8-15-1997

Providing Metainformation for World Wide WebLinks and Documents

Praveen Ramanathan
New Jersey Institute of Technology, praveen@homer.njit.edu

Michael Bieber
New Jersey Institute of Technology, bieber@cis.njit.edu

Ajaz Rana
New Jersey Institute of Technology, rana@cis.njit.edu

Providing Metainformation for World Wide Web Links and Documents

Praveen Ramanathan, Michael Bieber& Ajaz Rana
New Jersey Center for Multimedia Research
Hypermedia Information Systems Research Laboratory
Institute for Integrated Systems Research
New Jersey Institute of Technology
University Heights - Newark, NJ 07102 USA
praveen@homer.njit.edu; http://www.cis.njit.edu/~praveen
bieber@cis.njit.edu; http://megahertz.njit.edu/~bieber
rana@cis.njit.edu; http://www.cis.njit.edu/~rana

Abstract

Disorientation and efficiency loss are two major problems currently faced by WWW users. We counteract these through a system that would provides users with additional information about documents & links. This paper discusses motivation, the experiments to evaluate the system and our experimental measures.

Motivation

The World Wide Web has a large amount of information. A web site consists of a series of web-pages that contains information linked together in a meaningful way. As users browse through these Web pages, they generally try to grasp the information they encounter, as well as try to get a feel for what other information is available. In doing so they inevitably face a lot of problems.

One problem is *disorientation* [THH95]. As a result of traversing a large number of links, users can lose their sense of direction and end up feeling confused. Users getting lost in cyberspace generally use the browsers back key in order to get back to familiar territory. Maintaining a sense of how they arrived at a particular document might decrease the disorientation.

Another problem that users face is that while traversing a link they reach certain nodes which do not contain the information that they require. These users generally tend to return to the previous node and continue their search. If they could get a better understanding of where exactly a link leads, they may be able to avoid following unnecessary links.

Upon arriving at a document, users generally scan the entire document before they are able to assess the contents. Providing a summary or other characteristics of the document separately might allow users to decide whether to read a document fully, thus increasing user efficiency.

These are some aspects we hoped to address in the electronic version of the August 1995 special issue [BI95] on Hypermedia Design in the *Communications of the ACM* (hereafter called the "CACM system") [Bi97]. We had some additional motivation. First, we believe that experiencing papers about hypertext design within a hypertext environment help readers understand the concepts better. Second, we believe the paper links used by authors (e.g., "see section X for more details") are often vague and rarely point to exact positions within the destination paper. On-line cross-reference links help in pointing these out clearly. Third, we believe that many authors are not using the hypertext features of the WWW to their fullest advantage [BVA97]. To begin addressing these issues, our electronic version implements link labeling and document chunking.

The CACM System

The CACM system was developed at the Hypermedia Information Systems Research Lab at NJIT. It was designed to present information so that users could tell where they are, where they came from, and where they can go. Icons, semantic link labels and orienting graphics are used to provide the user with extra information about his or her current position.

But, unlike most electronic versions of documents, this was not developed to be a static system. Information presented could be filtered according to user preferences. Users are allowed to control: link information (link descriptions and icons reflecting the destination type: paper, footnote, figure, etc.); destination information (paper title, authors, destination type, and destination section number); and layout information (page layout, incoming link information, including links to previous/next sections, and table of contents). The system also places an "arrival icon" at the exact target point in the destination document. This reminds users of how they arrived at a document. For our experiments we have created several fixed environments using specific combinations of these types of metainformation.

The system is designed using browser frames as containers for different types of this metainformation. Frames allow us to design the system in a way that will allow regular document content to be show, but with a more organized feel to it. We keep metainformation out of the main document. For example, we place destination information in one frame and place arrival information in another frame. We used this model to develop the system.

The Experiments

If we can prove that providing metainformation allows users to interact with the Web more effectively, this would call for those developing browsers and new HTML standards to incorporate hooks to maintain and present metainformation. To this end we are designing a series of experiments. We have three basic hypotheses.

- (H1) Providing link or document metainformation makes users more effective than providing neither.
- (H2) Only providing metainformation about links makes users more effective than only providing metainformation about documents.
- (H3) Providing both link and document metainformation makes users more effective than providing just one type of metainformation.

The CACM system is broken down into the following frames:

- Frame 1 - the original document
- Frame 2 - current document information
- Frame 3 - arrival information
- Frame 4 - destination information
- Frame 5 - control information
- (copyright, etc.)

We have already designed and implemented the system. We currently are designing our experiments. During the experiments we will gather data through user logs, and through experimental activities (such as answering questions). Then comes data analysis. For the experiment we developed five versions of the CACM system with each version containing some combination of the frames. Each version is used to test a specific factor. For example, Environment 1 will be the control version, and will have no semantic (link) or structural (document) metainformation. We will assign groups of students to each version and measure their activity.

- **Environment 1** - This version is the baseline or control version and contains minimal metainformation. It has Frame 1 (the document itself) and Frame 5 (control information).
- **Environment 2** - This version contains only structural (document) information. There are five frames, but with Frame 2 (Current Document information), Frame 3 (Previous Document information), and Frame 4 displaying only structural information. Frame 1 has link icons reflecting the destination type instead of a plain icon.
- **Environment 3** - This version contains only semantic (link) information, with plain link icons embedded in the document within Frame 1. Frame 1 also has the arrival icon marking the location of where the previous link has led us to. Frames 3 and 4 only contain semantic information. Since Frame 2 only contains structural information, it will be removed.
- **Environment 4** - This version contains both semantic and structural information.

Web users browse for four different reasons. We will be testing for how semantic and structural information improves understanding and reduces disorientation in the latter three cases. Thus we will run twelve experimental conditions in all.

- 1. **Passing through**: Users sometimes pass by a Web page as part of surfing or on their way to a different page. We cannot effectively test whether structural or semantic information helps them in our system.
- 2. **Browsing for general knowledge**: Assessment of these users will basically involve one to two page essays and multiple-choice questions. Users must work within our system for six hours over a two week period.
- 3. **Browsing for deep knowledge**: Assessment of these users will basically involve one to two page essays and multiple-choice questions. Users must work within our system for six hours over a two week period.
- 4. **Specific Browsing for specific information** : Assessment of these users will be based on short answers to questions (specifically speaking two to three lines) and multiple-choice questions. These users will work under a time constraint.

For cases 2 and 3 we believe, a minimum of 6 hours will suffice over two weeks. We will confirm this through pilot studies. For all three cases users will be allowed to take notes. Allowing users to take notes should alleviate possible confounding concerning retention-related measurements.

Analysis of the user information will allow us to assess what permutations of information best serve each kind of browsing. Measurement therefore requires developing and defining constructs of interest, and operationalizing these constructs. Some of the constructs that we have identified so far are reading rate, accuracy of the reader, task completion time, presentation, degree of interest of the user and remembrance, which involves recall and memory [Re97]. The effectiveness of the hypotheses will concern some combination of these measurements. We currently are determining others. Reading rate analyzes how fast the user is able to grasp the information provided. Accuracy specifies the preciseness with which the user has been able to grasp the information provided by the system. Presentation factor deals with the arrangement of the web-pages in a particular fashion. The degree of interest of the user describes how appealing the web-page to keep the user interested in it. Finally, remembrance assess the amount of information that a user is able to retain and recall.

Conclusions

This paper discussed the CACM system and the various experiments aimed towards proving the need for providing and conveying metainformation about links and documents. By proposing features that improve people's ability to find and understand information and proving that these should be incorporated on the web in the future, we hope to help build a better cyberspace. At present we are investigating valid and reliable measures for operationalizing various constructs of interests.

Acknowledgments

This research is supported by the New Jersey Center for Multimedia Research, the New Jersey Commission of Science and Technology, the National Center for Transportation and Industrial Productivity at NJIT, the New Jersey Department of Transportation, the NASA JOVE faculty fellowship program, the AT&T Foundation, and the New Jersey Institute of Technology under Grant 991967. We would also like to thank Shaji Abraham for his great skill, energy and enthusiasm in developing the system. Readers may browse the CACM system (non-experimental version) at:

http://space.njit.edu:5080/cacm/overview.html.

References

[Bi97] Bieber, M. Advancing Information Comprehension through Hypertext. in Advances in Intelligent Hypertext. Mayfield, J. & Nicholas, C. (eds.), Springer-Verlag.
[http://space.njit.edu:5081/bieber/~pub/Bi97.html]

[BI95] Bieber, M. & Isakowitz, T. (eds.) (1995). Designing Hypermedia Applications. Special Issue of the Communications of the ACM, 38(8).

[BVA97] Michael Bieber, Fabio Vitali, Helen Ashman, V. Balasubramanian and Harri Oinas-Kukkonen, "Fourth Generation Hypermedia: Some Missing Links for the World Wide Web," forthcoming in the *International Journal of Human Computer Studies*.

[Re97] Michael Reed, "An Overview of Measurements in Hypermedia-Based Research," forthcoming in Ed-Media'97.

[THH95] Thüring, M., Hannemann, J. & Haake, J. M. (1995). Designing for Comprehension: A Cognitive Approach to Hypermedia Development. Communications of the ACM, 38(8), 57-66.