Applying Database Visualization to the World Wide Web

Masum Z. Hasan, Alberto O. Mendelzon and Dimitra Vista
{zmhasan,mendel,vista}@db.toronto.edu

Computer Systems Research Institute
University of Toronto
Toronto, Canada M5S 3H5

Abstract

In this paper, we present visualizations of parts of the network of documents comprising the World Wide Web. We describe how we are using the Hy$^+$ visualization system to visualize the portion of the World Wide Web explored during a browsing session. As the user browses, the web browser communicates the URL and title of each document fetched as well as all the anchors contained in the document. Hy$^+$ displays graphically the history of the navigation and multiple views of the structure of that portion of the web.

1 Introduction

World Wide Web browsers allow users to navigate the web in a hypertext style by clicking on the labels of the anchors of HTML documents[1]. However, as the browser session progresses, the user may quickly get out of context; the *lost in hyperspace* problem. Overview diagrams of hyperspaces (or parts of them) have been recognized as important aids to help both orientation and navigation through the space [Nie95, UY89].

In this paper, we show how we can generate overview diagrams easily, without having to significantly modify the web browser and without having to develop special tools for web visualization. We maintain a database with connectivity and other information about portions of the web and we use standard database functionality [CEH$^+$94] and standard visualization techniques [DETT94, Noi96] to produce the overview diagrams. In summary, these diagrams:

- are generated in a flexible manner through database querying;
- can be maintained using standard view materialization techniques;
- can be simplified by filtering and clustering to reduce clutter or demonstrate certain structural properties of the displayed information.

We have applied these ideas to create graphical presentations of two different parts of the web:

- *Visualization of the portions of the web visited during a browser navigation.* Browsing data are fed synchronously by the web browser to a database system. The database is updated with this information and all active visualizations, defined as views over the data, are updated immediately to reflect changes. These overviews provide a powerful history of the navigation that the user can refer to later on.
- *Visualization of server maps.* A *server map* is an overview diagram containing information about all the documents residing in a particular web server. In this case, the underlying data are generated either by a *robot* (a software system that traverses the web in order to collect information), or by executing a web query using some query language such as WebSQL [MMM96]. Server maps can be used by web site administrators to manage the structure of their site, or to help users navigate the site.

In this paper we emphasize the first kind of visualization. We describe how we modified and interfaced an early web browser, Mosaic 2.0, with the Hy$^+$ system, a general visualization system supporting database functionality. Section 2 discusses the interaction of the web browser with Hy$^+$, Section 3 presents some typical overview diagrams created by our system, and Section 4 discusses related work.

2 Interfacing Hy$^+$ to a Web Browser

The Hy$^+$ visualization system supports the visual presentation of structured data in the form of *hygraphs* [CEH$^+$94]. It provides a number of browsers and graphical editors to display and create the graphical presentations, and incorporates a layout engine that implements several different layout algorithms [Noi96]. Most important, Hy$^+$ supports a visual query language, called GraphLog [CM90], which allows visual query specification. GraphLog queries allow systematic filtering of data to reduce visual complexity, and support specification of new relationships among the data in much the same way as database views.

[1] An *anchor* is a pair of the label that can be clicked upon and the address, or URL, of the document associated with the label.

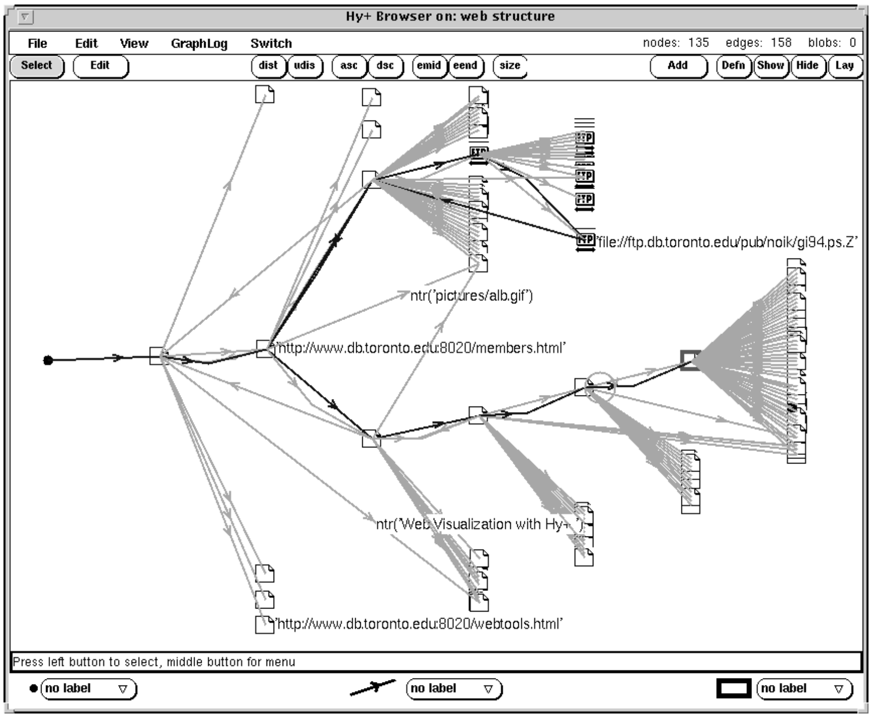

Figure 1: Web structure of a browsing session

We have interfaced the Hy$^+$ visualization system with the Mosaic 2.0 web browser [HMV95]. As the Mosaic session proceeds, the URL and title of each document fetched as well as all the anchors contained in the document are communicated to Hy$^+$. Hy$^+$, in turn, uses this information to update the different graphical views that it maintains about the sub-web seen so far. These views are internally specified through GraphLog queries and their number and content can be changed at any time. Queries on the complete database of all visited documents can also be issued at any time. In addition, all Hy$^+$ visualizations are active; by clicking on a node in a hygraph, the user can make the browser jump to the URL associated with the node.

3 Hy$^+$ Views

Figure 1 shows a Hy$^+$ generated graphical view of the sub-web seen by the user during a browsing session. In this view, there are two kind of edges, `ntr` (light color) and `tr` (dark color). An `ntr` edge is created from a visited document at URL `a` to any document with URL `b`, if `b` is cited in an anchor within document `a`; the label of the anchor is shown as the edge label, along with the type of edge (`ntr`). A `tr` edge is created from URL `a` to `b`, if `b` is visited immediately after `a`, either by clicking on an anchor inside `a`, or by clicking "back" or "forward", or by opening a new URL location while examining `a`.

The GraphLog query language can be used to restrict the set of documents displayed in this view, using very flexible criteria, including structural properties and regular expression matching on URL's, anchor labels, and document titles. Another advantage of this view is that the user has easy access to a document whose URL is referenced in an anchor within some visited document and can click on it directly, without having to fetch again the whole document that contains the anchor.

The history of visited nodes is shown in Figure 2. This view corresponds (to some extent) to the Mosaic history list. However, Mosaic and other web browsers do not keep any information corresponding to the `ntr` edges. They only maintain linear history lists that a user can refer to in order to return to a previously visited document. Moreover, these linear histories only provide a path from the origin of the navigation to the current document; all other branches are pruned. Our history graph is more general and shows more structure than just a linear list of visited nodes. This contextual information can be useful in identifying previously visited nodes that the user wants to revisit. In particular, since the browsing can be controlled from the visualization, clicking on any node causes Mosaic to jump to the associated document.

Figure 3 shows another view of the same web browsing session as above. This uses a feature of the data model supported by Hy$^+$, where in addition to associating a node with another node by an edge, any node can be associated with one or more labeled *sets* of

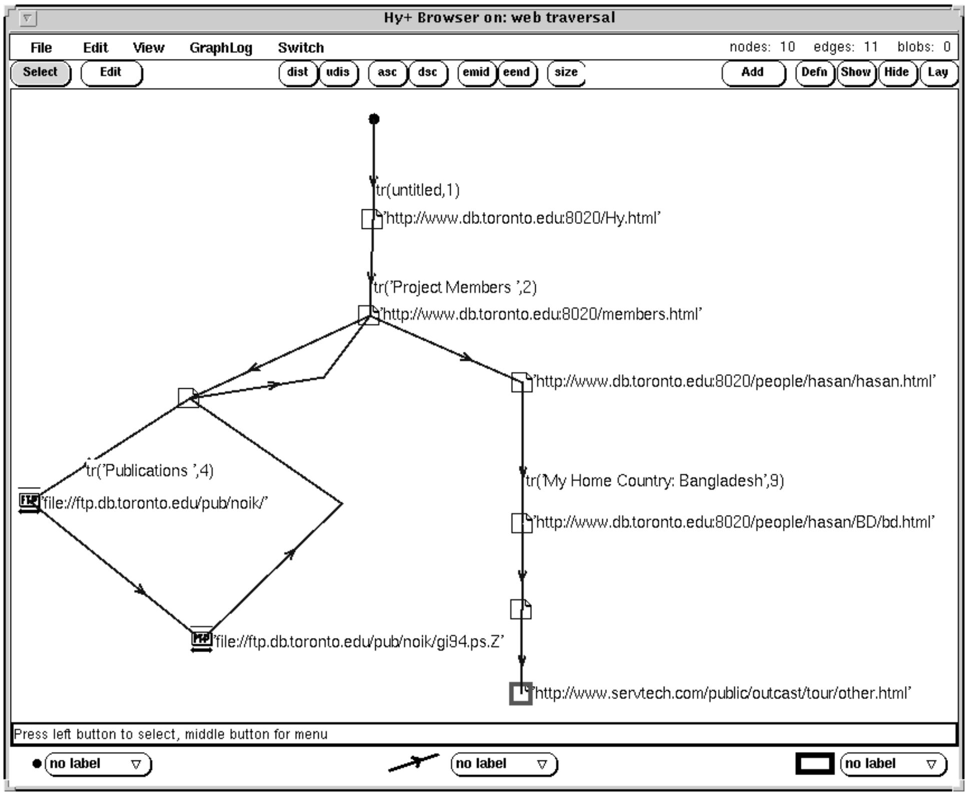

Figure 2: History of visited documents

nodes by placing these in boxes (called *blobs* in Hy$^+$ parlance). We use these boxes to represent HTML documents; the node at the top left-hand corner of a box is labeled with the title of the document, and each node inside a box represents an anchor (URL/label pair). An edge labeled `by_clicking` (dark color) is created from the anchor that was clicked to the document fetched as a result.

Boxes can be interactively collapsed to reduce cluttering. And different icons can be assigned to nodes based on predefined criteria to distinguish different kinds of nodes. For example, in our views different icons are used for different Internet access methods (http, ftp, gopher, etc.). Clustering as an aid to produce diagrams of the web has also been proposed by [MF95, MFH95].

Both the view definitions and the snapshots themselves can be stored and re-used in future sessions. An interesting direction, which we have not yet explored, is to merge views generated in multiple sessions to start building a global map of the portions of the web that are of interest to a user or group of users. Issues of scale, as discussed in [Men96], become more important in this context.

4 Related Work

The Navigational View Builder [MF95] supports overview diagrams of the web. However, these views do not correspond to any navigation of the web, and cannot be queried.

In the KJ-Editor [Nie95], the user can construct an overview diagram of the web, by deciding, for each page visited in the web browser, whether to add it to the overview.

WebViz [PK94] is a system for server access log analysis through visualization, developed at the Graphics Visualization & Usability Center at the Georgia Institute of Technology. It provides graphical information about document accesses and possible paths taken by users through a set of documents.

Other related work includes [Doe94, And95, AS95, GS95, TL95, HDWB95]. Some general principles for web visualization are discussed in [Men96].

Query languages for extracting subsets of the web, such as described in [LSS96, KS95, MMM96] could be used to produce interesting inputs for our views.

5 Conclusion

We have described a system that allows a user of a web browsing tool to get graphical overviews of his or

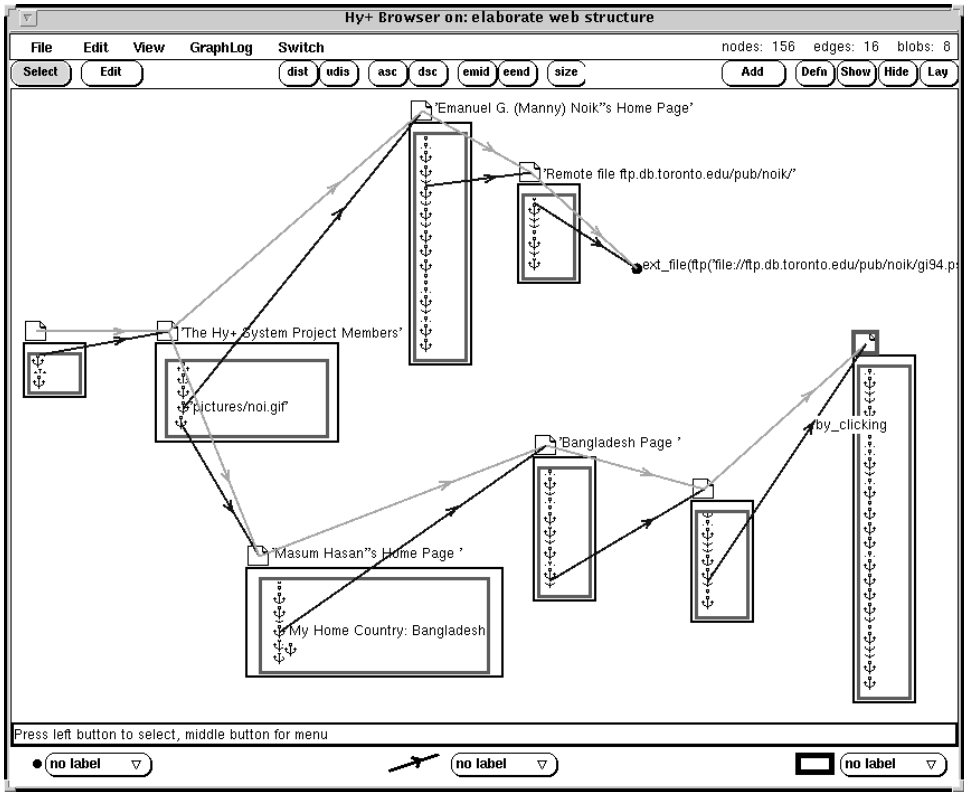

Figure 3: HTML documents viewed as blobs

her browsing session. The system provides views of not only the history of the navigation, but also the containment of links in visited documents, even if these are not followed in the navigation. These views are generated by executing queries of the graphical query language GraphLog. GraphLog queries can be used to further query the generated visualizations [HMV95]. All the views in the system are active, that is, the browser can be asked to fetch a document by clicking on the corresponding node in some view. Our system is unique in that it is based on a general purpose database visualization system instead of ad hoc techniques.

Acknowledgements

The work described was supported by the Information Technology Research Centre of Ontario and the Natural Sciences and Engineering Research Council of Canada.

References

[And95] Keith Andrews. Visualising cyberspace: Information visualization in the Harmony Internet browser. In *Proceedings '95 Information Visualization*, pages 97–104, Atlanta, October 1995.

[AS95] Eric Z. Ayers and John T. Stasko. Using graphic history in browsing the World Wide Web. In *Proceedings of the Fourth International World Wide Web Conference*, Boston, December 1995. `http://www.w3.org/pub/-Conferences/WWW4/Papers2/270`.

[CEH+94] M.P. Consens, F.Ch. Eigler, M.Z. Hasan, A.O. Mendelzon, M.G. Noik, A.G. Ryman, and D. Vista. Architecture and Applications of the Hy+ Visualization System. *IBM Systems Journal*, 33(3):458–476, 1994.

[CM90] M.P. Consens and A.O. Mendelzon. GraphLog: A Visual Formalism for Real Life Recursion. In *Proceedings of 9th ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, pages 404–416, 1990.

[DETT94] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. Algorithms for drawing graphs: an annotated bibliography. *Comput. Geom. Theory Appl.*, 4:235–282, 1994.

[Doe94] Peter Doemel. WebMap – a graphical hypertext navigation tool. In *Proceedings of the Second International World Wide Web Conference*, Chicago, October 1994. `http://-www.ncsa.uiuc.edu/SDG/IT94/Proceedings/-Searching/doemel/www-fall94.html`.

[GS95] Brain R. Gaines and Mildred R. Shaw. WebMap: Concept mapping on the Web. In *Proceedings of the Fourth International World*

Wide Web Conference, Boston, December 1995. `http://www.w3.org/pub/Conferences/-WWW4/Papers/134`.

[HDWB95] R.J. Hendley, N.S. Drew, A.M. Wood, and R. Beale. Narcissus: Visualizing information. In *Proceedings '95 Information Visualization*, pages 90–96, Atlanta, October 1995.

[HMV95] M. Hasan, A.O. Mendelzon, and D. Vista. Visual Web Surfing with Hy$^+$. In *Proceedings of the 1995 IBM CASCON Conference*, pages 524–535, 1995.

[KS95] D. Konopnicki and O. Shmueli. W3QS: A query system for the World Wide Web. In *Proc. of VLDB'95*, pages 54–65, 1995.

[LSS96] Laks V. S. Lakshmanan, Fereidoon Sadri, and Iyer N. Subramanian. A declarative language for querying and restructuring the Web. In *Proc. of 6th. International Workshop on Research Issues in Data Engineering, RIDE '96*, New Orleans, February 1996. In press.

[Men96] Alberto O. Mendelzon. Visualizing the World Wide Web. In *Proceedings of the International Workshop AVI '96*, pages 13–19, Gubbio, 1996.

[MF95] S. Mukherjea and J.D. Foley. Visualizing the World Wide Web with the Navigation View Builder. In *Computer Networks and ISDN Systems, Special Issue on the Third International Conference on the World Wide Web*, 1995.

[MFH95] S. Mukherjea, J.D. Foley, and S. Hudson. Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views. In *ACM SIGCHI*, 1995.

[MMM96] Alberto O. Mendelzon, George Mihaila, and Tova Milo. Querying the World Wide Web. In *Proc. PDIS '96*, December 1996.

[Nie95] J. Nielsen. *Multimedia and Hypertext, The Internet and Beyond*. AP Professional, first edition, 1995.

[Noi96] E.G. Noik. *Dynamic Fisheye Views: Combining Dynamic Queries and Mapping with Database Views*. PhD thesis, Dept. of Comp. Sci., U. of Toronto, April 1996.

[PK94] J.E. Pitkow and Bharat K.A. WebViz: A Tool for World Wide Web Access Log Analysis. In *Proceedings of the First International World-Wide Web Conference*, Geneva Switzerland, May 1994.

[TL95] Scott R. Tilley and Walter M. Lamia. Personalized information structures II: Hyperstructure hotlists. In *Proceedings of the 13th Annual International Conference on Systems Documentation (SIGDOC '95)*, pages 171–180, Savannah, Georgia, October 1995.

[UY89] K. Utting and N. Yankelovich. Context and Orientation in Hypermedia Networks. *ACM Transactions on Office Information Systems*, 7(1):58–84, 1989.