content of the individual pages. The absence of this supporting information may complicate the task of interpreting the usage reports.

Coordinated visualizations that simultaneously present multiple views of relevant information might be used to address both of these problems. This paper describes the use of Snap-Together Visualization (STV) [12] to manage tightly coupled displays of log data at different granularities, and in the presence of supporting contextual information.

2. MULTIPLECOORDINATED VISUALIZATIONS & STVs

Coordinated, tightly coupled displays have been shown useful in a number of domains [3,5,6]. Visualizations that present tightly coupled displays of web log data could assist in the process of inferring usage patterns. Possibilities include coordinated visualizations between aggregated and disaggregated data, and coordination between log displays and supporting external information such as site maps or web page displays. The use of coordinated visualizations for database exploration has become an active area of research in recent years [7,9,10,11].

Snap-Together Visualization (STV) [12] is an architecture that allows users to connect visualization tools such that actions of selections, navigation, and querying are coordinated. Furthermore, STV supports several visualization tools (including Spotfire), raising the possibility of coordinating starfield visualizations with tables, outline views, and web browsers. STV uses Microsoft's ODBC tools for database connectivity, so data preparation is straightforward: we import the data into Microsoft Access, and design appropriate database queries, using SQL or

Access' visual query editor. Previously, STV has been used to visualize aggregations of highway incident data [6]: this paper describes the application of similar techniques to web log data.

Using STV, we can create visualizations that provide coordinated views of multiple data sets, or multiple views of the same set. This coordination provides expressive power that goes beyond single displays of individual or aggregated web requests. This paper presents two possibilities: many others are possible.

3. AGGREGATIONS

Visualizations that present each web request as an individual point lead to densely populated displays that can be used – in combination with Spotfire's dynamic query tools – to infer patterns. This approach is fundamentally limited, as it does not account for aggregate counts that many site operators find useful. Visualizations of total number of hits by URLs, or hits counted by time of day, increase the expressive power of the visualizations.

Ideally, coordinated views based on aggregations would support moving between different levels of detail. By snapping a view of aggregations to a second window containing individual data points, users can move quickly between overview and detail analysis. Selection of an aggregate in the first visualization will lead to the display of the component requests in the second display, thus allowing users to “drill-down” to finer levels of detail.

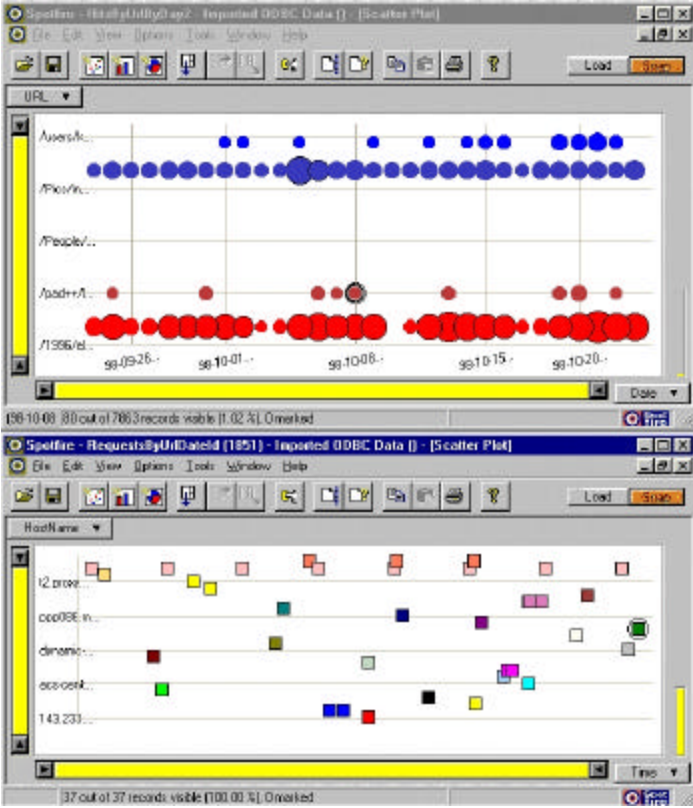


Figure 1: Two coordinated visualization windows: selection of an aggregate in the upper window leads to display of the appropriate constituent points in the lower window.

An example of this technique is shown in Figure 1. The aggregate display shows totals of the number of hits to a given URL (y-axis) on a given date (x-axis). Size coding displays the number of hits, so the larger circles indicate higher number of hits for the given URL on the given day. This visualization might be used to determine which pages are accessed most frequently, or how usage varies across dates (or days of the week). The individual data points found in a given aggregation can be displayed in a second visualization, which might present time on the x-axis and hostname of the requesting computer on the y-axis, presenting each request for a given URL on a given day

as a single point. The displays are tightly-coupled: selection of an aggregate point in the first visualization is selected, leads to display of the points found in that aggregation in the second visualization window.

4. INCREASED CONTEXTUAL INFORMATION

The data found in web logs is heavily context-dependent: the requests that are made, and the relationships between those requests are strongly influenced by a variety of internal and external factors. Perhaps most obviously, the paths that users take will be largely determined by the links that are provided. Given the crucial role that site design can play in influencing site usage patterns, it seems clear that consideration of site topology might be useful for interpretation of web log data. For example, a tool that provided site layout information alongside log data might help users build understandings that tie both data sets together. Unfortunately, log visualization tools often fail to support the integrated display of this useful information.

A simple example of the use of STV to coordinate web log visualization involves coordination between an outline view of site URLs, a browser window displaying a page from the site, and a Spotfire window displaying requests to a URL by time (x-axis) and hostname (y-axis) (Figure 11). When the user selects a URL in the outline view, data for that page is displayed in the Spotfire window, while the page itself is loaded into the browser window. This provides the user with additional context that would not be available in a single visualization. This added context may simplify the process of understanding patterns in the data.

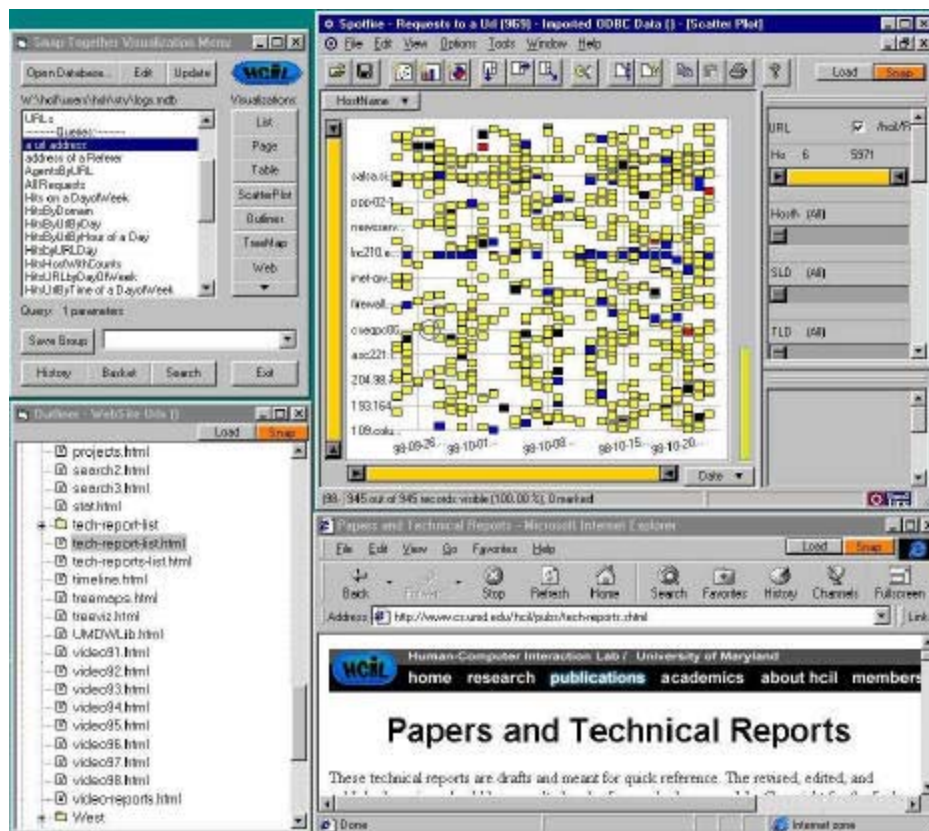


Figure 2: Coordinated visualizations for context: The outline window on the left-hand side provides a hierarchical view of URLs on the site, while the web browser window in the lower right corner displays a selected web page and the Spotfire display plots requests for a given URL, with time on the x-axis and hostname on the y-axis.

5. DISCUSSION & FUTURE WORK

By presenting two or more tightly coupled views, coordinated visualizations of web log data provide users with multiple perspectives which can be used to build interpretations and understandings in the appropriate context.

As STV provides a general-purpose, platform for visualization coordination, log data might be visualized alongside other relevant organizational data. For example, operators of e-commerce sites might construct coordinated visualizations that relate web log access patterns to customer purchase records.

These scenarios are just two applications of coordinated visualizations to web log data. As STV provides a general-purpose, database-driven, platform for visualization coordination, log data might be visualized alongside other relevant organizational data. For example, operators of e-commerce sites might construct coordinated visualizations that relate web log access patterns to customer purchase records. Furthermore, the flexibility provided by STV's use of a relational database provides the possibility of visualizing the results of arbitrary aggregations through SQL queries, side-by-side with "snapped" visualizations providing context and drill-down capabilities.

The utility of these coordinated visualizations might be improved by increasing the ease of constructing coordinated views and integrating these views with external data sources. Currently, development of coordinated visualizations using STV involves manual creation of appropriate SQL queries for aggregation and formatting the data. Appropriately designed tools could support the process of specifying queries and selecting the data sets to be coordinated and the tools used to generate the individual visualizations. Aggregation tools similar to Visage's outliner [10] or the Aggregation Eye [11] might simplify the process of specifying the desired aggregates. Further expressive power might be gained by increasing the range of visualization tools that can be used. Finally, tools that simplified the process of integrating the log data with external data sets might provide site operators with additional contextual information.

ACKNOWLEDGEMENTS

This research was supported by a grant from IBM's University Partnership Program. Thanks to Chris North for his assistance with Snap-Together Visualizations.

BIBLIOGRAPHY

- [1] Ahlberg, C., & Shneiderman, B. (1994) Visual information seeking: tight coupling of dynamic query filters with starfield displays. Proc. ACM CHI '94 Conference, ACM Press, New York, 313-317.
- [2] Chi, E., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., & Card, S. (1998). Visualizing the evolution of web ecologies. Proc. ACM CHI '98 Conference, ACM Press, New York, 400-407.
- [3] Chimera, R., & Shneiderman, B. (1994) An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents. ACM Transactions on Information Systems 12(4) October 1994, 383-406.
- [4] Cooley, R. Mobasher, B., & Srivastava, J.(1999). Data preparation for mining world wide web browsing patterns. Journal of Knowledge and Information Systems 1(1).
- [5] Cugini, J. & Scholtz, J. (1999). VISVIP: 3D Visualization of paths through web sites. Proceedings of the International Workshop on Web-Based Information Visualization (WebVis '99), in conjunction with DEXA '99 Tenth International Workshop on Database and Expert Systems Applications, 259-263.
- [6] Fredrikson, A., North, C., Plaisant, C., & Shneiderman, B. (1999) Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data Proc. of the Workshop on New Paradigms in Information Visualization and Manipulation, ACM Press, NY.
- [7] Goldstein, J., & Roth, S. F., (1994) Using aggregation and dynamic queries for exploring large data sets. Proc. ACM CHI '94 Conference, ACM Press, New York, 23-29.
- [8] Hochheiser, H. & Shneiderman, B (in press) Using interactive visualizations of WWW log data to characterize access patterns and inform site design. Journal of the American Society for Information Science, forthcoming.
- [9] Livny, M., Ramakrishnan, R., Beyer, K., Chen, G., Donjerkovic, D., Lawande, S., Myllymaki, J., & Wenger, K. (1997) DEVis: integrated querying and visual exploration of large datasets", Proc. ACM SIGMOD'97, ACM Press, New York, 301-312.
- [10] Kolojchick, J. & Roth, S. (1997) Information appliances in Visage. IEEE Computer Graphics and Applications, 17(4), July/August 1997, 32-41.
- [11] Mockus, A. (1998) Navigating aggregation spaces. Proc. IEEE Information Visualization Symposium 1998 Late Breaking Hot Topics Proceedings, IEEE Computer Society Press, 29-32.
- [12] North, C. & Shneiderman, B (2000). Snap-Together visualization: a user interface for coordinating visualizations via relational schemata. Conf. Proc. Advanced Visual Interfaces 2000, ACM Press, New York.
- [13] Spotfire. (1999). Spotfire [Online] Available at <http://www.spotfire.com> (Accessed June 16, 2000).