Research Directions

James Blustein

28 September 1999

My overall goal is to help people find and use information more effectively — a goal I believe can best be achieved through collaborative research.

Contents

---

1 Previous Research

Although hypertext and human-computer interaction are currently my main areas of interest, I have done significant work in other areas as well.

1.1 Information Retrieval Evaluation

Under the direction of my main Ph.D. advisor, Jean Tague-Sutcliffe, I wrote a program [3] to analyze results from the annual TREC IR trials as an early step in what was planned as a set of investigations into what makes for successful IR [10]. As part of the development of the program I created a library of bit manipulation routines [1]. The program was presented in a conference tutorial and made publicly available [11].

1.2 DNA Computing

As a student in the first course ever about DNA computing I gave a talk surveying methods of encoding data and instructions in DNA computers. An accompanying paper is expected to be in a forthcoming collection about molecular computers edited by David Wood at the University of Delaware.

1.3 Hypertext Generation and Evaluation

I became interested in hypertext as a way to help people use large numbers of documents as a graduate student. I developed a simple model of what I thought researchers would need to browse through a collection of electronic documents, used information retrieval (IR) techniques to generate links in a collection, and developed and tested an evaluation method. That work resulted in three publications [4–6]. The method I developed is being built upon by another Computer Science graduate student (Mark Staveley, University of Waikato) [8, 9].

As part of my Ph.D. studies I designed, tested, and implemented a study of people using automatically generated hypertext links to read scholarly journal articles [7]. It is significant that my experiments were conducted with people since so many studies have ignored the human element by concentrating instead on abstract qualities with the assumption that these qualities will somehow help users. I developed rules, based on semantic similarity of text passages, to make links between different parts of the articles (from sentences and phrases to sentences, paragraphs, and sections) and implemented those rules using two distinct information retrieval systems: SMART and latent semantic indexing. I also developed methods of presenting structural information about the articles and of making simplistic connections from the use of technical terms in the articles to their definitions elsewhere in the article.

To test how well the links worked I developed a formal summative evaluation method which can be used with similar hypertexts. People used the hypertext versions of the articles (two with all links available, and one with only basic structural links) in a limited time, and were asked to answer basic questions about the article and summarize it.

Their summaries and answers were taken as a measure of how well the different versions helped the readers in tasks they would be likely to do with journal articles. I developed and used a program to record which links people followed as they read the articles [2]. Readers rated a representative sample of links in each article they read to indicate how well they thought the link would help a first-time reader of the article find the information they needed to answer the questions and write the summary. Because people sometimes prefer systems with which they perform poorly and vice versa, I also collected information about the readers subjective experience with the articles.

In my experiment I found that factors other than pure link quality made the most difference to readers. No difference was detected between the two methods of making links although they perform very differently in standard IR tests.

In developing the experiment, I made use of a great deal of literature about the importance of interfaces suiting users' tasks. Several of my prototype interfaces had to be discarded for that reason. In developing a method to test how well automatically generated hypertext links could help readers of journal articles, I became interested in problems related to user interfaces. Ideally an interface to a program should serve the program's user as an extension of themselves — it should not be a barrier but rather a tool.

When writing about my previous work with hypertext evaluation using a simplified mathematical model of a user [5], I made the point that real testing can only occur with real users performing realistic tasks. To pursue my goal of helping people to find and use information I have moved more towards developing and testing interfaces.

2 Ongoing Research

I am an active member of the newly formed organization and implementation task force for HyNIC (a digital library project for the ACM's hypertext interest group). I plan to use that project as the motivation and testbed for many of my ideas.

2.1 Hypertext/WWW Usability Measurement

I am interested in developing a tool, or set of tools, to measure qualities of webpages that are most relevant to usability. I foresee an automated webpage and website analysis tool like CAST's Bobby[1]. A proposal for collaborative work is under consideration at this time.

2.2 Survey of Hypertext Link Evaluation

I have begun to produce a comprehensive survey of methods and techniques of evaluating hypertext links. Much has been published on this topic, and I am well acquainted with some of the methods. It seems to me that so much is written about it because people do not recognize that factors other than pure link quality matter most to readers of hypertext. I believe that a survey of methods will help to show people who are interested in developing methods of link generation where there are opportunities for improvement, while also pointing out the importance of other factors that contribute to successful hypertext (see Section 3.1 for plans related to one application of hypertext).

3 Future Research Goals

The following are interesting areas that I am already prepared to work in. I have broad interests and am willing to work in other areas as well.

[1]'Bobby [⟨URL:http://www.cast.org/bobby/⟩] is a web-based tool that analyzes web pages for their accessibility to people with disabilities' and was created by the Center for Applied Special Technology.

3.1 Electronic Journals

Electronic journals are certainly a part of our lives now, but I think that we still need to determine how best to present and use the articles. Electronic media provide us with some opportunities that are not available from print. However there are drawbacks to reading from computer screens, and interactive media demand a higher cognitive load. I am particularly interested in exploring how different user interfaces could help readers of electronically presented articles, and some factors that contribute to users' success in reading articles. I also believe that more research should be done on supporting the reader's own activity with articles. Specifically I would like to investigate the role annotation may play in helping readers. Some more specific goals are below.

3.1.1 New User Interfaces

We know that the implementation of hypertext features in today's most popular browsers are ill-suited to using electronic journal articles. We should be exploring the use of different techniques for presenting articles. Many of these techniques could be developed as prototypes within the browsers through use of their built-in scripting languages or server-side scripting. It is not enough just to write programs that will change the ways that people can read articles if we do not carefully test the new interfaces to see if they really do help.

I would like to investigate different methods of presenting links, specifically:

- presenting them in margins outside of the text, where more detail can be given about the destination of the link;
- with preview of the destination in a lower frame of the main window (this could be especially helpful with short destinations such as endnotes and citations).

Coupled with studies of how spatial reasoning affects people's reading of articles it will be interesting to consider different methods of making the structure of an article clear. In particular I suspect that dynamic fisheye view tables of contents and multiple windows (e.g., one for the Table of Contents and one for the main article) could help readers.

3.1.2 Determining Their Potential Utility

I am particularly interested in pursuing some of the future work from my previous experiments. Improvements in electronic journal interfaces could be brought about if we knew: (1) what human factors affect success with hypertext versions of journal articles; (2) if readers find links with to sentences, paragraphs or sections most useful; and (3) how advance organizers help readers find useful links.

3.1.3 Supporting User Annotation

We know from studies that hypertext linking is idiosyncratic. Furthermore, the results of one of my experiments (see Section 1.3) lead me to believe that software for reading journal articles should support readers in their task of making their own links and annotations. The first step in developing such a system will be an investigation of how people actually use articles. I know that others have studied this before so an experiment might not be necessary.

We need to determine the form or forms of annotation that should be supported. Would free-form annotation be appropriate? Would notes, questions, or both be helpful to readers? Should readers be able to make links to entire articles? Would it be better for them to make links to specific sections of articles, or to a list of such sections? If there is potential for annotation support to help readers then we need to be able to answer these questions confidently.

Once a system has been created to support the tasks that people already do with articles then we need to determine the criteria for success of such a system. I suggest that such a system would be a success if people felt comfortable using it and they could process articles faster using it than from a standard on-screen version.

3.2 Digital Libraries

I am particularly intrigued by the navigational challenges posed by the prospect of digital libraries. However as I have a background in information retrieval my interests are not limited to only those concerns. Based on results of earlier work I suspect that a sensible strategy might be to provide basic article level links (such as tables of contents, cross-references, citations) for all articles and allow people to make their own notes and links within and between articles and to supply powerful automatic linking and visualization tools at the level of document corpora.

References

[1] James Blustein. Implementing bit vectors in C. *Dr. Dobb's Journal*, 20(8), August 1995. For updates and additions see ⟨URL:http://www.csd.uwo.ca/%7ejamie/BitVectors/README.html⟩.

[2] James Blustein. Tracking links in a single HTML document. *The Perl Journal*, in press.

[3] James Blustein and Jean Tague-Sutcliffe. IR-STAT-PAK [computer program]. ⟨URL:ftp://ftp-nlpir.nist.gov/pub/irstat/⟩. Overview at ⟨URL:http://www.csd.uwo.ca/%7ejamie/PS/IRSP-one.ps⟩.

[4] James Blustein and Robert E. Webber. Using LSI to evaluate the quality of hypertext links. In Maristella Agosti and James Allan, editors, *IR and Automatic Construction of Hypermedia: A Research Workshop*. ACM SIGIR, 13 July 1995.

[5] James Blustein, Robert E. Webber, and Jean Tague-Sutcliffe. Methods for evaluating the quality of hypertext links. *Information Processing & Management*, 33(2):255 – 271, 1997.

[6] William James Blustein. An evaluation of tools for converting text to hypertext. Master's thesis, University of Western Ontario, London, Ontario, Canada, 1994.

[7] William James Blustein. Hypertext Versions of Journal Articles: Computer-aided Linking and Realistic Human-based Evaluation. Ph.D. thesis, University of Western Ontario, London, Ontario, Canada, 1999.

[8] Steve Jones. Automatic hypertext/my student. e-mail message, 4 March 1999.

[9] Mark S Staveley. Hypertext linking. e-mail message, 18 February 1999.

[10] Jean Tague-Sutcliffe and James Blustein. A statistical analysis of the TREC-3 data. In *Text Retrieval Conference*, Gaithersburg, MD, USA, November 1994. National Institute of Standards and Technology.

[11] Jean Tague-Sutcliffe, James Blustein, and Paul Kantor. What differences are significant? Statistical analysis of IR tests. Tutorial presented at SIGIR '95 conference, July 1995. Michael Nelson appeared in place of Tague-Sutcliffe.