Head-Tail Display: a Lightweight Approach to Query-Dependent Document Display

D. Berleant[1,2], J. Miao[3], M. Arvold[2,4], J. Brown[2,4], R. DeVries[2,4],
T. Drucker[2,4], L. Elkin[2,4], C. Gofron[2,4], and K.-H. Lim[2,4]

[1]Virtual Reality Applications Center, Information Infrastructure Institute, and [2]Dept. of Electrical and Computer Engineering, Iowa State University, 2215 Coover Hall, Ames, Iowa 50011; [3]Ricoh Innovations, Inc., 2882 Sand Hill Road Suite 115, Menlo Park, CA 94025; [4]former affiliation

berleant@iastate.edu, jhmiao@rii.ricoh.com

ABSTRACT

The value of showing important, yet separated, parts of a document simultaneously motivates head-tail display. 35% of Web documents tested benefit. A head-tail display provides a query-dependent view of a document using a split window. One subwindow shows the beginning of the document, a particularly important part of many documents. The other subwindow shows the query-relevant document "tail," starting from the first query term occurrence.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Theory and methods*.

General Terms: Documentation, Human Factors, Theory.

Keywords: Head-tail display, search engine interfaces.

1. HEAD-TAIL DISPLAY

Simultaneously presenting both the beginning of a document and its query-relevant tail is a reasonable design objective, because it makes sense to display important parts of a document simultaneously. Automatically informing users when query terms are not present also makes sense. Major Web search engines currently do not meet this set of requirements. Head-tail display is a lightweight, readily implemented way to provide these features for the significant fraction of Web documents that would benefit.

In a head-tail display, a document is divided into two parts, shown in adjoining upper and lower subwindows. The initial view in the upper subwindow is the beginning of the document. This subwindow also contains all other material in the document preceding the first query term occurrence, accessible via a scroll bar. In contrast, the initial view in the lower subwindow is query relevant, starting at the first query term occurrence. The lower subwindow also contains all material after that occurrence, and a scroll bar. Thus two parts of the document, both likely to be important but perhaps widely separated, are presented simultaneously (Figure 1).

A head-tail display is readily implementable in either a browser or a server. Our prototype system is server based. It outputs HTML

specifying a head-tail display, which is then presented by the user's Web browser. Because no commercial search engine outputs head-tail displays (at least not yet), the prototype is a proxy server that mediates between search engine and user.

To use the system, the user accesses the system's URL using a standard browser, types a query, and clicks to send it to a standard search engine. The search engine returns a list of URLs. Hyperlinks in the list are modified to point to a servlet to which is passed, as parameters, the user's query and the original hyperlink. *Example*: the hyperlink

    Nepal

is modified to:

Nepal.

The list of modified hyperlinks is then forwarded to the user's browser. Referring to the modified example, if the user clicks on the **Nepal** link, the **SplitWin** servlet is called with parameters **rigorousness** (the user's query), and the URL **http://asiatravel.com/nepalinfo.html**. The servlet downloads that URL and divides it into two parts, one preceding the first query term occurrence and one following it. The parts are then stored in head and tail files. To preserve original formatting in the tail, the tail file starts with appropriate HTML tags copied from the head, but with all head content removed. (Documents consisting of frames are passed through unmodified.)

HTML source specifying two frames, the upper one referencing the head file and the lower one the tail file, is then generated and sent to the user's browser. See Figure 1. (New hyperlinks within the document are also added to connect each query term instance with the next. This is just one of many features that can supplement a head-tail display foundation.) The boundary between the two subwindows is easily dragged to expand one of them, a property of the HTML frameset tag, if the user wishes.

2. EVALUATION

Head-tail display provides automatic, simultaneous access to both the document beginning and to a query-relevant tail. *This saves the user at least one mouse click compared to meta-data or toolbar based information access.*

Another aspect of head-tail display needing evaluation is its rate of applicability to real documents. We examined 200 Web documents consisting of the first ten returned for each of 20 diverse, typical queries. The 20 were composed of four groups: five 1-term queries as well as five each of 2-term, 3-term, and 4-term queries. One query in each group was manually given to Google, another to AltaVista, etc., resulting in testing of five search engines in all.

For 65% of the documents, the first query term occurrence was sufficiently close to the beginning to appear in a standard full-screen browser display using Internet Explorer with its default settings running on a Windows pc. In these cases, elementary methods like query term occurrence highlighting would probably be sufficient. However, in 26.5% of the documents the first query term occurrence was not visible on the screen without scrolling down to find it. In these cases, head-tail display would enable users to see the beginning and the query-relevant tail simultaneously. In the remaining 8.5% of the documents no query term was present. This percentage is significant considering the time and distraction a user can experience when presented with such a document without warning. Apparently those documents had changed since last crawled (or were mis-indexed). In these cases the tail is empty, so the lower subwindow is blank, making the absence of a query term occurrence immediately obvious. Thus the benefits of head-tail display applied to 35% (=26.5%+8.5%) of the documents. Thus, head-tail display applies to a significant percentage of Web documents. No clear differences emerged either among different search engines, or for different numbers of terms in a query.

3. RELATED WORK

Byrd [1] described using colored dots in a scroll bar to represent the locations of query terms in a document. While this provides interesting meta-data, head-tail display shows relevant content directly. Head-tail display and scroll bars enhanced with colored dots are complementary, however, and could be integrated into the same system. To provide even more guidance to users, a scroll bar can simultaneously contain both the shaded rectangular tiles of TileBars [2] *and* colored dots [4].

Kaugars' multiple fisheye views approach [3], like head-tail display, presents actual content. However multiple fisheye views, unlike head-tail display, can present the text surrounding *several* key term occurrences simultaneously. Unfortunately Kaugars' fisheye views, unlike head-tail displays, do not guarantee a readable initial display of the beginning of the document. Nevertheless, the beginning of the document *could* automatically be given special "fisheye" treatment. Similarly, head-tail display could be extended from 2 subwindows to several.

4. CONCLUSION & RECOMMENDATIONS

An interface approach was described which can present a document view not possible using standard interfaces. It is *prima facie* beneficial for displaying a sizeable subset of Web documents, as evidenced by 35% of our test corpus. Of course, a fuller investigation will need to also study actual users. Head-tail display is suited to adoption by Web search engines, with the caveat that head-tail display is only applicable to some documents, so systems should decide whether to use it to display any given document dynamically based on the properties of the document.

The reader may enjoy considering the many extensions that are possible based on a head-tail display core.

5. REFERENCES

[1] Byrd, D., A scrollbar-based visualization for document navigation, *DL '99*, Berkeley, CA, 122-129.

[2] Hearst, M., TileBars: visualization of term distribution information in full text information access, *CHI '95*, Denver, CO, 59-66.

[3] Kaugars, K.J., A hierarchical approach to detail + context views, dissertation, New Mexico State University, 1998.

[4] Berleant, D., Miao, J., Gu, Z., and Xu, D., Toward dialogues with documents: MultiBrowser, in *Int. Conf. on Inf. Tech.* (*ITCC* 2004), Las Vegas, Vol. 1, 287-294.

Figure 1. Head-tail display of a document. The upper subwindow initially shows the beginning of the document, thereby helping to communicate broad subject and context. The lower subwindow initially shows the query-relevant tail of the document beginning at the first query term occurrence, thereby helping to efficiently communicate query relevance.