Hypertext or Book: Which is Better for Answering Questions?

Barbee T. Mynatt, Laura Marie Leventhal, Keith Instone,
John Farhat and Diane S. Rohlman

Computer Science Department
Bowling Green State University
Bowling Green, Ohio 43403
(419) 372-2337 FAX (419) 372-8061
e-mail: mynatt@opie.bgsu.edu

ABSTRACT
An important issue in the evolution of hypertext is the design of such systems to optimally support user tasks such as asking questions. Few studies have systematically compared the use of hypertext to books in seeking information, and those that have been done have not found a consistent superiority for hypertext. In addition, designers developing hypertext books have few guidelines. In the present study, users performed information-seeking tasks and answered a variety of types of questions about Sherlock Holmes stories using either a conventional paper encyclopedia or a hypertext encyclopedia. The questions varied on the amount of information needed to derive an answer (fact or inference), the location of the question's key phrase in the hypertext (entry title or entry content), and the format of the information (text or map). Accuracy and time were recorded. The hypertext group excelled in answering fact questions where the information was embedded in a text entry. The book group excelled only in answering fact questions based on maps. In spite of having far more experience using books, the book group was not significantly faster overall and did not perform as well on an incidental learning task. Our results suggest that a hypertext book with a nonlinear structure and including a variety of navigational tools can equal or surpass conventional books as an information-seeking medium, even with minimal training.

KEYWORDS: experimental research, question answering, usability of hypertext, hypertext

INTRODUCTION
An important issue in evaluating the usability of any system is how well the system supports the user in attaining task goals. One task integral to many hypertext systems is answering questions. While many adults are highly skilled in using books to answer questions, hypertext systems can offer additional functions and alternative modes of interaction. These new approaches need to be systematically assessed in order to provide designers with empirically-based design guidelines. In general, little research has been done on the strengths and weaknesses of hypertext vis a vis more traditional media. McKnight, Dillon and Robertson (1989), for example, have suggested that there is a great need for experimental work to compare hypertext and paper media in a variety of situations.

The studies that have been done to date on the issue of books versus hypertext have not always painted a rosy picture for hypertext. An initial problem with hypertext is that users read textual material more slowly from screens than from paper media (e.g., Wright & Lickorish, 1983; Gould, Alfaro, Fin, Haupt & Minuto, 1987). Prior studies that compared hypertext books to conventional paper books have reported mixed results, with neither media emerging as uncontestably better. A variety of factors, including the form of computer implementation and the type of task, appear to influence usage patterns and user satisfaction. Marchionini and Shneiderman (1988) described studies that compared utilizing information in hypertext systems to conventional paper materials. In one study comparing a paper versus hypertext version of a maintenance manual, the paper version resulted in significantly faster times for 12 different tasks. A second study (also reported in Shneiderman, 1987) compared a paper versus a hypertext version of a database on the holocaust, developed in Hyperties. For simple fact retrieval questions, the paper version was faster. But as query complexity increased, the Hyperties system became equally fast. There was no difference between the groups in accuracy. In addition, users generally preferred the Hyperties system.

Egan, Remde, Gomez, Landauer, Eberhardt, and Lochbaum (1989) compared performance on a variety of tasks using a printed statistics book and three different versions of a hypertext facsimile of the same book, developed in SuperBook. The SuperBook version presented the text in multi-window format and supported word look-up and entry into the book from a table of contents. Subjects took part in a one-hour training session and five evaluation sessions. In a structured search task, the subjects attempted to answer questions that could be found by searching for key phrases in either headings only, text only, text or headings, or neither. With the improved hypertext versions, Egan, et al.
found that the SuperBook users were significantly more accurate. The paper book users were marginally more accurate only for searches where the key phrase was in neither the heading nor the text (e.g., key phrases using synonyms). Egan, et al. also found that search times were not statistically different for the two media overall in one version. However, SuperBook users had faster search times for questions whose keywords could be found in the text only. Users of the paper version were faster for searches where the key phrase was in neither the heading nor the text. The MiteyBook version, which included a number of refinements, resulted in 25 percent faster search times overall, compared to the paper version. In an open-book essay task, subjects from the SuperBook group obtained significantly better overall scores and factual content scores than subjects from the paper group. SuperBook users rated the documentation, the system, and affect for statistics as subjectively higher than the paper group. Finally, in an incidental learning task, SuperBook users recalled significantly more chapter headings than the paper group.

In this paper we report an experiment comparing the success users have answering different sorts of questions using either a hypertext encyclopedia, HyperHolmes™, or a paper version of the same encyclopedia. The goal was to determine the relative strengths and weaknesses of paper versus hypertext for this user task, and to determine, from a designers' perspective, what hypertext features make hypertext most effective for this type of application. More specifically, the current study addresses the following issues:

- The effectiveness of hypertext interface features in answering assorted types of questions. In prior research using a different sort of hypertext system, Egan, et al. varied the types of questions that subjects were to address. They found that SuperBook was faster for answering questions where the answer was in the text only, and book users were marginally more accurate for finding questions requiring synonyms. In our study, we explore a wider variety of question types, and utilize a different sort of hypertext system.

- The use of materials which are non-linear. In the Egan, et al. study, the statistics book represented in the SuperBook version was a standard format text book, organized into sequential chapters with subheadings. Thus, the naive reader would typically be required to go through the book sequentially in order to acquire a coherent understanding of the material. While it is clear that hypertext must be able to successfully present sequential material, it is perhaps a more natural medium for non-linear information with many interconnections. By non-linear information, we mean information that can be understood in isolation, and that does not assume that previous (non-general) information has been acquired. Reference materials, such as dictionaries, encyclopedias and cookbooks are common examples of non-linear information. Marchionini and Shneiderman (1988) did use non-linear information, but found no superiority for hypertext.

- The incorporation of more "hyper" features than previous hypertext versus book studies. In addition to using non-linear materials, the HyperHolmes system includes a number of features that were not used in the earlier studies. These include the use of graphics, backtracking, and searching based on links. More specific descriptions are provided below.

- The amount of training necessary to be an effective user. Tombough, Lickorish and Wright (1987) have shown that skill level and training can influence hypertext use. However, there are wide variations in the amount of training given to users in previous studies. Egan, et al. used a one-hour training session, and then examined performance across 5 additional hours of interaction with the system. In contrast, Marchionini and Shneiderman used a 7 to 8 minute training period. In the present study, a modest amount (approximately 30 minutes) of training was used.

A DESCRIPTION OF HYPERHOLMES

HyperHolmes: The Electronic Encyclopaedia Sherlockiana™ incorporates information from Jack Tracy's Encyclopaedia Sherlockiana (1977). The Encyclopaedia Sherlockiana is an alphabetically arranged list of keywords and key phrases related to the Sherlock Holmes stories. Each keyword or key phrase has an accompanying description, and possibly references to other entries in the encyclopaedia. The book is not indexed and entries are found strictly alphabetically. Tracy's book is considered by Sherlocchians to be the definitive encyclopedia of information on Sherlock Holmes as depicted in Sir Arthur Conan Doyle's stories.

The HyperHolmes system was implemented in HyperCard 1.2.5, using the Dialoger and aGHAS tools (Instone, 1990). The system contains over 3200 cards. A single card consists of an entry (as copied from the encyclopedia) of text, graphics or a combination of the two. Any entry may contain one or more direct references to other entries. Direct references to other entries are completely capitalized. Each entry is a window on a card. In addition to the entry, other icons and information are present on the card.

HyperHolmes provides a number of navigation tools and methods. The simplest way to move among the cards is to click on any word or phrase in the current entry that is displayed in all capital letters. (See sample screen in Figure 1.) Every capitalized string is a button that takes the user to the indicated card.

A backtrack tool (represented by the bent arrow icon) takes the user back to the most recently visited card. Repeated use of the backtrack tool steps the user back through prior cards in reverse historical order. An alphabet slide bar (seen on the right side of Figure 1) approximates the use of a thumb index followed by page turning in a book. A sliding box
inside the bar can be moved towards the top to select from cards at the beginning of the alphabet, and towards the bottom to select cards later in the alphabet. When the slide box is released, the box contains the letter of the alphabet which matches the title of the card displayed in the main window. Once in the approximate area, clicking on an up or down arrow moves a single card at a time up or down alphabetically.

The search tool (represented by a magnifying glass icon) provides text-searching capabilities. The user types in a target word or phrase and selects the desired search space. The search space can be either the entry titles or the entire textual contents plus titles. (In the current version, the text on maps and drawings is not part of the search space.) Once activated, the search tool displays a list of the titles of cards containing a match, and the contents of one matching card are displayed in the main window. The user may then select any of the displayed titles, select the displayed card and/or dismiss the search tool.

Finally, HyperHolmes includes the Incoming Links and Outgoing Links tools. Clicking on the Incoming Links tool produces a list of all cards in HyperHolmes which reference the current card. The Outgoing Links tool lists all cards referenced by the current card (this is equivalent to the set of buttons associated with the current card). To our knowledge, HyperHolmes is the first hypertext system to incorporate such tools.

In HyperHolmes, several specialized types of cards are used to provide organization, way-finding information and help. These include the Holme card, overview cards and help cards. The Holme card is the first card displayed when the system is initially activated. It contains the title, a large logo (silhouette of Sherlock Holmes) and several buttons, including Getting Help and Overviews. The Holme card can be directly reached from any other card in the system by clicking on a button that is always present and consists of a miniature version of the logo. Thus the Holme card serves as both the entry point and home base. The Getting Help cards describe how to use the system and include visual "snapshots" of examples. The Overview button produces a table of the various overviews (seven total) that are available. Each overview is similar to a table of contents, but focused on one topic. For example, there are overviews of all the stories in publication order, Sherlock Holmes, London, maps and so on. Cards cited in the overviews are directly accessible from the overviews.

**METHOD AND PROCEDURES**

**Experimental Design**

The experiment used a combination of between- and within-subjects experimental design and consisted of three independent variables: question type (described below), medium (book vs. hypertext) and trials. Dependent measures included the accuracy of users' responses and their speed in answering the questions. In addition, the extent of incidental learning was assessed.

**Materials**

Five types of questions were used: Fact/Text/Title, Fact/Text/Contents, Fact/Nontext, Inferential/Text, and Inferential/Nontext. (There were no Fact/Nontext/Contents questions because the system did not provide the ability to search on textual material in nontext cards.) The questions varied in the amount of information needed to answer them (a single fact or an inference drawn from combining two or more facts), the location of a matching keyword (entry title or entry content) and physical format of the target entry (text or nontext). To answer a Fact question, information from only one entry was needed. Title questions were questions whose answers could be found if the user searched the entry titles using a keyword from the question. An example of a Fact/Text/Title question is "What is a spud?" If the user did a search through entry titles using the keyword "spud," the search would find and display the text entry entitled "Spud," which contained a definition of the term. Contents questions were questions whose answers could only be found by searching for a keyword embedded in the text of an entry (a search on title would not yield any hits). To answer an inferential question, information from more than one entry was needed. An example of an inferential question is: "Name two countries Holmes visited in The Final Problem." To answer this question, the user must find at least two entries related to "The Final Problem" which mention Holmes traveling through a particular country. The incoming links tool is potentially quite useful in answering inferential questions. In the "Final Problem" question, a user might go to the node about "The Final Problem." Clicking on the incoming links button brings up a list of the nodes which reference this story.
Included in this are several countries (like Luxembourg and Italy), which the user can learn more about by selecting the name from the list. A total of fifteen questions, three of each type, were used in this study. Each user received one of four different random orders of the questions. The orders were restricted in such a way that blocks of five questions, one question of each type, were created within each order.

Following the questions, users also completed an incidental learning task. The incidental learning task consisted of a list of titles of entries. Half of the titles were ones the user must have seen if she or he answered the questions correctly. Half were titles that were probably not seen. The users were asked to circle the titles they remembered seeing.

**Users**
Novice users were studied because we felt that an important aspect of many hypertext systems is that they be accessible even to novices with a minimum of training. Twenty-nine college students enrolled in Introductory Psychology who were in their first to third year of college were recruited. As a prerequisite for participation, all users had some experience using computers. Fifteen used HyperHolmes and 14 used the paper version of Encyclopaedia Sherlockiana.

**Training**
Because users in both groups answered the task questions on a Macintosh II computer, all were given an introduction to the Macintosh II. A guided tour was used to review mouse skills such as pointing, clicking and dragging. The users in the HyperHolmes group also received instruction on working with more than one window, scrolling a window and closing windows. Following the Macintosh training, the experimenter read a tutorial to the users that explained the basic features of either the paper version of Encyclopaedia Sherlockiana or HyperHolmes. The paper tutorial explained the organization of the book and gave examples demonstrating how the entries were related. The training for the paper book lasted approximately 15 minutes. The HyperHolmes tutorial explained the organization of the software, the Holme card, the Overviews, and how to work the buttons. Each tool was demonstrated and explained. The users practised using each tool while it was explained. Next, they were asked to complete eight exercises using the tools. These exercises reviewed how each tool worked and gave them an opportunity to practice using the tools. The exercise did not give any practice in answering questions. The combined training and exercises with HyperHolmes took approximately 30 minutes.

**Procedure**
The users assigned to the HyperHolmes condition first received the Macintosh and HyperHolmes tutorials. They were then given a booklet containing the three blocks of questions and asked to search for the answers. They were told to use the remaining time of the two-hour session. When an answer was found, the user clicked on an "Answer" icon on the screen. This caused a text window to appear where the answer was to be entered. If the user was unable to find the answer after searching for 15 minutes, (s)he entered "I don't know" in the answer space and went on to the next question. The users using the paper version of Encyclopaedia Sherlockiana first received the Macintosh and Encyclopaedia tutorials. They were then given the three blocks of questions to answer. The answers were entered into the computer using the same answer icon/text window as the video users. These users were also told to type "I don't know" in the answer space if they could not find the answer after 15 minutes of searching. After completing the questions all of the users completed the incidental learning task.

**RESULTS**

**Accuracy**
Every answer was scored using a 0 to 2 rating scale. A score of 0 indicated the answer was incorrect or missing. A score of 1 indicated partially correct. Only the answers to questions requiring a two part answer (i.e., Inferential/Text and Inferential/Non-text) could potentially receive a score of 1. A score of 2 indicated a completely correct answer.

The mean accuracy scores for the five different types of questions as a function of medium type are shown in Figure 2. Question type had a significant effect on accuracy of responses, $F(4,104) = 20.62$, $p < .0001$. Scores were highest overall for the Fact/Text questions (mean = 1.61) and lowest for the Fact/Text/Content questions (mean = 0.74). The interaction between question type and medium was also significant, $F (4,104) = 8.81$, $p < .0001$. Figure 2 illustrates this interaction and shows that Fact/Text/Content questions were harder for both media groups, but were extremely hard for the Paper group. The Fact/Nontext questions were particularly easy for the Paper group, but only of average difficulty for the HyperHolmes group. Overall, the scores were higher for the group using HyperHolmes than for the Paper group (means = 1.38 vs. 1.23). This difference was marginally significant, $F(1, 26) = 3.41$, $p < .08$. 

![Figure 2: Mean accuracy scores as a function of type of question and medium type](image-url)
Thus, performed the problem-solving version. The question is below."

HyperHolmes points to overlooked advantage of books. Incidental information is much more a "fact of life" with books, and is typically not a feature of hypertext. (However, note our findings in regard to incidental learning discussed below.) The HyperHolmes search tool made finding Fact/Text/Content answers easy, if the user understood and used the content-searching mode. The fact that the HyperHolmes group performed worst on this type of question suggests that using the content-searching mode was not always intuitive. This is not surprising, because it is perhaps the content-search feature that potentially makes an electronic version of a book most different from a paper version. Searching line-by-line through text is simply not feasible in a paper book, and so is not a frequently-used problem-solving approach.

The finding that paper book users performed best on Fact/Non-Text questions while the HyperHolmes users performed only at an average level on such questions was unexpected. Fact/Non-Text questions involved answering a single, factual question based on a map. One explanation for this outcome is that in the paper book the maps were inserted alphabetically in with the rest of the materials. Thus, if the user looked up "India", (s)he would find not only a text entry, but one or more maps in the same physical location. In HyperHolmes, the search facility listed "Map of India" as one of possibly several matches. The user had to specifically select "Map of India" in order to view the map; this is an additional step that is not required with the book. Although there was an overview of all maps available in HyperHolmes and there was no such listing in the paper book, this added information did not overcome the apparent handicap of the extra steps necessary to access a map. A second possible explanation for the somewhat poorer performance of the HyperHolmes group on the map questions is related to screen size and acuity. The maps represented on the screen were physically smaller than those on paper. Furthermore, a number of the maps from the book had to be divided into two or more screens. If this occurred, arrow buttons were added to the map to quickly take the user to the adjacent map. Nonetheless, division reduced the amount of information simultaneously available.

Interestingly, the HyperHolmes users did not have improved accuracy for the inferential questions, in spite of having the incoming links tool. We speculate that this is due to a lack of familiarity with this function, because books do not have this function available.

The use of three repeated sets of questions allowed us to look at learning effects. Although there was a significant effect of trials, F (2, 52) = 4.19, p < .03, there was no interaction with medium type. Performance on the last trial (mean = 1.45) was the best, while performance on the first trial (mean = 1.29) and second trials (mean = 1.20) was poorer. There was also a significant interaction of question type and trials collapsed across medium type, F(8, 208) = 5.78, p < .0001. These data are shown in Figure 3. Perhaps the most noticeable effect contributing to the interaction is the great difference in accuracy on Trial 1 between the Fact/Text/Title questions and the Fact/Text/Content questions. All subjects in both conditions on Trial 1

![Figure 3: Mean accuracy scores as a function of type of question and trial collapsed across medium type](image)

Figure 3: Mean accuracy scores as a function of type of question and trial collapsed across medium type

![Figure 4: Mean speed in answering questions as a function of medium and type of question](image)

Figure 4: Mean speed in answering questions as a function of medium and type of question
Incidental Learning

A Mann-Whitney U test comparing the number of correct answers on the incidental learning task showed that the HyperHolmes group exhibited significantly more incidental learning, $U = 46.5$, $p < .02$. That is, the HyperHolmes group correctly identified more previously-seen entry titles, and correctly rejected unseen titles compared to the Paper group. Incidental learning is of interest because it indicates that unmotivated, nonintentional learning is taking place. That is, if incidental learning does occur, the user is getting "bonus" knowledge with no intentional expenditure of effort. Heller (1990) points out that incidental learning in hypertext-assisted instruction is a desirable feature. Our finding of greater incidental learning from the hypertext version is also of interest because although the book users actually were exposed to more information, the hypertext users apparently learned more.

CONCLUSIONS

The results of this experiment highlight how the functionality offered by a system, the task to be performed and characteristics of the user all interact (Eason, 1984). For some types of tasks, such as Fact/Text/Content questions, HyperHolmes users clearly had advantages over the paper group, because this type of search is more difficult or impossible in a paper book. However, on the Fact/Text/Content questions the performance of HyperHolmes users was still worse than their performance on other question types. In this case, the problem was that the users did not effectively use an available tool that would have made the task easy. An issue for further exploration is how to draw the user's attention to appropriate tools at appropriate times. A tutorial system or knowledge-based system would probably be necessary. The performance on the inferential questions was much the same across the two medium. These result suggest that even when the system supports a type of access, if it is an unfamiliar strategy, the users probably will not use the strategy as effectively as more familiar ones. Our results further suggest that question types that deal with visual and non-textual information are challenging for hypertext book users. In HyperHolmes, maps apparently need improvement to become at least equally accessible as the same information in the paper book.

It is perhaps not amiss to point out again that the book users were highly practiced in using books, while the hypertext users had only one-half hour of training. Considering this large discrepancy in training, our results suggest that well-designed hypertext has great potential for giving users enhanced functionality at little cost to the user.

REFERENCES

2. Eason, K. D. Towards the experimental study of usability, Behaviour & Information Technology, 3(2), 1984, 133-143.


