Patterns of Reading and Organizing Information in Document Triage [1]

Soonil Bae

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112

Catherine C. Marshall

2Microsoft Corporation, One Microsoft Way, Redmond, WA 98052

Konstantinos Meintanis

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112

Anna Zacchi

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112

Haowei Hsieh

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112

J. Michael Moore

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112

Frank M. Shipman (corresponding author)

Center for the Study of Digital Libraries, Texas A&M University, College Station, TX 77843-3112 shipman@cs.tamu.edu

People engaged in knowledge work must often rapidly identify valuable material from within large sets of potentially relevant documents. Document triage is a type of sensemaking task that involves skimming documents to get a sense of their content, evaluating documents to assess their worth in the context of the

current activity, and organizing documents to prepare for their subsequent use and more in-depth reading. We have performed a study of document triage by collecting multiple forms of qualitative and quantitative data to characterize how 24 subjects read about a new topic and assessed and organized a set of 40 relevant Web documents. Our results indicate that there are multiple strategies for document triage, each involving different styles of reading, interacting, and organizing. Common strategies include: 1) focused reading early in the task, relegating the organizing until later in the process; 2) skimming performed in tandem with organizing, which relies on gaining an incremental understanding of the topic; and 3) metadata-based organizing, a strategy that stresses working with document surrogates to minimize the time spent reading. The findings suggest ways applications may better support the intertwined nature of the browsing, reading, and organizing activities in document triage.

Introduction

The growth of the Internet and digital libraries as information resources, combined with advances in global indexing techniques, provides people with a vast amount of information to sift through as they engage in research-intensive activities. Many searches return thousands, if not millions, of matching documents or Web pages. Winnowing down the relevant documents to a tractable collection of useful material and interpreting this collection in a task context is often onerous and time-consuming.

Document triage is the portion of the research process that involves collecting material, reading or skimming it to get the gist and evaluate its worth, and organizing the culled documents into a personal resource. This type of research has traditionally been the province of analysts, professional researchers, and reference librarians; but more recently, people in many walks of life find themselves performing document triage. In the process, people may re-read the material they have collected, progressively refine its organization, and share the results of this work with others. If the topic is unfamiliar, document triage also includes a period of incrementally forming a better understanding of the topic and the material that has been gathered.

Prior analysis of document triage (Bae et al. 2005) demonstrates that users frequently switch among searching, reading and organizing documents. People use search engine metadata such as page title, URL, or a document snippet to predict the relevance of given documents. This prediction influences subsequent reading and organizing. Most dramatic is the case in which a person organizes documents solely based on the metadata without consulting the actual content. However, more frequently, people create and refine categories incrementally as they read documents and decide what to do with them. Thus

the categories that have already been defined influence subsequent searching and reading just as reading affects searching and organizing. In this way, searching, reading and organizing are closely intertwined in document triage, as anticipated by Bates's Berrypicking model of search (1989). The apparent interaction among these three subtasks suggests that there is potential value to examining the relationships among them more closely and to characterizing their interdependences. For example, does the availability of good metadata cause people to read the actual documents less often and instead rely on the surrogate?

The study presented in this paper explores the patterns of different types of reading and organizing during document triage as well identifying triage strategies. We have used both quantitative and qualitative data to more fully understand the activity. This understanding has the potential to improve the design of systems that support document triage and inform related work in the areas of digital libraries and Web tools.

The next section describes prior and related work on document triage, sensemaking, and knowledge work more generally. Following this is a description of the document triage task and the study design. The study results are presented in two sections: one primarily concerned with reading and one describing organizing. The paper closes with a discussion of these results and their implications.

Related Work

We look to two streams of related work in studies and systems development to guide our study design and to explore the potential implications of our findings. Studies have examined various aspects of the sensemaking process, as well as the activity in its entirety. Systems have also been developed to support all or part of the activity. Document triage is necessarily a kind of sensemaking, but it hinges on partial understandings and preliminary category creation (sorting things out), rather than being aimed toward a finished product like a written document that synthesizes other material.

Prior Studies

Previous studies examined the impact of system design on document triage. A 1997 paper by Marshall and Shipman (1997) compared the within-collection search and organization practices that occur in paper-based triage with triage performed using two versions of the spatial hypertext system VIKI, one with a means of constructing a hierarchy of workspaces and the other providing a single flat space for organizing documents. This study found that participants opportunistically used visual features, spatial layout, hierarchic groupings, and annotations to organize the materials and to communicate the results of the triage to others. The study also showed that participants used visual objects

as surrogates for the documents themselves; hence people did not like to use multiple references to the same underlying document.

Shipman et al. (2004) looked more specifically at document triage with Web materials. This study compared document triage with two different software tools, a Web browser and either the Visual Knowledge Builder spatial hypertext system or a text editor of the participant's choosing. The results indicated that spatial hypertext users felt better able to express themselves and more confident that their expressions would be understood by others. It also showed that a system-applied visualization reduced the participants' willingness to use visual attributes (e.g. color, border width) to express their own interpretations, since they worried about interfering with or losing the system-generated visualization.

Slaney and Russell (2005) asked subjects to access, explore and organize a collection of documents using three presentation tools, a bound paper printout of the documents, a temporal display, and a semantically-clustered overview. The study showed that understanding increases throughout the triage process and that while the paper-based tool was more effective in the knowledge building process that occurs during the first five minutes of the task, after fifteen minutes subjects performed similarly using all three tools.

In a study conducted by Uren et al. (2006), students used three different approaches to perform a sensemaking task: a form-filling interface for building claim networks; a model-building application with support for search, information visualization and discovery; and a novice interface to the same search and information exploration services. Results showed that the claim network approach promoted comprehension by helping people form their own opinions about the topic. However, the study also showed that people can handle information in written format faster than they could when using the claim network.

To identify distinct activities that need to be supported in a digital library environment, Paepcke (1996) interviewed customer service center engineers and a variety of workers at a computer printer company. The interviews revealed five general and strongly related activities: discovery, retrieval, interpretation, management, and sharing. By analyzing each activity, Paepcke identified the need to support post-retrieval interpretation by providing tools for summarization, clustering, statistical analysis, ranking, and alternative visualizations. The results also emphasized the necessity of allowing users to construct flexible structures that can be easily changed to meet different contexts or user interests.

Related Prototypes

A number of research prototypes have been developed to support large-scale sensemaking tasks; these systems are based on a broad understanding of the activities that constitute sensemaking such as gathering related information, forming categories, and collection management, activities that are central to triage.

Based on the results of studies of analysts, Jonker et al. (2005) designed TRIST, a retrieval and triage component for the analytical environment nSpace. TRIST supports information retrieval with a visual representation of search results, document comparison through difference visualization, vector-based clustering, and a workspace for organizing documents.

Qu (2003) describes an integrated workspace for information gathering and sensemaking. His framework provides tools for advanced search and a topical tree-structure representation of the gathered information. It also employs machine learning techniques that support automatic information seeking, classification, and clustering, as well as methods for presentation-structure generation and manipulation.

Schwarzkopf (2004) presents an infrastructure for collecting, organizing, sharing, and sensemaking. The key element in this client-server architecture is the "document manager", a client-side application that consists of a zoomable workspace where information items take the form of index cards. Users can associate items by grouping and defining relations between them, forming concepts in the system's ontology. Concepts are used then by the document manager for the addition of relevant information items as well as the discovery of new relations among the items of the group.

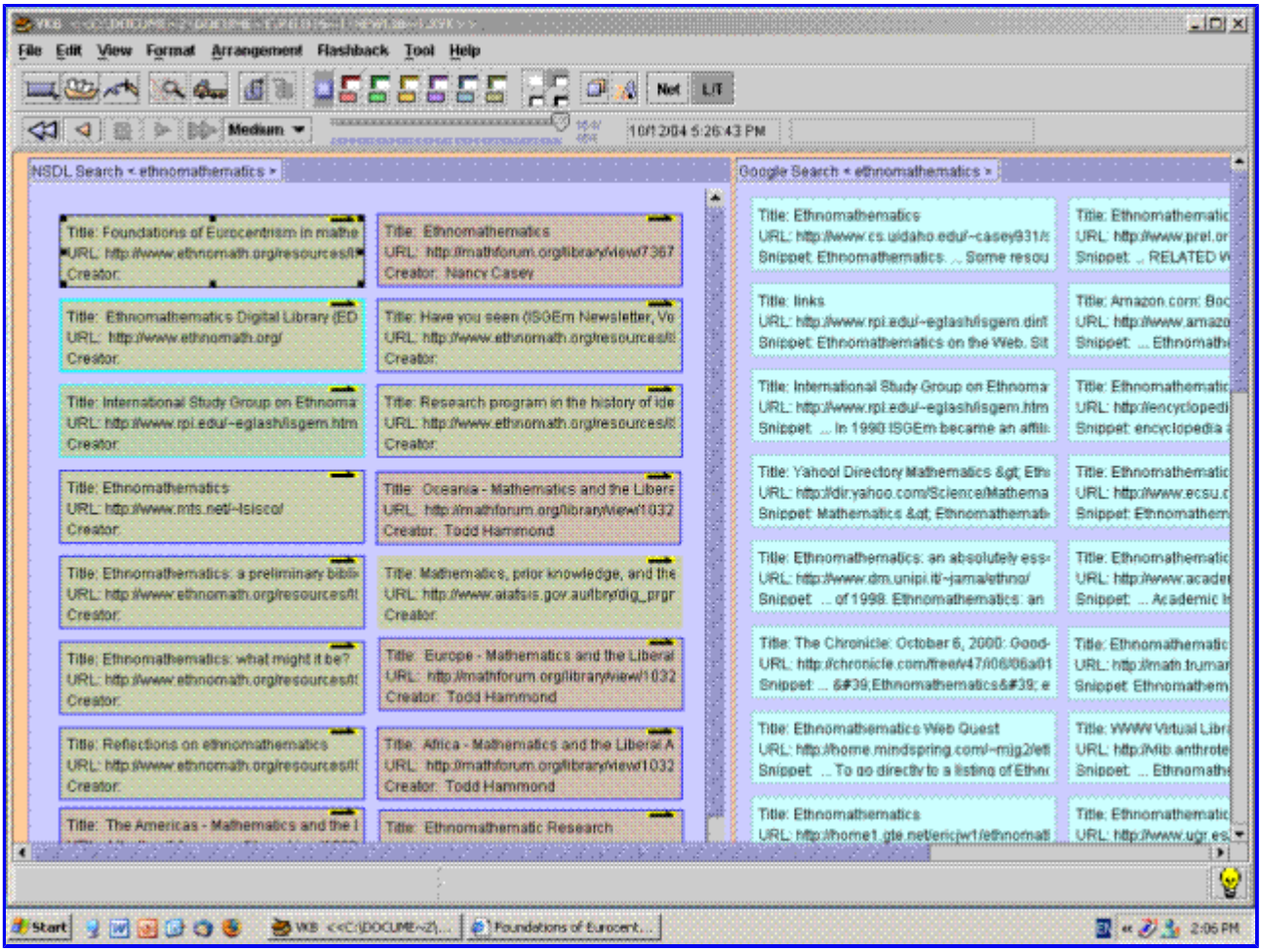
Unlike the systems we have described thusfar that emphasize retrieval and organization, Bier et al. (2004) have created a family of applications to address in-depth long-term reading of many documents. These applications support gathering, managing, and associating information and include a tool for retrieving related documents given the references from a seed document and a tool for collection exploration. In the latter, users zoom in or out to get more detailed or more comprehensive views of the collection.

Much of the related work we describe focuses on a portion of document triage rather than the activity as a whole. For example, there have been a number of studies of organizing (e.g. Bowker & Star (1999) and Uren et al. (2006)), but fewer detailed studies of how people read during document triage. On the other hand, most reading studies are lab studies focused entirely on reading (e.g. Dillon et al. (2004) & Tyrrell & Leibowitz (1990)), not on the interaction between reading and organizing or on reading that is more fragmented and less comprehension-oriented (e.g. skimming). Other studies that examine the interpretive aspect of reading generally assume a single document-centered view (e.g.

Fu et al. (2005) and Quyyum & Bilykh (2005)). Similarly, system design generally assumes retrieving and organizing are relatively independent from reading. Thus we have elected to explore the interaction between reading and organizing that is central to triage.

Study Design

The document triage study was conducted at a large university. 24 subjects (19 males) were recruited within the Computer Science department for the study via flyers and mass e-mail. Ages ranged from 18 to 40. 23 out of 24 of the subjects had used computers for more than 5 years. While 80% of subjects answered that they access informational web pages frequently (i.e. short newspaper articles, reviews, magazines etc.), only 38% answered that they read long documents (e.g. a 20 page paper) on the screen.



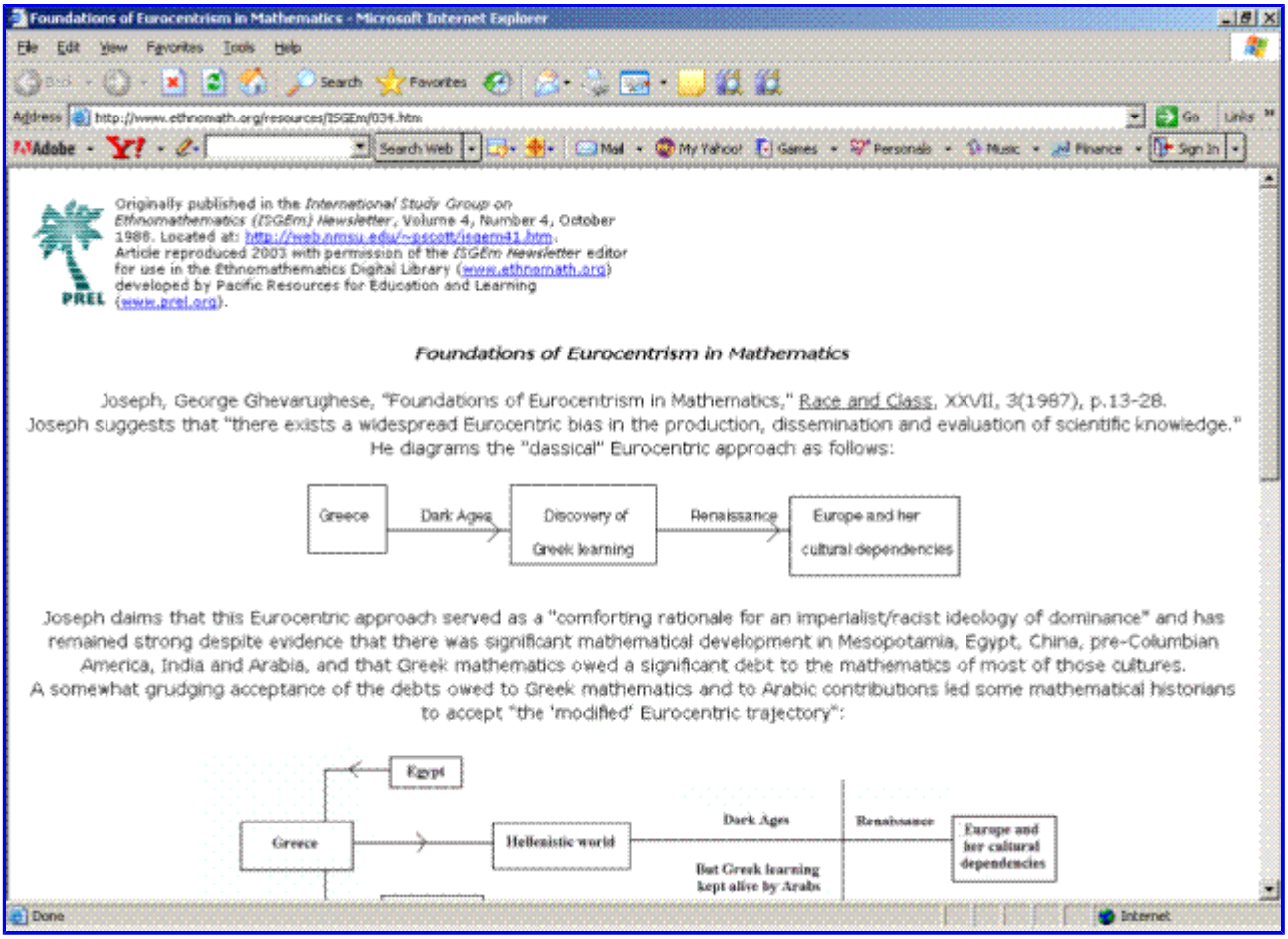


Figure 1. Subjects used a VKB space for searching & organizing (left) and Internet Explorer for reading (right).

This study duplicated the setting and task described by Shipman et al. (2004). Subjects were asked to act in the role of a reference librarian and select and organize material in for a high school teacher preparing a class on ethnomathematics. This relatively unknown and difficult topic ensures that the participants would know relatively little about the domain at the outset of the study. Subjects started with 40 documents relevant to the topic of ethnomathematics, 20 documents returned by the National Science Digital Library (NSDL) search facility and 20 documents returned by Google. Subjects worked from an overview of the document set in the Visual Knowledge Builder (VKB), (see http://www.csdl.tamu.edu/VKB). Each document in VKB was represented in the overview by a surrogate or document object that linked to the full text; the surrogate displayed the document's title, URL, and creator for the NSDL results, and the title, URL, and a document snippet for the Google results. Double-clicking on the *document object* opened the document in a Web browser (Internet Explorer - IE). Figure 1 shows the initial overview of the documents in VKB on the left and one document opened for reading in IE on the right.

While our emphasis was not on evaluating the VKB software, it provided a representative

environment for performing the triage task. VKB is a spatial hypertext tool that offers an integrated environment for searching and organizing documents from the Internet and other sources. VKB enables users to create a hierarchy of two-dimensional workspaces called collections that can contain document objects or other collections. Users organize and interpret documents by placing document objects in collections and by changing the document object's visual attributes, such as color or border width. Subjects learned how to use VKB during an unrelated 15 minute training exercise prior to the actual document triage task.

The study task asked subjects to organize the documents in a VKB workspace that would be given to the teacher. Subjects used two displays for the task (see Figure 2), with the thought that this configuration would give them ample room to see both VKB overview and individual documents while they were performing the triage.

Subjects were videotaped from an over-the-shoulder angle to conceal their identities and to record the focus of their attention (see Figure 2); at the same time, screen capture software produced a movie of what was happening on both displays. After completion of the triage task, subjects were asked to fill in a questionnaire about aspects of the task and to choose the five most and least useful documents. We also conducted short semi-structured interviews to elicit additional information about the subject's approach to the activity, to clarify any confusing answers on the questionnaire, and to elaborate how they had assessed document utility. In addition, system logs were kept to track events of interest related to reading and organizing (for example, scrolling events and link selection in IE and interactions with VKB objects and collections).

Studies of reading and organizing, even when they are apart, involve balancing concerns of rigor and realism. The known variability of individual reading and organizing practices suggests that we approach this balance creatively, giving participants a uniform task, corpus, and technology, yet allowing them to go about the task flexibly. The resulting individual variability made the data unsuitable for full-on quantitative analysis. Instead, we took a qualitative approach to identifying patterns among subsets of our participants, triangulating among our quantitative and qualitative data sources; we found this triangulation very helpful in resolving interpretive ambiguities. The over-the-shoulder video served as a bridge between quantitative data (e.g. activity logs) and qualitative data (e.g. interviews); rough focus of attention could be determined and interpretations of the other sources verified.

The next section describes the findings related to subjects' reading and skimming. It is followed by a complementary discussion of how subjects went about organizing the workspace.

Patterns of Reading-Related Activity

At the outset, we were interested in how a multi-document task of this sort affected reading, since subjects had neither time nor motivation to read everything carefully. Would they skim selectively? Would they read any of the documents? How much would they trust metadata to guide them in their assessment of a document's ultimate value or content?

To investigate these questions, we compared user activity at different points during the triage task. Since the subjects were given no firm time constraint, the total time they spent on the task varied (although on average they spent a little over an hour reading and organizing the documents). For our analysis, we normalized the total experiment time and partitioned the entire time period into fifths. We then examined user activity during the different segments. We considered reading to be focused attention on the content displayed in the IE window, since subjects also read metadata in the overview; we found that observing and recording their attention in this way struck a balance between our data collection needs and interfering with the subjects' performance of the task.

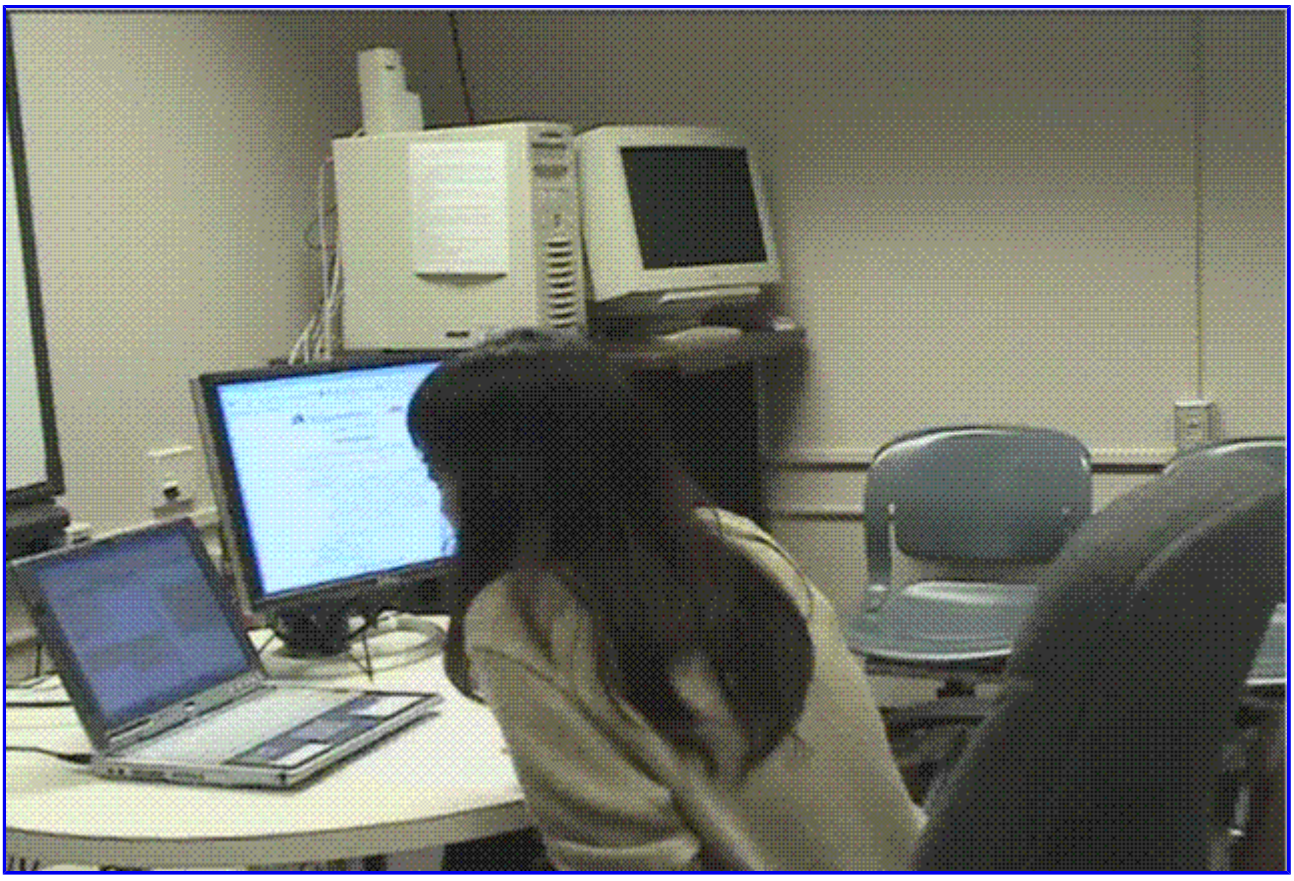


Figure 2. A subject performing document triage in the study

Differences in Reading Styles

The videotapes revealed differences among individual reading styles. User events collected in the system logs confirmed that there were differences in when and how much

different subjects read during the course of the task. Instead of finding one or two canonical patterns of reading activity, we found variations on four common themes, with a few outlying individuals exhibiting singular activity patterns. Figure 3 shows four representative patterns selected from individual subjects; the dotted line indicates the average time the individual subject spent reading throughout the task. While these four patterns do not fully represent all 24 subjects, they do characterize the majority.

Pattern 1 shows a subject who spends less reading time than his or her average during the initial time period (the first 20% of the total task period), more reading time during the middle time periods (about 21~80% of the total task period) and substantially less reading time during the final period (the final 20%). This is the most common reading pattern among the subjects; seven followed this pattern of activity.

Pattern 2 is somewhat similar to Pattern 1, except the subject has spent more reading time during the initial period, exceeding his or her average until the task approaches its final period, when reading drops off precipitously. We can think of the general case as having a roughly flat distribution of reading time during the first 80% of the task and a sharp tail-off during the last 20%. Five subjects follow this pattern.

Pattern 3 shows a higher than average time reading during the time period 1 (0-20% of the total task period), a reduced reading time during the second period, a greater than average reading time during period 4 (61-80% of the total task period) and a reduced time spent in the final reading period (81-100%). Four subjects follow this cyclic pattern.

Pattern 4 starts with a higher reading time during periods 1 and 2 (0-40%) than shown by the individual average, but declines thereafter. Three subjects follow this pattern.

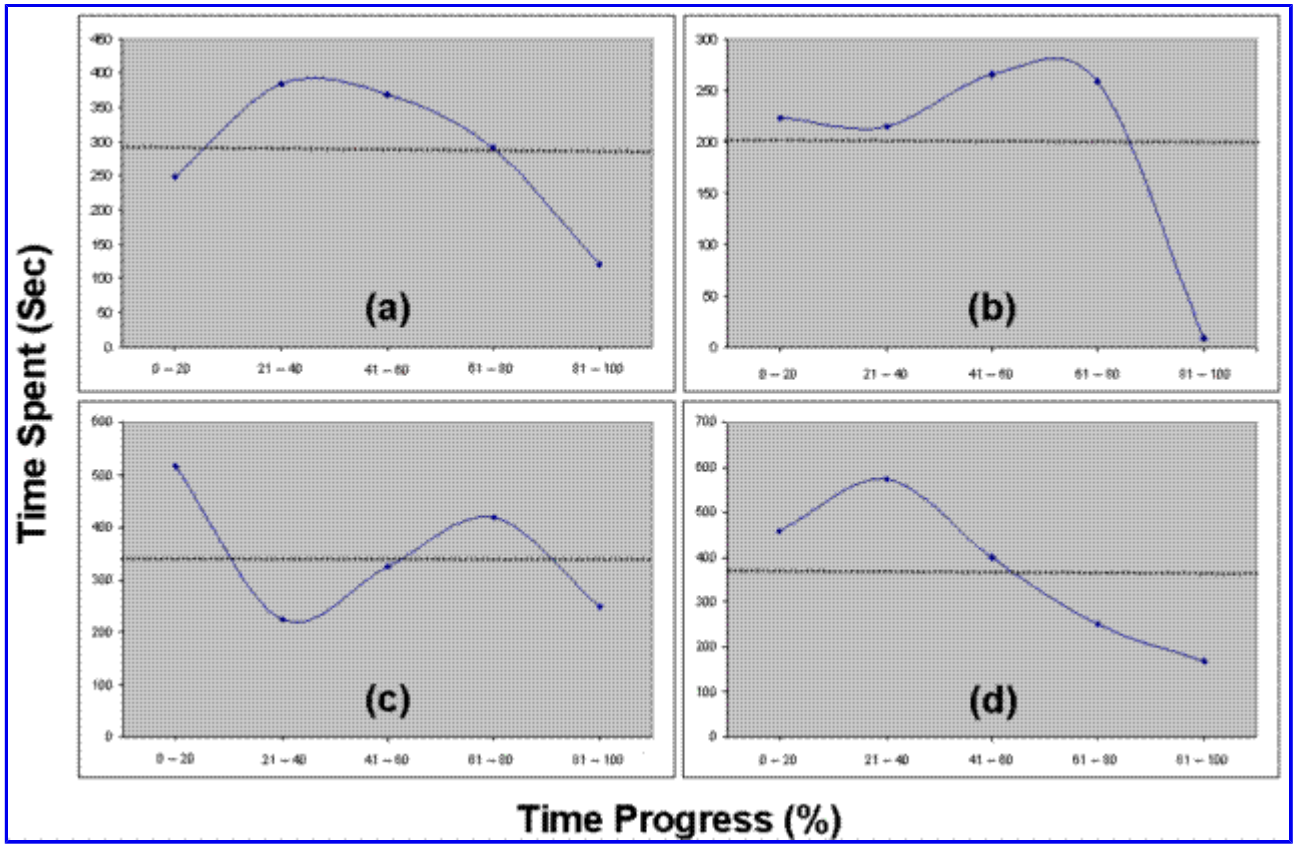


Figure 3. Four patterns of the time spent in reading: (a) Pattern 1 (subject 21); (b) Pattern 2 (subject 11); (c) Pattern 3 (subject 19); (d) Pattern 4 (subject 14)

One characteristic that is common to all four patterns is that reading time is not constant during the triage period: some subjects spent more time reading at the start of the activity while others spent more time reading in the middle. However, there is almost invariably a downward trend at the end: 20 out of 24 subjects spent less than half their time reading during the final 20% of the total task period. It is most likely that subjects were more focused on organizing their results for the hypothetical teacher during that period.

A question that remains is how deeply the subjects read the material and when they were most likely to be looking in detail at the material. Observation revealed that some subjects would alternate between focused attention and quick scrolling, as reported by subject 3:

> "For the most part, I didn't want to get too bogged down looking at specific stuff, because I was trying to get an idea of the subject rather than in depth. So I was trying to look for more like key, interesting things that might be mentioned. Like, several of them mentioned different societies, and how this applied to them. So you can look for a subsection - scrolling through, I'd see the subsection being mentioned, rather than having to look paragraph by paragraph to get to that point."

To characterize the variability of reading with the quantitative data, we can examine how

many documents the subjects visit (how many they open in the reading interface) and when they visit them. Figure 4a shows how many documents subjects visit during the triage period. Subjects tended to visit fewer documents in the first and last time periods (0-20% and 81-100%), while they tended to visit more documents in the middle time periods (21~80%). Individual patterns reveal that 17 out of 24 subjects visited fewer documents in the first time period (0~20%), while 15 out of 24 subjects visited fewer documents in the last time period (81-100%).

If we combine this data with the data shown in Figure 3, which shows that subjects spent less time reading during the last time period, we can infer that subjects read fewer documents, but read more intensively during the first time period, and simply read less during the final period. The data confirm this trend (see Figure 4b).

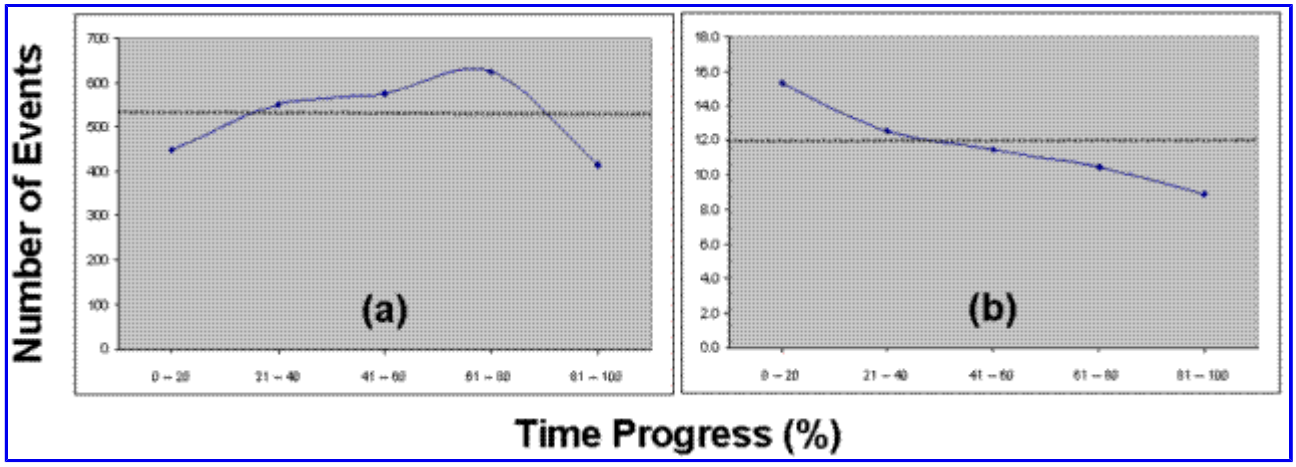


Figure 4. (a) Average number of visits to documents over the task duration; (b) Average reading time per visit

Figure 4b shows the average time spent per visit, which decreases almost linearly from the first time period to the last one. This graph confirms that subjects tended to switch back and forth less frequently when they read documents during the initial stage of the task, but shift their attention among applications more rapidly as the triage activity progresses.

These signs of more focused attention at the task's outset may be partially due to the subjects' lack of familiarity with ethnomathematics; thus they may be reading to understand the topic. For example, in describing his strategy, subject 5 told us, "Firstly, I didn't know the topic." However, the interviews and videos reveal other reading goals as well. During the initial phase, subjects may also be reading to ascertain the scope and content of the collection they were given to work with; as subject 4 explained, "What I set out to do was I looked and saw, okay, what do all these things have in common?" During the later stages of triage, reading often means re-reading: subjects quickly re-read the

documents they read earlier to remind themselves of the content while they refine its organization. Were the triage task on a more familiar topic, the profile of reading behavior might be somewhat different, but subjects would probably still need to discover the collection's scope during the early phases and still need to re-read during the later phases.

Effect of Document Assessment on Reading Time

Subjects were asked to identify the five most useful and five least useful documents in the collection. Based on these responses, we can examine differences in the characteristics of subjects' interactions with their most and least useful documents (see Figure 5). The most striking aspect of this data is how important indicators - reading time and number of visits - greatly differ during the initial phases of the triage (the first 40% of the task period) and tend to converge in the latter phases (the final 60%). As we would expect, at the outset subjects spend more time with the documents they judged to be useful and they scroll them more often too.

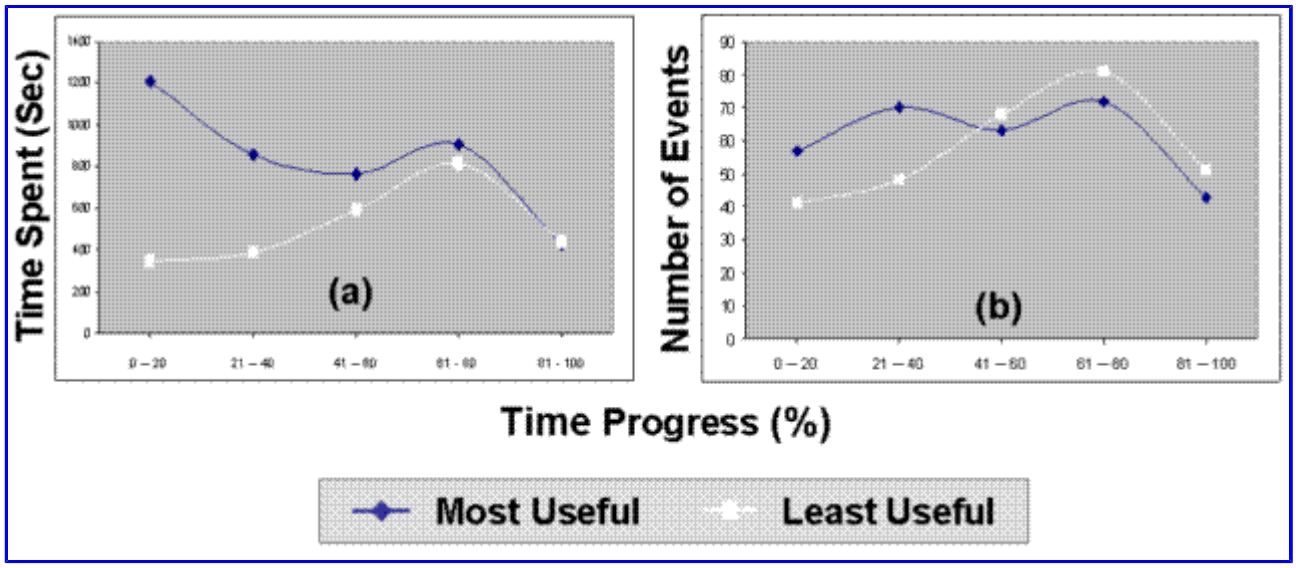


Figure 5. (a) Reading time for documents judged useful/not useful; (b) Number of scrolls for useful/not useful documents

When we look at the documents assessed as most and least useful and couple this judgment with the interview data, we find that the preferred documents tend to appear authoritative and dense with information (or were a point of easy access to authoritative documents), without obvious advertising (including self-interested references) or distracting visual elements. Subject 4, for example, told us, "The ones I chose [as most useful] had a lot of text. It was just text. I could read through it and find the information that I want." Subject 3 elaborated:

> "A couple of them were papers by the original person himself. So I thought those were

> pretty interesting, because he seems to have a lot of enlightenment in the subject itself. And I think one or two others may have been the libraries where you can search for documents themselves. And I thought those were very useful."

There is a difference, however, between portals to authoritative collections, and a page that is simply a naïve set of uncurated links. Subject 6 explained, "There was one page where there is only a few links. And I think that was useless… I was looking for content, not just links." Subject 15 referred to one of the documents judged less useful as "cheesy" and preferred to see who the author was; another document was noted as using an unfamiliar language.

This initial divergence and later convergence corroborates our earlier explanation that subjects tended to perform relatively deep reading in the early stage of triage and ranked the documents they used to understand the domain highly. Then, once the subjects reached the middle and later stages of document triage, refining the organization becomes the focal task and the desire to read individual documents declines. In addition to examining reading time and number of visits, we have explored other user events such as clicks, scrolls, text selection and reading time per visit and found that there are very different patterns between how the documents judged as most and least useful are used in triage. It is clear that the subjects' judgment of and interest in individual documents affects how they interact with them during triage.

Activity Patterns during Document Triage

The videos and user event logs show that subjects went through different phases of reading-related activity as they performed the triage task. These phases varied in ways that reflected individual differences among subjects such as a subject's level of knowledge of related topics, a subject's interest in the topic, his or her engagement with the task and mastery of the VKB software, and simple differences in styles or preferences.

However, in spite of these differences, we identified some rough patterns of user activity. The following is a summary of these patterns according to our normalized breakdown of the document triage period.

Time periods 1 and 2 (0-40%): During this initial 40% of the triage period, subjects tended to perform any relatively deep reading, to spend a greater proportion of their time reading (especially the most useful documents), and to visit a greater number of the most useful documents. The subjects' judgment of document utility influenced the time they spent reading them, the number of visits, and the reading time per visit. Subjects may have the most interest in understanding the topic or the collection's scope during these early phases.

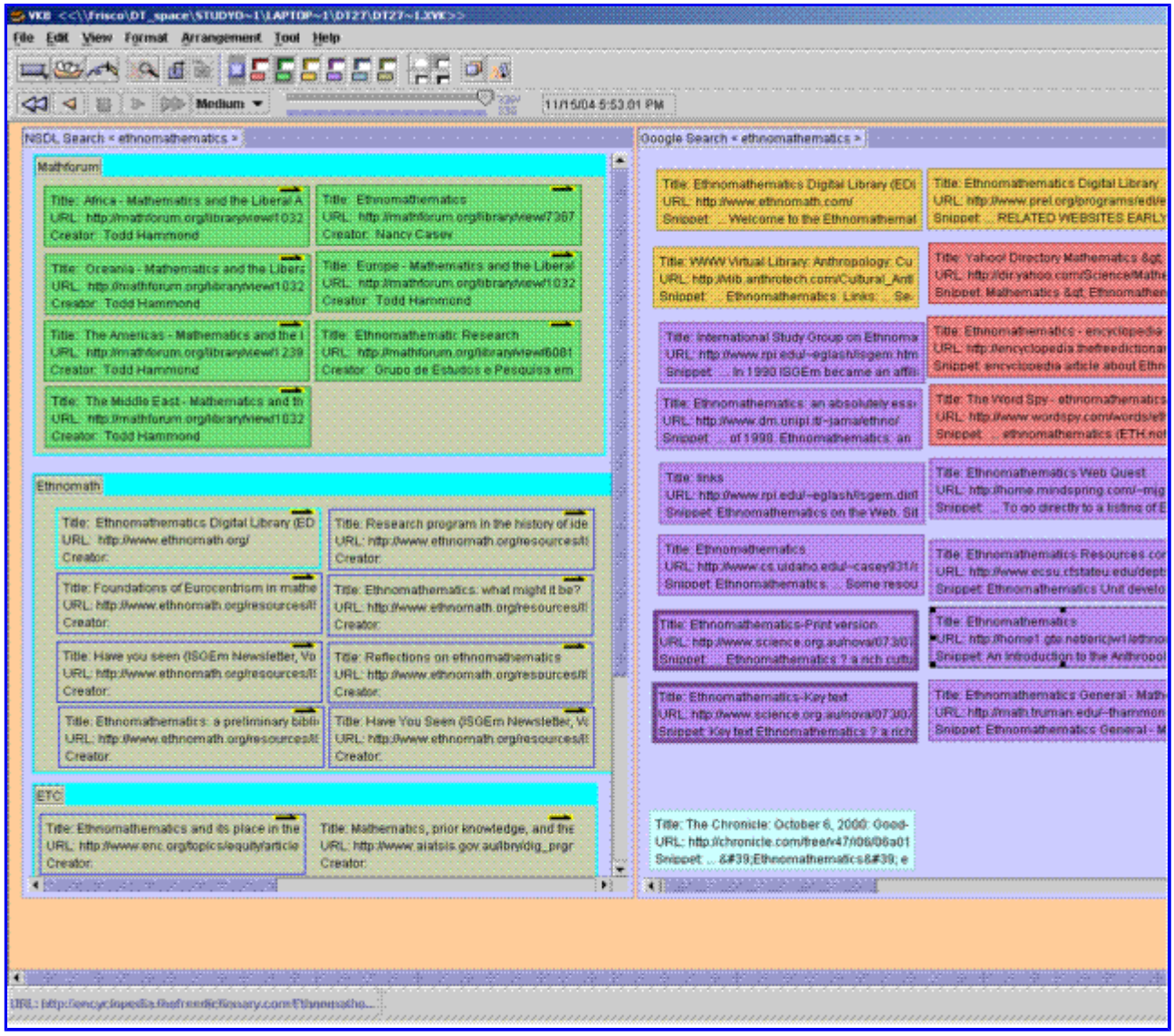
Time period 3 (41-60%): Although there is no drastic change in the time subjects spend reading and the number of times they visit documents during the third time period, subjects are spending less time reading per visit. This trend toward briefer time with individual documents indicates a change in how subjects are reading. Thus, the middle time period might be characterized as a shift from the early "reading-in" style to the scanning and reminding types of reading that are associated with the later interpretive or organizational periods.

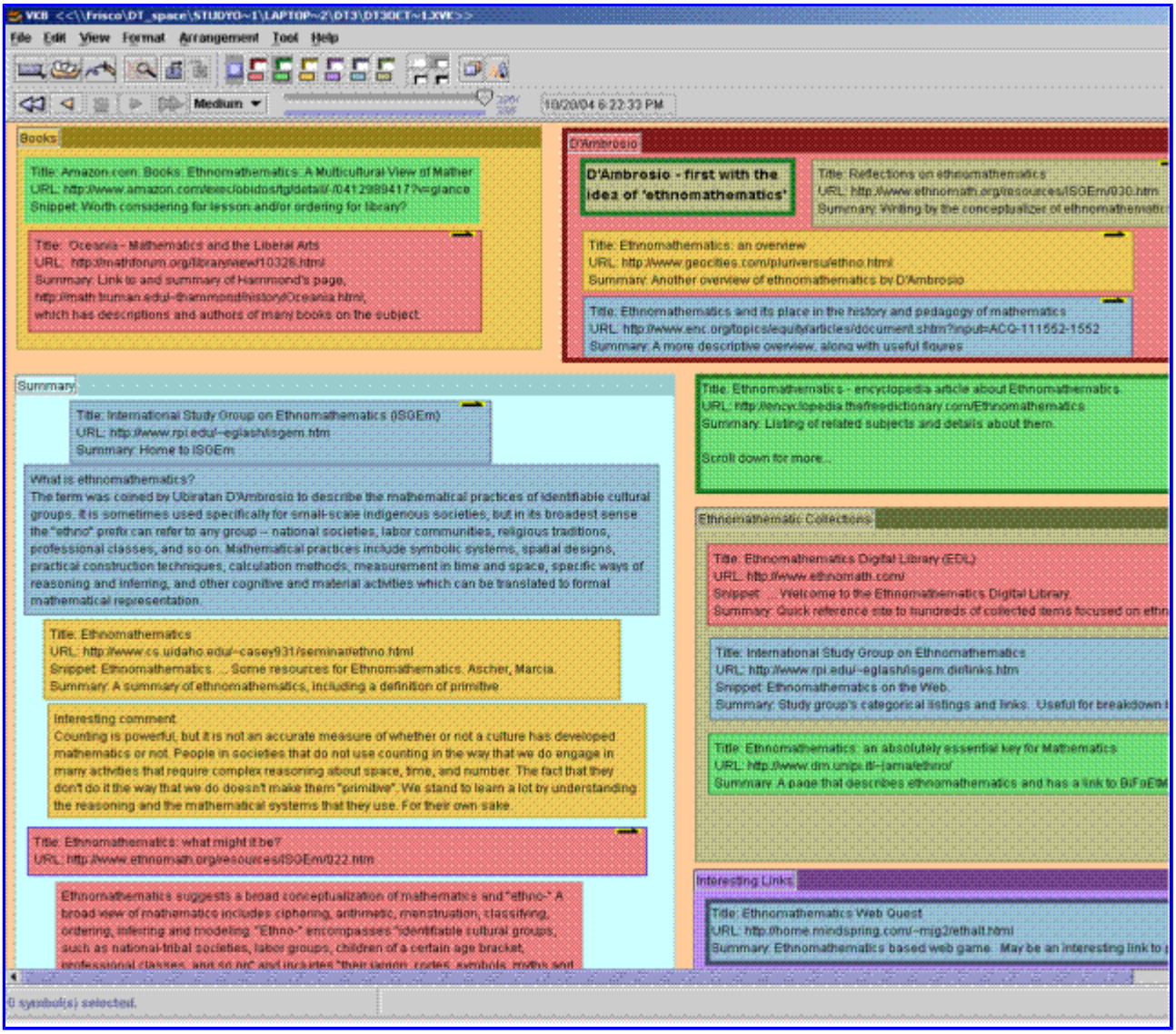
Time period 4 (61-80%): Although reading time remained stable through this period, the number of visits and the number of scrolls increased. This may imply that the form of reading has changed to some extent. Reading is probably no longer directed toward seeking new information or developing a categorization strategy, but rather toward revisiting and re-reading to confirm and refine the resulting structure or to populate portions of the structure with other documents.

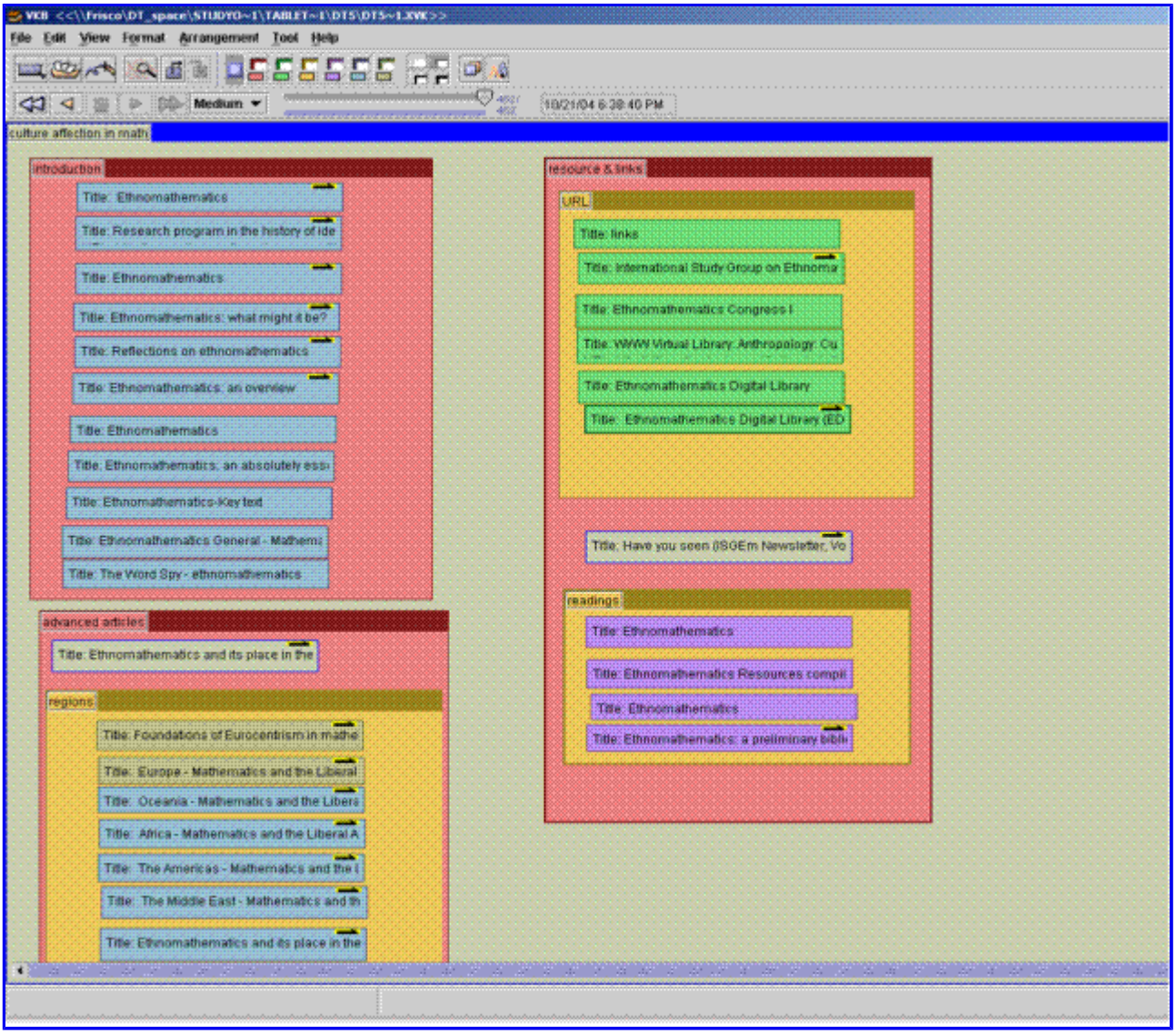
Time period 5 (81-100%): Evidence from the event logs suggests that there was much less reading activity during the final period of the triage task: less time spent reading, fewer document visits, and fewer user events in general. Subjects focused on organizing during this period. Any reading manifested itself in the form of short visits to documents (re-reading or scanning) to complete the task. Videos reveal that some subjects quickly checked documents to make sure they were categorized correctly or to confirm their ideas.

Patterns of Organizing-Related Activity

As subjects read materials in IE, they used VKB to organize the documents. They created and labeled VKB collections to categorize and subcategorize the documents; they arranged the documents in these spaces to indicate further categorization and document properties; and they expressed document properties using visual attributes such as color. The use of different data sources - VKB logs, continuous screen capture, over-the-shoulder video, questionnaires, and interviews - helped us establish the subjects' patterns of organizing-related activity.


File Edit View Format Arrangement Tool Help
Medium
11/15/04 5:53:01 PM
NSDL Search « ethnomathematics »
Mathforum
Title: Africa - Mathematics and the Liberal A
URL: http://mathforum.org/library/view/1032
Creator: Todd Hammond
Title: Ethnomathematics
URL: http://mathforum.org/library/view/7367
Creator: Nancy Casey
Title: Oceania - Mathematics and the Liberz
URL: http://mathforum.org/library/view/1032
Creator: Todd Hammond
Title: Europe - Mathematics and the Liberal
URL: http://mathforum.org/library/view/1032
Creator: Todd Hammond
Title: The Americas - Mathematics and the L
URL: http://mathforum.org/library/view/1239
Creator: Todd Hammond
Title: Ethnomathematic Research
URL: http://mathforum.org/library/view/6081
Creator: Grupo de Estudos e Pesquisa em
Title: The Middle East - Mathematics and th
URL: http://mathforum.org/library/view/1032
Creator: Todd Hammond
Ethnomath
Title: Ethnomathematics Digital Library (ED
URL: http://www.ethnomath.org/
Creator:
Title: Research program in the history of ide
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Foundations of Eurocentrism in mathe
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Ethnomathematics: what might it be?
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Have you seen (ISGEm Newsletter, Vo
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Reflections on ethnomathematics
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Ethnomathematics: a preliminary bibli
URL: http://www.ethnomath.org/resources/t
Creator:
Title: Have You Seen (ISGEm Newsletter, Vo
URL: http://www.ethnomath.org/resources/t
Creator:
ETC
Title: Ethnomathematics and its place in the
URL: http://www.enc.org/topics/equity/article
Creator:
Title: Mathematics, prior knowledge, and the
URL: http://www.aiatsis.gov.au/bryd/ig_prgr
Creator:
Google Search « ethnomathematics »
Title: Ethnomathematics Digital Library (EDI
URL: http://www.ethnomath.com/
Snippet: ...Welcome to the Ethnomathemat
Title: Ethnomathematics Digital Library
URL: http://www.prel.org/programs/edu
Snippet: ...RELATED WEBSITES EARLY
Title: WWW Virtual Library: Anthropology: Cu
URL: http://lib.anthrotech.com/Cultural_Anth
Snippet: ...Ethnomathematics Links: ...Se...
Title: Yahoo! Directory: Mathematics >
URL: http://dir.yahoo.com/Science/Mathe
Snippet: Mathematics > Ethnomathe
Title: International Study Group on Ethnoma
URL: http://www.rpi.edu/~eglash/isgem.htm
Snippet: ...In 1990 ISGEm became an affili
Title: Ethnomathematics - encyclopedia
URL: http://encyclopedia.thefreedictiona
Snippet: encyclopedia article about Ethn
Title: Ethnomathematics: an absolutely essi
URL: http://www.dm.unipi.it/~jamalethno/
Snippet: ...of 1990. Ethnomathematics: an
Title: The Word Spy - ethnomathematics
URL: http://www.wordspy.com/words/eth
Snippet: ...ethnomathematics (ETH.not
Title: links
URL: http://www.rpi.edu/~eglash/isgem.dir/
Snippet: Ethnomathematics on the Web. Sit
Title: Ethnomathematics Web Quest
URL: http://home.mindspring.com/~mig
Snippet: ...To go directly to a listing of E
Title: Ethnomathematics
URL: http://www.cs.uidaho.edu/~casey931/t
Snippet: Ethnomathematics ... Some resou
Title: Ethnomathematics Resources co
URL: http://www.ecsu.ctstateu.edu/dept
Snippet: Ethnomathematics Unit develo
Title: Ethnomathematics-Print version
URL: http://www.science.org.au/nova/073/0
Snippet: ...Ethnomathematics ? a rich cultu
Title: Ethnomathematics
URL: http://home1.gte.net/enc/w1/ethno
Snippet: An introduction to the Anthropol
Title: Ethnomathematics-Key text
URL: http://www.science.org.au/nova/073/0
Snippet: Key text Ethnomathematics ? a rich
Title: Ethnomathematics General - Math
URL: http://math.truman.edu/~thammon
Snippet: Ethnomathematics General - M
Title: The Chronicle: October 6, 2000: Good-
URL: http://chronicle.com/free/v47/i06/06a01
Snippet: ... 'Ethnomathematics' e


File Edit View Format Arrangement Tool Help
Medium
10/20/04 6:22:33 PM
Books
Title: Amazon.com: Books: Ethnomathematics: A Multicultural View of Mathem
URL: http://www.amazon.com/exec/obidos/tg/detail/-/0412989417?v=glance
Snippet: Worth considering for lesson and/or ordering for library?
Title: 'Oceania' - Mathematics and the Liberal Arts
URL: http://mathforum.org/library/view/10326.html
Summary: Link to and summary of Hammond's page,
http://math.truman.edu/~dhammond/history/Oceania.html,
which has descriptions and authors of many books on the subject.
D'Ambrosio
D'Ambrosio - first with the idea of 'ethnomathematics'
Title: Reflections on ethnomathematics
URL: http://www.ethnomath.org/resources/ISGEm/030.htm
Summary: Writing by the conceptualizer of ethnomathematic
Title: Ethnomathematics: an overview
URL: http://www.geocities.com/pluriversu/ethno.html
Summary: Another overview of ethnomathematics by D'Ambrosio
Title: Ethnomathematics and its place in the history and pedagogy of mathematics
URL: http://www.enc.org/topics/equity/articles/document.shtm?input=ACQ-111552-1552
Summary: A more descriptive overview, along with useful figures
Summary
Title: International Study Group on Ethnomathematics (ISGEm)
URL: http://www.rpi.edu/~eglash/isgem.htm
Summary: Home to ISGEm
What is ethnomathematics?
The term was coined by Ubiratan D'Ambrosio to describe the mathematical practices of identifiable cultural groups. It is sometimes used specifically for small-scale indigenous societies, but in its broadest sense the "ethno" prefix can refer to any group -- national societies, labor communities, religious traditions, professional classes, and so on. Mathematical practices include symbolic systems, spatial designs, practical construction techniques, calculation methods, measurement in time and space, specific ways of reasoning and inferring, and other cognitive and material activities which can be translated to formal mathematical representation.
Title: Ethnomathematics
URL: http://www.cs.uidaho.edu/~casey931/seminar/ethno.html
Snippet: Ethnomathematics: ... Some resources for Ethnomathematics: Ascher, Marcia.
Summary: A summary of ethnomathematics, including a definition of primitive
Interesting comment:
Counting is powerful, but it is not an accurate measure of whether or not a culture has developed mathematics or not. People in societies that do not use counting in the way that we do engage in many activities that require complex reasoning about space, time, and number. The fact that they don't do it the way that we do doesn't make them "primitive". We stand to learn a lot by understanding the reasoning and the mathematical systems that they use. For their own sake.
Title: Ethnomathematics: what might it be?
URL: http://www.ethnomath.org/resources/ISGEm/022.htm
Ethnomathematics suggests a broad conceptualization of mathematics and "ethno-" A broad view of mathematics includes ciphering, arithmetic, menstruation, classifying, ordering, inferring and modeling. "Ethno-" encompasses "identifiable cultural groups, such as national-tribal societies, labor groups, children of a certain age bracket, professional classes, and so on" and includes "their jargon, codes, symbols, myths and
Title: Ethnomathematics - encyclopedia article about Ethnomathematics
URL: http://encyclopedia.thefreedictionary.com/Ethnomathematics
Summary: Listing of related subjects and details about them.
Scroll down for more...
Ethnomathematic Collections
Title: Ethnomathematics Digital Library (EDL)
URL: http://www.ethnomath.com/
Snippet: ... Welcome to the Ethnomathematics Digital Library
Summary: Quick reference site to hundreds of collected items focused on ethn
Title: International Study Group on Ethnomathematics
URL: http://www.rpi.edu/~eglash/isgem.dir/links.htm
Snippet: Ethnomathematics on the Web.
Summary: Study group's categorical listings and links. Useful for breakdown
Title: Ethnomathematics: an absolutely essential key for Mathematics
URL: http://www.dm.unipi.it/~amalethno/
Summary: A page that describes ethnomathematics and has a link to BiFoEth
Interesting Links
Title: Ethnomathematics Web Quest
URL: http://home.mindspring.com/~mjg2/ethalt.html
Summary: Ethnomathematics based web game. May be an interesting link to
0 symbol(s) selected.


File Edit View Format Arrangement Tool Help
Medium
10/21/04 6:38:40 PM
culture affection in math
introduction
Title: Ethnomathematics
Title: Research program in the history of ide
Title: Ethnomathematics
Title: Ethnomathematics: what might it be?
Title: Reflections on ethnomathematics
Title: Ethnomathematics: an overview
Title: Ethnomathematics
Title: Ethnomathematics: an absolutely essi
Title: Ethnomathematics-Key text
Title: Ethnomathematics General - Mathemz
Title: The Word Spy - ethnomathematics
advanced articles
Title: Ethnomathematics and its place in the
regions
Title: Foundations of Eurocentrism in mathe
Title: Europe - Mathematics and the Liberal
Title: Oceania - Mathematics and the Liberz
Title: Africa - Mathematics and the Liberal A
Title: The Americas - Mathematics and the L
Title: The Middle East - Mathematics and th
Title: Ethnomathematics and its place in the
resource & links
URL
Title: links
Title: International Study Group on Ethnoma
Title: Ethnomathematics Congress I
Title: WWW Virtual Library: Anthropology: Cu
Title: Ethnomathematics Digital Library
Title: Ethnomathematics Digital Library (ED
Title: Have you seen ISGEm Newsletter, Vo
readings
Title: Ethnomathematics
Title: Ethnomathematics Resources compil
Title: Ethnomathematics
Title: Ethnomathematics: a preliminary bibli

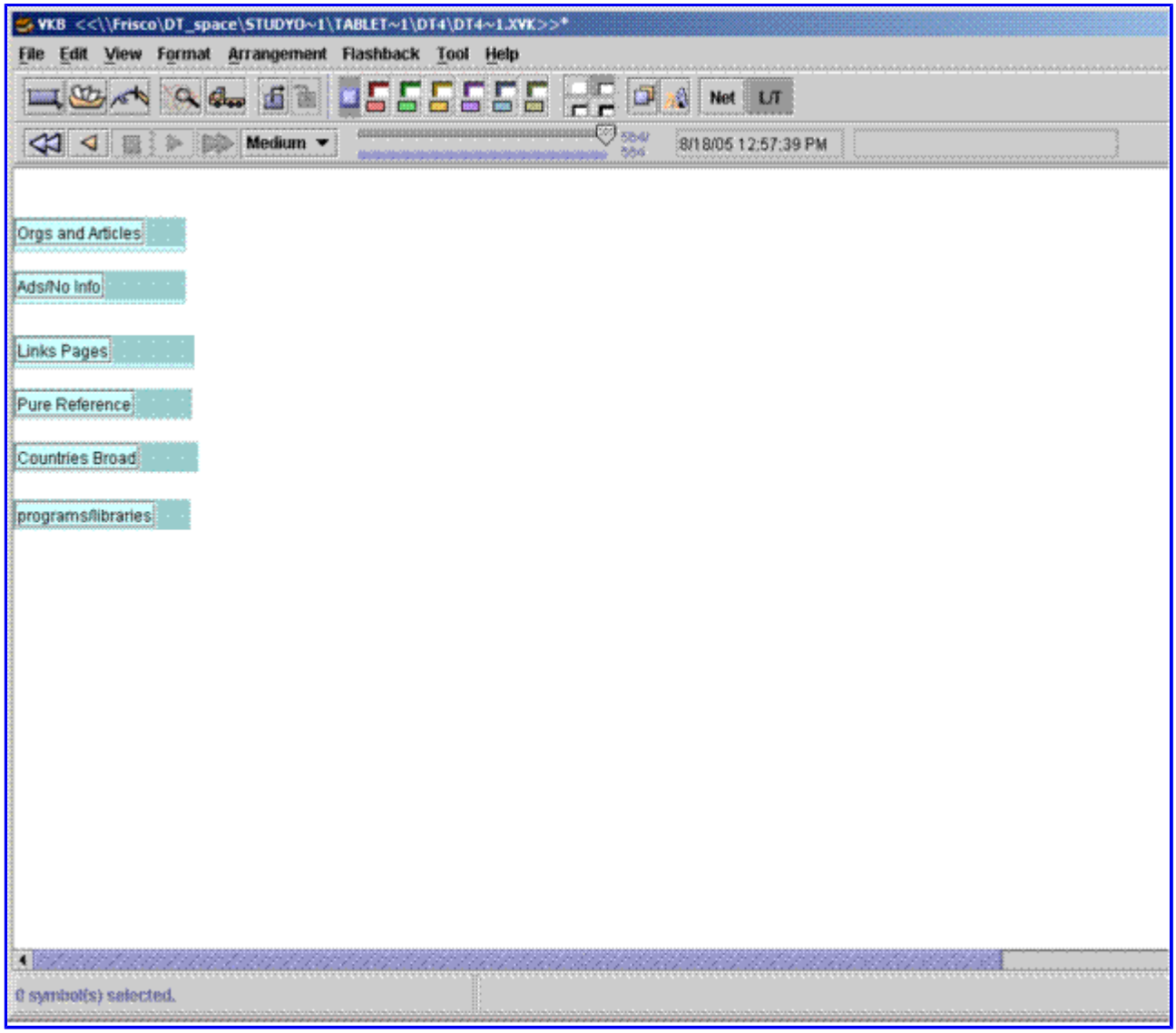


Figure 6. VKB workspaces collected from 4 different subjects

Figure 6 shows how four of the subjects organized the 40 ethnomathematics web pages they were given. These workspaces represent four very different approaches to expressing document categories and characteristics. As predicted by prior studies of document triage (Bae et al. 2005; Marshall & Shipman 1997; Shipman et al. 2004), subjects used a combination of visual attributes of document objects and placement of document objects in collections to express categories at various levels of abstraction.

User Events Associated with Organizing Activity

The subjects' VKB event logs recorded fourteen different types of user activity (i.e. atomic actions) during their interaction with the workspace. Seven of these event types were common to all 24 subjects: Add Object, Move Object, Resize Object, Delete Object, Change Content, Change Background Color and Change Border Width. To be consistent with our analysis of reading-related activity, we normalized the total experiment duration and

report activity within five uniform time periods.

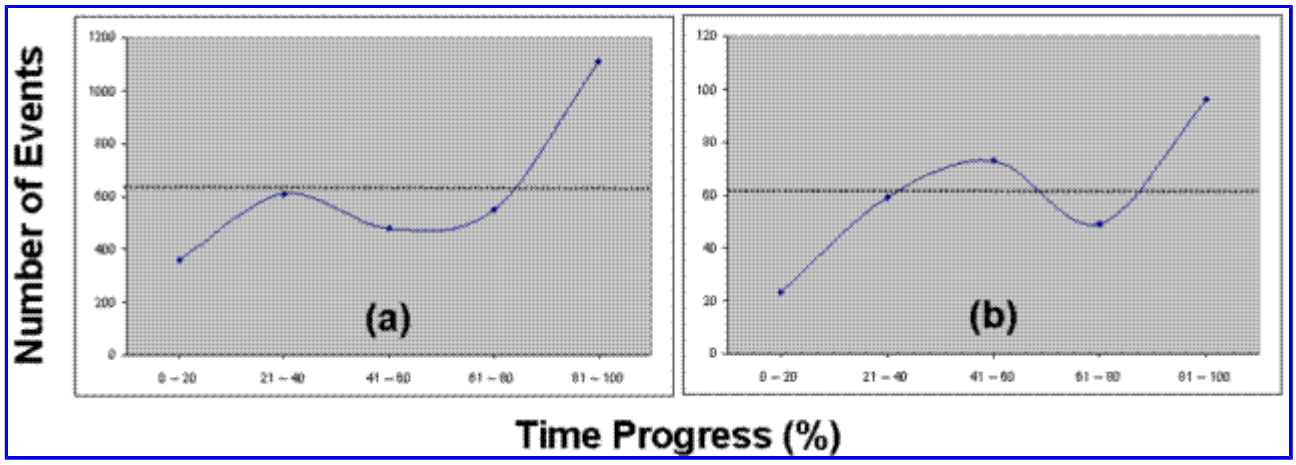


Figure 7. (a) Average number of organizing-related interactions over time; (b) Pattern for a representative subject

Figure 7a shows the average number of organizing-related user interactions (e.g. resizes, color changes, or moves) that were logged for all subjects during each of the five time periods. But is this composite representative of the individual subjects? We can readily see from Figure 6 that organizing strategies varied considerably among the study participants. Individual patterns reveal that only 12.5% of the subjects generated more user events than the average number of user events for each individual during the first time period, while 75.0% of the subjects generated more user events during the final time period. Figure 7b shows a representative pattern selected from among the individual subjects; 50% of the subjects show a similar pattern. This evidence supports the interpretation that subjects engaged in fewer organizing-related activities during the initial time period (the initial 20% of the triage period) and appreciably more during the final time period (the last 20%). This complements the patterns of reading-related activities we discussed in the previous section.

While the temporal pattern of organizing-related activities is similar when we aggregate the various types of events as we did in Figure 7, there are differences in the temporal patterns associated with different types of user events (Figure 8). So, for example, we can see that adding objects may be more characteristic of early-phase triage activities, while moving and resizing objects may be associated with later phases. These differences may imply that there are some distinguishable phases to organizing.

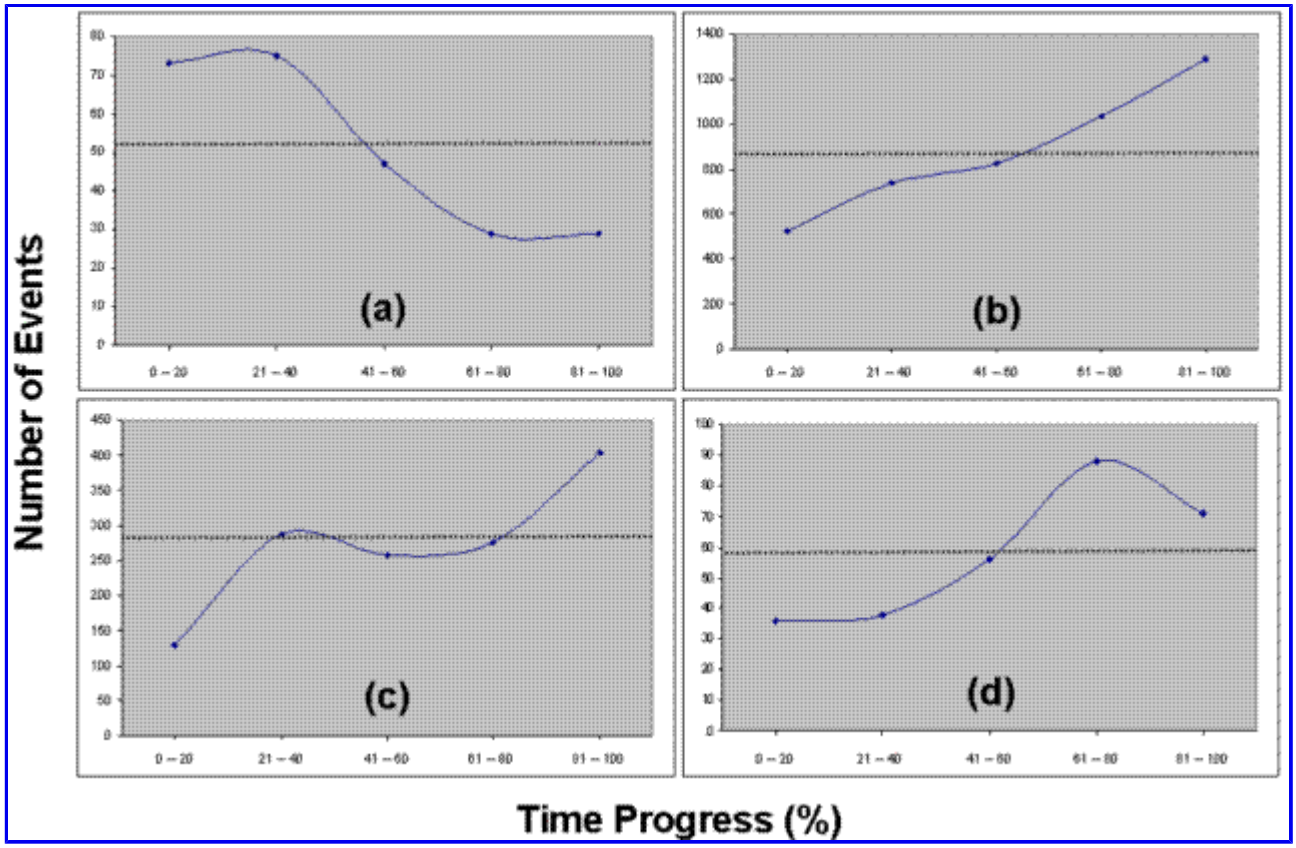


Figure 8. Organizing-related interactions over time: (a) Add Object; (b) Move object; (c) Resize object; (d) Delete object

Figure 8a shows the pattern of Add Object events over the duration of the triage task. For this task, Add Object events are generally associated with the creation of VKB collections, an explicit indicator of category creation. They can also indicate the creation of text notes in the workspace, although analysis of the resulting organizations and the videos of activity shows few notes being created. The data shows that category creation occurred in all time periods, which may imply that the subjects built structure incrementally as they learned more about the topic and the document set. However, subjects created many more categories during the first two time periods (0-40%). To substantiate this inference from the aggregated data, the data for individuals shows that 62.5% (for the first time period) and 75.0% (for the second time period) of the subjects created a higher than average number of collections, while only 20.8% of the subjects created a higher than average number of collections during the last two time periods (61-100%).

Figures 8b-d show that the Move Object, Resize Object, and Delete Object events are more common during the final time period (81-100%). These modes of expression were used more to refine the organization, to develop subcategories, and to make the organization more intelligible to others. We have examined the individual patterns of organizing-related user events to establish that Figures 8b-d are reasonably representative. Only 12.5% of

the subjects generated more Move Object events during the first time period (0-20%), while 83.3% of the subjects generated more Move Object events during the last time period (81-100%). Similarly, only 4.2% of the subjects generated more Resize Object events during the first time period (0-20%), while 75.0% of the subjects generated more Resize Object events during the last time period (81-100%). Only 25.0% and 20.8% of the subjects generated more Delete Object events during the first and second time periods (0-20% and 21-40%), while 58.3% and 54.1% of the subjects generated more Delete Object events during the last two time periods (61-80% and 81-100%).

In summary, subjects created categories throughout the triage process, organizing the information in part as they encountered it. The earlier phases of document triage - periods when the subjects tended to read more material more deeply and certainly to encounter more new information and notice more new topics - included more category creation. Other types of organizing-related activities, such as moving and resizing objects or changing their background colors - interactions usually associated with expressing more implicit characteristics of the documents - show different temporal patterns. This illustrates how reading and organizing patterns may be intertwined: category creation may be related to the relatively focused reading that takes place in the early stage of triage and category refinement may be related to the relatively faster form of reading that occurs later in triage.

Effect of Spatial Location on Triage

VKB objects display metadata describing the documents they refer to: in this case, title, URL, and creator or summary in a manner comparable to search engine results. In a prior study of triage behavior (Shipman et al. 2004), subjects evaluated 19% of the given documents using metadata alone (i.e. without looking at the document itself). This displayed metadata necessarily formed the linchpin to the subject's decision whether to read the document or not and what to do with it in the context of the triage task; factors such as authority, topicality, and document role were all gleaned from the displayed metadata. For example, one subject described his organizational approach using URLs:

> "I started with just schools. And then I was like, wait, there's a lot of orgs here too, and that's kind of formal too, so I kinda stuck that in the same thing."

Thus, if subjects are working from metadata in this manner, it seems important to examine the effect of how the objects are initially presented to the user on the display. Are objects examined in the order in which they are presented? Are objects that are not immediately visible more apt to be ignored, despite plenty of clues that more documents are available?

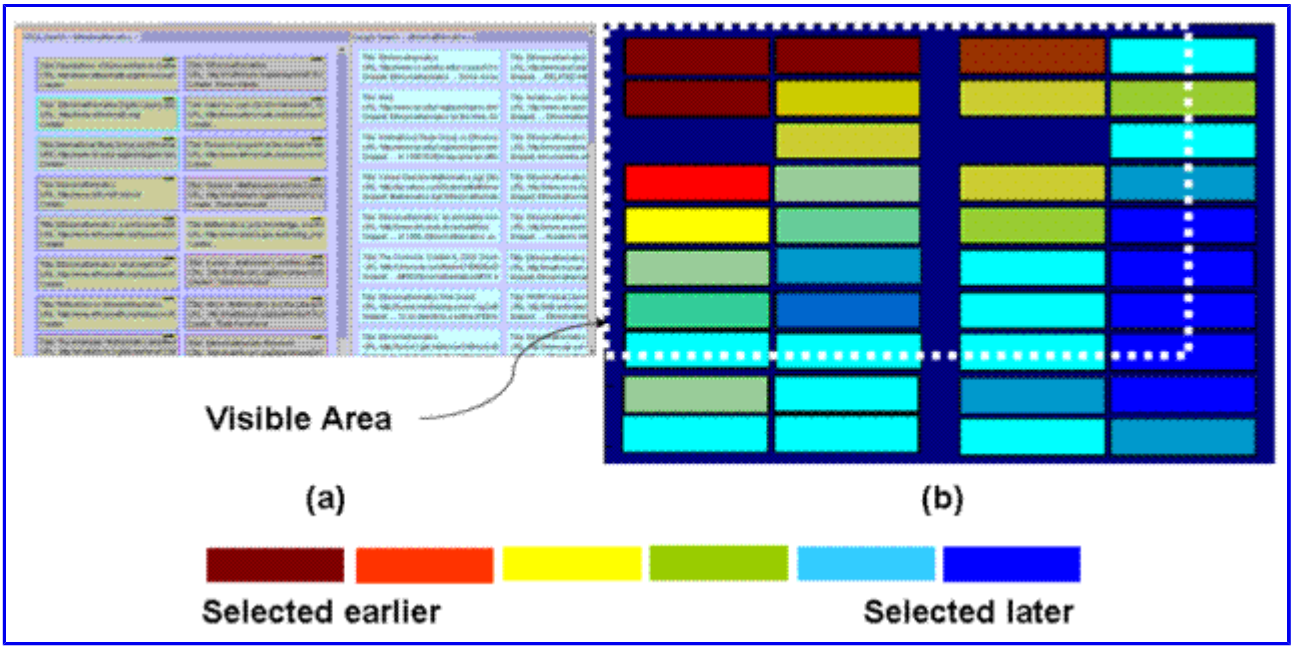


Figure 9. (a) Original location of document objects in the VKB workspace; (b) Order that the subjects first read the corresponding documents

Figure 9 illustrates the order that the 24 subjects first read documents: Figure 9a shows the initial layout that the subjects encountered - the documents from each search (NSDL and Google) are ordered in two columns with the first column listing the top ten results and the second column documents ranked 11 through 20. Figure 9b shows the order in which the subjects first read the documents. The gaps in the third position from the top in the first and third column indicate a repeated document for which the data from the log file could not be analyzed. Figure 9b indicates that the spatial position of document objects has an effect on reading order. Subjects tended to go through the document objects from those in the top-left corner to those in the bottom-right corner.

Earlier we mentioned that subjects usually progressed through different reading phases during the triage, even though the phases depended on individual styles. This implies that when during the triage process a document is encountered may have a decisive effect on *how* it is read. From this perspective, the initial layout of document objects in the workspace can affect a person's reading in very fundamental ways.

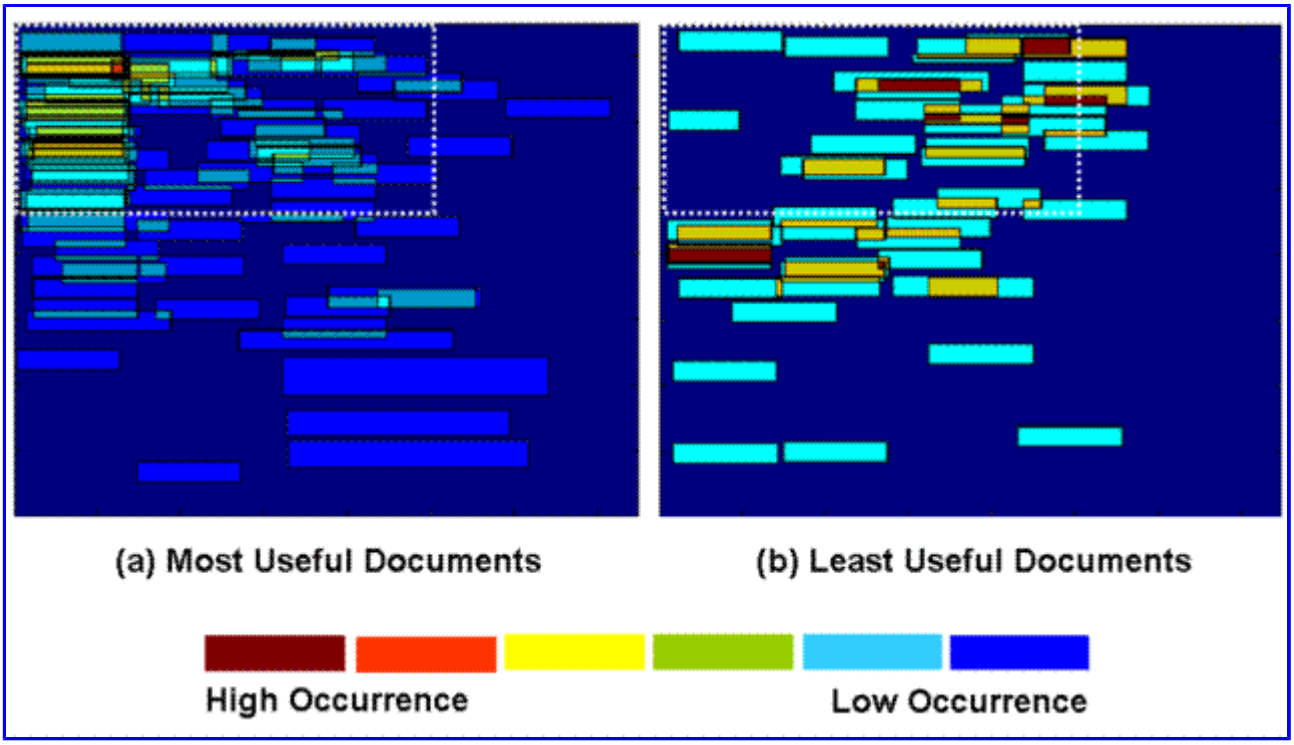


Figure 10. The post-triage locations of document objects judged to be the most and least useful - lighter colors indicate more items found in a location.

Figure 10 illustrates the relationship between the final location of document objects after triage and the 24 subjects' stated assessment of the documents. Figure 10a indicates the post-triage distribution of the subjects' most useful documents (top and left), while Figure 10b indicates the post-triage distribution of the subjects' least useful documents (more widely scattered, avoiding prominent top and left positions). The difference between 10a and 10b indicates that many subjects used space to express utility. Thus, the spatial location of document objects both affects how the documents are read and forms a key, but implicit, expressive feature of the workspace.

Discussion

The study shows that there are a variety of ways that people approach document triage and that considering reading and organizing separately limits the potential for computer support. Instead, reading and organizing need to be thought of as interacting with each other in very fundamental ways. A unified record of triage-related events may be used as the basis for adapting the multiple reading and organizing applications that come into play when a person performs this complex activity.

The study identified a need for organizing tool to include different kinds of document surrogates that readers may quickly switch among (such as overviews that display document thumbnails and overviews that present selected metadata elements). For

example our subject who wanted to see documents dense with text rather than libraries of links might like a thumbnail view of the documents, while the subject who worked from metadata might want to see just the URLs, since they were the salient characteristic he or she used to organize the documents.

People also need tools for skimming long lists of documents and through the content of long documents; these tools can aid them in locating content relevant for their current task without sacrificing context. Popout Prism (Suh et al. 2002) and TextTiling (Hearst 1997) are example of tools to aid in this activity by using term frequencies or retrieval terms to highlight certain words or portions of the document. However, term frequencies are task-independent and most users' minimalist approach to specifying query terms limits query terms' usefulness as distinguishers. Instead, triage tools need to bring out what people find salient about the document for their particular task. In other words, text analysis is not enough; the task and user preferences have to come into play. Mining reading activity for indications of user interest, such as the recognition of important annotations in XLibris (Shipman et al. 2003), has the potential to enhance support for retrieval and organization as well as directing later reading or re-reading. Similarly, mining organizing activity for indications of interest can enhance subsequent retrieval, reading, and categorization.

Given the intertwined nature of reading and organizing in triage and the fact that no one application is likely to support all aspects of the triage task, there is an overarching need for infrastructure for the applications involved to share information about the user's prior activity and the current state of the task. Our initial design of such infrastructure and models that combine reading and organizing to recognize user interest is reported in (Badi et al. 2006).

Because multiple applications are typically used during triage, the deeply intertwined nature of reading and organizing suggests a need to support the rapid switching between applications. In the study, the use of two displays provided screen space for the arrangement of multiple windows. Even so, moving windows and switching focus between displays kept such transitions from being effortless.

The study showed that the two dimensional spatial layout of search results influences the order in which they are read and that reading order, in turn, influences the triage outcome. Systems need to lay out search results in a way that makes any ordering clear to users; furthermore, the top ranked results should be visible without user effort. Also, the initial layout must balance the desire to show as many results as possible with the understanding that users need a workspace for organizing and interpreting the documents as triage proceeds.

In triage tasks on less familiar domains, as was true in this study, support for topic familiarization becomes important. Even if the domain is familiar, users need to gain an understanding of what is in the collection of results. Tools that aid image collection understanding (Chang et al. 2004) are a step towards supporting this.

Support for the overall process of document triage is complicated by the variety of work practices observed in the study. Rather than embedding an expected work flow or triage strategy, systems must be adaptable, and perhaps adaptive, to reflect the styles of individual users. The differing proportions of reading and organizing activity at different stages of the task indicate the potential value for systems to attempt to recognize such changes and adapt accordingly. Such recognition is complicated by the fluid transitions observed between periods of reading and periods of organizing.

Conclusion

The observed variety of approaches and strategies to performing this increasingly common task reveal that it would be unwise to build monolithic document triage applications or to embed assumptions about how best to go about this complex set of intertwined reading and organizing activities. Instead, we come away from this study with a small set of overarching concerns that will guide our future system development efforts:

- The tight integration of reading and organizing suggests a need for inter-application communication.
- Participants' varying uses of metadata in lieu of the complete document suggests that document surrogates should be adaptable to support different triage practices.
- Patterns of triage practice suggest that tools can use activity-based indicators of readers' interests to help them work through lists of relevant documents more effectively.

In short, the integrated nature of reading and organizing, coupled with the variable approaches to triage, should have important implications for how we design the many types of information systems we bring together when we perform document triage.

Notes

[1] This research was supported by a grant from Microsoft Corporation. *Back*

References

Badi, R., Bae, S., Moore, J.M., Meintanis, K., Zacchi, A., Hsieh, H., Shipman F., & Marshall, C.C. (2006) Recognizing User Interest and Document Value from Reading and Organizing Activities in Document Triage *Proceedings of ACM Intelligent User Interfaces*

pp. 218-225

Bae, S., Badi, R., Meintanis, K., Moore, J.M., Zacchi, A., Hsieh, H., Marshall, C.C., & Shipman, F.M. (2005) Effects of Display Configurations on Document Triage *Proceedings of IFIP INTERACT Conference* pp. 130-143

Bates, M. (1989) The Design of Browsing and Berrypicking Techniques for the Online Search Interface *Online Review* 13(5), 407-424

Bier, E., Good, L., Popat, K., & Newberger, A. (2004) A document corpus browser for in-depth reading *Proceedings of Joint Conference on Digital Libraries* pp. 87-96

Bowker, G. & Star, S.L. (1999) *Sorting Things Out: Classification and Its Consequences* MIT Press: Cambridge, MA

Chang, M., Leggett, J., Furuta, R., Kerne, A., Williams, J.P., Burns, S., & Bias, R. (2004) Collection Understanding *Proceedings of Joint Conference on Digital Libraries* pp. 334-342

Dillon, A., L. Kleinman, R. Bias, G. Choi, & D. Turnbull. (2004) Reading and searching digital documents *Proceedings of ASIS&T* pp. 267-273

Fu, X., Ciszek, T., Marchionini, G., & Solomon, P. (2005) Annotating the Web: An Exploratory Study of Web Users' Needs for Personal Annotation Tools *Proceedings of ASIS&T*

Hearst, M. (1997) TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages *Computational Liguistics* 23(1), 33-64

Jonker, D., Wright, W., Schroh, D., Proulx, P., & Cort, B. (2005) Information Triage with TRIST *International Conference on Intelligence Analysis* MITRE

Marshall, C. & Shipman, F. (1997) Effects of Hypertext Technology on the Practice of Information Triage *Proceedings of ACM Hypertext* pp. 124-133

Paepcke, A. (1996) Digital Libraries: Searching Is Not Enough *D-Lib Magazine* May 1996

Qu, Y. (2003) Sensemaking-Supporting Information Gathering System *ACM CHI 2003 Extended Abstracts* 2003, 906-907

Qayyum, A. & Bilykh, I. (2005) Navigational Characteristics of E-document Readers *Proceedings of ASIS&T*

Schwarzkopf, E. (2004) Enhancing the Interaction with Information Portals *Proceedings of ACM Intelligent User Interfaces* pp. 322-324

Shipman F., Hsieh, H., Moore, J.M., & Zacchi, A. (2004) Supporting Personal Collections across Digital Libraries in Spatial Hypertext *Proceedings of Joint Conference on Digital Libraries* pp. 358-367 http://www.csdl.tamu.edu/~shipman/papers/jcdl04.pdf

Shipman, F., Price, M., Marshall, C.C., & Golovchinsky, G. (2003) Identifying Useful Passages in Documents based on Annotation Patterns *Proceedings of European Conference on Digital Libraries* pp. 101-112

Slaney, M. & Russell, D. (2005) Measuring Information Understanding in Large Document Collections *Proceedings of Hawaii Int. Conf. on Systems Sciences (HICSS)* p. 105

Suh, B., Woodruff, A,. Rosenholtz, R., & Glass, A. (2002) Popout Prism: Adding Perceptual Principles to Overview+Detail Document Interfaces *Proceedings of CHI* pp. 251-258

Uren, V., Buckingham Shum, S., Bachler, M. & Li, G. (2006) Sensemaking Tools for Understanding Research Literatures: Design, Implementation and User Evaluation *International Journal of Human-Computer Studies* 64(5), 420-445

Tyrrell, R. & Leibowitz, H. (1990) The relation of vergence effort to report of visual fatigue following prolonged near work *Human Factors* 32(3), 341-357

Whittaker, S. & Hirshberg, J. (2001) The character, value, and management of personal paper archives *ACM Transactions on Computer-Human Interaction* 8(2), 150-170