

Spatial Hypertext and the Practice of Information Triage

Catherine C. Marshall¹ and Frank M. Shipman III

Department of Computer Science
Texas A&M University
College Station, TX 77843-3112
(415) 812- 4288
marshall@parc.xerox.com, shipman@cs.tamu.edu

ABSTRACT

Information triage is the process of sorting through relevant materials, and organizing them to meet the needs of the task at hand. It is a practice that has become increasingly common with the advent of “at your fingertips” information resources. To explore the characteristics of information triage and its interaction with spatial hypertext, a medium we claim supports the process, we have studied subjects engaged in a time-constrained decision-making task using a large set of relevant documents. We use the study task to investigate information triage under three different conditions: one in which the participants used paper documents, and two others in which the participants used variants of VIKI, a spatial hypertext system. Our findings suggest that during information triage attentional resources are devoted to evaluating materials and organizing them, so they can be read and reread as they return to mind. Accordingly, hypertext tools to support the practice should facilitate the rapid assimilation and assessment of new material, aid in the creation and management of a fluid category structure, allow readers to track their own progress through the information, and use minimum-effort methods to promote the intelligibility of results.

KEYWORDS: information triage; analysis; information workspaces; qualitative study; gathering interfaces, spatial hypertext; VIKI; digital libraries.

1 INTRODUCTION

An analyst fires off a query to a Web search engine or other on-line information resource and retrieves the 100 top-ranked documents. Most of them appear to be relevant to the task that prompted the search. What does the analyst do next? In this situation, it is common practice to print the relevant retrieved materials, mark them up with highlighters and the briefest of marginalia, and put them into piles [9]. In this way, the analyst begins to make sense of the materials,

to understand how they add up, and to explore alternative explanations, evaluations, or designs -- to produce the expected results that motivated the query in the first place.

We are investigating how spatial hypertext can support this activity as well as other aspects of information-intensive work. The process we describe appears to be a central part of using materials retrieved from large-scale information resources like the Web.

Information triage is the process of sorting through (the possibly numerous) relevant materials, and organizing them to meet the needs of the task at hand. The term “triage” seems particularly appropriate for this activity, since it is often time-constrained, and requires quick assessment based on insufficient knowledge. It is increasingly true that people do not have enough time to read all the information that they collect -- information that might have always existed, but is now at their fingertips. They must develop and apply strategies to scan, locate, skim, organize, and evaluate, instead of engaging in focused reading. They rely on what Joyce refers to as “successive attendings” to the same materials [6], rather than on scholarly reading and note-taking. If a retrieved article is important, it will be read again and again for new levels of detail as the reader’s understanding changes. As Joyce points out, “The value of what we collect is not as much embodied in what it is as in how we found it and why we keep it.”

Because many types of information are becoming readily available through facilities like digital libraries, and because there is an increasing emphasis on the rapid production of results, computer-based tools for information triage (and other interpretive activities) are vital. We can see this trend realized in Web workspaces like VIKI [12], Eastgate’s Web Squirrel [1], or the Web Forager [2]. The reader is transformed into a gatherer, as Rosenberg suggests [16].

To better understand the activity of information triage, and the interaction between computer-based tools and the process and results of information triage, we have conducted

1. Author’s current address: Xerox PARC,
3333 Coyote Hill Road, Palo Alto, CA 94304 USA

a qualitative study of people given a fictitious, but realistic, task. In the study described in this paper, we have presented subjects with a task that is modeled after an actual analysis performed in an industrial setting. In the study, subjects use a document corpus formed by culling a representative subset of the materials used in the real analysis.

The aspects of the task that we are focusing on are: (1) Strategies for dealing with too much relevant material over a constrained time period; (2) Methods of manipulating documents to make sense of them, including informal mechanisms for category creation and ways people keep track of their own activities; (3) Ways people manage the complexity of an evolving information space; and (4) How people create intelligible organizations (that is, organizations that can be apprehended by other readers). We plan to use these results in our future development efforts.

Our study explores the practice of information triage in two different ways. First, we investigate how people use paper documents while they are sorting through and evaluating information. Second, we use two variations of an experimental spatial hypertext tool, VIKI, to provide a technological substrate for the activity.

VIKI is based on a spatial notion of hypertext; it provides users with visual and spatial affordances for organizing and interpreting information [12]. The tool's design is based on previous studies of how people organize information in space [13]. Using this tool, people manipulate individual lexia (articles, notes, and other locally coherent units) as objects, and are able to create hierarchically nested spaces as collections. The relational structure remains largely implicit, and is available through heuristic recognition [17].

Our study differs from other studies of how people organize materials in their offices [8, 10, 11], since those studies are not based around a focused task, but rather aimed at general "office life." This study also differs from past studies of information-organizing tasks such as [14] in that we are departing from a scholarly notion of analysis, and investigating the exigencies of a world with fragmented attention, plenty of on-line resources and search tools, and a reported more urgent need for results [4].

In this paper, we describe our study approach and discuss our qualitative results. We conclude by discussing how hypertext tools for information triage might interact with the existing practice.

2 DESCRIPTION OF STUDY

Before describing the study in detail, we will briefly describe VIKI, the spatial hypertext system used by many of the subjects in the study. The description is a condensed version of the document used to train the subjects in the use of the system.

2.1 Relevant Features of VIKI

VIKI is a tool for organizing and interpreting information. VIKI allows users to create and manipulate visual objects in

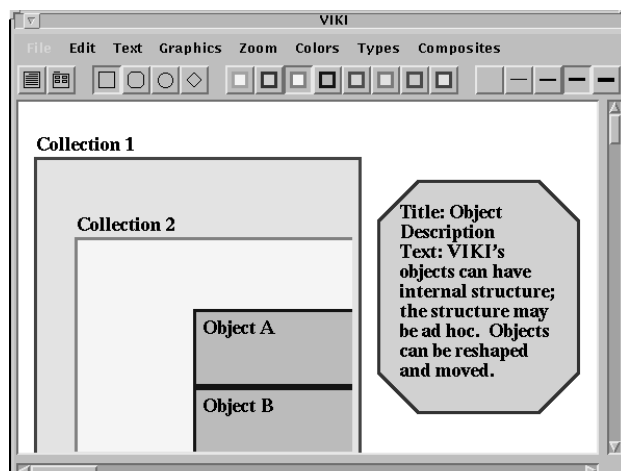


Figure 1: VIKI screen showing objects and collections.

a 2 dimensional space much like drawing or presentation software (i.e. MacDraw, PowerPoint). Objects can have their color, border width, and geometric shape changed using buttons below the menubar. (See Figure 1.²)

Each object contains textual materials -- in this case, the initial set of objects that subjects saw contained articles from an on-line service. Double-clicking with the mouse on an object presents the complete contents of the object in a separate window. Users can make virtual copies of objects -- new visual symbols that refer to the same textual content, but which may have different visual properties.

VIKI also allows users to create new spaces inside the current space. These new spaces are called *collections*, since they collect objects together. Objects and other collections may be put inside collections, and collections may be moved, colored, and reshaped like objects. In addition collections can be named.

By double clicking on any visible collection border, the user maximizes, or navigates into, the collection. Like Boxer [3], this causes the collection to fill the space available in the window. Double clicking on a maximized collection's border navigates back out of that collection.

Other facilities in VIKI helpful to the information triage study task are Search and Zoom. Search performs a full text search and displays the objects it finds in a dialog window, ordered according to a heuristic assessment of their relevance. Zoom scales objects in the space. The "Show all" zoom option resizes all of the objects and collections in the space to fit within the current window, so that no objects or collections are hidden (unless, of course, the scale is such that they are beyond screen resolution).

2. In this figure and subsequent figures, the colors in VIKI are represented as distinct shades of gray.

2.2 Subject Selection

The 15 subjects of this study were students just beginning an undergraduate Computer-Human Interaction class in a computer science department. The subjects ranged from 3.5 to 17 years of experience with computers, and 3.5 to 11 years of experience with windowing systems. All subjects had at least some prior experience with drawing programs like MacDraw, MS PowerPoint, or CorelDraw.

2.3 Study Task and Conditions

Subjects were given a scenario in which they had to recommend a machine translation package for a corporation with operations in Korea, Japan, Mexico, and the U.S. To help them make this recommendation, they were given 75 articles of mixed length related to machine translation, retrieved from a commercial service. Subjects worked alone and had 45 minutes to look through the articles, organizing them as they saw fit. At the end of this time, the subjects were asked to recommend a translation system to their manager. The subjects were also told that their manager would look at their notes and workspaces to better understand their recommendations.

The subjects were randomly assigned to one of three conditions: (1) those using the VIKI system with collections; (2) those using the VIKI system without collections; and (3) those using paper and standard office supplies. After completing the task, the subjects were given a questionnaire asking about their recommendations, the task itself, and prior experience with computer systems.

We gave subjects using either version of VIKI a 10-15 minute supervised training session prior to the task. Training involved going through a brief Web document that explained the relevant features of the version of VIKI that they would be using. Subjects were encouraged to try out all of the features described in the documentation during training. After training, the subjects were given the written scenario. When they completed reading the scenario, we started VIKI. Initially, the 75 objects appeared in a list down the left-side of the window; the "Title" field appeared on each object's symbol. Figure 2 shows this layout; the list of articles extended beyond the limit of the window, so subjects needed to scroll or zoom to see the entire list. We then left the subjects alone for 45 minutes, interacting only to restart the system when it crashed or to warn them when the 45 minutes were almost up.

Subjects using paper and office supplies sat at a clean 5' by 2.5' desk; off to one side of the desk were assorted fasteners post-its, pens, and highlighters. After subjects had read the printed scenario, we gave them the 2/3" tall stack of trade press articles and instructed them to use the desk and floor, if necessary, to complete the task. Again we left the subjects alone except to warn them when the 45 minutes were almost up.

2.4 Questionnaire

We gave each subject an exit questionnaire at the conclusion of the study task. The questionnaire confirmed the subjects'

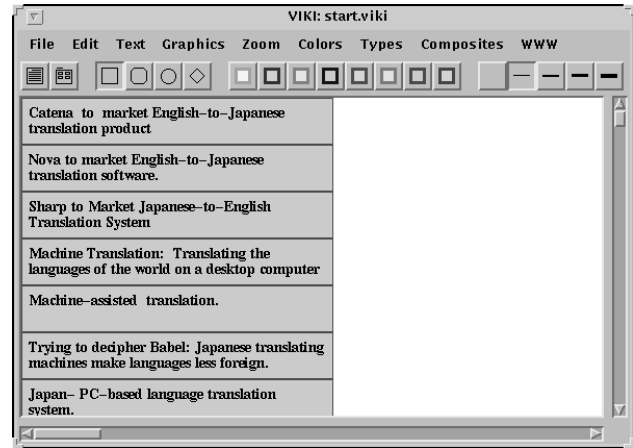


Figure 2: Initial layout subjects using VIKI saw.

choice of translation software and how they would support such a decision. The questionnaire was also designed to help us assess how the subjects felt about the task (whether they were confident in their recommendations, whether they thought that more time would have helped them complete the task, and whether they thought more time would have changed the task's outcome).

The questionnaire included a few questions about the subjects' past experiences with computer systems. This allowed us to determine their familiarity with computers, windowed interfaces, drawing and presentation software, idea processing and brainstorming software, and whether they had used any commercial software they felt would help them in this particular task.

2.5 Other Data Collected

We collected other data that resulted from each subject performing the task, including the way in which they had organized the articles. In the paper condition, we used a digital camera to record an image of the desk, and fastened together any piles of papers for later examination. In the two conditions using VIKI, we saved the subject's final layout. During the course of the task, VIKI created a file containing the order of "undoable" actions (move, create, delete, change visual properties, change text, ...) performed by the subject.

3 RESULTS

After 45 minutes of work, each subject was able to recommend a machine-translation system for his or her hypothetical corporation. Many of them recommended the same system (NEC's Pivot) to their managers, although there were several others described in the set of articles that would have met the task's specifications; a few subjects recommended these variants. Few were very confident with their recommendation, given the sharp time constraints in which the task was performed. This result was constant across the three conditions, although in general, the group using paper felt more confident in their recommendations than the two using the computer-based tool.

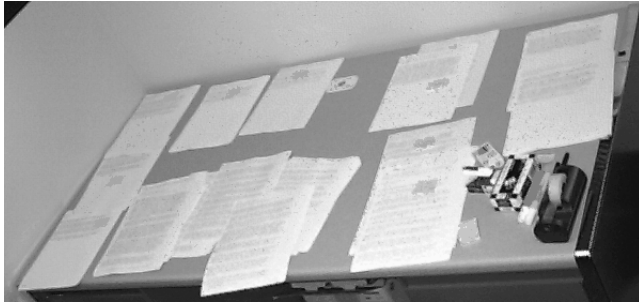


Figure 3: Desktop after P1 completed task. Piles of paper around sides were used to categorize articles while center pile in front describes recommendation.

Example results from each of the three conditions are shown below. In the discussion, P<n> refers to the five subjects who used paper copies of the articles; NCV<n> refers to the five subjects who used VIKI with objects only, no collections; and FV<n> refers to the five subjects who used VIKI with both objects and collections.

Figure 3 shows P1's desktop after completing the task. The subject has piled materials into different stacks, and has indicated that a small stack of papers pertaining to his recommendation are for his manager (the papers that are partially hanging off the front edge of the desk). Figure 4 shows NCV4's screen at the close of the task. Once again, the material has been organized, and a small set of articles pertaining to the two candidate translation packages the subject has located are left in evidence. Figure 5 shows FV1's final screen layout. Objects representing the bulk of the materials remain on the left side of the window. The subject has created collections to show his recommendation and some general articles that his manager might read to better understand machine translation.

Some characteristics of task performance held across the three conditions (paper and the two system variations); others were specific to the medium. Individual differences in

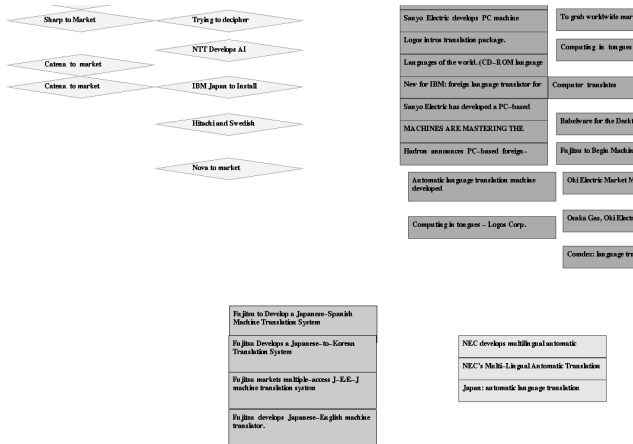


Figure 4: NCV4's organization at end of task.

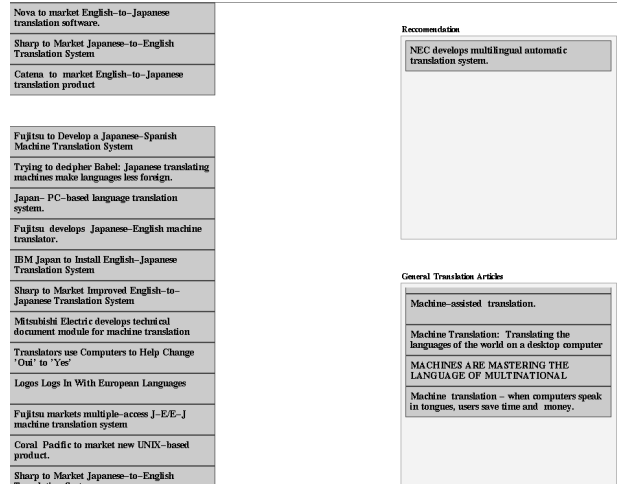


Figure 5: Final layout for FV1.

strategy were also apparent. The common elements of the three use situations were:

- (1) The articles were organized according to the nature of the task, rather than by other strategies suggested by their content. For example, many articles were press releases for product announcements or evaluations of specific software packages, yet our subjects often did not choose to cluster them this way (by system or software package name, or by vendor). Pragmatic and rhetorically-oriented categories were also created under each of the three conditions.
- (2) Subjects found ways to create categories and subcategories within the confines of the particular medium they were working in. They used the affordances of the medium creatively to transcend its limitations.
- (3) Subjects devised a variety of ways to communicate to their hypothetical manager about the information spaces they created. A few left explicit notes, but others relied on the structures themselves as being readily intelligible to their expected reader.

3.1 Task-Based Organization

Most of the subjects attempted to organize materials into semantic categories based on the constraints of the task, rather than the topic of the article. As Kidd has noted, a knowledge worker's ability to organize information depends greatly on the role it will fill [8]. Our subjects created three kinds of task-based organizations: semantic categories (derived from article content); pragmatic categories (based on ultimate expected usefulness of the article in performing the task, especially within the 45 minute time period); and rhetorical structures (based on the explicit requirement to communicate and support their decisions).

The semantic categories the subjects created are similar in all three conditions. Subjects created categories based on the language pairs that the corporation was posed as needing (English --> Korean, for example) and sorted articles

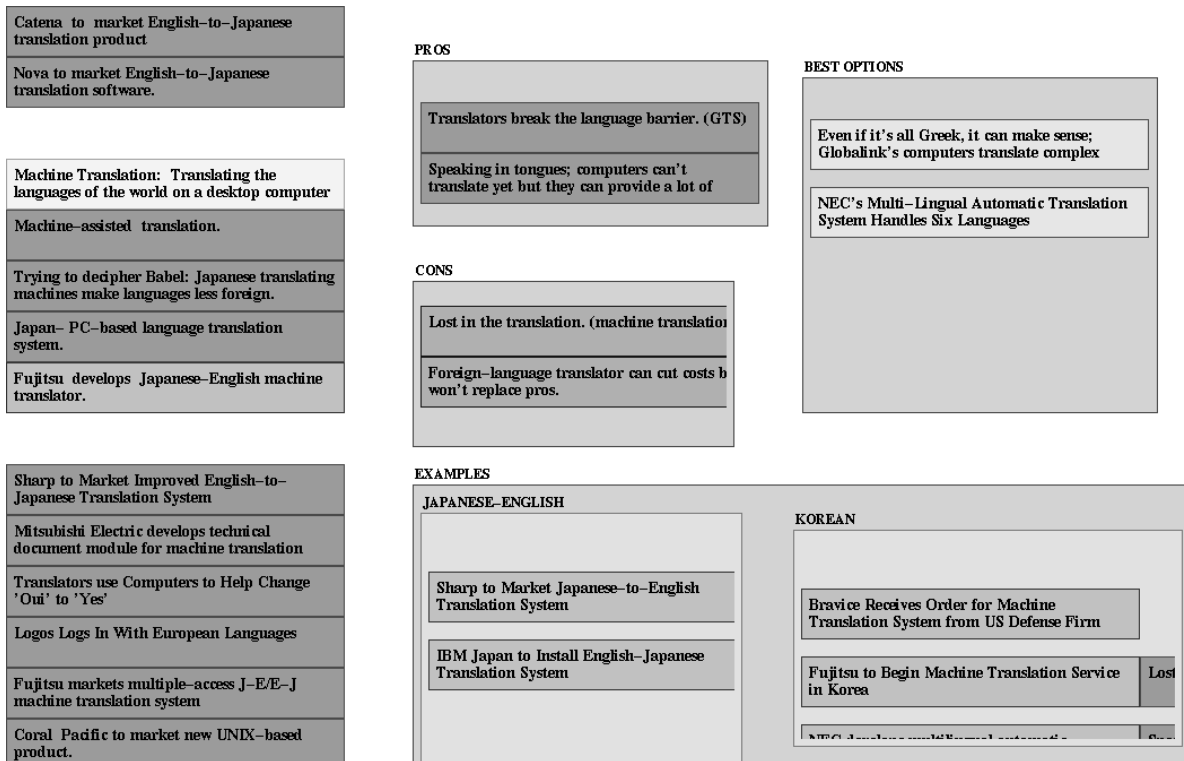


Figure 6: Final organization from FV3.

accordingly. Several also created subcategories, based on source language, then target language. One subject called his higher-level category “Japanese --> ?” and the subcategories beneath that “Japanese --> English”, “Japanese --> Korean”, and so forth. Another created categories for language pairs, and subcategories based on additional languages handled.

We observed pragmatic categorizations under all three conditions. Most subjects created special categories to contain the longer, multi-topic articles; these categories were labelled “General Translation Articles”, “background information”, and “general info” for example. Articles that described systems handling extraneous language pairs found their way into categories like “Unneeded translations”. Several subjects considered the age of the materials as a relevance, and grouped them accordingly. One subject, using VIKI without collections, color-coded the older information gray, while several others pushed it off to the side. One of the subjects using paper tossed these articles into a pile on the floor.

Subjects outlined simple rhetorical structures to address the directions they were given, to organize the materials well enough that their manager would be able to look through them so she could learn more about their recommendations. One subject using the full system created a presentational structure with PRO and CON collections which both contained longer articles, and a BEST OPTIONS collection that contained a list of relevant articles. (See Figure 6.) Interestingly, the titles of the articles in the PROS and CONS collections also reflect the positive or negative qualities the

subject has ascribed to their categories, and the titles of the BEST OPTIONS selections both mention the software providers. Others used an even simpler strategy of simply making a list of the articles of interest or, in the paper case, a distinct pile, set apart from the rest of the materials.

3.2 Using the Affordances of the Medium

How were categories created? They were created, regardless of the study conditions, but here we observed some differences based on the affordances of the various media. In the paper condition, subjects were given leave to use the physical space of the desk and floor, a variety of fastening equipment (paper clips, tape), marking devices (pens and highlighters), and post-its. VIKI, used without collections as it is in the NCV condition, allows users to change the color, shape, text style, line width, size, and spatial positioning of objects. With collections, VIKI also allows users to create new subspaces, with varying visual properties of line width, color, and size, which can also be positioned in the space that contains them. Thus each medium provided our subjects with different ways of creating categories -- many of which were used.

The subjects using paper copies of the articles generally sorted materials into piles. Several subjects labeled the piles with post-its to describe the contents of the individual piles. To mark “articles to return to”, one of the subjects used post-its as a bookmark or signal. Subjects also used the highlighting pens (or, in one case, underlines) to emphasize the words or phrases they used to put the material into the piles. Beyond the post-it labels and within-article

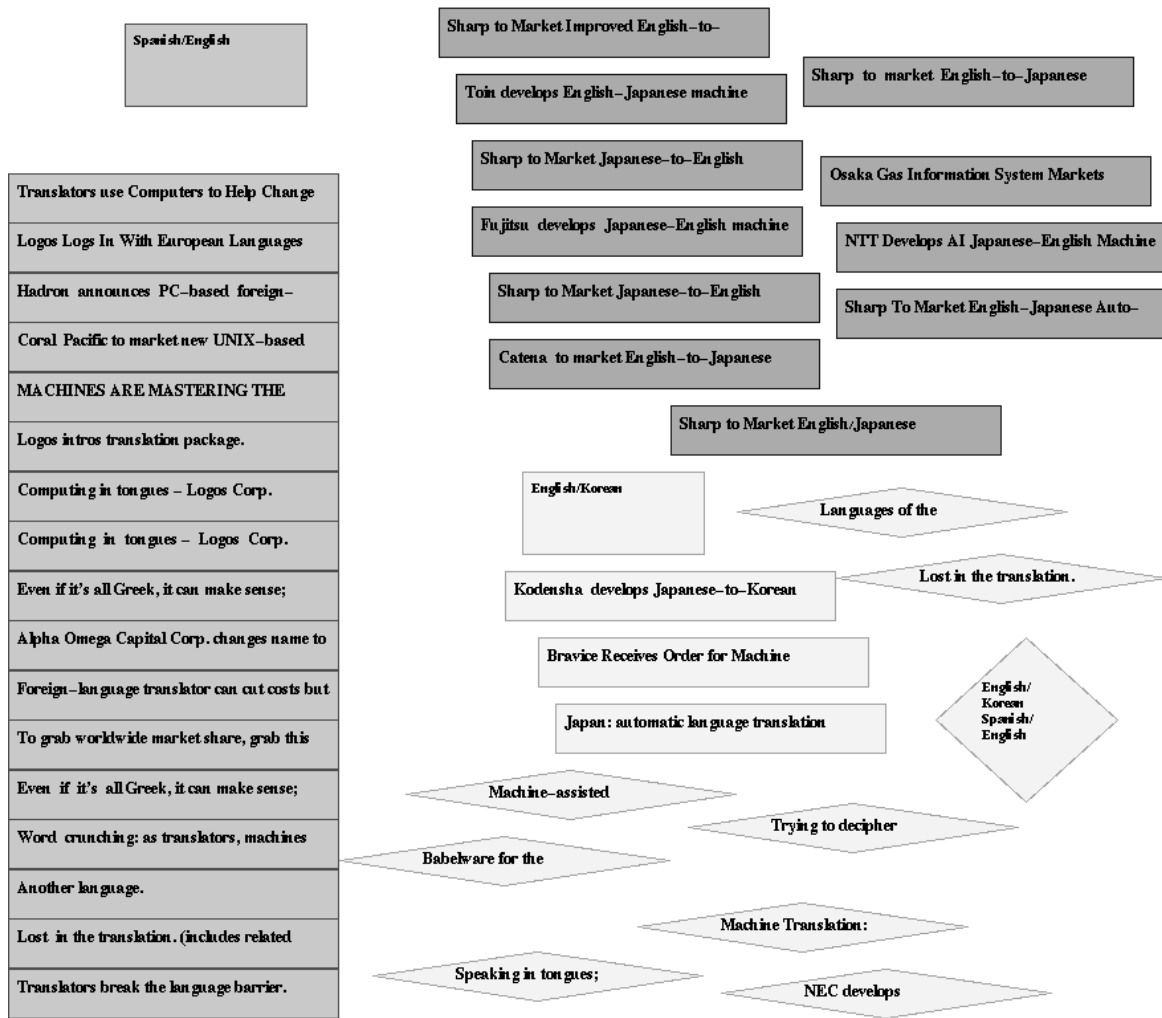


Figure 7. Categories and subcategories using color and shape.

highlighting, few annotations were observed; the subjects did not take additional notes, nor did they scribble marginalia.

The subjects using VIKI without collections, the NCV condition, made heavy use of spatial structures and color for basic categorization. Each of the five subjects performed the categorization in a slightly different way, but most took advantage of the ability to use color and to manipulate objects in space. A few used object shape to encode category information. Figure 7 shows a portion of a space created by subject NCV5. Like the other VIKI subjects, this subject has used lists to organize the space. In the lower left corner of the window, the subject has used several other category-creation conventions. First, he has color-coded the articles to match the category label. The article whose title begins “Kodensha develops Japanese-to-Korean...” is the same color and shape as its category label “English/Korean” (the article mentions the English/Korean translation capabilities of the system). The subject has also created subcategory elements that are diamond shaped to match their label “English/Korean Spanish/English”. Each of the articles that is this shape and

color describes a software package that handles these language pairs. We saw similar strategies using shape and color in three of the five subjects’ work; the other two made extensive use of unlabeled lists to sort the articles.

One of the subjects, NCV3, uses his exit questionnaire to explain the organization of the space:

“grayed out documents are older (before 1990)
columns are: multilanguage E/J J/K E/S not useful
important documents are in red, orange, or green”

A footnote to the subject’s questionnaire adds:

“should have made all documents that talk about Fujitsu and/or Atlas the same color (all blue documents pertain to Sharp)”

Once again, under NCV conditions the subjects did not really take notes, although during the training each of the subjects demonstrated that they could create new objects and type into them freely. The two out of five subjects who created new objects did so to label their categories. Thus

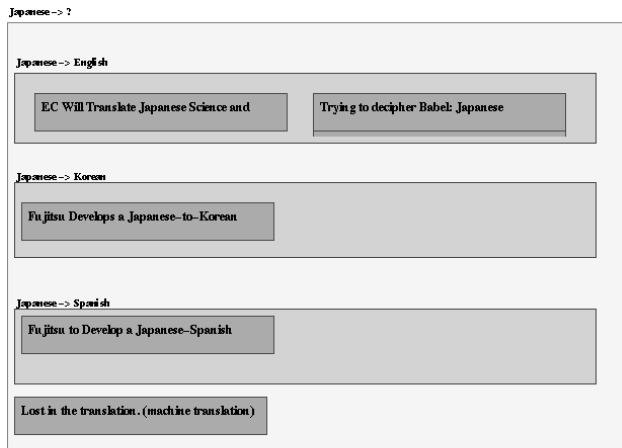


Figure 8. A category/sub-category structure using collections

their use of new objects was directly analogous to the use of post-its by the subjects under the paper condition.

The FV condition offered the subjects the ability to create collections, which they invariably did, to categorize materials. The categories were similar in type to the kinds created under the other conditions. Figure 8 shows an interesting contrast to the category/sub-category structure exhibited by Figure 7. Here hierarchical nesting of collections (spaces) is used to create something analogous to Figure 7's complexly-coded structure.

The subjects' organizational strategies changed over the task's duration as they became more familiar with the materials. Partial codings were in evidence in many of the VIKI conditions -- for example, an object or two is turned orange because the article covered English/Spanish translation, but then the coding strategy is discontinued as other aspects of the task come to light. Such a change in strategy is evident in the quote from NCV3 (reported above). Many other subjects reported that, given more time, they would have changed their organizing strategy.

3.3 Communicating about the Organization of Materials

The subjects all responded to the directions that they communicate to their manager not only their decision about which software package to buy, but also about how their materials were organized so she could understand the decision. In fact, the scenario poses that the subject will be at the corporation's facility in Japan during the software purchase (and will be called in the middle of the night if there are questions).

The subjects took three different approaches to this communication. The first was to organize the space (whether physical or computer-based) with an eye toward its future intelligibility; the structure itself is an embodiment of the communication. We see evidence of this strategy in the use of post-its, labelling objects, and named collections; in fact, the ability that collections offer to help manage space (the collections can be resized to hide constituent objects) seems

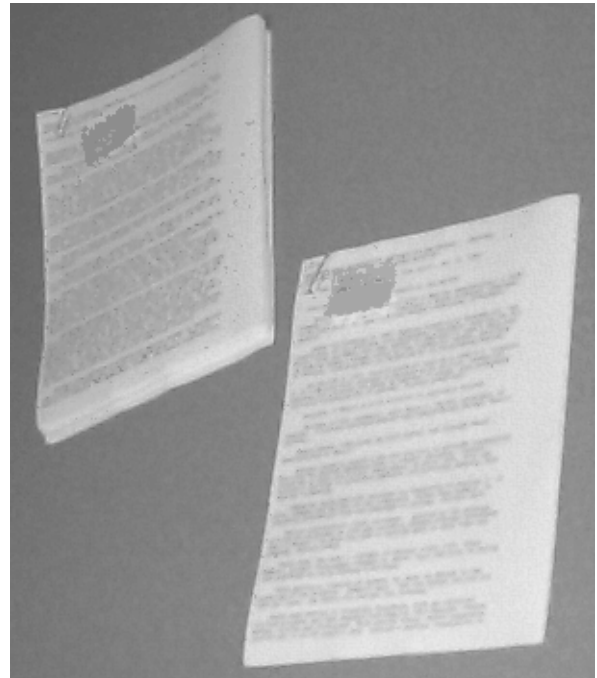


Figure 9. Communication strategy in paper condition

to lead to neater, more visually intelligible structures in the FV condition. Figures 3, 7 and 8 show examples of communication through structure.

In a second approach, subjects left explicit notes for their manager, with the assumption that the process they used to sort the materials and find the appropriate system(s) were not of particular interest. Subject P2 left his materials sorted into two piles on his desk -- the materials ancillary to his decision, and a stack with a post-it note on top "Recommend we check into these." Figure 9 shows P2's communication strategy; Figure 5 above shows the analogous strategy in VIKI with collections (FV1).

Finally, some of the subjects described their codings after the fact (in the questionnaire), as a step toward making their organizations readable. This strategy seemed the most prevalent in NCV, since color and shape were apt to be used as tacit codings. When asked what notes he would leave his manager, Subject NCV5 reported, "A key describing the color coding I used and the shape coding for identifying the important points."

4 DISCUSSION

This study gives us some new insight into both the practice of information triage and technology to support it. Informal discussions with people who do information-intensive work tells us that the work is changing. A description of "A Day in the Life of..." an information analyst who is responsible for daily report reveals intense time pressures, interruptions, and an hour or two to read through roughly as many (to twice as many) articles as we have given our subjects [4]. Information triage is a real outgrowth of new information technologies. What are the changes it embodies, and what are the tools it demands?

4.1 Organization vs. Reading

Information triage implies changes in reading and attention. Tools to support information triage may change the nature of reading even more, by making quick, partial interpretation of content possible, and multiple readings the order of the day.

In general, our subjects who used paper articles and physical space as their organizing medium read more. This result is unsurprising in some ways, since it is still easier to read paper documents than documents on the screen. But we might expect this effect to simply manifest itself in paper subjects getting through more articles. Instead, we observed that organizing became more central than reading when the subjects used computer-based tools; they relied extensively on appropriate title (which is what they could see, after all), and on the full text search facility the tool provided.

What would subjects have done, given additional time? This is a key question for information triage, since time is assumed to be in short supply, especially in comparison to the sheer volume of information it is possible to obtain at the outset of a task like this. Across all three groups, subjects reported that they thought about supplementing the comparison, or completing what they were doing when time ran out. However, there were some notable differences among the three when it came to how they described further activities. On their exit questionnaires, the subjects who used paper more often reported that they would read for additional detail:

P1: "Read more detailed info. about currently available products (the BYTE article, p. 177) And find out about hardware requirements, software capability, and cost"

P2: "Organize the material and look through it again. This time I would read more and look for detail, about each product in order to organize them better and make a more informed choice."

P3: "Read the info that I selected as critical more carefully and perhaps highlight some important text for my boss to help support my decision"

By contrast, subjects who used the two variations of the VIKI system (NCV, the version without collections, and FV, the unmodified version) focused on searches and structure-building. The subjects who used NCV generally worried about how the documents were organized. When they were asked how they would spend additional time, three of the subjects came up with similar responses:

NCV2: "Finish the organization, then select the best one of each group"

NCV3: "better job of re-organizing the documents: I spent my time coming up with a recommendation, not organizing the documents"

NCV4: "reorganize things into companies & then read all the independent studies."

Subjects who used the complete system (FV) focused even more closely on the structure-building activities suggested by the task. The following three responses to the "extra time" question all seem to arise from the use of the system's collection mechanism to organize the materials:

FV1: "Try to develop a convenient + easy to follow filing system for the articles -
1-1 translation
1-2 "
2-1 "
many-many "

FV2: "Have a box with recommended packages"

FV5: "I would organize each big collection into smaller collections and possibly change some of the names that I currently have on the big collections. Also I would look a little more carefully at some of the articles as some might be misplaced."

Of course, one of the objectives of the task was to organize the materials, so it is clear that some of the responses arise out of a desire to follow our directions. Yet the paper group seems to have a much clearer focus on reading and the task itself (hence the desire read for more detail), while the computer-based groups have focused on the tool, using full text search capabilities, and producing the visually apparent organization of materials.

The group who used paper documents had the advantage of many more cues about the scope of the task. They could easily ascertain the length of individual documents, as well as their progress through the stack of articles. They also could manipulate the documents with far greater facility -- for example, they could highlight text, an activity the computer-based tool doesn't support. Yet, when we demonstrated the system to the subjects who used paper, they seemed to feel that it would have been useful in performing the task.

In all three conditions, the longer articles were retained as potentially valuable, but they were less often used or read. The short, single-topic articles were more often attended to and used in the subject's decision. Unsurprisingly, many of the subjects felt that additional time would have been valuable, and would have allowed them to take a closer look at these articles (although several felt it would not have changed their decision, only their ability to support it to their manager).

In the set of articles we gave the subjects, information about several of the systems was distributed among many articles; for example, a subject would need to integrate information from multiple articles to know that Fujitsu's Atlas translation package addressed translations among the scenario's four languages. Few subjects had the time to do this kind of integration. Thus distributed information was discarded in favor of localized information. Over half of the subjects (8/15) recommended the NEC system, the only one

that was described in short articles that mentioned all the languages in question. Subject P1 reported:

“Most of the information did not affect my decision. It was simply based on two pieces of information: the NEC articles and the Byte articles.”

Traditional methods of scholarship and analysis would assume a focused reading; all candidate software packages would have been discovered and evaluated in such a regime. Instead, information triage practices more closely follow pragmatic notions of sufficiency.

4.2 System Design Considerations

From our observations of subjects engaging in information triage, we arrive at a set of system design refinements for supporting this kind of process. We know from this study, as well as previous studies, that rapid assimilation of new material, category creation and the management of complex information spaces, and easy ways of promoting intelligibility are all vital to the process. How can the additional details of this study's results be translated into design refinements, given a gathering-style hypertext system as our base.

Rapid Assimilation of New Material.

The rapid assimilation of new material relies on techniques that both draw on and diverge from information retrieval; three of our “Paper” subjects noted that they would have liked to use software that would let them search. Unlike most information retrieval settings, we assume that the user is already working within a task-based corpus; materials have been winnowed down to just those that are potentially relevant to the work at hand. In our observations, most searches that people did were motivated by a desire to cluster documents into specific semantic categories. For example, the searches people performed mostly involved individual language names, language pairs, or system names. Heuristic approaches based on partial representations of domain concepts and instances would help users create such clusters. Subjects also expressed a desire to constrain searches to a designated portion of the space.

Since longer, multi-topic articles proved to be such a problem in a time-constrained situation, techniques for topic/subtopic segmentation would support a user's ability to locate individual text segments of interest in longer articles. Some work, such as Hearst's TextTiling has already been done in this area [5]. Better facilities for within-document interpretive marking -- analogous to VIKI's whole-document marking capabilities -- may help practitioners set off and record impressions of these text segments once they have been identified.

Category Creation and the Management of Complexity.

The second design aspect, category creation, poses some interesting questions. In VIKI, we assume that structure is hypertextual -- that is, users will refer to content from many

different contexts. VIKI's data model implements this by distinguishing between object and symbol, where the object is the underlying content, and the symbol is a visual reference. In analysis, this distinction is intended to support multiple interpretations and multiple uses of the same material. In information triage, this distinction may not serve the task as well as it does subsequent interpretive activities, since people use the references as feedback for “how far along” they are -- how much they've looked at already and how much is left. To take off from Rosenberg's discussion, they are using VIKI to track their own activity as reader-gatherers [16]. Hence they are managing the gathering and interpretation process itself using the tool.

We observed that few subjects distinguished between objects and symbols, instead preferring to treat the object and symbol as a single unit (to them, the purple box “is” the article). To this end, few even reshaped or changed the visual characteristics of very many of the symbols that referred to the trade journal articles (although some color-coding was observed, and a few changed their shape). Instead, they were used as is -- dragged around in space, or put into collections.

Collections, on the other hand, proved to be very useful in managing scale and the complexity of a multiple category triage process. One of the subjects working in VIKI without collections reported that “For me, it might have been easier to use a filing system (preferably under OS/2, but Windows File Manager would have worked)”. When subjects could construct a hierarchy (the FV condition), they did. In fact, collections seemed to support the use of symbol-as-object.

How can we interpret these results? We might argue that if symbols are to be regarded as fixed representations of objects, then it might be best to use them to reclaim some of the visual cues lost in an electronic environment. Indications of article length and age would have helped the subjects according to their actions and the questionnaire results. Monty has found that such cues are important to information systems users [14].

On the other hand, information triage is only the start of a more extended interpretive process, one that is highly fluid in nature. Even during their brief tenure as systems analysts, our subjects experienced meaning changes midstream, as their understanding of the task evolved and familiarity with the materials grew. As Kaplan and Moulthrop point out, semantic spaces need not be limited to the same degree of fixity as architectural spaces [7].

Intelligibility of Organizations.

Rapid performance of tasks does not bode well for requesting extraneous work in documenting either process or product, or in building complex argument structures. Our subjects left only the briefest indications of how they had organized materials. An automated way to meaningfully document process, such as Reeve's embedded history [15],

would have helped our hypothetical manager decode the subjects' organizations.

Structure-building mechanisms that support categorization and space management (such as collections) turned out to be a great boon to intelligibility in our study. The meaning of color codings was not always made perspicuous in the space, but collections were, for the most part, self-documenting.

5 CONCLUSIONS

In the final regard, what did we learn from performing this study?

We expected a strong influence of hypertext technology on the practice of information triage. One of the strongest influences arises before our technology even enters the picture. The original analysis task -- the task our study was modeled after -- took place in 1992 and used a commercial service. At the time this paper was submitted, a complex query that used Alta Vista to gather like materials returned more than 1000 Web pages³; most look at least as informative as the collection gathered in 1992. By the time we revised this paper in January, 1997, the same query returned about 6000 pages; a more carefully tuned version returned 2000. Our subjects felt overwhelmed and frustrated with 75 articles; experienced analysts cite 100-150 documents as their limit in a similar situation. The ready availability of the materials and the exigencies of the practice are compelling.

Naturally, people take advantage of the medium they are working in, be it paper or a gathering system like VIKI. As Joyce predicted, their main concern is how to assess the materials they have, and how they return to them as the materials return to mind. Although our paper-using subjects did more reading, reading is less central to the practice of triage than quick assessment; this subset of our subjects immediately warmed to the idea of a hypertext gathering tool for manipulating and structuring the documents.

We expect to see more such hypertext gathering tools in the future. Our study confirms that it is important that they include affordances for fluid recording of interpretation, ways of dealing with the great variety of pragmatic issues that arise (longer documents, for example), and the means for triage practitioners to track their own activities within the space.

REFERENCES

- [1] Bernstein, M. Web Squirrel. <http://www.eastgate.com/squirrel/>
- [2] Card, S., Robertson, G., and York, W. The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. *Proceedings of CHI '96*, Vancouver, B.C. (April 14-18, 1996).

- [3] diSessa, A., Abelson, H. "Boxer: A Reconstructible Computational Medium." *Communications of the ACM*, 29, 9, (1986), pp. 859-868.
- [4] Gifford, D. "Life in the Trenches." A presentation given at the ARPA Information Science and Technology Superlink Workshop, Arlington, VA, July 19-20, 1995.
- [5] Hearst, M.A. Content and Structure in Automated Full-Text Information Access. Ph.D. Dissertation, Dept. of Computer Science, Univ. of California, Berkeley, 1994.
- [6] Joyce, M. "The lingering errantness of place." A talk given at the ACRL/LITA Joint Presidents Program, American Library Association, 114th Annual Conference, Chicago, June 26, 1995.
- [7] Kaplan, N. and Moulthrop, S. "Where No Mind Has Gone Before: Ontological Design for Virtual Spaces." *Proceedings of ECHT '94*, Edinburgh (September 18-23, 1994), pp. 206-216.
- [8] Kidd, A. "The Marks are on the Knowledge Worker." *Proc. of CHI '94*, (Apr. 24-28, 1994), pp. 186-191.
- [9] Levy, D.M., Marshall, C.C. "Going Digital: A Look at the Assumptions Underlying Digital Libraries." *Communications of the ACM*, 38, 4, (1995), pp. 77-84.
- [10] Malone, T.W. "How do People Organize their Desks? Implications for the Design of Office Information Systems." *ACM Transactions on Office Information Systems*, 1, 1 (Jan., 1983), pp. 99-112
- [11] Mander, R., Salomon, G., Wong, Y.Y. "A 'Pile' Metaphor for Supporting Organization of Information." *Proc. of CHI '92*, (May 3-7, 1992), pp. 627-634.
- [12] Marshall, C.C., Shipman, F. M. "Spatial Hypertext: Designing for Change." *Communications of the ACM*, 38, 8, (1995), pp. 88-97.
- [13] Marshall, C.C., Shipman, F.M. "Searching for the Missing Link: Discovering Implicit Structure in Spatial Hypertext." *Proceedings of Hypertext '93*, Seattle, WA (November 14-18, 1993), pp. 217-230.
- [14] Monty, M.L. *Issues for Supporting Notetaking and Note Using in the Computer Environment*. Ph.D. Dissertation, Department of Psychology, Univ. of California, San Diego, 1990.
- [15] Reeves, B.N. *Supporting Collaborative Design by Embedded Communication and History in Design Artifacts*. Ph.D. Dissertation, Dept. of Computer Science, Univ. of Colorado, 1993.
- [16] Rosenberg, J. "The Structure of Hypertext Activity." *Proceedings of Hypertext '96*, Washington, D.C. (March 16-20, 1996), pp. 22-30.
- [17] Shipman, F.M., Marshall, C.C., Moran, T.P. "Finding and Using Implicit Structure in Human-Organized Spatial Layouts of Information." *Proc. of CHI '95*, (May 7-11, 1995), pp. 346-353.

3. The query was of the form: ((language near translation) and software) and (Spanish or Japanese or Korean).