Effects of Link Annotations on Search Performance in Layered and Unlayered Hierarchically Organized Information Spaces

Landon Fraser and Craig Locatis
National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894. E-mail: locatis@nlm.nih.gov

The effects of link annotations on user search performance in hypertext environments having deep (layered) and shallow link structures were investigated in this study. Four environments were tested—layered-annotated, layered-unannotated, shallow-annotated, and shallow-unannotated. A single document was divided into 48 sections, and layered and unlayered versions were created. Additional versions were created by adding annotations to the links in the layered and unlayered versions. Subjects were given three queries of varying difficulty and then asked to find the answers to the queries that were contained within the hypertext environment to which they were randomly assigned. Correspondence between the wording links and queries was used to define difficulty level. The results of the study confirmed previous research that shallow link structures are better than deep (layered) link structures. Annotations had virtually no effect on the search performance of the subjects. The subjects performed similarly in the annotated and unannotated environments, regardless of whether the link structures were shallow or deep. An analysis of question difficulty suggests that the wording in links has primacy over the wording in annotations in influencing user search behavior.

Introduction

The Internet and the World Wide Web have grown to the point where they have become a major resource for accessing information. Web-accessible information is composed largely of hypertext documents. These documents have links to subunits of information within them, or to information in other documents. The earliest hypertext systems predated the World Wide Web, and one problem with hypertext since its inception has been navigation and way finding (Hammond, 1988; Kerr, 1986). Because information is presented in interlinked fragments that can be accessed nonlinearly, many users become disoriented (Marchionini & Shneiderman, 1993). The problem becomes magnified when users attempt to find information in a global hypertext environment as vast as the Web. The way information is represented and linked in hypertext can either alleviate or exacerbate the way finding problem. This study builds on previous research related to information linking and representation by testing the effects of link annotations on user ability to search hierarchically organized hypertext under conditions where the hierarchical structure has intervening link layers and conditions where it does not.

Background

There is no single way to design Web pages, but the way information is presented affects a user’s ability to navigate its contents (Khan & Locatis, 1998b; Leader & Klein, 1996). Existing style guides present different strategies for designing user-friendly Web pages (e.g., Horton & Lynch, 1999; Nielsen, 1999a). Such guides address the retrieval and disorientation problems people encounter when navigating hypertext documents, like those found on the Web, but many style guide recommendations are conflicting and not empirically verified. Because hypertext documents are nonlinear and have links connecting their various parts as well as other related documents, information can be accessed out of context. Moreover, the size of the document/information space being traversed and its boundaries are not always apparent. These and other factors contribute to disorientation (Hammond, 1989).

Being able to navigate the Web meaningfully is a special problem because of the volume of information on the Internet. Prior knowledge about a subject, search strategies, and retrieval tools can greatly affect the ability to search through information (Hill & Hannafin, 1997; Leader & Klein, 1996). Information retrieval also depends on the mental model user’s create of the system that they are searching, and these mental models are greatly affected by system functions and features (Marchionini & Shneiderman, 1993; Voithofer, 1999). There is a significant relationship between a user’s information processing style and
navigational methods (Lee & Harvey, 1999). Lack of navigational support (Nielsen, 1998) and absence of controlled descriptions of the available resources (Powsner & Roderer, 1994) affect users’ navigational ability in a negative way. Problems can be partially addressed through the implementation of navigational support systems within hypertext documents (Locatis, Letourneau, & Banvard, 1989). Menus are important for navigation, and the users of hypertext documents should be kept in mind when developing navigational aids. Two major concerns are accurate portrayal of the content that users will choose and the physical layout of the information (Frick, Kisling, Cai, Yu, Giles, & Brown, 1999a; Frick, Monson, Xaver, Killie, Conley, & Warney, 1999b). Concern for user interaction should extend to how the content is perceived in its entirety, not just individual elements (Rogers & Erickson, 1998).

The primary method of navigation on the Web is through links. There are three main types of links: referential, organizational or hierarchical, and keyword (Strasma, 1995). Referential linking allows parts of a document or different documents to be accessed in any order, regardless of how they are logically related. Hierarchical linking imposes restrictions because links are only made from the higher levels of a document (or document collection) and its broader subjects to its lower levels and narrower topics. Keyword linking results from the use of search engines. Keywords entered into forms on Web pages retrieve links to documents or document fragments containing the words. These different link types allow Web designers considerable latitude in providing mechanisms for navigating their sites.

If the information being navigated is inherently hierarchical, then hierarchical linking should enable users to develop better mental models regarding its organization. There are problems, however, with hierarchical linking. Hierarchical linking requires users to traverse menus of links to information at lower levels. This can add link layers, but adding link layers can decrease the accuracy and increase the number of errors a user makes as well as increase the amount of time it takes a user to search (Wallace, Anderson, & Shneiderman, 1993). Although people are less likely to become disoriented when navigating hypertext with fewer link layers, the shallower structure is usually achieved by increasing the number of links displayed at higher levels. Instead of displaying fewer link choices on consecutive menus, more links are presented at once. Unfortunately, research indicates that presenting too many links also can encumber search performance because there are more options than users can meaningfully process (Welsh, Murphy, Duffy, & Goodrum, 1993). Moreover, adding links to a document’s upper levels can increase Web page length, making it necessary for users to scroll down the page to view all of the options. Links located below a Web browser’s window are less likely to be chosen, because research indicates that many users fail to scroll and may be unaware they even have the option to do so (Neilsen, 1999b). Even when users do scroll, it takes time and can add to the overall search effort.

When looking at the Web, users tend to scan through text, selecting only highlights and key words, and only reading selective paragraphs (Neilsen, 1999b). This creates a need for a different approach to writing specifically for the Internet environment. Links and their titles are important cues for helping people navigate through the web (Nielsen, original date 1996). Links highly descriptive of the content referenced are helpful because they can provide information telling where the link leads. Strong correlations between the wording in a link and in user queries have been shown to dramatically enhance search performance (Khan & Locatis, 1998b), and presenting links in lists rather than embedding them in paragraphs and limiting the number of links also has positive effects (Khan & Locatis, 1998a). Providing cues that links are internal or external to a single document to be searched have little effect on performance (Khan & Locatis, 1998a) at least in situations where users are not specifically told the meaning of the internal and external cues in advance (Khan & Locatis, 1998b).

This study extends previous research on linking and retrieval by specifically examining the effects of adding descriptive summary annotations to links. Many Web developers annotate links. In addition, many Web search sites add annotations when links are retrieved. These may be based on metadata provided in the Web resource or on other information found in the body of the document. If link query correspondence helps users find information, adding annotated information may help users decide whether to pursue a link because there is more data that could potentially map to the query. On the other hand, annotations add information that users must process and the strategy might be prone to the same problems as embedding links in paragraphs.

Research Design and Variables

Three research questions were addressed in this study. First, will adding information to a link in the form of an annotation that further describes the content to which the links points help users find answers to queries? Second, are the effects of link annotations different when the hypertext structure is shallow, without intervening link layers, or deep, with intervening link layers? Shallow hypertext is easier to navigate, and it may be that the additional information annotations provide would only be useful when the information space is deep, with many link layers. Finally, are the effects of link annotations different when queries are easy or difficult: for example, with high or low degrees of correspondence between the wording in queries and the wording in links? Studies show correspondence has a powerful effect on user performance. Summary annotations would presumably be most helpful when correspondence is low because users have to make inferences about the content in the target being linked rather than rely on matching words in the link itself. The additional information that annotations provide may help users make better inferences about content.
The independent variables in this study were (1) annotated and unannotated links, (2) deep and shallow hypertext structure, and (3) search question difficulty. The dependent variables were accuracy and efficiency of search performance. Accuracy was determined by whether searchers could identify the correct section of a hypertext document containing the answer to a query. Efficiency was determined by the number of separate mouse clicks, or links traversed, to perform the search and the elapsed time from when the query was presented to when searchers indicated they had identified the answer. The subject had to traverse through paragraphs of information to locate the answers to the queries given. The research design for the treatment conditions is shown in Figure 1. Subjects randomly assigned to each cell received identical questions of three levels of difficulty—easy, medium, and difficult.

**Research Methods**

Four activities were conducted prior to doing the actual research: determining the search space, developing the hypertext, identifying the search queries, and specifying the research protocol.

**Determining the Search Space**

To do the research, it was necessary to define the search space. A document that met certain criteria was chosen for the search. First, it needed to be of reasonable length, not too long or short, for purposes of the study. Short documents might prove to be too easy and very large documents too difficult, causing everyone to perform well or poorly regardless of treatment. Moreover, if a large document were used it might expand search time beyond what would be reasonable to observe in the study. It was decided to use a document of a size approximating that of longer journal articles or book chapters. The document used for the study was 25 single-spaced pages. Second, its content needed to be obscure so the subjects who participated in the research study would know very little about the topic, because prior knowledge might affect the ability to search. The document still had to be of a topic of interest. The document employed in the study concerned adolescent development, separation anxiety, anorexia, and suicide—subjects the high school students who were study participants would find of interest but about which they were not likely to be knowledgeable.

**Developing the Hypertext**

The document identified for the study was linear, and had four main topic sections: adolescent development, separation anxiety, anorexia, and suicide. Because two of the four topics comprised most of the document, three more documents containing information related to the other two topics were integrated into the original document to expand each section to approximately the same size. Four different versions of the document were created. A layered version displayed the titles of the four main sections on a main menu that linked to submenus, each having three links to other submenus. These submenus each had three additional links to primary content. An unlayered version was created in which links to the 48 sections were displayed using a single table of contents, nested within topical headings and subheadings identical to those used in the layered version. Diagrams showing these link structures appear in Figures 2 and 3. Two additional versions of the document were created by adding annotations to the links in the layered and unlayered versions. These provided additional information about the content being linked. Figures 4 and 5 depict two of the annotated and unannotated treatments.

Having 48 links was important because it allowed three or four links to be displayed on each menu in the document’s layered versions. Having too few links on each menu layer might have oversimplified the search environment and made it artificial. Without a minimal number of links per

---

**FIG. 1.** Research design.

**FIG. 2.** Shallow link structure.
menu, there would be no logical reason to subdivide content into intervening link layers. Moreover, having fewer menus with more links in the layered version would have eroded differences between the treatments. Care was taken to develop link wording that was descriptive of the content behind the link. The annotations further amplified these descriptions into one or two summary sentences. Some contained identical or synonymous words to those in queries subjects were given, and the degree of similarity was used to calibrate the difficulty of the search. Navigation links were created for each document version. The layered versions had the name of the previous link layer menu in bold at the top of each page and links to either the main menu or previous menu, with the exception of the main page, which had the main menu in bold at the top. The lowest level of the layered version containing the primary information had links to both the main menu and the previous menu. The unlayered versions had a single link at the top back to the main menu.

**Identifying Search Questions**

Search queries were devised for each of the document’s 48 sections and classified as easy, medium, or difficult. Questions classified as easy contained a word corresponding to a word appearing in only one link. Questions classified as medium contained a common word, or a close synonym, but the word appeared in multiple links. Question classified as difficult had no direct correlation with the wording in any of the links. When annotations were added, they followed the same standards. If a word appeared in only one link, then it would also appear in only that link’s annotation. If a word appeared in multiple links, it would appear again in each corresponding annotation. If a synonym appeared in a link, then it would appear again in the annotation. A total of 16 easy, 20 medium, and 12 difficult questions were formulated. Twelve queries were drawn from the query pool for pilot testing—three easy, six medium, and three difficult. These were grouped into question sets containing one easy query, two medium queries, and one difficult query, each with answers in different major sections of the document. The idea was to spread the questions out to avoid giving the impression that all the answers were in one or two places. Each question set was tested three times with a different subject, and the search times were averaged and ranked. Subjects also were encouraged to comment on each question and to think aloud as they searched.

Some of the questions were found to be easier or more difficult than originally classified, while others (those having no correspondence to wording in links) were so challenging that subjects could not find the answers to them. The latter were discarded, and a new classification schema was devised, based on an analysis of the search times for the remaining questions. In this schema, easy queries were still defined as those queries having a word in them corresponding to a word found only in a single link. These queries had the shortest search times. Queries of medium difficulty were redefined to be queries that had a word in them corresponding with a word occurring in multiple links. These queries had the next shortest search times. Difficult queries were redefined to be queries that had synonyms to link wording. Although their search times were long, subjects were able to answer them. These queries were tested in another pilot

---

**FIG. 3.** Layered link structure.

**FIG. 4.** Unlayered/unannotated linking.
study using the same procedures as the first, and were afterward grouped together into sets having one easy, one medium, and one difficult question, each with answers in a different document section. In a third pilot study each question was tested a total of five times in layered and unlayered hypertext environments. Means and standard deviations of query search times were computed to select the three questions that were used in the study. The questions selected were those that had search time means and standard deviations that overlapped the least and that also had answers in different sections of the document.

**Specifying the Research Protocol**

A computer was chosen in a place containing few distractions and which was readily available to the researcher. The high school students participating in the study were read an orientation to the study’s procedures, indicating that they would be given three consecutive search questions, that they were to identify the section in the document that they thought contained the query’s answer, that their performance would be timed, and that if they did not identify an answer after 10 minutes they would be told to go on to the next query. They were also told that the design of the document, not their individual search performance, was being researched, and that even though their performance was not being tested that they should try to answer each query as best as they can. They were then given a search question and allowed to click on the computer screen to start. A start page was developed so that the subjects would not be able to see the testing environment before they began their test. The order in which each subject was provided easy, medium, or difficult search questions, was randomized, and each subject was randomly assigned to one of the four treatments: layered.annotated, layered/unannotated, unlayered/annotated, unlayered/unannotated. Twenty sub-

| Table 1. Figure means/standard deviations of dependent variables. |
|-----------------|-----------------|-----------------|
| Environment     | Easy            | Medium          | Hard            |
| Unlayered annotated | 1.4/0.60        | 2.0/1.75        | 2.5/1.54        |
| Mouse clicks    | 41.32/30.15     | 85.72/75.75     | 129.71/84.25    |
| Time in seconds | 1.00/0.00       | 0.85/0.32       | 0.40/0.50       |
| Accuracy        |                 |                 |                 |
| Unlayered unannotated | 1.4/0.50        | 2.35/2.58       | 2.85/1.79       |
| Mouse clicks    | 50.77/35.23     | 102.74/84.76    | 113.84/74.05   |
| Time in seconds | 0.95/0.22       | 0.75/0.44       | 0.55/0.51       |
| Accuracy        |                 |                 |                 |
| Layered annotated | 3.8/1.54        | 6.4/3.17        | 5.5/2.63        |
| Mouse clicks    | 60.63/61.99     | 144.52/72.85    | 128.83/82.43   |
| Time in seconds | 0.95/0.22       | 0.80/0.41       | 0.45/0.51       |
| Accuracy        |                 |                 |                 |
| Layered unannotated | 5.2/4.02        | 6.1/2.43        | 6.8/4.35        |
| Mouse clicks    | 77.49/70.52     | 107.73/59.26    | 138.51/118.17  |
| Time in seconds | 0.90/0.31       | 0.70/0.47       | 0.40/0.50       |
| Accuracy        |                 |                 |                 |

**Table 2. Percent accuracy of search performance.**

<table>
<thead>
<tr>
<th>Environment</th>
<th>Easy</th>
<th>Medium</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlayered annotated</td>
<td>100%</td>
<td>85%</td>
<td>40%</td>
</tr>
<tr>
<td>Unlayered unannotated</td>
<td>95%</td>
<td>75%</td>
<td>55%</td>
</tr>
<tr>
<td>Layered annotated</td>
<td>95%</td>
<td>80%</td>
<td>45%</td>
</tr>
<tr>
<td>Layered unannotated</td>
<td>90%</td>
<td>70%</td>
<td>40%</td>
</tr>
</tbody>
</table>

objects were assigned to each treatment and there were 80 subjects overall. Each subject gave written consent prior to participation in the study.

**Results**

Means and standard deviations of the number of mouse clicks and search times for each treatment condition are depicted in Table 1, while those for accuracy appear in Table 2. A multivariate analysis of variance, MANOVA, was conducted on the three dependent variables of search performance. MANOVA indicated significant main effects for layering, $F(61,736)$, $df = 3$, $p < 0.0005$, and for question difficulty, $F(14,310)$ $df = 6$, $p < 0.0005$, but not for annotations, $F(1,940)$ $df = 3$, $p < 0.124$, or for interactions between layering and annotations, $F(1,140)$ $df = 3$, $p < 0.334$. There were no significant interactions between difficulty and annotations, $F(0.649)$ $df = 6$, $p < 0.691$, difficulty and layering, $F(0.49)$ $df = 1$, $p < 0.815$, or difficulty, layering, and annotations, $F(0.720)$ $df = 6$, $p < 0.633$. Subsequent univariate analyses showed that the main effects for layering were significant for number of mouse clicks, $F(119,153)$ $df = 1$, $p < 0.0005$, and search time, $F(5,408)$ $df = 1$, $p < 0.021$, but not accuracy, $F(0,927)$ $df = 1$, $p < 0.337$. Subsequent univariate analysis of difficulty showed that its main effects were significant for all variables: mouse clicks, $F(7,927)$ $df = 2$, $p < 0.0005$, search time, $F(19,380)$ $df = 2$, $p < 0.0005$, and accuracy, $F(31,821)$ $df = 2$, $p < 0.0005$.

**Discussion**

The results show that the link annotations had a minimal effect on search performance, regardless of whether the search space was shallow or deep. There were large variations in search performance. Some subjects guessed correctly, and went directly to the section that contained the answer, while others search widely before they found it, if they found it at all.

Observations of their search performance confirm observations by Nielson (1999b) that users tend to scan Web documents for key words. Few subjects took time to read annotations. Moreover, they appeared to focus on the key words in links rather than the wording in the annotations.

There was great variation in performance based on search difficulty, which was expected. This variation also helps explain why annotations may not have affected performance and why subjects may have relied more on key-
words. The easy query had a word that matched to a word present in only one link and, if it had an annotation, only one word in the annotation as well. The medium question matched words in multiple links and, in the annotated condition, the corresponding annotation as well. The difficult question used synonymous wording, which had by far the worst performance. The easy and medium questions had keyword matches; the difficult question did not. The fact that there were no interactions between difficulty and annotations further supports the possibility that subjects tended to focus on links.

Question difficulty has other implications as well. The fact that even synonymous words had poor results, and that questions used in a pilot study that did not even have this level of correspondence had to dropped because they proved too difficult, has implications for the design of information systems. Word matching was controlled in the phrasing of the search queries for this study. In the real world, there is no such control, because users of information systems have idiosyncratic ways of expressing their searches. In the real world, adding more words in annotations may increase the chances that the ways individual search questions are expressed might match the words descriptions provided, but given the user variability in query expression, this is unlikely. Moreover, adding annotations takes time, and it may be better to invest extra effort in providing search mechanisms instead of annotations. If these search mechanisms are to be powerful, they need to go beyond keyword matching and include thesauri that can map user queries to synonymous words in the information environment.

The layering outcomes observed in this study confirm earlier research. Although there were no significant differences in search accuracy between the layered and unlayered versions of the search space used in this study, those given the layered treatment invested more effort and took longer to either succeed or fail in answering queries. Shallow information spaces were easier to navigate.

**Conclusion**

This study confirms previous research that shallow link structures are better than deep (layered) link structures and also shows that annotations have no effect on search performance. The subjects performed similarly in the annotated and unannotated environments, regardless of whether the link structures were shallow or deep. Analysis of question difficulty suggests that the wording in links has primacy over the wording in annotations in influencing user search behavior.

Future research on link strategies and searching is needed that uses undefined search tasks and larger and more diverse subject populations, as well as additional link types and kinds of annotations. It should also address expert/novice differences and problems in measuring search performance.

**Undefined Search Tasks**

Subjects were given specific search queries to which they were asked to find the answers. The questions were quite specific, sometimes resulting in the subject thinking that they found the correct answer when the answer they found was related to the answer of the question but was not the designated answer. How the annotations and the layering affect the search performance in environments where the search task is not specifically stated has not been researched. When a user is trying to locate information they often do not have tasks as specific as what was given to them in this study. Giving the user a larger search space may also affect how they perform.

**Larger and More Diverse Sample Size**

The methods various age groups use when they search through information spaces vary. A more diverse sample needs to be tested to see the effects on the general population. This study only observed the effects of the annotations on high school students and, although a total of eighty people were tested, only 20 people were tested in each environment. An older, more diverse population might use annotations differently and a larger sample size might produce more reliable outcomes.

**Different Link Types**

Only hierarchical links were tested in this study. The effects of annotations on keyword and referential links were not tested. Annotations may have a completely different effect on search performance in information spaces and on the Web where links presented are not hierarchical. The effect of different types of links needs to be looked into further.

**Various Types of Annotations**

The annotations tested in this study were considered true annotations in that they summarized the information that they pointed to. Not all annotations found on the Web are true summaries. They also can be keyword annotations and excerpts taken directly from the document, perhaps by extracting the first sentence or sentences appearing in the body of the text. The presentation of these annotations and their effect on a user’s search performance is currently unknown.

**Expert/Novice Differences**

Whether annotations affect search performance of experts and novices differently was not controlled in this study. Novices having less knowledge of the subject matter may rely more heavily on annotations than experts knowing the content and its relationships. Although it was assumed the subjects in the study were unfamiliar with the content, this may not have been the case. The subjects gave no
indication of content familiarity during the experiment, and this, combined with the study’s results, would tend to rule out the contention that content novices might be more prone to use annotations when doing a search. More intriguing, however, is the possibility of expert/novice differences based on experience using the Web. Novice Web users might rely on annotations more than experienced ones, but the subjects in this study generally had substantial Web experience. They attend a high school that is home to a district magnet science and technology program, and that has significant computing resources. The majority of subjects were enrolled in the program.

Multiple Tests of the Same Subjects

The subjects were given one set of search questions. Although the questions were carefully selected, additional questions would produce a more accurate measure of their performance. In addition, performance over time should be observed. Given more time to search or the opportunity to search one or more information spaces on multiple occasions, links might have been used more and the outcomes might have been different. The actual performance of a subject could be better determined.

References


