



**HAL**  
open science

# Detecting lost users: empirical studies on browsing hypertext

Carolyn L. Foss

► **To cite this version:**

Carolyn L. Foss. Detecting lost users: empirical studies on browsing hypertext. [Research Report] RR-0972, INRIA. 1989. inria-00075587

**HAL Id: inria-00075587**

**<https://inria.hal.science/inria-00075587>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**INRIA**

UNITÉ DE RECHERCHE  
INRIA-SOPHIA ANTIPOLIS

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
B.P. 105  
78153 Le Chesnay Cedex  
France  
Tél.: (1) 39 63 55 11

Rapports de Recherche

N° 972

*Programme 8*

**DETECTING LOST USERS:  
EMPIRICAL STUDIES ON  
BROWSING HYPERTEXT**

**Carolyn L. FOSS**

**Février 1989**



\* RR - 8972 \*

# Detecting Lost Users: Empirical Studies on Browsing Hypertext

## Etudes Empiriques sur l'Exploration des Documents Hypertextes

Carolyn L. Foss  
INRIA  
Sophia - Antipolis  
2004 Route des Lucioles  
06565 VALBONNE - FRANCE

### Abstract

Browsing large hypertext networks often causes users to feel disoriented. Since it is difficult to remember the connectivity of arbitrarily structured networks, users often have trouble navigating through and finding information in hypertext documents. Although disorientation is a major problem for users of hypertext systems, little is known about its causes, symptoms, and how to prevent it. The purpose of the two studies reported in this paper is to determine the conditions under which disorientation occurs, its *observable* symptoms, and user's strategies for avoiding it. The first study, which uses Hypercard, focuses on problems associated with reading and comprehending nonlinear documents. The second study, which uses NoteCards, examines problems arising from the high cognitive demands placed on users while reading and analyzing hypertext networks. An interesting outcome of these studies is that the *pattern* of a subject's search through the network (e.g., repeated return to previous cards, closing and then reopening cards) can be used to automatically detect inefficient browsing strategies without knowing the semantic content of the network. Both of these studies provide new directions for the development of hypertext browsing tools.

### Résumé

Les utilisateurs qui parcourent un large réseau d'Hypertext se sentent souvent désorientés. Comme il est difficile de se rappeler la connectivité, les utilisateurs ont souvent du mal à retrouver les informations contenues dans des documents d'Hypertext. Alors que la désorientation est un des problèmes majeurs pour les systèmes d'Hypertext, peu est connu sur ses causes et symptômes et comment y remédier. Le but des deux études contenues dans ce papier est de déterminer les conditions de désorientation, symptômes *observables*, et les stratégies des utilisateurs pour l'éviter. La première étude, conduite en Hypercard, se concentre sur les problèmes associés avec la lecture et la compréhension de documents non-linéaires. La seconde étude, faite sous NoteCards, examine les problèmes soulevés par d'intenses demandes cognitives sur les utilisateurs, pendant la lecture/analyse du réseau d'Hypertext. Une intéressante conclusion de ces études est le fait que la séquence d'actions du sujet à travers le réseau (par exemple des retours répétés à la même carte) peut être utilisée pour détecter mécaniquement les stratégies de parcours inefficaces, et cela sans connaissance du contenu sémantique du réseau. Ces deux études fournissent de nouvelles directions pour le développement d'outils d'aide à la navigation pour systèmes Hypertext.

### Acknowledgments

This research was partly supported by Xerox France Inc. (Paris) who lent us a workstation and put the NoteCards environment at our disposal.

### Remerciements

Cette recherche a bénéficié du soutien de la compagnie Xerox France, qui nous a prêté une station de travail et a mis à notre disposition l'environnement NoteCards.

## Introduction

As hypertext technology becomes widely available, many online linear documents will be converted to hypertext format with the aim of improving their usability. Documents that are stored in a hypertext network allow users to access, view, and compare its various parts, which do not necessarily share adjacent locations in the underlying database. Hypertext networks can contain text, pictures, animations, and sounds that can be easily browsed. (See Conklin, 1987 for a general review of Hypertext systems.) Unfortunately, browsing through nonlinear networks of frames often leaves people with a general feeling of disorientation, being lost, or of losing context. Conklin (1987) has described the disorientation problem as not knowing where you are in the information space or not knowing how to access something you believe exists there.

Although the problem of "getting lost" is well known to hypertext users and is commonly mentioned as a major drawback of hypertext systems, little is known about its causes, symptoms, and how to alleviate it. This paper will describe two empirical studies that address these questions: What strategies do skilled hypertext users use to avoid disorientation? What strategies do users employ to reduce mental load while browsing through hypertext networks? When does nonlinear text interfere with comprehension? We hope that the results of these studies will move us beyond subjective descriptions of disorientation and its associated problems, and towards an understanding of its causes and observable outcomes. This better understanding of user-disorientation will allow the 1) prevention of disorientation via good design of hypertext networks; 2) computerized diagnosis of user-disorientation, and 3) development of browsing tools that alleviate the symptoms of disorientation.

Typically, hypertext browsing problems have been described in terms of subjective outcomes (see Frisse, 1987; Conklin & Begeman, 1987; Yankelovich et al., 1985) (e.g., being "lost," losing context, forgetting what parts of the network have been explored) rather than in terms of externally observable behaviors. For purposes of discussion, we have categorized common browsing problems in terms of their cognitive outcomes (see Table 1). Our aim is to eventually supplant these subjective descriptions of browsing problems with a set of observable behaviors that co-occur with these subjective states, but for now we describe only the cognitive outcomes of disorientation.

One group of problems arises from unfamiliarity with the structure and conceptual organization of the document network. These problems are usually designated "the disorientation problem" or "being lost" (e.g., Conklin, 1987) and are inherent in the very nature of hypertext since documents often contain hundreds of linked nodes forming a complex structure, yet only a small portion of them can be visible on the screen at once. These problems include retrieval failures, trouble navigating towards or jumping to desired neighborhoods in the network, as well as uncertainty about the extent to which the relevant portions of the network have been examined or whether something important might have been missed. Not knowing the extent of a network or what proportion of relevant items remain to be examined has been referred to as lack of "closure" (Oren, 1987).

A second group, Task Management problems, originates from the cognitive demands placed on the users of hypertext documents. Readers have to decide which path to take through the network, but this can be difficult when users notice too many relevant topics to explore in parallel or too many interesting things that distract attention away from a main task. Thus users often digress from a main task but forget to return, or they move to some point in the network, but forget why they went there.

A third group of problems stems from the lack of discourse cues (e.g., chapter headings, topic sentences, previews, summaries, etc.) in hypertext documents. Authors of traditional linear documents use either implicit or explicit discourse cues to help readers recognize important points, assumptions, the structure of their arguments, among other things. When such cues are absent, users can have difficulties remembering, consolidating, and understanding the semantic content of the nodes that have been examined. The main cognitive outcomes are lack of detailed memory for any particular item that was examined, the inability to summarize what was learned from a browsing session, or inability to analyze or critique the semantic content of the hypertext network. Such problems are not unique to hypertext applications, but they seem to be exacerbated by the very nature of online nonlinear documents.

As mentioned above, our aim is to move away from subjective descriptions of browsing problems and towards an account that is based on observables. The observable behaviors associated with disorientation are of two types: *symptoms* and *user response*. Symptoms refer to observable outcomes of browsing problems. For example, "looping" (i.e., repeatedly returning to the same node or series of nodes) is an observable event that often indicates disorientation; Messy or disorganized

screen tops (e.g., more open cards that a user can attend to at once) often indicates difficulties with Task Management; and, poor search strategies (e.g., depth-first when breadth-first would be better) or very short reading times often co-occurs with poor Learning or Search Strategies.

---

**Browsing Problems:  
Causes and Cognitive Outcomes**

**Disorientation**

Cause: large, complex network, only partly visible

Outcomes: retrieval failures, navigation difficulties, lack of closure, not knowing if following a link will bring up sought after information

**Task Management**

Cause: too much information available at once

Outcomes: neglecting to follow or return digressions, forgetting why you moved to a particular place,

**Search/Learning Strategies**

Cause: lack of discourse cues

Outcomes: Inability to summarize browsed information, failure to remember details, inability to analyze information in network

Table 1

---

User response refers to what users do once they realize they're having problems. For example, disoriented users often return to familiar parts of the network to reorient themselves. Users who are having Task Management problems might stop, close most of the cards on their screen, then adopt a more systematic strategy for arranging the cards on their screens. Users who aren't learning enough from browsing might start taking longer to read through the cards. Detection of these symptoms and user behaviors do not depend on knowing the semantic content of the hypertext network, only the relative patterns of subject behavior. Of course, drawing inferences about users' mental states, given only observable behaviors, is difficult. But this is at least a start towards a better understanding of the problems users have while browsing.

The first study examines the problems people have while reading nonlinear text. The subjects in this experiment were given a small network of linked text that they were to attempt to read completely in the allotted time. The nonlinear format caused problems for the subjects that would have been absent had they been reading a traditional linear text. The fact that they had to concentrate on navigating around the network, in addition to trying to comprehend the text, caused a number of the problems associated with Disorientation and poor Learning/Search strategies mentioned above. The second experiment focused more on the set of problems associated with Task Management. Subjects in this experiment had to analyze the contents of a hypertext network, but since many different lines of comparison were possible and screen space was limited, subjects had to strategically select which lines of analysis they would pursue. Both of these experiments produced browsing situations that caused subjects problems -- the idea was to set up situations where Disorientation, Task Management, and Learning/Search strategy problems would occur so that their observable symptoms and what users did in response to them could be observed.

### **Study 1: Reading and Comprehending Nonlinear Text**

A hypertext representation is clearly advantageous for large documents containing many cross references, footnotes, and bibliographic notes, or whose use would require a lot of page turning to figures, references, or dictionary definitions. In fact, several large reference works that previously occupied numerous large files or heavy volumes have been successfully converted to hypertext format with high levels of user acceptance. These include software documentation (Walker, 1988; Remde, Gomez, & Landauer, 1987), medical handbooks (Frisse, 1988), and the 41 million word Oxford English Dictionary (Raymond & Tompa, 1988).

Since users of reference materials such as these often desire information related to fairly specific topics, the optimal form of information retrieval usually combines fast search with browsing. Users can first jump (via search) to areas of the network likely to contain sought-after information and then browse around until they find what they are after (see e.g., Frisse, 1988). Since this process involves viewing more than simply the target information, users are exposed to other related concepts, which can lead to refinement of their original questions while getting a sense of how the network is organized. These kinds of hypertext documents combine

the best aspects of retrieval from databases with perusal of paper documents.

Information retrieval from other types of documents, however, might not benefit much from a conversion to hypertext format. Texts containing carefully structured logical arguments, narratives, research reports, or instructional materials that presume knowledge of one concept before discussing another are usually intended to be read in specific orders. The logic of a carefully-planned argument can crumble when read in a different order from what the author originally planned. Thus Nonlinear representations of such texts, which grant users latitude to chose their own reading paths through the document, might hinder comprehension, especially by users who are unfamiliar with the topic area.

It turns out, however, that many favorable claims have been made about the educational benefits of presenting just these kinds of documents as hypertext (e.g., Yankelovich, Meyrowitz, & van Dam, 1985; Bolter & Joyce, 1987). For instance, Beeman et al. (1987) has argued that the inherent nonlinearity of hypertext will lead to the development of new "non-linear" cognitive styles. They claim that the connectivity of hypertext will help students think about concepts from a variety of perspectives. Landow (1987) has also noted that seeing connectivity between diverse subjects such as literature, history, and politics is crucial for a good education--and that hypertext makes the seeing of these connections possible. It has also been suggested that hypertext can facilitate comprehension of complex arguments by displaying an abstraction of the relationships between premises, assumptions, and conclusions (Conklin & Bergman, 1987; Marshall, 1987; Smolensky, et al., 1987 ). An assumption underlying many of these claims is that presenting information nonlinearly is beneficial because readers will have to "un-linearize" it anyway in order to integrate it into their own internal semantic networks (cf Smith, Weiss, & Ferguson, 1987).

Although these advantages seem plausible, Nonlinear text also presents a lot of problems for readers. These problems range over the entire set that was presented earlier, but we will focus on those associated with Learning Strategies, which arise from the lack of discourse cues in hypertext. Discourse cues are used by writers to give structure to a text so that its themes and logical structure can be understood by readers. Cues such as chapter divisions and section headings let readers organize and abstract the important parts of a text. The placement of the topic sentence at the beginning of a paragraph helps users integrate the remaining sentences. At the sentence level discourse cues give a document coherence (e.g., "next we discuss X").



Readers use such cues to control their reading strategies: decisions about which parts of the text should be read carefully or skimmed, or how particular points fit into the general theme--and if they are missing, comprehension will suffer (see Charney, 1987; Smith et al., 1987). Most Nonlinear documents segment a unified text into parts and these parts can be read in arbitrary orders; thus, they lack the conventional structure and discourse cues that readers have come to depend on for guiding their reading strategies. A reasonable assumption is that with careful design, such as careful segmentation, the tagging of important points, or the provision of preplanned "tours," comprehension problems can be avoided. Unfortunately, we do not yet have enough experience with authoring Nonlinear text to satisfactorily provide the kinds of cues readers need and expect.

I experienced some of these problems directly while reading a Hypercard stack (Apple, 1987) that was based on a 9-page printed report about the Hypertext '87 conference held in Chapel Hill, NC in November 1987. The printed report (Nielsen, 1988) summarized the important papers that were presented at the conference and outlined some of the important issues related to hypertext. I enjoyed the printed version very much and after reading it, felt that I had a good overview of what happened at the conference. However, the Hypercard version, which I read two months later, was rather difficult to assimilate.

The Hypercard version of the conference report (also authored by Nielsen) contained exactly the same text as the printed version except that it was segmented into 14 sections, each containing buttons pointing to other sections. Each section contained between one and four screens of text and some of the sections contained subsections. The stack structure was Nonlinear; i.e., all the cards were linked together in a network (via the use of buttons) and it was not possible to page through the stack sequentially, from the first to the last card.

This nonlinearity caused problems, most of which arose from not knowing the layout of the stack. I often wondered how much of the stack I had already read and how much remained. Furthermore, I worried about missing something important and found the stack impossible to read systematically. In the experiment described below, many other problems with reading Nonlinear text were observed.

The purpose of this experiment was to document the problems users have when reading Nonlinear text and the conditions under which they occur; in particular, their observable symptoms and users' strategies for coping with them. It

is hoped that experiments like this will lead to the development of computerized support for reading hypertext and to guidelines for designing readable hypertext documents.

In this experiment, reading comprehension, user preference, and browsing data (cards visited, buttons pressed, and times) were compared between two different versions of the Hypercard conference report described above. One group of subjects read a Nonlinear version, very similar to the Nielsen's original stack. A second group read a Linear version of the same stack. The Linear version was identical to the Nonlinear stack except that all buttons that linked the cards together in a network were removed and "next" buttons were added; i.e., the Linear version was essentially just a sequentially-ordered stack of cards, each linked only to its immediate predecessor and successor.

### **Method**

**Materials.** The Nonlinear stack contained 24 text cards (see Figure 1, top), 5 bibliography cards, and 5 overview maps (see Figure 2); the Linear stack contained 26 text cards (see Figure 1, bottom). Figure 3 displays another text card from the Nonlinear stack. There were two types of buttons in the Nonlinear stack, which distinguished it from the Linear stack: Note buttons and Reference buttons. The Note and Reference buttons usually appeared as boxed text (see Figure 1, top). Note buttons would bring up a little window containing a footnote-like comment; Reference buttons, on the other hand, would close the current card and bring up a card from another section of the stack. In Figure 1, the button labeled "20 years ago" was a Note button, and pressing it caused a little pop-up window to appear over the text card. In the Linear version, the text from the pop-up windows was added to the bottom of the main text along with an asterisk (see Figure 1, bottom) where the Note button appeared in the Nonlinear stack. The button labeled "Ted Nelson" in Figure 1 is a Reference button, and pressing it would move the user to the section of the stack that discussed Ted Nelson's work. Occasionally Reference buttons appeared outside of the main text (e.g., "Definition of Hypertext" in Figure 3), but usually both Reference and Note buttons appeared as boxed text.

NONLINEAR

**Is Text Data ?**  
**(Andy van Dam)**

Andy van Dam (Brown University) gave the opening speech giving his perspectives on hypertext from a historical point of view. They had invented the first hypertext system at Brown in 1967 - i.e. 20 years ago. The University of North Carolina had been the first place outside Brown itself to use the system so this made it fitting to have this workshop at UNC. First, van Dam wanted to acknowledge two trailblazers in the HT field: **Doug Engelbart** and **Ted Nelson**. Engelbart invented outlining/idea processing and did office automation before the word even existed, while Nelson coined the word "hypertext" itself and tried to put some early discipline into the links and associations in HT with his concept of gradually expanding stretchtext.

The original HT editing system ran in an 128 K machine timeshared with other users on a computer which was slower than a Macintosh.

This 20 year time lack between laboratory invention and commercial realization seems to be quite typical of many breakthroughs in computer science which is not moving so quickly as some may think. There was also about 20 years between Engelbart's invention of the mouse in 1963 and the use of it on a popular PC [Lisa] in 1983. Cf. also the comment by Tom Landauer that a rule of thumb at Bell Labs was that it would take 15 years from Lab concept to Real World application [ cited in my CHI'86 Trip Report , IFIP INTERACT Newsletter No. 17, pp. 3-10].

media and not to have size limitations on anything. Additionally, they would not have the size of something impact its performance. The system included bidirectional links: Not just goto, but also come from.

LINEAR

**Is Text Data ?**  
**(Andy van Dam)**

Andy van Dam (Brown University) gave the opening speech giving his perspectives on hypertext from a historical point of view. They had invented the first hypertext system at Brown in 1967 - i.e. 20 years ago. The University of North Carolina had been the first place outside Brown itself to use the system so this made it fitting to have this workshop at UNC. First, van Dam wanted to acknowledge two trailblazers in the HT field, Doug Engelbart and Ted Nelson. Engelbart invented outlining/idea processing and did office automation before the word even existed, while Nelson coined the word "hypertext" itself and tried to put some early discipline into the links and associations in HT with his concept of gradually expanding stretchtext.

The original HT editing system ran in an 128 K machine timeshared with other users on a computer which was slower than a Macintosh. The original sponsor was IBM but they could not tell them at that time that they were using the expensive graphics terminals just for displaying words. Later IBM did accept the idea of displaying words interactively on a graphics terminal and sold the system to

\* This 20 year time lag between laboratory invention and commercial realization seems to be quite typical of many breakthroughs in computer science which is not moving so quickly as some may think. There was also about 20 years between Engelbart's invention of the mouse in 1963 and the use of it on a popular PC [ Lisa] in 1983. Cf. also the comment by Tom Landauer that a rule of thumb at Bell Labs was that it would take 15 years from Lab concept to Real World application [ cited in my CHI'86 Trip Report , IFIP INTERACT Newsletter No. 17, pp. 3-10].

Figure 1: Examples of cards from Nonlinear and Linear stacks used in Experiment 1.

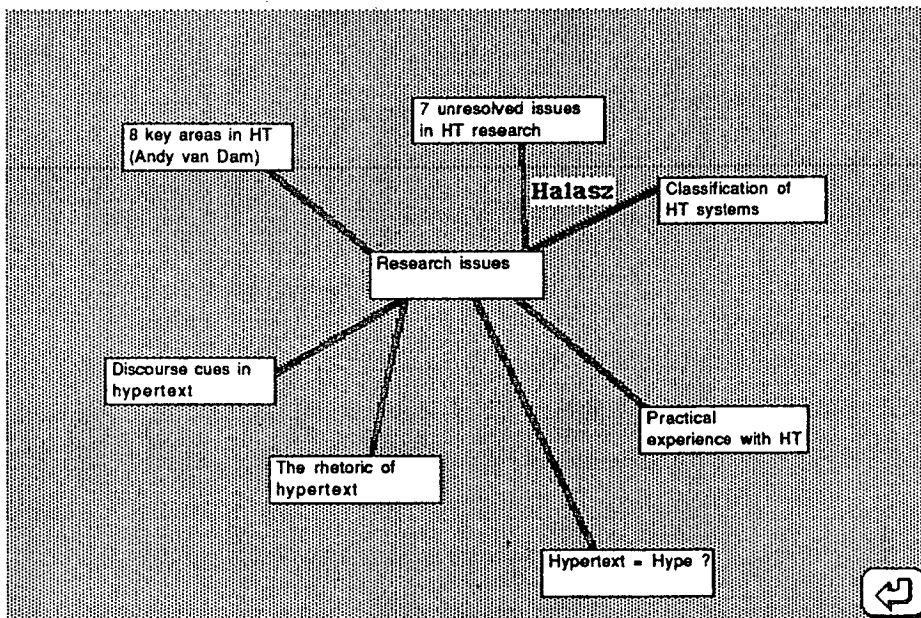
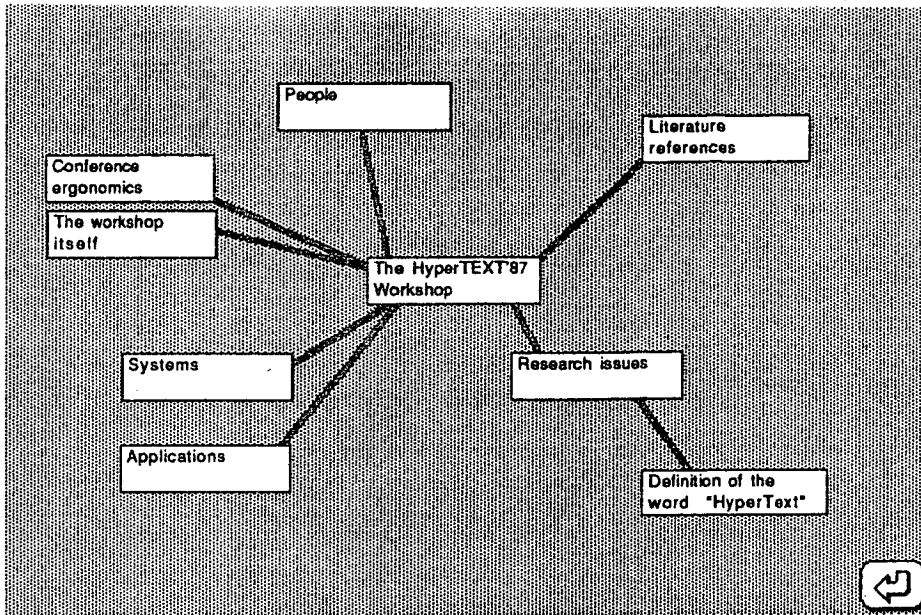


Figure 2: Examples of Overview Maps from Nonlinear stacks used in Experiment 1.

Both stacks contained Next buttons, which appeared as right or left arrows. Every card in the Linear stack had two (see Figure 1, bottom), whereas they only appeared in the Nonlinear stack to link together cards within a subsection. Finally, all the cards in the nonlinear version had a Return button (curved arrow in top card of Figure 1) that returned a user to whichever card was previously open. If a Reference button moved a user to a remote section of the stack, the Return button would bring him back to where he was before.

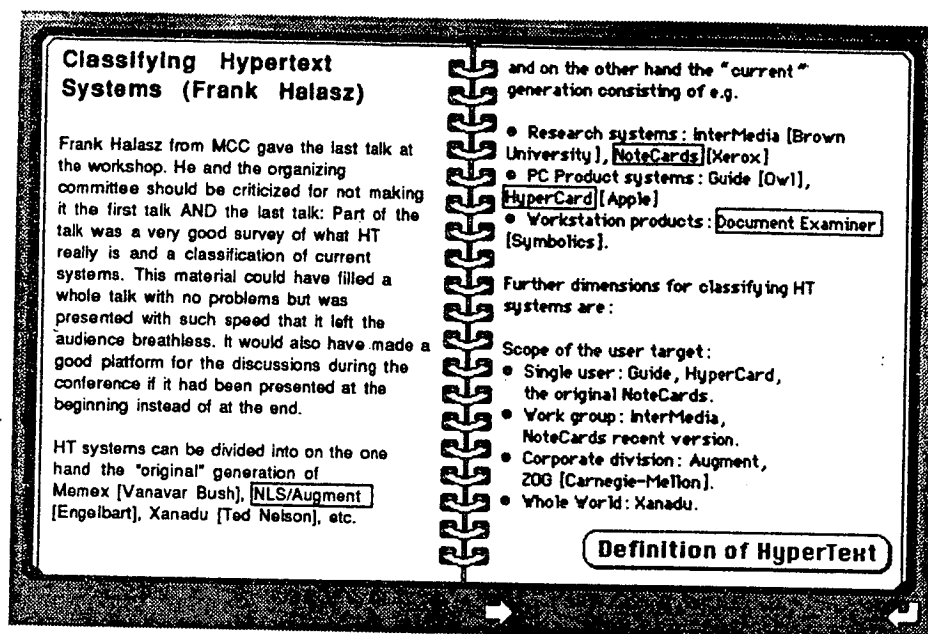


Figure 3: Example of text card from Nonlinear stack used in Experiment 1.

The Nonlinear stack also contained 5 overview maps (see Figure 2) that could be used to move to various sections within the stack. The top card in Figure 2 shows the top level overview map, which names all the main sections of the stack. Pressing one of the buttons in the top level map either brings up another overview map or moves a user to the section in the stack named by the button. For example pressing the button labeled "Research issues" in Figure 2 (top map) brings up another overview map centered on "Research issues" (bottom map) and pressing the button labeled "Classification of HT systems" brings up the first card in that section (see Figure 3).

The Nonlinear stack was designed so that the only means of navigation was via the use of these overview maps or by following Reference buttons from section to section. Navigation in the Linear stack was only possible by using the Next buttons.

**Subjects.** Ten subjects participated in this experiment; five read the Nonlinear version of the Hypercard stack and five read the Linear version. All were experienced

computer users but none were familiar with hypertext. Only minimal instruction in the use of Hypercard was required, however. More importantly, none of the subjects were familiar with the issues and researchers discussed in the Hypertext conference report, which was the basis for the two stacks.

**Procedure.** Subjects received the same instructions in both the Linear and Nonlinear conditions. They were told they had 45 minutes to read through the stack and that they would be given a short test of their comprehension afterwards. Pilot testing showed that the average time need to read all the cards in the stack was 60 minutes. Thus no subjects were expected to read all the cards and no one had time to read and study the text carefully. This was to ensure that a ceiling effect would not be observed. No notetaking was allowed. After the 45 minutes was up subjects were given a 15 item comprehension test. They could take as much time as necessary to complete the test; the average completion time was 12 minutes. Finally, they were given a short questionnaire, which asked what they liked or disliked about the stacks followed by an interview.

All card events (cards opened, buttons pressed, and times) were automatically saved, so they could be replayed for later analysis, but the subjects were unaware of this.

## **Results**

**Interviews and Questionnaires.** The subjects in the Nonlinear and Linear groups differed greatly in their evaluation of the stacks they read. Subjects in the Linear group had few complaints about the organization and presentation of the report other than the criticism that the font used in the stack was too small. Subjects in the Nonlinear group, however, encountered numerous problems while reading their version of the stack. In particular, three things about the Nonlinear stack were unknown, which lead to the problems described below: its overall structure (e.g., how the cards were organized into a network and the size of the subsections); the relative importance of cards and sections; the optimal path through the sections.

First, they wondered what the optimal order was for reading the sections in the stack. Since they were unfamiliar with the reading topic, they would have liked some guidance about which sections should be read first and which sections should be read later or last. But the Nonlinear stack lacked discourse cues for directing subjects to read the cards in a particular order. The rotary Overview maps (see Figure 2) did not indicate any obvious starting or finishing points, so subjects had to

use their own judgment for deciding where to start and what to read next. Furthermore, subjects had no guidance for deciding which Reference buttons to follow (which would move them to a different part of the stack) and which buttons to ignore. In contrast, the very structure of the Linear stack was an implicit discourse cue that indicated the optimal ordering for reading the cards; subjects could have chosen to deviate from the Linear ordering somewhat by paging forward or backward through the cards in the stack, but subjects realized that such skipping around was a deviation from the reading order that was intended by the author.

A related problem was the lack of discourse cues indicating the relative importance or difficulty of the various sections. Some sections described fundamental hypertext concepts whereas other sections described rather esoteric issues and applications. In the Linear stack the ordering of the cards was one discourse cue subjects could use to determine the importance of a card: the fundamental concepts appeared in the beginning cards and the more advanced concepts were described in the last cards.

A third problem for the readers in the Nonlinear group was that it was difficult to systematically read through all the sections of the stack, due to its network structure. It was difficult to systematically visit every card in the stack without a map that depicted how the cards were interconnected. The Overview maps did not assist in systematic exploration as they were "descriptive fictions" that showed the semantic relationships between the stack's sections, but not actual connectivity of the cards in the stack. Thus subjects wondered if they were missing some parts, did not know what percentage of the stack they had read, and ended up just roaming around the stack by following Reference buttons without a well defined reading plan.

**Problems with buttons.** The appearance and functionality of the buttons also caused some problems for the readers of the Nonlinear stack. Two problems arose from the way the buttons appeared in the stack. First, the Note and Reference buttons were both represented as boxed text (see Figure 1). Since both types of buttons were represented identically in the stack, subjects never knew if pressing a button would simply pop up a note or throw them to another part of the stack, thereby disrupting their current line of reading. A second problem was that the buttons' appearance were confused with emphasis. Boxed words could be interpreted as serving the same function as italicization.

In addition to problems arising from the appearance of buttons, three other problems arose from the functionality of the Reference buttons. One problem was

that the intended destination of the buttons was ambiguous. In Hypercard, as in most current-generation hypertext systems, the source of a button that links two cards together can be a single point, but the destination is an entire card. Thus, pressing a Reference button moved a reader to a card without indicating exactly the intended destination of the button. It was the readers' responsibility to determine what part of the destination card was relevant. This problem was exacerbated when the destination of a button was the first card of a section because then it was unclear if the intended link was to simply the first card of the section of the entire section.

A second problem related to the functionality of the buttons was that the destination card was often disappointing to the readers. Pressing a Reference button often took readers to a section of the stack that seemed irrelevant to what they were previously reading. Landow (1987) has pointed out that when hypertext readers follow a link, they expect the destination to be obviously and meaningfully related to the source of the link. When pressing a button moves a reader to something that is trivial or unrelated to the current line of reading, users get frustrated.

Finally, the Reference buttons frequently took readers to parts of the stack they had already visited. Upon arrival at Reference button's destination, subjects would have to first determine whether they had already seen the card, since it wasn't immediately obvious to them.

**Comprehension Test.** Given these problems that the readers of the Nonlinear stack faced, we assumed that their comprehension of the material would be hindered when compared to the Linear group, who did not have to cope with such problems. It turns out, however, that no significant differences in performance on the 15 item comprehension test were found between the Linear and Nonlinear groups.

The comprehension test contained 15 questions: 5 Integration questions, which required subjects to synthesize information that occurred in different sections of the stack; 5 Fact questions, which simply asked about something that was directly stated in the stack; and 5 Comparison questions, whose correct answer required the comparison of two things that were explicitly linked by buttons in the Nonlinear stack (but not linked together in the Linear stack). It was hoped that we would observe some performance differences between the Linear and Nonlinear group on these questions. We thought that the Nonlinear group would perform better on the Comparison questions, since these questions drew on information that was explicitly linked in the Nonlinear stack, and that the Linear group would perform better on the



Fact and Integration questions, since the Nonlinear group would not have to cope with the problems described above.

Unfortunately, the individual differences in the reading ability of the subjects was greater than the differences produced by reading the Linear versus the Nonlinear stack. Recall that this experiment had a between-subjects design; i.e., 5 subjects read the Linear stack and 5 subjects read the Nonlinear stack. A better way to have designed the experiment would have been to give each subject a Linear and Nonlinear stack to read and to make within-subject comparisons. But due to a shortage of potential subjects, this initial experiment had to be designed to minimize participation time.

**Browsing Data.** Although no differences were observed in comprehension performance, the Linear and Nonlinear groups did differ markedly in their reading strategies. A Hypercard script was written that automatically and unobtrusively saved subjects' reading behaviors. Every time a card was opened or a button was pressed, its name along with the event time was recorded and saved for later analysis. After the subjects finished reading the stacks, the experimenter had a record of card events and times that could be used to replay and analyze the subjects' reading behaviors. The main idea was to find *observable* sequences of card events that coincided with the subjects' browsing problems.

The main results of these analyses are displayed in Table 2. The first comparison made between the Linear and Nonlinear subjects is the total number of cards visited during the 45 minute reading session. The proportions include all cards that were opened during the session, even if they were not open long to read completely. These proportions give an idea of how thoroughly the two stacks were explored by the subjects. The Linear group visited an average of 24 of the 30 cards in the stack (i.e., .80 of the stack) while the Nonlinear group visited an average of 21.2 of 29 cards (i.e., .73). These numbers reflect the different characteristics of the two stacks and the different reading strategies adapted by the two groups of subjects. Three of the 5 subjects in the Linear group started by paging through all 30 cards of the stack without reading them, just to get an overview. The remaining two subjects simply started reading the first card of the stack and continued reading card after card until the time was up. In the Nonlinear group, an average of only 21.2 cards were visited. This is because the only means of navigation was through the Overview maps or by following Reference buttons making it difficult to systematically visit every card in the stack. Also, some of the cards were more likely

to be opened than others; some cards had many Reference buttons pointing to them while other cards had fewer entry points.

The second comparison between the Linear and Nonlinear groups is the mean number of cards read. This measure includes only those cards that were opened for more than 1.5 minutes as pilot testing showed that the average time to read a card was just about 2 minutes. By this measure, the Linear group read more 50% more cards than the Nonlinear group (see Table 2). The Linear group read an average of 19.2 cards while the Nonlinear group read an average of 12.2 cards. This difference is partly explainable by the tendency of subjects in the Nonlinear group to press Reference buttons, which might have appeared in the middle of a card, which would move them to a new card before reading completely the source card.

#### Browsing Data

	Non linear	Linear
Cards visited	.73	.80
Cards "read" (visited > 1.5 mn)	.42	.64
Looping (visited > 3 times)	.36	.18

-----  
 Note: Mean proportions appear in table. Number of Cards: Nonlinear = 29 (text); Linear = 30 (text and reference)

Table 2

The final comparison made between the two groups is the tendency to "loop" through the cards in the stack; i.e., to repeatedly visit the same cards or set of cards. Subjects in the Nonlinear group complained that the Reference buttons often took them to previously visited parts of the stack, which was annoying. We measured looping by totaling the number of cards that were visited more than 3 times (see Table 2). We chose 3 as the threshold for looping because it was, in fact, reasonable to open the same card 2 or 3 times. For example, if a subject in the linear group wanted to overview the stack by paging through the cards from beginning to end and back to the first card, each card would have been opened twice before even starting to read. But opening a card 4 or more times can be assumed to be pathological. As can be seen in Table 2, the Nonlinear group was far more likely to visit a card more than three times.

Finally, other aspects of subjects' behavior can be used to make inferences about reading difficulties. For instance, if a subject presses a reference Button then immediately returns to the source card without enough time to read through the destination card, one can assume that the subject was disappointed with the results of the button push.

### **Discussion**

The results of this experiment suggest that observables such as the temporal and spatial characteristics of users' trail through a hypertext network can be used to make inferences about their intentions or about whether they are disoriented. User behaviors such as flipping rapidly through a series of cards, pressing a button then immediately returning to the button's source card, and short reading times can imply that a reader is getting an overview, is disappointed with the destination of a button, or is getting distracted by the numerous buttons to press from the main task of reading the cards in the stack. Of course inferences about a user's mental state must be made with caution and, ideally, would be corroborated with other evidence. The main point here is that some aspects of user-disorientation can be diagnosed without knowledge of the *semantic content* of a hypertext network, simply by observing the pattern of a user's trail through the network (see Canter, Rivers, & Storrs, 1985).

We described many problems with the design of the nonlinear stack. After the fact, the problems described above seem obvious and easily avoided. However, it shows that we don't yet have enough experience with designing nonlinear text -- conventions for creating readable hypertext have yet to be agreed upon. And in particular, unnecessary segmentation should be avoided.

Many of the problems that users had reading the nonlinear stack could have been avoided with simple changes in its basic design. For example, the Reference and Note buttons could have had different appearances, and buttons should never be confused with emphasis. Also, the Reference buttons could have indicated whether the destination card was already visited (e.g., by automatically turning grey once the user visits the destination card). Furthermore, each individual card could have been marked as examined so that readers could have easily seen if they had already opened it without having to reread parts of it.

Another remedy for the observed problems would be to improve the Overview maps. As mentioned above, the Overview maps did not depict the real structure of the stack and instead imposed a logical organization. An improvement would be to

have dynamic Overview maps that depict the actual connectivity of the cards (or of the sections) and that become greyed-out as the reader moves through the stack to indicate which portions have been visited and which portions have yet to be examined. Another point is that although readers of hypertext do want an idea of the extent of a stack, simple solutions sometimes surpass complicated ones. For example, some subjects in the Linear group got a satisfactory overview of their stack by merely paging through all the cards. This took less than one minute and gave the subjects a satisfactory idea of the length of the stack. One can imagine creating an reader-controllable overview movie for a hypertext network that quickly pages through the cards as a way of giving an idea of the extent of the stack. Other solutions would have to be designed for extremely large networks, however.

### **Study 2: Analysis by Browsing**

A major problem faced by hypertext users is that of task management. An advantage of hypertext is that many different kinds of information are accessible at once. At any point, there are many opportunities for digressing from a main task and users are often tempted to follow many paths in parallel, which can cause them to lose their place or forget exactly why they chose to follow a link in the first place. Thus, experienced hypertext users devise task-management strategies that enable them to pursue multiple paths without getting lost.

The purpose of this experiment was to gather data on the task management and search strategies that novice hypertext users invented while working on a task that required the comparison and analysis of information in a hypertext network. A secondary purpose of this experiment was to collect data on browsing problems that arose from poor strategies for task management. Because the network used in this experiment had a rather simple structure, the problems we observed could not be attributed solely to unfamiliarity with the structure of the document network.

This experiment used an encyclopedia that was scanned into the Xerox NoteCards hypertext system (see Halasz, Moran, & Trigg, 1987 for details about NoteCards). The encyclopedia, called the World Fact Book, contains information about 188 countries with respect to 7 dimensions: Land, Water, People, Economy, Communications, Government, and Defence. For the experiment, a NoteCards

network (i.e., Notefile) was created that contained descriptions of only 10 of these countries. In this Notefile, however, all the names of the countries were removed and replaced by a simple identifier (e.g., "Turkey" was designated country "C"). The subjects were given the names of the 10 countries whose descriptions appeared in the Notefile and their task was to guess which of the 10 country-names matched which of the 10 entries in the encyclopedia.

This task has several nice properties. First, the task has a well defined goal state: the number of countries correctly identified. A structured task with an obvious measure of success is preferable to an open-ended task for the study of browsing because the strategies subjects use to perform the task can be objectively evaluated. Second, successful performance on this task required the comparison and analysis of many dimensions from many countries at once, but only a subset of the cards could be viewed on the screen at once, and the time to complete the task was limited to 45 minutes. Thus in order to complete the task on time, subjects had to adopt effective strategies for both managing the layout of the cards on the screen and for selecting the dimensions that would best discriminate between the countries. Finally, as will be described below, the structure of the Notefile was simple, so any symptoms of being "lost" (i.e., disorientation) could not be attributed solely to unfamiliarity with the structure of the hypertext network but instead to difficulties with task management. Conklin (1987) calls browsing problems arising from task management difficulties the Cognitive Overload Problem.

## **Method**

**Materials.** The Notefile containing the 10-country encyclopedia was organized as two trees containing common leaves. The leaves were 70 text cards that contained text on 7 topics for each of the 10 countries (see Figure 4, right). The root node of the first tree was a Filebox card labeled "Countries" (top left in Figure 4), which contained link icons pointing to Fileboxes for the 10 countries (A, B, C, etc.). Each of the 10 Fileboxes contained links to the seven topics (see card "A" in Figure 4) for that particular country. Selecting the link icons labeled "Land," "Water," "People," etc. would bring up a text card containing the information for that country. These cards typically contained one paragraph of text, but ranged in size from one sentence to about one typed page. The root node for the second tree was the Filebox card titled "SubTopics" (middle left in Figure 4).

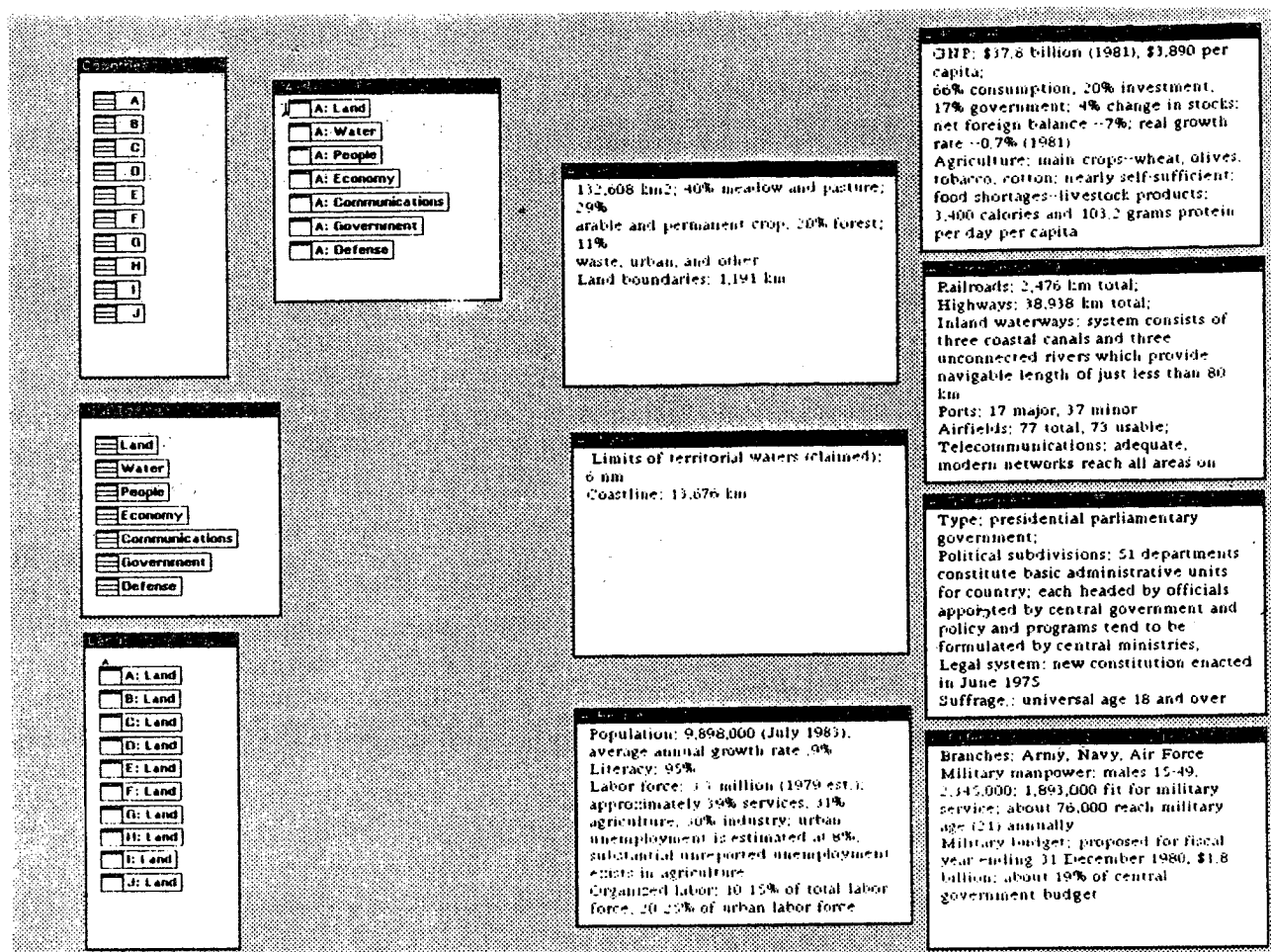


Figure 4: Screen image showing structure of NoteCard's encyclopedia used in Experiment 2.

This card contained link icons pointing to Fileboxes for the 7 topics. A topic Filebox contained pointers to the text cards which describe that particular topic for all 10 countries (see card "Land" lower left, Figure 4). Note that the card titled "A: Land" could be accessed either from the "Land" Filebox or the "A" Filebox.

Altogether this Notefile contained 89 cards: 2 root Fileboxes, a Filebox for each of the 10 countries, a Filebox for each of the 7 topics, and 70 text cards (10 countries x 7 topics).

The functionality of the NoteCards system was simplified for this experiment. No new cards or links could be created, cards could be read but not modified, and the set of normally available window commands was limited to CLOSE (removes a window from the screen), MOVE (changes the position of a window), SHAPE (changes the size of a window), and SHRINK (turns a window into an icon). Subjects could have as many of these cards opened on the screen as they wanted and they could be arranged and sized in any way.

**Browsing Task.** The subjects were given a sheet of paper listing the 10 countries whose descriptions appeared in the Notefile (see Figure 5). Their task was to determine match the country names with the appropriate letter by analyzing and comparing the information in the Notefile. Subjects could take notes (with paper and pencil) if they wished.

This task involved making many *relative* comparisons and deductions over several dimensions for each country. For example, the statistic of 17K sq km for the area of a country might not mean much to someone who doesn't know what the average size of a country is. But after determining the area of several countries (by examining their Land cards), one can get an idea the size that country compared to the others in the Notefile. Also, two countries might be very similar with respect to size, economy, and literacy rates, but have different governmental systems (e.g., communist vs. democratic systems). Thus, the correct identification of a country might depend on comparing it over several dimensions with several other countries - and the most discriminable dimensions might change from country to country.

---

China =

---

Denmark =

---

Egypt =

---

Greece =

---

India =

---

Iran =

---

Portugal =

---

Romania =

---

Switzerland =

---

Turkey =

---

Figure 5: Example of answer sheet used in Study 2

**Country List:**

- A
- B
- C
- D
- E
- F
- G
- H
- I
- J

**Country C Data:**

- Population: 45,851,000 (July 1981); average annual growth rate 2.8%; Literacy: around 44%
- Labor force: 13.4 million; 45% agriculture, 13% industry, 11% trade and finance, 26% services and other; shortage of skilled labor
- Organized labor: 1 to 3 million

**Country D Data:**

- Population: 49,135,000 (July 1981); average annual growth rate 3.2%; Literacy: 62%
- Labor force: 17.14 million; 58% agriculture, 29% service, 13% industry; surplus of unskilled labor (1980)
- Organized labor: 10-15% of labor force

**Country H Data:**

- Population: 42,490,000 (July 1983); average annual growth rate 3.1%; Literacy: 50%
- Labor force: 12.0 million, est. (1979); 33% agriculture, 21% manufacturing; shortage of skilled labor substantial

**Country H Summary:**

- Type: republic
- Political subdivisions: 23 provinces, subdivided into districts, subdistricts, 2 counties, and villages
- Legal system: the new constitution codifies Islamic principles of government
- Suffrage: universal, age 15 and over

**Country D Additional Data:**

- and 76.5 grams protein per day per capita (1977)
- Major industries: textiles, food processing, chemicals, petroleum, construction, cement
- Exports: \$4.0 billion (c.i.f., 1981); crude petroleum, raw cotton, cotton yarn and fabric, rice, onions, potatoes, chemicals, cement
- Imports: \$6.8 billion (c.i.f., 1981); foodstuffs

**Country H Additional Data:**

- Crude steel: 1.8 million tons produced (1981); 27 kg per capita
- Exports: \$4.701 million (c.i.f., 1981); cotton, tobacco, fruits, nuts, metals, livestock products, textiles and clothing
- Imports: \$8.931 million (c.i.f., 1981); crude oil, machinery, transport equipment, metals, mineral fuels, fertilizers, chemicals

**Country D Additional Data:**

- Major industries: crude oil product (2.3 million b/d in 1982) and refining, textiles, cement and other building materials, food processing (particularly sugar refining and vegetable oil production), metal fabricating (steel and copper)
- Exports: \$18.4 billion (est., 1982); 97% petroleum; also carpets, fruits, and nuts
- Imports: \$ 10.6 billion (est., 1982);

Figure 6: Screen image illustrating analytic task from Experiment 2.

Figure 6 illustrates one exemplary strategy for successfully identifying the 10 countries. This hypothetical subject started by examining the People cards for all 10 countries, which contains information about population and literacy rates. On the basis of population, he made a rough division of the 10 countries into 3 groups: high, medium, and low population. Next, he noted that the three countries that had a medium population also had fairly low literacy rates (about 50%). After scanning the list of 10 possible countries, the subject hypothesized that countries C, D, and H were probably Iran, Egypt, or Turkey, but he didn't know which was which. At the top of Figure 6, this subject opened the Fileboxes for countries C, D, and H, and then he examined the text cards for People and Economy for those countries. He noted that both countries D and H exported oil but not country C, and so he hypothesized that C was Turkey (he knew that Turkey did not export oil), and that D and H were Iran and Egypt. To further discriminate between D and H, he opened the Government card for H and saw that it had an Islamic legal system. Thus he decided that H had to be Iran, D was Egypt, and that C was Turkey.



**Subjects.** The subjects were 10 research scientists and students from INRIA and Xerox PARC. None of the subjects had prior experience with hypertext systems, but all were workstation users and therefore familiar with the use of a mouse and windows.

**Procedure.** First, subjects were given a basic overview of NoteCards: how to open cards, follow links, position cards on the screen, etc. Since NoteCards was simplified for this experiment, subjects did not require much training in its use beyond a simple five-minute overview. Next, subjects were given instructions for the identification task. After they were given the answer sheet, which listed the 10 countries, they were told to determine the names of the countries in the Notefile and to place the appropriate letter beside the country name on the answer sheet. They were also told they could take notes if they wanted. Finally, they were told they had 45 minutes to complete the task.

While the subjects worked on the task, all card events (e.g., open card, close card, move card, reshape card) and event times were automatically saved. After the experiment was finished, the experimenter could then play back a "movie" of the experimental sessions for analyses of the subjects' browsing strategies.

## **Results**

The primary data for this experiment were the card events that were saved while the subjects worked on the task, think-aloud protocols, and post-session interviews.

To start, the strategies used by the most and least successful subjects will be contrasted. In this task, subjects had develop strategies to contend with two limitations. First, since there was not enough time to examine all seven topics for the 10 countries in the Notefile, subjects had to make comparisons between what they thought would be the most discriminating topics. For example, land mass and population-size discriminates China and India from the other countries, but not from each other. Type of government would discriminate between China and India, but not between Portugal and Greece. The inability to perform an exhaustive search is an inherent limitation for most information retrieval and browsing tasks, so users must adopt adequate search strategies to ensure success.

Second, subjects had to contend with screen space limitations, which are endemic to most computerized tasks. Since NoteCards sets no limit on the number of cards open at once, the screen can easily become cluttered with lots of open cards,

many of which might be unnecessary for the current task. Once the screen becomes cluttered, users have to spend time recovering from "Window Faults" -- the unnecessary reshaping, burying, unburying, and moving of cards. With some foresight, the number of Window Faults can be reduced. A problem related to this is faced by many people who keep a cluttered desk in their office. Many people position stacks of papers on their desks to act as reminder cues for tasks that need immediate attention. Papers for less immediate tasks might be placed at the back of the desk or on the bottom of a pile. Furthermore, the layout of papers and books on a desk can help people reestablish context when they return to work after an interruption. Once the desk becomes piled with papers and disorganized, important jobs and agendas can be lost. In the same way, users can use the placement of windows on a computer screen to remind them to tasks to be done or to help them reestablish context after interruptions. The second problem for subjects was to devise strategies for maintaining an effective organization of the open cards on the screen: limiting the number of open cards without losing context. These problems will be referred to as Task Management problems.

***Separation of Task Management and Search Strategies.*** The card event data that was saved during the experiment was played back by the experimenter for analysis of the subjects' strategies. A separation of Task Management and Search strategies could be seen in the data. The 10 subjects fell into three groups: 3 subjects had good Task Management strategies but poor Search strategies; 4 subjects had good Task Management strategies and good Search strategies; and 3 subjects had poor Task Management strategies but good Search strategies. This rough grouping can be seen in Figure 7, which plots the average number of open windows (i.e., cards) by the number of correctly classified countries.

To determine the average number of open windows, a count of open windows was taken from the the card-event data at 3 minute intervals, for a total of 15 samples from the 45 minute experiment, which was then averaged.

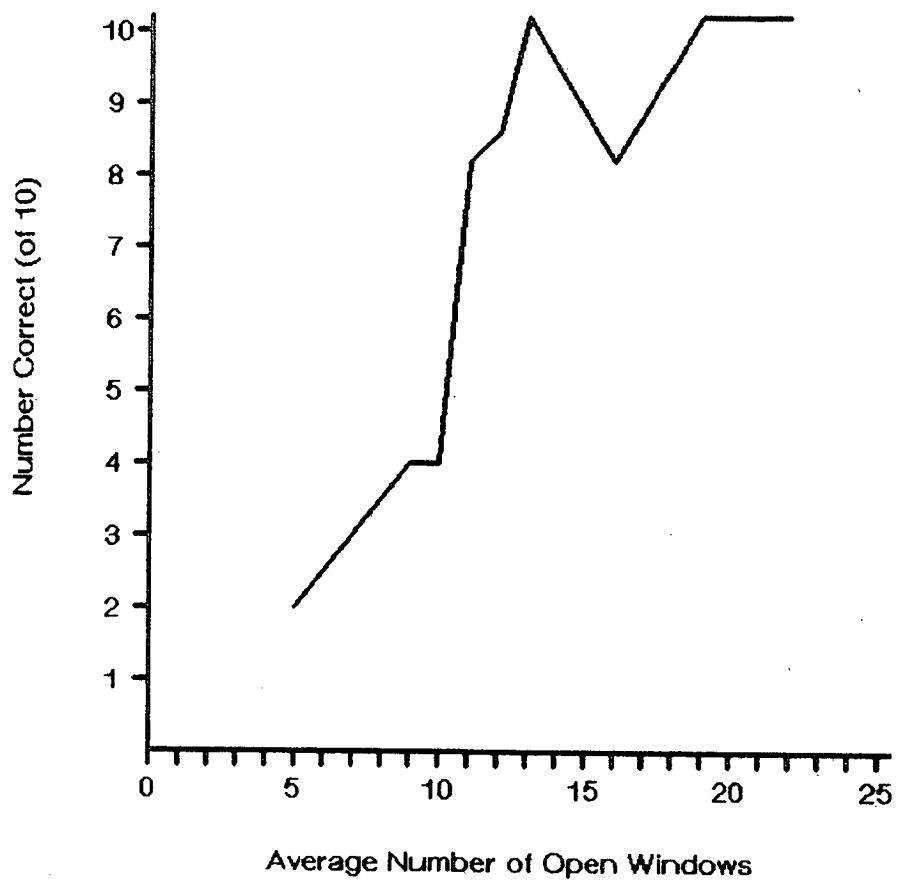


Figure 7: Relationship between number of open windows and performance in Experiment 2.

As can be seen in Figure 7, the 3 subjects who the lowest average number of open windows performed the poorest on the classification task. These subjects had good Task Management strategies but poor Search strategies. A screen image from one of these subjects is shown in Figure 8. This subject had an average of 9 open windows throughout the course of the experiment. His approach to the classification task was to get a "gestalt" of each country by opening the Land, People, and Government cards without making comparisons with the other countries. He would close all the text cards for one country before opening the cards for another country. Although this subject had very few Window Faults, his search strategy was ineffective because very few between-country comparisons were made. Two of the subjects in this group limited themselves to only three or four topics while ignoring the others. Furthermore, much of the available screen space that could have been used for reminder cues or task tracking was wasted. Instead, these subjects took extensive notes, copying down the statistics for the various countries on his answer sheet, rather than grouping the cards with crucial information somewhere on the screen.

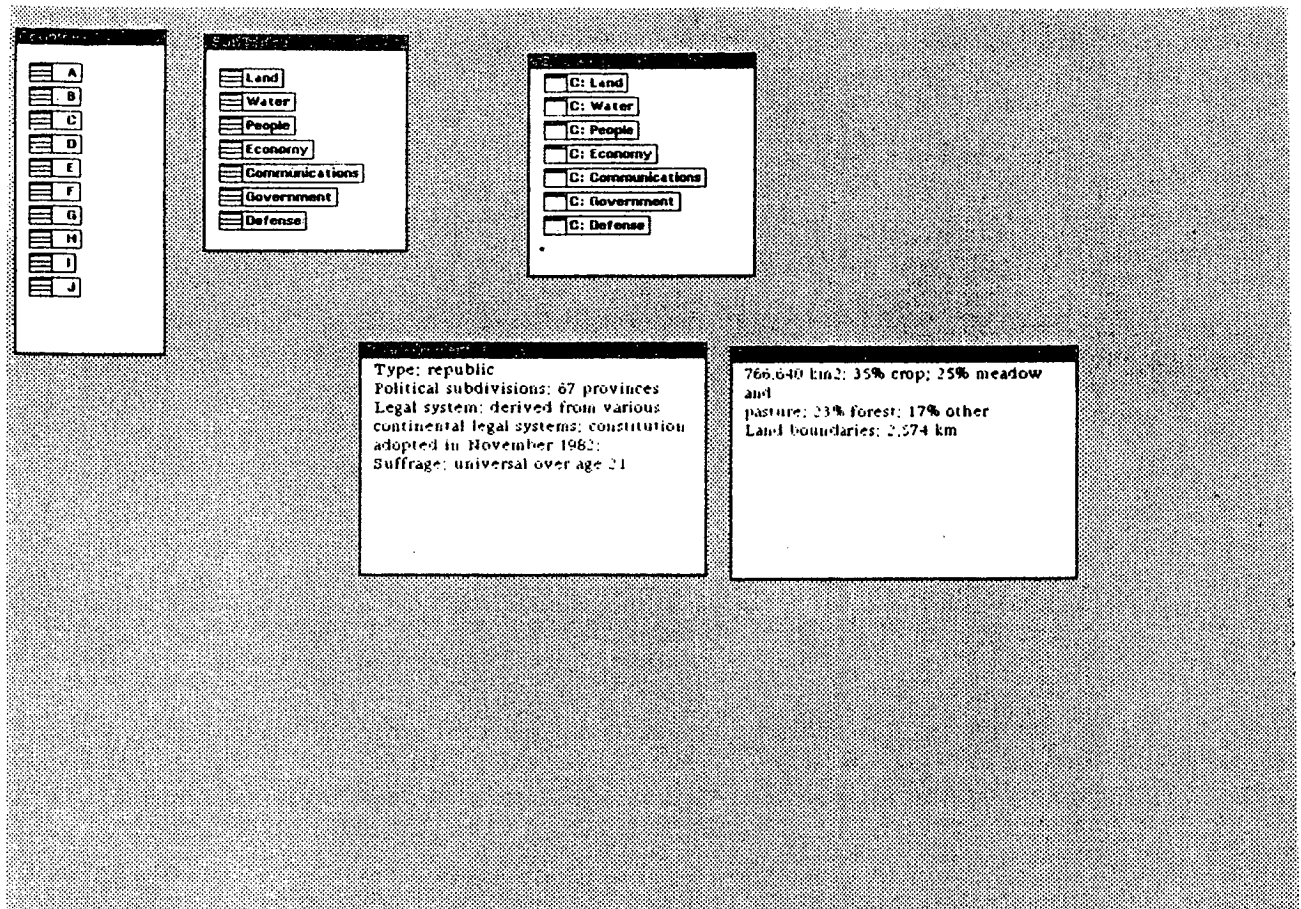


Figure 8: Screen image illustrating a subject's screen who had good Task Management strategies, but poor Search strategies in Experiment 2.

Even though these subjects had a very low average number of open windows, they examined a greater proportion of the cards in the Notefile (see Figure 9) and had a longer average solution time per correct classification (see Figure 10).

Next, we discuss the 3 subjects who had poor Task Management Strategies but good Search Strategies. Figure 11 shows a screen image taken from a subject who had the highest average of open windows (24) during the experiment. There are 29 open windows in this figure. His approach to the task was similar to the 2 other subjects who averaged 20 or more open windows. Interestingly, although this subject had considerably more Window Faults than the other subjects, he managed to correctly classify all 10 countries. As can be seen in Figure 11, the subject had many cards open on his screen but a good proportion of them were buried.

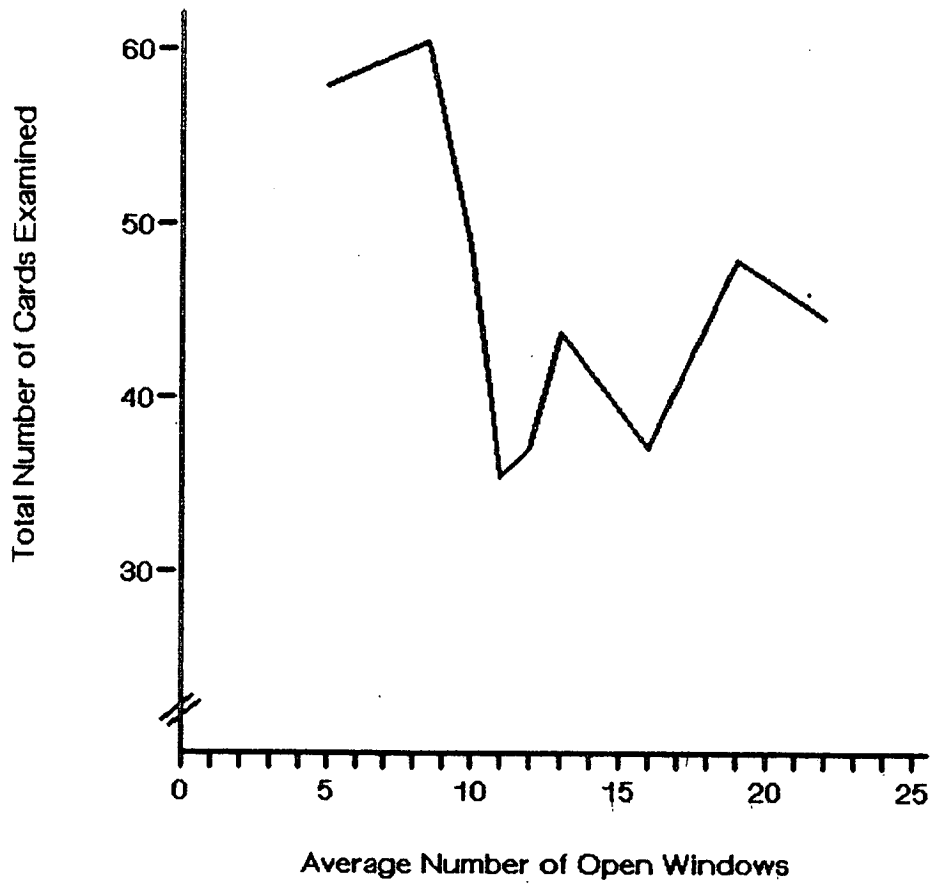


Figure 9: Graph displaying the relationship between average number of open windows and the extent to which the cards in the Notefile were examined (data from Experiment 2).

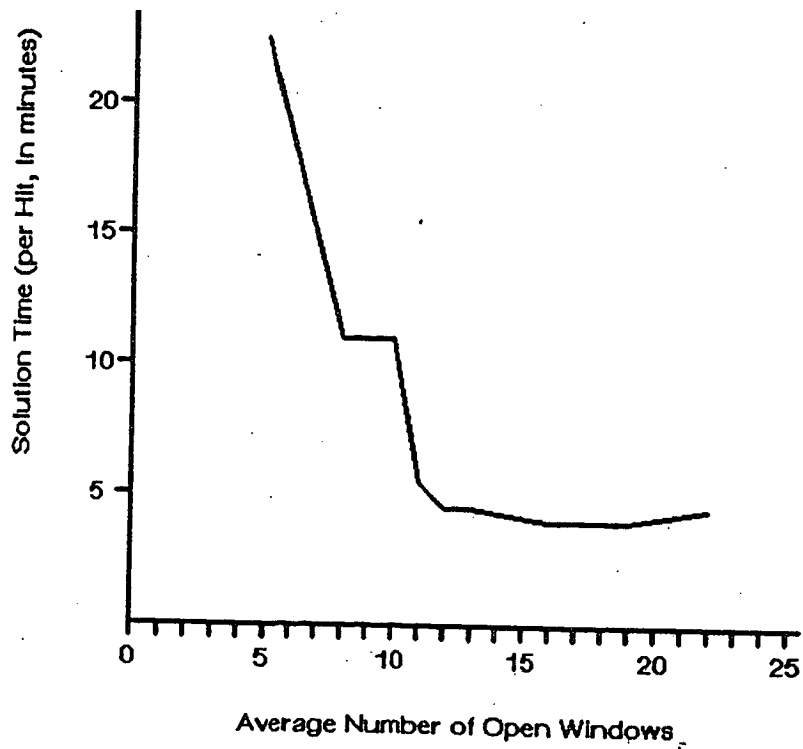


Figure 10: Graph displaying solution time as related to the average number of open windows in Experiment 2.

In contrast to the subjects who kept a very neat screen, this subject rarely closed cards when he was finished with them. Due to his cluttered screen, he would forget which cards were already opened, and due to his unsystematic search strategy, he would return to cards he had already examined.

This subject would also get sidetracked a lot. For example, while he was trying to determine whether country A or country B was Greece, he opened a Water card for the first time (prior to that he had focused on Government, Land, and Economy). After noticing that the Water topic contained information about the length of a country's coastline, he decided to examine the Water cards for all 10 countries in order to check his earlier hypotheses before returning to the task of assigning Greece to A or B. But since making a large number of comparisons between countries and topics was critical for success in this experiment, the 3 subjects with the most open windows performed well. Although some time was wasted on window faults, the large number of open windows allowed these subjects to maximize the number of comparisons made and to notice incidental information that would have been missed had they kept a neat screen.

Figure 12 shows a screen image taken from a member of the group with a moderate average number of open windows (14). This group, containing 4 subjects, had both good Task Management strategies and good Search Strategies. The subjects in this group structured the cards on the screen to simplify the comparison process, to act as reminder cues, and to keep track of running hypotheses. They also tended to close cards after they determined which country corresponded to them rather than leaving them open on the screen. They also took a more systematic approach to the classification task than the subjects with the highest number of open windows. In Figure 12 for example, some topic cards for countries A, B, and C were opened and neatly arranged to facilitate comparisons. This subject hypothesized that A, B, and C were probably Greece, Portugal, and Turkey, but he was not sure of the exact mapping; he stayed with the task of determining the exact assignments before moving on to another subtask.

Figure 13 displays an example of the use of card placement for reminders. Before moving to the task of differentiating countries C, D, and H, this subject placed the cards A: People and B: People in the upper right corner of his screen to remind him to return to categorizing countries A and B after the current task is finished. This Task Management strategy typifies how the 4 subjects in this group placed cards on the screen to facilitate comparisons, to keep track of their hypotheses, and as memory cues.

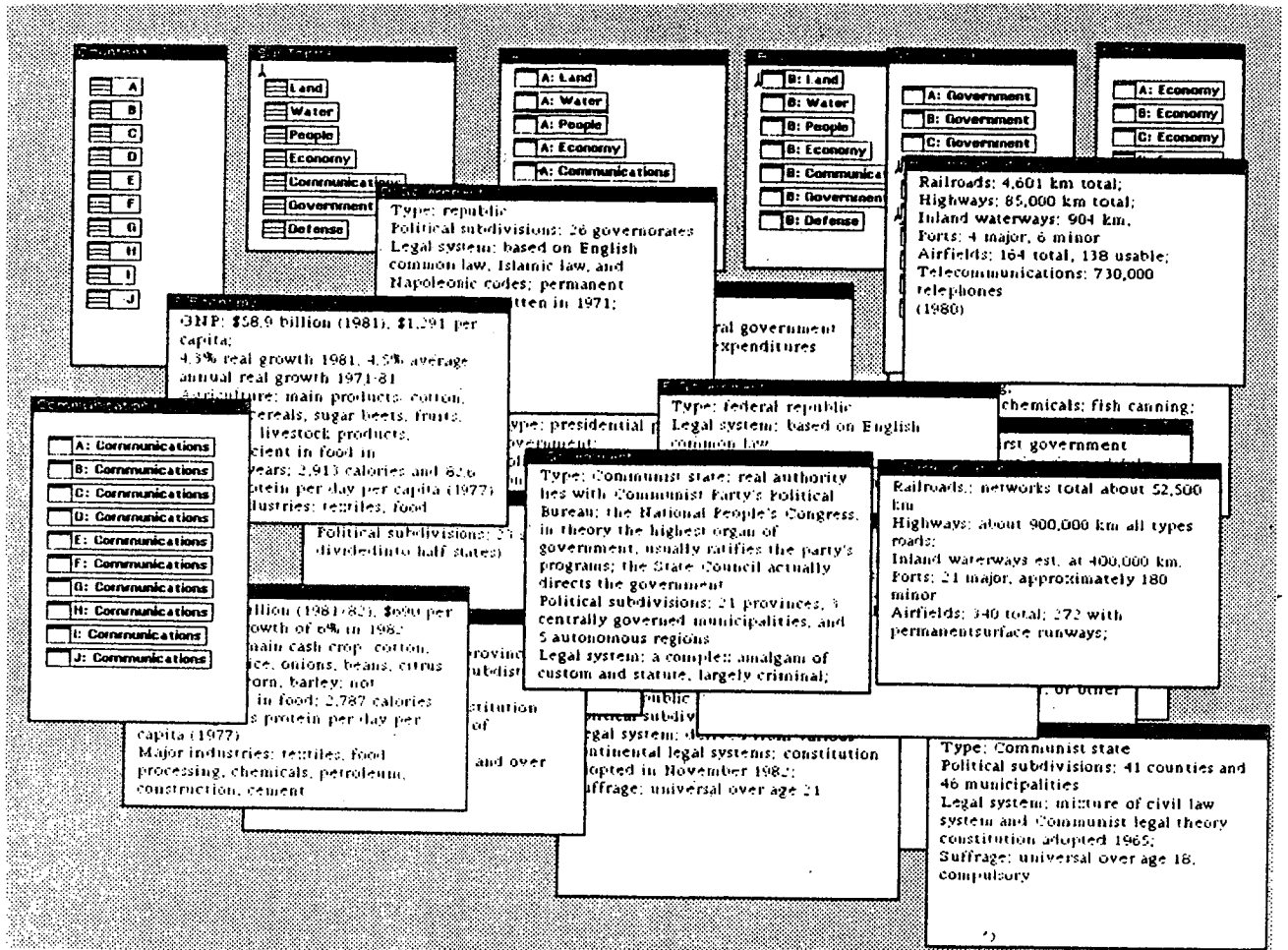


Figure 11: Screen image from subject with poor Task Management strategies but good Search strategies. 29 cards are open in this screen snapshot (image from subject in Experiment 2).

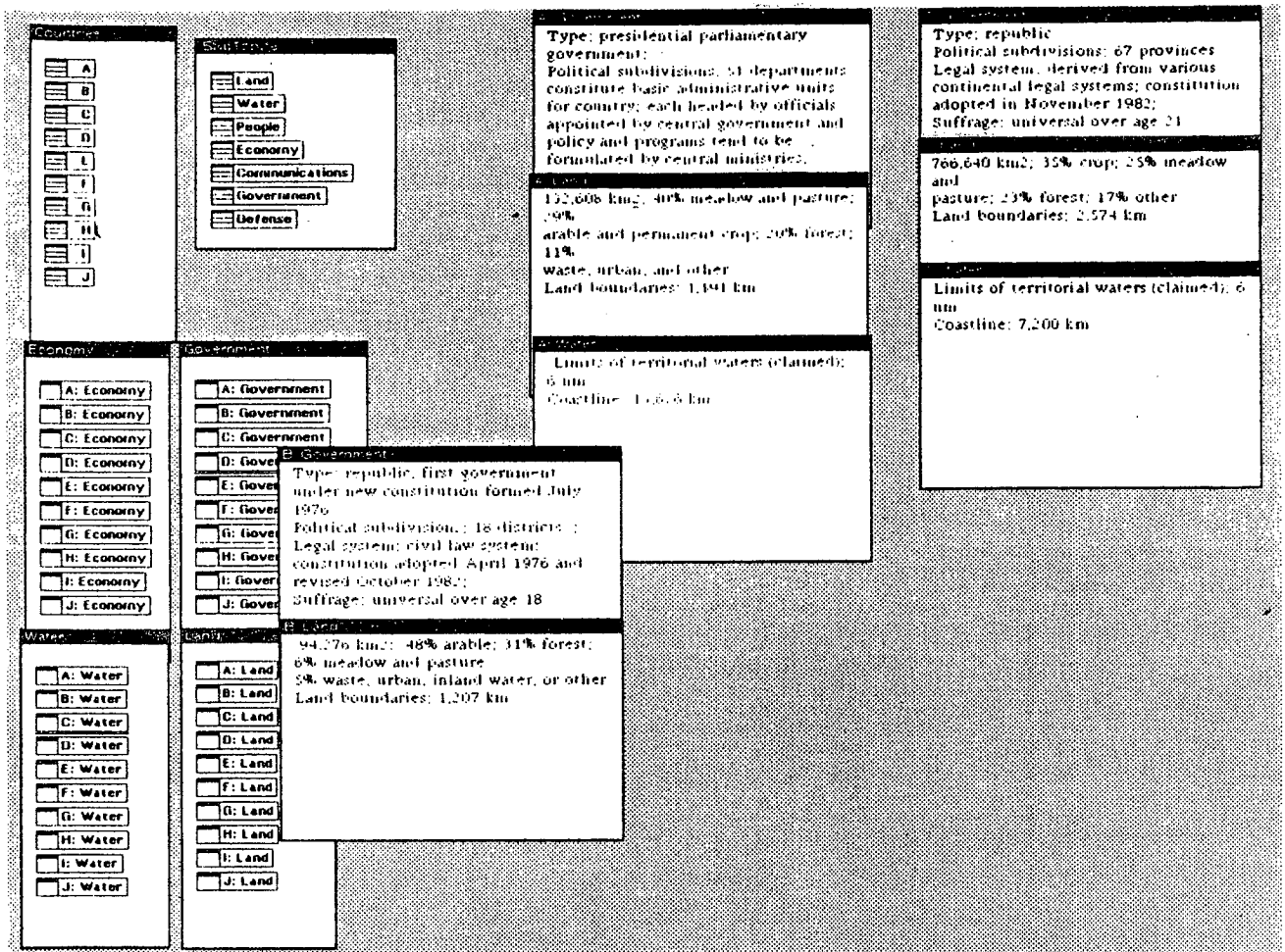


Figure 12: Screen image from subject in group with both good Task Management strategies and good Search Strategies in Experiment 2.



The image shows a complex interface with several windows and data boxes:

- Left Panel:** A vertical list of categories A through J, each with a small icon.
- Top Row:**
  - Land:** Land, Water, People, Economy, Communications, Government, Defense.
  - Water:** Land, Water, People, Economy, Communications, Government, Defense.
  - People:** Land, Water, People, Economy, Communications, Government, Defense.
  - Economy:** Land, Water, People, Economy, Communications, Government, Defense.
  - Communications:** Land, Water, People, Economy, Communications, Government, Defense.
  - Government:** Land, Water, People, Economy, Communications, Government, Defense.
  - Defense:** Land, Water, People, Economy, Communications, Government, Defense.
- Right Side:**
  - Country 1:** Population: 9,698,000 (July 1983), average annual growth rate 0%.
  - Country 2:** Population: 10,008,000 (July 1983), average annual growth rate 0.4%, Literacy: 70%, Labor force: (1980) 4.0 million; 36% services, 7% agriculture, 27% industry; unemployment is now more than 15%. Organized labor: about 50% of labor is organized.
- Bottom Row:**
  - Left Panel:** A vertical list of categories A through J, each with a small icon.
  - Country 3:** Type: republic. Political subdivisions: 26 governorates. Legal system: based on English common law, Islamic law, and Napoleonic codes; permanent constitution written in 1971.
  - Country 4:** Type: republic. Political subdivisions: 67 provinces. Legal system: derived from various continental legal systems; constitution adopted in November 1982. Suffrage: universal over age 21.
  - Country 5:** GNP: \$28.9 billion (1981), \$1,291 per capita; 4.5% real growth 1981; 4.5% average annual real growth 1971-81. Agriculture: main products: cotton, tobacco, cereals, sugar beets, fruits, nuts, and livestock products; self-sufficient in food in average years; 2913 calories and 82 grams protein per day per capita (1977). Major industries: textiles, food processing, chemicals, petroleum, construction, cement.
  - Country 6:** Type: republic. Political subdivisions: 23 provinces, subdivided into districts, subdistricts, counties, and villages. Legal system: the new constitution codifies Islamic principles of government. Suffrage: universal, age 15 and over.
  - Country 7:** Population: 42,490,000 (July 1983), average annual growth rate 1.1%, literacy: 50%. Labor force: 12.0 million, est. (1979); 13% agriculture, 21% manufacturing; shortage of skilled labor substantial.

Figure 13: Screen image displaying use of reminder cues in Experiment 2. This subject placed 2 cards in upper right-hand corner to remind himself to return to the task of analyzing those countries.

## Discussion

Our original hypothesis was that the subjects with poor Task Management strategies would waste too much time organizing their cluttered screens and therefore perform poorer than those who kept neat screens. But as can be seen in Figure 10, the subjects with the largest average number of open cards are also the ones who categorized the countries fastest. Even though they may have spent some time uncluttering their screens, the large number of open cards enabled them to view much information at once, which facilitated their performance in the classification task.

Our results show that the main problems that the subjects in this experiment faced were: poor Search Strategies (by those subjects who kept neat screens but failed to make enough between-country comparisons); losing track of unsubstantiated hypotheses; forgetting which evidence was used to make particular classifications; and finally, too many tedious window operations (e.g., opening a series of cards, arranging them in a matrix, then closing them).

Now that we know exactly what kinds of problems users will have, it is possible to design various tools that will fix these task-specific problems. Typical browsing tools such as graphic maps, bookmarkers, and tables-of-contents do not provide adequate support for analytic browsing tasks such as this. Our subjects faced problems that differed from the standard disorientation problems that most browsing tools were designed to alleviate. (See Foss, 1989 for a discussion of how task-specific browsing tools can surpass global maps for alleviating user disorientation.)

Figure 14 displays a History Matrix, designed specifically for alleviating the problems subjects had while working on the classification task in Experiment 2. This browser is not yet implemented, but is an instance of a task-specific browser. This specialized browser features:

*Overview of Notefile:* This matrix represents an overview of the 70 text cards and 17 main Fileboxes in the Notefile used in this experiment. The rectangles are active buttons that open the cards in the Notefile.

*Task management assistance:* The buttons in the top row and first column access the Topic and Country Fileboxes. For example, pressing the button labeled People opens the People Filebox (Figure 14, lower left). Pressing the button labeled "E" opens the Filebox for country E. Individual cards can be accessed by pressing one of the buttons in the main matrix; e.g., pressing the button under Land - A opens the A: Land card.

*History:* Each card that has been opened is shaded grey in the matrix. For example, this subject examined the People cards for all 10 countries. Furthermore, each individual card that has been opened is marked as examined (with a "+"). This helps users notice which cards or types of cards they have yet to examine and helps them keep track of which cards they've already opened.

*Dependency tracking:* The black buttons in the matrix mark the pivotal information that the subjects based their classifications on. For example, the matrix indicates that the subject decided that country J was China on the basis of the information in J's People and Government cards. Note that the subject only examined one other card for Country J: Land (shaded grey). Subjects keep track of these dependencies themselves, by double-clicking on the appropriate button, which blackens it.

*Progress tracking:* As users make their classifications, they can type their country assignments to the left of the main column, which automatically darkens the button for that country's Filebox button. One can also keep track of current hypotheses.

The idea behind the History Matrix, which was designed specifically for this task, is that a task-specific browser can be easily designed to support tasks that a more generalized browser cannot. Users could be provided with tailorable browsers or browser toolkits, which could be used to customize browsers and support tools for their own specific needs. In addition to the History Matrix, other tools could be built such as a timeline that reinstates previous screen layouts or a screen manager that automatically lays out a matrix of open cards by Country and Topic. A direction for future research is the development of browser toolkits that allow users to specify a customized arrangement of browser primitives (e.g., history mechanisms, overview mechanisms, screen managers) for their particular task and hypertext document.

		Land	Water	People	Economy	Comm	Gov't	Defense
	A			■			■	
	B			■			■	
	C			■			■	
	D			■			■	
Romania	E	■		■			■	
India	F			■			■	
	G			■				
	H			■				
	I			■				
China	J	■		■			■	

People	
A: People	
B: People	
C: People	
D: People	
E: People	
F: People	
G: People	
H: People	
I: People	
J: People	

E	
E: Land	
E: Water	
E: People	
E: Economy	
E: Communications	
E: Government	
E: Defense	

J	
J: Land	
J: Water	
J: People	
J: Economy	
J: Communications	
J: Government	
J: Defense	

Figure 14: A History Matrix, an example of a task-specific browser that was designed for the analytic task in Experiment 2.

## Conclusions

In the introduction, a classification of browsing problems according to cognitive outcomes was presented. The results of these studies allow us to augment this characterization of browsing problems. In Table 3, we have supplanted the subjective descriptions of browsing problems with observable behaviors. *Symptoms* refer to the externally observable behaviors that accompany Disorientation, Task Management, and Search/Learning strategy problems. For example, disoriented users might take a long path to a node when a more direct route is possible, or they might open cards and then immediately close them after seeing that they've already visited them. Symptoms of Task Management problems usually pertain to how cluttered the screen is. Users who arrange their screen-tops carefully, arranging cards to act as reminder cues, are less likely to lose their place and to more likely to resume tasks after interruptions and digressions. Symptoms of subjects who have difficulty learning from browsing might include relatively short reading times, which indicates only cursory examination of the contents of the network, or very restricted search paths. Of course, symptoms such as these cannot be used alone to determine whether a user is having browsing problems. Other factors such as the expertise of the user, the task he is trying to accomplish, and the characteristics of the hypertext network must be taken into account.

*User responses* refer to what users *do* once they realize they're having problems. Disoriented users might return to a known part of the network, definable as a portion of the network that has been visited the most frequently during a browsing session. Users who realize that they are having trouble completing subtasks, which are a part of a larger task, might suddenly most of the windows on their screens then maintain a much lower average of open windows, at the same time reducing the number of Window Faults. Users who aren't comprehending what they're reading might suddenly increase their reading times.

With a better understanding of the precise symptoms of disorientation and other problems, and the conditions under which they occur, we can develop browsing environments that support the kinds of tasks users wish to accomplish while avoiding common pitfalls of browsing large networks of information.

**Browsing Problems:  
Observables**

**Disorientation**

Symptoms: looping, inefficient paths to target, query failures

User Response: return to familiar neighborhood, reformulate queries

**Task Management**

Symptoms: window faults and messy screen; repeated states, paths,  
and relationships; inverse operations

User Response: close many/all windows, rearrange screen layout

**Search/Learning Strategies**

Symptoms: poor search strategy, very short reading times, probe  
questions

User Response: shift in card/link types examined, shift in reading  
times, shift in search strategies (more depth or breadth)

Table 3

### References

- Apple Computer, Inc. HyperCard user's guide. Cupertino, CA; 1987.
- Beeman, W. O., Anderson, K. T., Bader, G., Larkin, J., McClard, A. P., McQuillan, P., & Shields, M. Hypertext and pluralism: From lineal to non-lineal thinking. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Bolter, J. D. & Joyce, M. Hypertext and creative writing. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Canter, D., Rivers, R., & Storrs, G. Characterizing user navigation through complex data structures. Behaviour and Information Technology, 4(2): 93-102; 1985.
- Charney, D. Comprehending non-linear text. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Conklin, J. Hypertext: An introduction and survey. Computer, 20(9):17-41; 1987.
- Conklin, J. & Begeman, M. L. gIBIS: A hypertext tool for team design deliberation. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Foss, C. L. Some tools for browsing and reading hypertext. In press, Information Processing and Management, 1989.
- Frisse, M. E. Searching for information in a hypertext medical handbook. Communications of the ACM, 31(7): 880 - 886; 1988.
- Halasz, F. G., Moran, T. P., & Trigg, R. H. Notecards in a nutshell. Proceedings of the ACM Conference on Human Factors in Computing Systems; Toronto, Canada; April 1987.
- Landow, G. P. Relationally encoded links and the rhetoric of hypertext. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Marshall, C. C. Exploring representation problems using hypertext. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Nielsen, J. Trip report: Hypertext '87 Chapel Hill, North Carolina, 13 - 15 November 1987. SIGCHI Bulletin, 19(4): 27 - 35; 1988.
- Oren, T. The architecture of static hypertexts. Proceedings of Hypertext '87, Chapel Hill, NC; November 1987.
- Raymond, D. R.; Tompa, F. W. Hypertext and the New Oxford English Dictionary. Communications of the ACM, 31(7): 871 - 879; 1988.
- Remde, J. R., Gomez, L. M., & Landauer, T. K. SuperBook: An automatic tool for information exploration - Hypertext? Proceedings of Hypertext '87, Chapel Hill, NC; 1987 November.

- Smith, J. B., Weiss, S. F., & Ferguson, G. J. A hypertext writing environment and its cognitive basis. Proceedings of Hypertext '87, Chapel Hill, NC: 1987 November.
- Smolensky, P., Bell, B., Fox, B., King, R., & Lewis, C. Constraint-based hypertext for argumentation. Proceedings of Hypertext '87, Chapel Hill, NC: November 1987.
- Walker, J. Supporting document development with Concordia. *Computer*, 20(1): 48-59; 1988.
- Weyer, S. A. The design of a dynamic book for information search. *International Journal of Man-Machine Studies*, 17: 87-107; 1982.
- Yankelovich, N., Meyrowitz, N., & van Dam, A. Reading and writing the electronic book. *Computer*, 17(10): 15-30; 1985.



