

# Integration of Browsing, Searching, and Filtering in an Applet for Web Information Access

Kent Wittenburg\*, Eric Sigman  
Bell Communications Research  
445 South St.  
Morristown, NJ 07960-1910  
{kentw, erics}@bellcore.com

## ABSTRACT

Improvements to information access on the World Wide Web has to be considered one of today's strategic challenges. In this paper we present a Java applet called AMIT (Animated Multiscale Interactive TreeViewer) that integrates fisheye tree browsing with search and filtering techniques. Used in combination with a web walker, a search server, and a tree server, it shows promise as a scalable solution to information access in configurable web spaces.

## Keywords

Information access, information visualization, search, browsing, filtering, animation, fisheye, World Wide Web.

## INTRODUCTION

The World Wide Web has brought on an unprecedented information glut. Despite the appearance of massive indexing and cataloging efforts, the current state of information access is a stark indicator of the difficulty in dealing with vast unstructured information spaces. Here we discuss AMIT, the client-side interface of a software suite for Web information access. AMIT (Animated Multiscale Interactive TreeViewer) integrates search, browsing, and filtering techniques. Font scaling and tree pruning are used to provide multifovea fisheye views [3] of tree structures over a customized Web domain. AMIT is designed as part of a software suite that includes web walking and metadata gathering as well as indexing and search. We are now deploying it for access to a webspace of approximately 12,000 documents distributed over 4500 hosts (<http://community.bellcore.com/mbr/sailing-page.html>).

Interfaces for information access on the Web today tend either to use search engines to match queries and rank-order results or else to manually construct HTML-based hierarchies of Web domains. The search services are faced with the problems of how to manage and structure the often huge hit sets returned by their queries, how to include some form of quality control, and how to focus the queries and follow-ups with easy-to-use interfaces. The manual cataloging efforts are limited by the amount of labor required to clas-

sify and maintain such large and rapidly changing document collections. Also, it seems evident that multiple organizations of web spaces are desirable to support different tasks and users and that methods are needed for focusing within and across given hierarchies and/or domains. More recently, there has been interest in community-based filtering for Web information access. Our experience [4] suggests that community-based metadata is probably most useful as a scoring function. Browsing and searching technologies are still needed.

## AMIT

The AMIT client applet integrates hierarchical browsing with content-based search. It incorporates community-based filtering and/or alternate scoring functions through multi-scale fisheye visualization. Animation is used to provide transitions between the many customized views that users can construct.

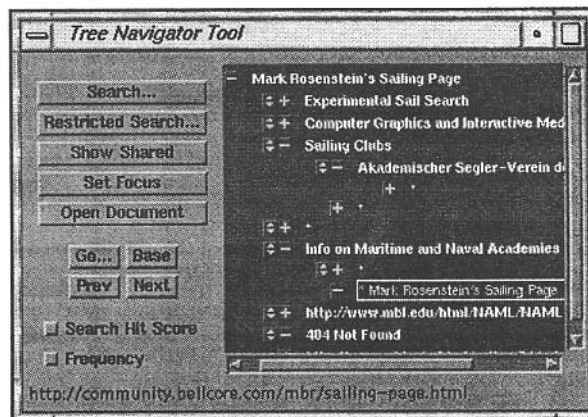


Figure 1. A tree-based view of a web space.

## Browsing

Figure 1 shows an AMIT view of a collection of web documents. Specification with an off-line web walker determines the underlying web space. The tree structure is imposed on the underlying directed graph link structure by picking a starting root node and including outgoing links of each document only once. Duplicate references to the same document appear as differently styled nodes -- the selected node in Figure 1 is an example. Users can navigate to a view of

\* First author's current address: GTE Labs, 40 Sylvan Rd., Waltham, MA 02254, kwittenburg@gte.com.

the space that includes all occurrences of a document by selecting a node and clicking on the "Show Shared" button.

We have expanded the set of usual tree manipulation gestures because of the size and topology of Web link trees. Nodes with no text represent a reduction of a contiguous set of sibling nodes. Interactively, users can reduce siblings by clicking on the double-arrow icon. An interactive slider appears that reflects the possible space of the reduction and feedback is given through graying out of nodes. Another navigational move is afforded with the "Set Focus" button. Users can select one or more nodes that they want to remain in the succeeding view, and the system automatically reduces the rest of the tree.

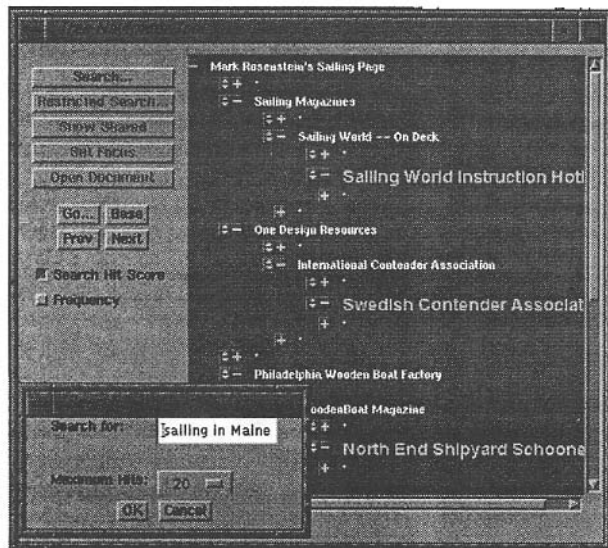


Figure 2. A query and a view of its result set.

### Search

The off-line web walker that gathers the link metadata underlying the tree view also gathers text for content-based indexing, for which we use Latent Semantic Indexing (LSI) [1]. At runtime, when a user makes a query, an intermediate server forwards the query to an LSI search engine, and hit results along with relevance scores are returned. AMIT then generates a view in which all hits up to a configurable threshold appear in the tree (with a different color). As shown in Figure 2, the tree is pruned with an algorithm that retains all paths to the root of the tree and reduces all remaining nodes either as subtree or sibling collapses. The size of a node is a function of its score.

We suspect that the most significant feature of the tool is these hit set views. One of the primary difficulties users have in dealing with the large hit sets returned from Web searches is the problem of quickly evaluating the relevance of the hits. As is evident in Figure 2, AMIT shows hits in context. Users can quickly see that some hits have to do with one-design boats, others with wooden boats, etc.

Users can (re)focus the query by selecting one or more nodes of the tree and picking "Restricted Search," an indication that only those hits that appear in the subtrees of the

nodes selected are to be returned. The result tree views can be manipulated through the threshold or scoring functions.

### Filtering

Scoring functions from most content-based search engines, LSI included, are opaque to users. AMIT offers the possibility of using a variety of scoring/filtering functions. By choosing a different scoring function, a user can specify succeeding views in which node text is rescaled and the tree is repruned. In the current implementation we support the choice of frequency-based scoring (which is a measure of how often a given document is referenced in the collection as a whole), LSI relevancy scoring, or a combination of the two. Community-based usage, recommendations, or server log information could easily be added.

### Animation

All tree manipulations and view transitions in AMIT are animated. A navigable dialog history is maintained. Our hypothesis is that not only will users be better able to maintain context visually with animation, but they will be better able to understand the effect of different scoring functions and threshold manipulations, thus directly supporting the task at hand.

### CONCLUSION

AMIT is an interface that integrates three methods for Web information access: search, browsing, and filtering. There is already evidence that information access is improved by posting search hits against an interactive tree structure [2]. AMIT also affords the hypotheses that querying against a Web space can be improved through tree-based focusing and that multiscaling combined with animation is an effective method for visualizing alternative evaluation functions.

### ACKNOWLEDGMENTS

We thank Joel Remde, Larry Stead, and Mark Rosenstein for web walking, Sue Dumais and Todd Letsche for LSI tools, and Louis Weitzman for related work. This research was supported in part by ARPA grant N66001-94-C-6039.

### REFERENCES

1. Deerwester, S., Dumais, S., T., Landauer, T. K., Furnas, G. W. and Harshman, R. A. Indexing by latent semantic analysis. *Journal of the Society for Information Science* 41, 6 (1990), 391-407.
2. Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., and Lochbaum, C. C., Formative design-evaluation of SuperBook, *ACM Transactions on Information Systems* 7, 1 (January 1989), 30-57.
3. Furnas, G.W. Generalized fisheye views. In *CHI '86* (Boston MA, April 1986), ACM Press, 16-23.
4. Wittenburg, K., Das, D., Stead, L., and Hill, W. Group asynchronous browsing on the World Wide Web. In *Proceedings of Fourth International World Wide Web Conference* (Boston MA, December 1995), O'Reilly, 51-62. [<http://www.w3.org/pub/Conferences/WWW4/Papers/98/>].