A Study of Measures for Research in Hypertext Navigation

D.G. Hendry, T.T. Carey and S.T. TeWinkel

Dept. of Computing and Information Science
University of Guelph, Guelph Ontario Canada N1G 2W1

Abstract: The research described here investigated the strategies people use to navigate a hypertext document in reading comprehension tasks. We present the results from experiments where people were initially instructed to browse and later to study a document containing expository information on 35 mm. cameras. The users interacted with the document through one of two presentation methods: traditional page sequences or access by hypertext links. A variety of measures were used to illuminate users' navigation strategies: Each of these measures contributes in a different way to our overall understanding of users' navigation (and raises additional questions).

1.0 Introduction

Documents presented with hypertext systems offer the potential of much more flexible access to information than traditional linear, paper-based documents. But flexible access requires choice—the reader must make decisions about what to read next. The process of choosing an access sequence is often described as analogous to navigating around a physical space.

Navigating in spatial environments can be both frustrating and rewarding—frustrating when you get disoriented, rewarding when it leads to increased understanding. Consider, for example, the experience of driving yourself somewhere to which you have previously been driven by others. You can be suddenly aware of how little you have absorbed as a passenger about the structure of the neighbourhood. Navigation can be an important learning process in its own right.

There are numerous reports of hypertext navigation being disorienting [e.g. Edwards and Hardman 88]. But our knowledge of navigation is limited, particularly in terms of its potential to be a learning experience in itself rather than just a means to get to document content.

Understanding navigation is an important research goal for understanding hypertext: how do people navigate through hypertext documents, how does that differ from their navigation through traditional documents, how the navigational strategy is formed, and what impacts—positive and negative—it has.

This paper reports on a research study of hypertext navigation. A variety of research measures are necessary to gather information about this process: we report here on the role of different data gathering techniques in filling in our picture of hypertext navigation. Section 2 describes the experiment, Section 3 presents the results from the various measures used. Section 4 compares the roles of the various research measures and discusses further ongoing work.

2.0 The Navigation Experiment

2.1 Choice of task

Research has been done on how people retrieve information from hypertext systems (e.g., [Campagnoni and Herrlich 89]) and how well people use systems to answer open-book essay questions [Egan et al 89]. However, despite obvious implications [Charney 87], little research has been done on how hypertext systems affect our comprehension for the content of a text. Since many hypertext systems are used for more that just information retrieval, we decided to investigate how a hypertext system affects comprehension outcomes, resulting from various reading tasks. Further, it was decided that we would initially investigate the essence of hypertext—making decisions about where next to read—without the influence of navigational aids such as maps [Memik et al 89] or more elaborate methods for signaling document structure [Conklin and Begeman 88].

In the experiment, subjects were given two reading tasks: browsing and studying. Subjects always browsed first, then studied. For browsing, subjects were instructed to "quickly browse through the document, trying to determine its organization and the nature of its contents". For studying, subjects were instructed to "learn as much as possible" and to expect a test. The instructions did not specify the nature of the test, but did describe a scenario in which students might use the information.
2.2 Choice of document

We decided to use expository text passages, which had been used in previous experiments [Bromage Mayer 81]. These passages explained the operation of a 35mm camera and the experiments examined what good problem solvers learn from them. Two reasons led us to this decision. First, the text passages were designed for subjects from a subject-population similar to the one available to us, namely undergraduate students. While the passages contain complex expository information, (as do many commercial hypertext documents) no specialized background knowledge was needed of the topic. Second, cued-recall questions for measuring subjects’ comprehension had already been prepared; thus, using these text passages gave us a method for measuring comprehension. Using existing expository text passages and cued-recall questions gave us time to concentrate our efforts elsewhere, as well as the potential to compare results and methods.

2.3 Converting to hypertext

To develop the hypertext document, we first decided upon a strategy for transforming the expository passages into standard topologies of nodes and links. These logical structures (i.e., the topologies) together with the physical structure (i.e., the manner in which titles, passages, and links are shown) constitute a rhetorical convention for organizing the hypertext. This rhetorical convention can be summarized by describing the topologies, the methods for connecting the topologies, and the physical structure.

Three topologies are used in the document: simple-, sequenced-, and complete-collection. Each topology consists of a topic node and several descriptive nodes. A topic node introduces the subject matter for its topology. How the descriptive nodes are linked together (i.e., how they are related) is what distinguishes the three types of topologies. In the simple-collection, each descriptive node has a link to-and-from the topic node. In the sequenced-collection, the descriptive nodes are sequenced in some order, and the topic node has a link to the first descriptive node and the last descriptive node has a link back to the topic node. In the complete-collection each descriptive node has a link to-and-from the topic node and all descriptive nodes have links to each other. An illustration of these topologies appears in Figure 1: the vertical line on the left side links together a simple collection, the box clusters together a complete collection, and the loop on the right side links a sequenced collection. These topologies were chosen because: (a) they reflect common methods for structuring text [Meyer Freedle 84] and (b) similar topologies have been used to create topologically complex but purposeful hypertext documents [Bernstein 87].

Two methods can be used to connect topologies together. First, collections can be embedded in each other, by making the topic node for one collection a descriptive node in another. This method is used to create hierarchies. Second, topologies can be connected together by what are called associative links. These links provide efficient access to nodes that are otherwise far away (i.e., several links have to be followed up and down a document) and they allow an author to show important relationships between nodes that, for expository reasons, are most effectively placed in different topologies.

Figure 1 shows the nodes titles and topology which make up the document hierarchy for the experiment. All links are two-way unless explicitly marked with an arrow. Figure 2 shows the 26 associative links. The document consists of 39 nodes, 62 links, and 2088 words.

Figure 3 shows the physical appearance of a node. All nodes contain a thematic title that is elaborated by a text passage. Links (i.e., titles of other nodes) are shown beneath text passages. The positions of the links are random, because a systematic method for placing links (e.g., all links going down the hierarchy are shown on the left) might induce a navigation bias, a complication we did not want to face in our initial work.

Experience in creating hypertext documents is showing that strategies for linking nodes (i.e., structuring text) are often subjective, taking the form of authors' personal (nonstandard) tastes [Alschuler 80]. The rhetorical convention, described above, was intended to provide a general approach for creating a class of documents, all with the same style. Criticism about the idiosyncrasies of the hypertext document can thereby be deferred, to some extent at least, to the rhetorical convention. But, the convention only reduces the possibility of ad hoc document structures, it does not eliminate it. On the one hand, decisions about choice of topology can usually be based on the nature of the ideas in the text. These decisions are relatively straight-forward. For example, passages on the steps of taking a picture naturally fit into a sequenced-collection, whereas the parts of a camera, having no inherent sequence and being relatively independent, fit best into a simple-collection. On the other hand, decisions about when to use associative links are not straight-forward. (An informal exercise showed that when given the hierarchical structure of the document as a basis, and asked to add about 30 associative links, two people experienced with hypertext had about 50% of their proposed links in common.)
2.3 Procedure

The measures and results discussed in the next section are part of a larger experiment, reported elsewhere [Hendry 85]. This section summarizes the procedure for that experiment.

Forty-four undergraduate students, who reported no familiarity with operating 35mm cameras, were randomly assigned to experimental or control groups. The experimental group subjects read the hypertext document by navigating through links, whereas control subjects read a linearized version of it by paging. (The control and experimental groups were in fact subdivided further, but we do not discuss those results here.)

After subjects completed a tutorial, they were instructed to browse the document (6 min). Then, they were asked to recall as many nodes titles as possible. Next, they were asked to study the document (15 min). Finally, they were again asked to recall as many node titles as possible, and to then answer several cued-recall questions, which appraised their comprehension of the material. The experiment was thus a mixed design: task instructions were an incomplete within-subject variable, and user-interface for reading was a between-subject variable.
3.0 Measures and Results

In this section the details of the experiment are summarized; in the next section we discuss how the various measures of performance contribute to our understanding of users' navigation.

3.1 Measures of comprehension

The results show that reading by navigating neither enhanced nor hindered comprehension for the material, as measured by the cued-recall questions. However, subjects who read by navigating recalled more node titles than subjects who read by paging (33% vs. 20%, F(1,40)=19.1, p<.001). Similar results were found in an experiment where subjects were instructed to find information using either SuperBook or a page-bound manual: users of SuperBook recalled more section titles than users of the page-bound manual [Egan et al 89]. Taken together, the results from this experiment suggest that navigation enhances the learning of incidental information but does not affect learning for content.

3.2 Measures of node access and coverage

Subjects who read by paging covered more of the nodes than subjects who read by navigating (92% vs. 82%, F(1,30)=29.1, p<.001), apparently because it is harder to assess one's coverage and strive for complete coverage when reading a hypertext document. This result reminds us of the importance of providing navigation aids that help people obtain a sense of closure. As expected, there was a positive correlation between scores on comprehension questions and coverage of the document (Pearson r=.57, F(1,31)=13.0, p<.001).

When subjects browsed the hypertext document they generally had shorter node visiting times than when studying (19 vs. 23 sec, F(1,30)=7.2, p<.025) and the standard deviations of node visiting times were shorter when they browsed than when they studied (10 vs. 16 sec, F(1,30)=26.7, p<.001). This data suggests that the task instructions did affect the way subjects read. Browsing might be characterized by relatively short visit times, with low variance, whereas studying might be characterized by relatively longer visit times, with high variance.

3.3 Comparisons with original study (paper-based text)

We did not duplicate the experimental conditions of Bromage and Mayer, so that comparisons between our results and the original study were not straightforward. For one thing, while we reproduced their comprehension questions, their marking scheme was not recorded so we had to estimate how the scoring was done. For factual questions this was simple, but for more complex questions we adopted a conservative scoring which may have varied considerably from the original. On the factual questions, the scores were very similar (mean of 7.1 in our study vs. 7.3, out of a possible score of 17 in each case). On more difficult questions, our subjects were considerably lower. We did not attempt statistical comparisons— it is not clear whether these differences were due to differences in experimental conditions, subject populations or scoring methods.

A second difference in experimental methods occurred in the instructions to subjects. The browse/study distinction was not presented in the original experiments. Subjects were instructed there to "read the document through once", in preparation for answering questions. This does not translate easily into the hypertext condition, in which a number of nodes may be revisited before the entire document has been read. We attempted to achieve the same effect by asking subjects in the browse condition to try to cover the whole document and restricting the time allowed based on pilot studies. In the actual experiments the subjects did not always achieve complete coverage of the document. This illustrates the way in which media considerations complicate comparisons between conditions.

3.4 Measures of navigation patterns

The sequences of nodes that users create as they navigate is an important source of data about the usability of hypertext systems. Analyzing these sequences seems to have been first proposed and explored, in an experimental setting, by Canter and his colleagues ([Canter et al 85 & 86]), has been used informally in practical settings to evaluate the usability of both help systems ([Campagnoni & Ehrlich 80] [Campbell 88] and walk-up-and-use information systems [Mountford 89]).

For this research we adapted some indices for characterizing navigation [Canter et al 85] so they could be used to characterize subjects' navigation through the hypertext document. We computed the extent that subjects covered all the nodes in a topology before moving on to another one; the extent that subjects moved through the document by going up and down, instead of following associative links; and the extent that subjects made loops back to a topology in the cases where an associative link was followed. We anticipated, as have others (e.g., [Canter et al 85] [Bernstein et al 88]), that indices such as these could be used to discriminate paths through hypertext documents under varying task conditions.

Contrary to our expectations, the indices failed to show any systematic differences in the way subjects sequenced nodes during browsing and studying. Furthermore, when subjects were divided into two groups, subjects scoring low in comprehension and those scoring high (19 vs. 28 points, F(1,20)=57.4, p<.001), there were no significant differences in the navigation indices across the two groups. It appears that a particular sequence for visiting nodes is as likely to occur during browsing as it is during studying, and that good-comprehenders do not sequence their access to nodes (as characterized by the indices) any differently than poor-comprehenders.
3.5 Verbal protocols and debriefing

To supplement the controlled experiment, several additional subjects were presented with the hypertext condition. They were encouraged to verbalize while they read, and questioned afterwards about aspects of their strategies. We were particularly interested in their perceptions of their strategies during the browse and study stages. We also wanted to assess whether the recall of node titles was an aid to searching in the document, so subjects were asked to search through the document to find answers to several of the original comprehension questions.

During the browse condition, subjects focused on the goal of covering the document, skimming the text of nodes and concentrating to some extent on titles. This was particularly evident towards the end of the time period allotted for browsing. During the study condition, subjects' goal was to remember the document contents that they considered "testable". The instructions, which indicated they would be tested, seemed to have biased them away from information suited for their own use, such as photo processing.

While subjects did indicate these different strategies for the browse and study conditions, they did not differ in directness. This is in line with the navigation path data described above. Our experimental conditions did not induce the kind of undirected browsing we were seeking, which may explain in part why the navigation indices did not differ. To create an undirected browsing condition, we would have to remove the goal statement. The best way to do this appears to create a "waiting" condition, in which subjects are ostensibly waiting to participate in an experiment and are given the opportunity to use the hypertext system to fill in their time. All the people we debriefed indicated that they found using the medium enjoyable, so that the hypertext environment would probably be attractive enough to engage subjects outside an explicit experimental instruction to use it.

When asked to search for the answers to questions, navigation became the primary task rather than studying for recall. People experienced more frustration using the document for this purpose, because they wanted to home in on the desired information but were often unable to quickly find it. They did frequently verbalize node titles as search objects. The rhetoric conventions adopted to guide reading in the hypertext document were counter-productive for searching, e.g., the sequenced-collection topology. The presence of associative links was noticed more during search, where it tended to diminish subjects' initial perceptions of the document as solely hierarchical.

Subjects also differed considerably in their search strategies. One subject deliberately returned to the initial node to start each search, but tended to use associative links during traversal more than other subjects. Another person, who had focussed in the browse and study stages on learning where information rather than details of content, performed very efficient searches on general questions but had difficulties on specific questions (where the answer was in a single node).

4. Conclusions and further work

Each of the measures employed supplied complementary information about the experiment. The comprehension measures demonstrated that hypertext users were not at a significant disadvantage over readers of the paged text, despite the need to navigate explicitly through the document. The recall of node titles suggested that hypertext users might be able to use titles to navigate in search tasks, which was confirmed by the verbal protocol sessions. The node coverage statistics showed that coverage of the document was uneven, and that people who scored higher on the comprehension measures had been more successful at exploring the whole document.

The verbal protocol sessions showed that different strategies were used in the browse and study stages, but the navigation data showed that these strategies did not differ overall in terms of node access patterns. The verbal protocol data also demonstrated how the task instructions influenced the navigation. This suggests aspects of the experiment not considered in the original Bromage and Mayer study. For example, the original study noted that subjects scoring higher on comprehension questions also scored higher on general problem-solving ability. They did not consider whether the differences in recall were due to more accurate guesses about the "testables" which the experimenters would ask about. The verbal protocols make it clear that subjects' perceptions of the probable questions influenced their navigation strategy.

A number of extensions to our study are now underway. The most obvious addition is a map of the hypertext document. Other studies have demonstrated the positive effects of a visual overview to guide navigation [Mosk et al 85, Hammond & Allison 87 & 88], and the verbal protocols of subjects searching for information indicate frustration at the absence of such an aid. During browse and study stages the same comments are not frequent, although there are other indications that subjects would have appreciated a map.

For example, some subjects were distracted by a node title "Overview" which occurred at the second level of the document hierarchy (below "Introduction", see figure 1). In the browse stage, they failed to return up the hierarchy past this point, thinking that "Overview" was the main topic node rather than a subnode of "Introduction". As a result, they failed to cover the several parts of the document until the study stage—when they were returned to the top node automatically and remarked on what they had missed.
We are particularly interested on the effect of a map on navigational patterns, compared with the data on use without a map. Does a map change the navigational strategy or just facilitate it? We currently have map prototypes with different appearances (topological network vs. table of contents) and different features (static map, map which records current position, map which allows direct access). Our concern at present is not to choose an optimal map format, but to first understand the effects of various map structures in different conditions.

A second area of further work is the rhetorical conventions used to organize reading for comprehension. Our subjects received no introduction to these conventions, and the verbal protocols indicate discovery about the role of hierarchical vs. associative links occurring after considerable navigation. We want to compare the impact of instruction in the conventions against the impact of a map (and against both a map and instruction). We also need to study how the rhetorical conventions affect navigation during search, and to what extent people are disoriented by an associative access which bypasses the organizational structure. We also want to observe the search process on parts of a document which have not previously been browsed, to determine the role which browsing plays in formation of a navigational strategy.

Acknowledgements

This research was supported by Cognics Inc., the Ontario Technology fund and the Natural Sciences and Engineering Research Council of Canada.

References


