WebJournal: Visualization of a Web Journey

Bipin C. **Desai**
Stan **Swiercz**
Department of Computer Science
Concordia University
Montreal, CANADA
{bcdesai, stan}@cse.concordia.ca
http://users.encs.concordia.ca/{~bcdesai/, ~staff/stan/}

Introduction

An increasing number of research institutes, universities and business organizations are currently providing their reports, articles, catalogs and other information resources on the Internet in general and the Web[BERN, BERN3] in particular. This is now becoming the accepted method of disseminating and sharing information resources in hypermedia. At this time a number of information sources, both public (free) and private (available for a fee), are available on the Internet. They include: text, computer programs, books, electronic journals, newspapers, organizational local and national directories of various types, sound and voice recordings, images, video clips, scientific data. Also, private information services such as price lists and quotations, databases of products and services, and speciality newsletters are available.

The current trend in the use of hypermedia and many of the documents accessible via the Internet using various browsers are great showcases for the features of the system. The practicality, however, of using links to related documents is both over utilized and a cause of annoyance to the users. A link in a document being perused leads the reader from one document to another, with in turn more links to tertiary documents. The user eventually gets lost and feels trapped in the web.

A number of browsers have been developed which allow ease in accessing this information on the Web. Mosaic is one of the more popular ones. Users, while accessing such information resources, follow links from one document to another. Without a good method of recording their travels on this information highway, they could get "lost". Some of these browsers, record the users path in a list and allow the user to return to any of these pre-visited documents. However, the relationship among the pre-visted documents is not preserved. Nor is it possible to select and save the pointers to only a subset of these pre-visited documents except in a hot-list which may become too large.

As such, there is a need for a system which allows the visualization of a journey through the Web and furthermore provides ease in selectively marking documents for later review or sharing with colleagues.

The WebJournal[DESA] is such a graphical system which provides users with a pictorial representation of their search of resources on the Internet and for allowing random access to these visited resources. It keeps track of links including figures and postscript documents. In addition, the WebJournal provides direct access to any previously visited resources. Marking interesting resources and collecting these and storing them in a named file for later perusal or sharing is also possible. The marking of selected documents for printing is also possible.

Journaling a Web Journey

WebJournal, in the current implementation, works in conjunction with the Mosaic browser[MOSA]. The latter is modified to allow it to share information with the Webjournal and provide with a method of displaying required documents. The WebJournal provides a two dimensional tree of visited hypermedia documents. The nodes could be labelled with the order of its visit, the URL[BERN1] of the resource visited or the title of the resource visited[BERN2]. The user can select any of these labelling schemes. In the latter two cases, however, the nodes tend to be large and the visualization of the journey is lost due to the limited size of the screen.

Figure 1 shows the status of the WebJournal beginning with a local home page(node 1). The next node(node 2) is one that is reached when the user directly enters an URL. The third node is reached by following a pointer in the second document. The fourth node is reached from the third node which in turn leads to a fifth node: a Postscript figure in the document corresponding to the fourth node. This Postscript would have been opened by the browser using an external application program such as Ghostview, or XV.

Layout and Semantics of WebJournal

WebJournal displays each visited hypermedia resource (represented by an URL) as a node in a tree. This tree "grows" from left to right. There are three methods of labelling the displayed modes. The default mode is to label each node with its ordinal number. The other two labelling possibilities are to use the node's actual URL or node's title . However, as the URL's and titles can be rather long, this results in the nodes being rather large. Large nodes use up much of the

screen allowing one to view only a small part of a Web journey for a given size of the WebJournal window. As such it is preferable to leave the display in the default mode. The default mode can be changed by selecting the Display menu option in the Options menu.

Besides labelling the nodes, the user can also control how certain nodes are shown. URL's can have hyper-links to other sections of the same document. By default WebJournal does not display these sections as separate nodes even though the URL's are different (there are "#" found in the section URL's to indicate what the anchor is). If the user wishes to display all distinct URL's as nodes then they can do so by selecting the Show Self Links menu button in the Options menu.

It is possible that two different looking documents have the same URL. This happens with CGI executable that send confirmations. The titles may be different but the URL's are identical. WebJournal only takes into account the URL. This implies that the confirmation will not show up in WebJournal.

It is also possible for the same document to have two different URL's. This happens when one WWW site has different names for itself or the document is found in more than one location on the site. As well, in the case of a site home page, the file name may be missing in one instance and if the user selects the "back to home document" anchor, a URL with the file name may arrive. WebJournal will treat these URL's as different nodes.

Semantics of Nodes

The node whose URL and title are displayed in the top part of the WebJournal has its label displayed in bold. Single clicking on a previously visited node will change the label to bold and the URL and title information displayed at the top part of the WebJournal will change to show the information of this node. This allows the user to search for a URL or title they had visited earlier and would want to have the corresponding resource displayed again. Single clicking does not change the contents of the browser window.

The node for the URL (active node) that is currently being displayed in the browser window is "highlighted" by using dark label characters on a light background. Previously visited nodes (inactive nodes) are displayed using light label characters on a dark background.

Double clicking on a node will "highlight" the node and the characters will appear in bold. This will cause a message to be sent to browser (Mosaic in the current implementation) requesting a reload of the URL associated with that node. Note that Mosaic at this point does

not check if the URL is in its cache. It will retrieve the document. In this sense, the graph of the browser's visit of URL's is not the same as that displayed in the WebJournal.

URL's that use an external viewer (gifs, jpegs, mpegs, sound files, Postscript documents, binary ftp'd files for example) do not display in Mosaic's window and thus the nodes associated with these URL's do not get highlighted. The node characters are displayed in bold. Double clicking on such a URL will not highlight it even though Mosaic will be sent the request. The node for the URL displayed in Mosaic's window will remain highlighted.

Semantics of Edges

The tree created by WebJournal is linked together with solid lines. There are also four different types of links denoted by dashed lines that link nodes outside of the regular tree structure. These dashed links "emerge" from the bottom of a node (to indicate where the link started) and go to the top of another node (the destination or new node).

The first type indicates that a user has selected an anchor in a URL to a URL that has already been seen (in the same Mosaic window).

The second type shows that the user has entered in a URL manually using either the "Open Url" or "Open Local" buttons in the Mosaic's File menu.

The third type shows that the user has selected a URL from Mosaic's hotlist.

The fourth type (clone link) displays that the user has opened another Mosaic window using either the Clone button or using the middle mouse button to select an anchor. If the Clone button was pressed the link will be from the node with the same URL as the new node. If the user pressed the middle mouse button then the link will be from the node where the anchor was selected. If the user pressed the middle button and a node already existed for the new URL in the branch of the original Mosaic window then the link will originate from that node.

URL's that are the same but in different Mosaic windows are distinct. Links of the first three types are only created for nodes in the same Mosaic window.

Handling Multiple Browser Windows

Browsers such as Mosaic allow a user to open multiple windows to view different parts of the same document or a number of different

documents. Such cloning of the browser windows in Mosaic is done by using the browser's Clone button or by using the middle button of the mouse. Multiple windows allow subsequent independent travel to different parts of the hypermedia hyperspace. Multiple windows are represented in the WebJournal by designating a highlighted node for each window. The most recent resource visited is represented by displaying the contents of the corresponding node in bold. The cloning is represented by a "clone" edge.

Managing the WebJournal Real Estate

WebJournal tries to keep all the branches for a Mosaic window together. If the user for example manually enters a URL and there already exists clone branches, WebJournal will move the clone branches down to make room for the new node.

Double clicking on a node will result in the URL being displayed in the window where it originally appeared. In a case where the window has been deleted, the URL would appear in the "oldest" Mosaic window. A warning message would appear in the WebJournal window. A new node would be created in the branch of the tree corresponding to the "oldest" window unless a node with the same URL already exists in the "oldest" window. In this case, that node will be highlighted.

1) light character & dark background	visited node
2) dark character & light background	a node in this pruned branch is currently displayed
3) same as 1 but the first node is "highlighted"	visited node that has its URL and title displayed at top
4) same as 2 but the first node is "highlighted"	a node in this pruned branch is currently displayed and has its URL and title displayed at top

As the Web journey continues, the tree in the WebJournal grows in size. Such a growing tree can take up a lot of space and the user may find it hard to visualize the recent path. At such points, the user can prune the tree by using the Prune/Expand button found at the bottom of WebJournal window. The label in the Prune/Expand button indicates whether the current node (indicated by the characters of its label being in bold) can be pruned or expanded. The pruned part of the tree, though not displayed, is stored internally and can be recalled by using the "Prune/Expand" button. The pruned branch of the tree is replaced by a graphic representation of a tree as shown in Figure 2.

There are four types of pruned nodes as noted above. They correspond to the state of a regular node.

Single clicking or double clicking on a prune node acts on the root of the prune branches.

If a display node is set to Show Self Links and a node, which is a self anchor to another node, is pruned and then Show Self Links is turned off, that node should disappear but it might have other URL's "under" it. In such a case, WebJournal displays the pruned node but it becomes insensitive to clicks. The lines drawn in this button are fuzzier. Also the Prune/Expand button also becomes insensitive if this node button is current. The only way to expand such an insensitized node is to set Show Self Link and then expand the insensitized node using the it Prune/Expand button.

Marking Nodes, Saving, Reloading and Printing

A node may be "marked" for later use by using the mark button found at the bottom of WebJournal's window. This button works on the current node (i.e., whose label is in bold). The mark button is insensitive to clicks when the current node is pruned. When a node is marked its label changes to a slanted font. All occurrences of that document will be marked. If there are any nodes that have the same "root" URL (i.e., if a node is a section of the document or the same URL in a different Mosaic window) they will be considered the same document.

A list of URLs of marked nodes can be saved in a file using the Save Marked button of the File menu. This file can then be loaded at a future date using the Load button of the File menu. When URLs are loaded a node is created with a label consisting of an asterisk. This is to indicate that what follows originally came from a WebJournal file. This node is linked with a dashed line indicating a manually entered URL. This node is associated with the "oldest" Mosaic window. There can only be one asterisk node per Mosaic window. With the deletion of Mosaic windows it is possible to have more than one asterisk node.

An example of such a file list is shown in Figure 3. The order of URL's in the file is the order in which nodes were marked and it is in the form of a linear list. The tree structure is lost.

Marking nodes and collecting them can also be used for later printing the contents of the resources represented by the nodes. This feature would alleviate the current practice of some suppliers of over utilizing hyper-links and making up a document using small sections. Such a practice is also used by many converters. A user wanting to

read such a document, off line can mark these nodes and save them in a file. This file can be used to download the corresponding documents and print them.

Marked Nodes and Printing

Marked nodes can be printed from WebJournal by selecting The "Print Marked" File option. A dialog will popup where you can modify the print command and the format to use to print the documents. Once set up the user selects "Ok" and journal opens a new Mosaic window to do the printing. If the URL is for an internal Mosaic document then once the URL is retrieved WebJournal will tell Mosaic to print the document. Webjournal will then instruct Mosaic to fetch the next URL. The user is kept informed with an information popup window.

In the case where the URL requires an external viewer a window pops up telling the user that they should use whatever commands available in the external viewer to print what it is displaying. Once the user is done with the external viewer they should press the "Ok" button in this popup so that WebJournal can continue. This is done so that the user's screen is not filled with several xv, ghostviews, xdvi, etc. Note that if a marked URL is for a sound file the use should select Ok right away. For xdvi documents there is no print option available. The user however should see the name of the temporary file where the dvi file is found. They could copy this file to another file and use whatever dvi converter they have available.

Once all the marked documents are printed the clone Mosaic window used for printing is automatically closed.

Exploding a Node

When the Explode button, found at the bottom of WebJournal, is pressed the current URL is examined for the number of hyperlinks found within it. Note that the Explode button is only sensitive when a node is both highlighted and current (i.e. the font used is bold). WebJournal examines the source of the selected node to determine the links in it and the number of distinct URLs, therein. Some of the links could be to anchors to different sections of the same document. WebJournal will consider all these links as a single distinct URL. Once the number of distinct URLs are determined, WebJournal informs the user of this number by a pop-up window. If there are some links that point to the same document then the number of distinct nodes is also shown. If all the URL's are from one distinct document then WebJournal will inform the user of it. If the one distinct document happens to be the node that is being exploded, this fact is also mentioned to the user. In this case there is no point in exploding the node.

The popup window then asks whether the user wishes to explode the node given the information provided. The user can either select the yes or no button. If the use selects yes then another window pops up asking whether the nodes
coming from the explosion (fragment nodes) should be marked or not. One reason to mark these nodes would be to print out the exploded set of documents.

This feature should be used with great care. Exploding a node can cause the creation of several hundred new nodes that will slow down the display. Exploding one of the "fragment nodes" could produce additonal edges represented by dashed lines to existing fragments. Such dashed which would clutter the WebJournal window and could make it difficult to visualize the Web journey.

This feature was designed so that a user who finds an online manual and wishes to print it, has an easy way of bring up the individual URLs and marking them instead of requiring a manual visit to each URL. Such manuals visitation for documents and manuals having a large number of sections and chapters would be a long tedious task.

WebJournal Internals.

WebJournal is written in Motif 1.2.3. The widget used to draw the nodes and edges is a modified version of the Tree widget described in [YOUN][1]. Extra code was added to this widget to draw the dashed lines.

Communications between Mosaic and WebJournal is through 2 *pipes*[BACH]. One pipe is for communication from Mosaic to WebJournal while the other one handles communication in the reverse direction. Instructions between the processes are very simple instructions such as fetch this URL or such a URL has arrived. X input event handlers are created in both programs to handle their inputs.

Conclusions

Our experience of using the web as an information discovery tool time and again and being "trapped" in it, resulted in the conception of our version of a safety net to avoid such a trap. In this paper, we describe how the WebJournal helps its users find their bearings. The additional benefits of this tool are to record and share the highlights of an interesting journey. The WebJournal's beta version will be publicly released and available from:

[1]A highly recommended text for those intrested in X Window applications.

ftp://ftp.cs.concordia.ca/pub/bcd/WebJournal/

References

[BACH] Bach, Maurice J., "The design of the Unix Operating System", Prentice-Hall, 1986

[BERN] Berners-Lee, T., & Cailliau, R., WorldWideWeb: Proposal for a HyperText Project http://info.cern.ch/hypertext/WWW/Proposal.html

[BERN1] Berners-Lee, T. UR* and The Names and Addresses of WWW objects,
http://info.cern.ch/hypertext/WWW/Addressing/Addressing.html
see also RFC 1738,

[BERN2] Berners-Lee, Tim, Connolly, Hypertext Markup Language, Internet working draft,
http://info.cern.ch/hypertext/WWW/MarkUp/HTML.html

[BERN3] Berners-Lee, T. Wide Web Initiative: The Project,
http://info.cern.ch/hypertext/WWW/TheProject

[DESA] Desai, Bipin C., WebJournal: Visualization of Web Journey, August 1994, http://www.cs.concordia.ca/WebJournal.html

[MOSA] NCSA Mosaic
http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html

[YOUN] Young, Douglas, "The X Window System, Programming and Applications with Xt", 2nd Edition, Prentice-Hall, 1994.