Information organization and retrieval: A comparison of taxonomical and tagging systems

Richard Pak[1], Steven Pautz[1], and Rebecca Iden[2]

Clemson University

[1]Department of Psychology, [2]Department of Industrial Engineering

Pak, R., Pautz, S., & Iden, R. (2007). Information organization and retrieval: An assessment of taxonomical and tagging systems. *Cognitive Technology, 12(1),* 31-44.

Abstract

The purpose of the following two studies was to examine performance and subjective workload differences between two types of information organization systems. Organizing information by using hierarchical categories or by labeling (i.e., tagging) each has their benefits and drawbacks. However, no empirical work has examined the performance advantages of one system over the other. In Experiment 1, taxonomies and tagging were compared in terms of how they affected performance and workload at the information organization stage. The results showed that users encountered greater frustration when organizing information in the tagging system. However, they also organized information more quickly in the tagging system compared to the taxonomic system. Experiment 2 examined performance and workload at the information organizational *and* retrieval stage using a computerized interface. The results showed that taxonomic systems resulted in faster times to organize information, but also resulted in more mouse clicks at retrieval. Overall, the results show that each information organization scheme may lead to different performance outcomes.

With the ubiquity of information technologies, a large portion of our time spent online is looking for information (Pew Internet and American Life Project, 2006). It has been estimated that workers spend between 15% and 30% of their time searching for information (Feldman, 2004). At home, looking for information is the predominant activity when we are researching a health condition, searching for a friend‟s email address, or looking for the best cellular phone plan. The success of our information search activities has a direct effect on our productivity at work and the quality of the decisions we make (e.g., losing a potential customer, deciding on a Medicare drug plan).

In addition to *retrieving* information, we spend a large portion of our time *organizing* information with the intent of retrieving it later. Just imagine the most frequent task of your typical day: a large majority of the tasks probably involve filing information away for later retrieval (e.g., documents, emails, and photographs) either electronically (e.g., on a hard disk) or otherwise (e.g., file cabinets). The strategies that we use to initially organize electronic documents may contribute to our ability to retrieve this information at a later time and with the least amount of error and frustration.

Existing psychological literature on information search and retrieval has been focused on examining how user characteristics (e.g., cognitive ability, search strategies) influence success or failure in the information search task (e.g., Stronge, Rogers, & Fisk, 2006; Vicente & Williges, 1988). The results have shown that there are many factors to consider when trying to predict success or failure. For example, spatial ability has been consistently shown to be an important ability predictor of user‟s ability to navigate information structures (e.g., Chen & Rada, 1996; Stanney & Salvendy, 1995; Pak, Rogers, & Fisk, 2006).

Another ability factor that has been shown to be important in information retrieval tasks is working memory capacity. Measures of working memory have frequently been found to be related to a wide range of technology-related tasks (programming, Davies, 1993; computer data entry, Czaja & Sharit, 1998; internet use, Morrell & Echt, 1996; telephone menu systems; Sharit, Czaja, Nair, & Lee, 2003; basic computer skills, Echt, Morrell, & Park, 1998). Indeed, the psychological processes required when interacting with different forms of technology (keep a primary task goal in mind as well as sub-goals while trying to remember previous information) seem consistent with the definition of working memory.

One limitation of previous studies may be that the participants searched through *hierarchically* organized systems. In hierarchical systems, information is organized into general categories that may have parent/child relationships with each other (e.g., the parent category of "files" may contain the sub-category of "project 1"). To efficiently search through hierarchically organized information one must be able to keep track of where they are in the hierarchy which requires adequate working memory and spatial ability (e.g., Pak, Rogers, & Fisk, 2006; Sharit, Czaja, Nair, & Lee, 2003).

One way that the system can potentially mitigate some of the cognitive ability demands of the task would be to re-organize the hierarchical structure of the system to include more *breadth* and less *depth* (Jacko & Salvendy, 1996; Zaphiris, 2001). An information system with more breadth and less depth would be less reliant on working memory capabilities and spatial abilities. Would people be better able to retrieve information if the organization of information was not hierarchical (thus rely less on working memory and spatial ability)?

One example of a non-hierarchical information organization system is tagging systems. In tagging systems, data are assigned descriptive labels, or tags, instead of being placed in discrete spatially organized locations (e.g., a document located in the subfolder "project" under the "Work" folder). In a tagging system, users would be presented with a list of labels as the search interface. If an e-mail was labeled as "work", "project A" and, "important" selecting any of these labels would call up that email (similar to the Google web-based e-mail service). Tagging systems may be less demanding of cognitive abilities such as working memory and spatial abilities because these systems are non-hierarchical and may not require the user to keep track of the spatial relationships between categories.

Existing literature on tagging has stressed the benefits of tagging in collaborative settings (e.g., Furnas et al. 2006; Golder & Huberman, 2006, Gordon-Murnane, 2006). In these settings, other people would share in the burden of applying descriptive tags to data. However, collaborative tagging is not practical for personal data where privacy is an issue such as with e-mails or documents. It may also be extremely effortful if a large quantity of information must be organized. Nonetheless, tagging systems, represent an alternative to taxonomical systems when users are organizing their personal documents—and one that may place less or different demands on abilities.

There is much enthusiasm about the purported benefits of tagging compared to more traditional ways of organizing information (e.g., Furnas et al., 2006; Sinha, 2005). However, this has not been supported by empirical evidence showing usability or performance benefits compared to traditional methods of information organization (Morville & Rosenfield, 2006).

Characteristics of Taxonomy & Tagging Systems

Organizing information into taxonomies or by labeling with tags could be considered strategies (ways of organizing information) or systems (imposed by the computer system). This distinction is not important for the current discussion and we will subsequently use the term "system" to mean both. Taxonomical and tagging systems are notable for the differences in cognitive operations required at organization (i.e., using the system to organize information) and retrieval (i.e., using the system to retrieve earlier information).

Taxonomy. Taxonomical systems involve organizing information hierarchically. Hierarchically organized menus are ubiquitous in modern computer systems (e.g., Windows Start menu). For example, in a computer drop-down menu, the options of "Save" and "Print" would be contained under the main menu "File" because saving and printing are operations that can be carried out on files. Many web sites also arrange their content in a hierarchical fashion. The vast majority of the information search tasks investigated in previous studies have involved the user searching through a hierarchically organized information system (for a review see Pak, Rogers, & Fisk, 2006).

When users are organizing information into a taxonomical system, they are creating a model of the information in an ad-hoc fashion. Taxonomically organized information is also similar to the cognitive constructs of schemas (Fiske & Taylor, 1991), and mental (situation) models (Morrow, Stein-Morrow, Leirer, Andrassy, & Kahn, 1997). Schemas are knowledge frameworks that contain information arranged hierarchically (Taylor & Crocker, 1981). Schemas play an important role in allowing us to organize and remember a large amount of information and interpret new information. Organizing information into these kinds of knowledge structures may enhance the ability to recall information at a later time.

The depth of the hierarchy, however, can introduce difficulties for information retrieval. Navigating particularly deep hierarchical information structures can also be cognitively demanding (e.g., Pak, Rogers, & Fisk, 2006). Users can get lost in the system (Edwards & Hardman, 1989; Vicente, Hayes, & Williges, 1987) leading to decreased navigational performance (Seagull & Walker, 1992). An alternative information retrieval interface is one that that has a "flatter" organizational structure (more breadth, less depth).

Tagging. Tagging involves the application of word labels to individual documents or pieces of information. There is no explicit organization such as in taxonomical systems, instead, the labels, or tags, applied to the files are used to succinctly describe the contents of the file. Tagging is a popular way to organize information online (e.g., "Flickr" for picture organization, "deli.cio.us" for internet book mark organization).

The purported benefits of tagging are thought to lie in the ease with which people are able to come up with "tags" or keywords to associate to bits of information (Sinha, 2005). This is assumed to be an easier process than trying to organize information into discrete categories because categorization involves a process of deciding the optimal category (selection of concepts and then decision) while tagging does not (selection only). Another benefit of tagging is the inherent social computing aspects (i.e., tagging online documents for others). However, as a first step, the current research is limited to examining individual-level tagging (i.e., tagging documents for self; but see Pitman & Payne, 2006).

The costs of tagging systems lie in the need to tag each piece of information. To come up with a descriptive label, one must deeply process or engage each item of information. This is contrasted with taxonomical systems where item-specific information is not usually generated; instead the goal of categorization is to determine how the to-be-organized item is similar to a generic category. This requirement to generate item-specific labels, versus generic categories, may lead to a higher level of workload during information organization, especially if the item is difficult to describe. Additionally, good verbal ability (i.e., accumulated knowledge) is required to be able to create descriptive tags. However, this deeper level of processing may prove beneficial for retrieving information. Extra effort due to the need to generate descriptive labels at the organization phase may produce a more durable memory trace for later retrieval (e.g., Slamecka & Graf, 1978).

From the perspective of information retrieval, tagging allows an object (e.g., a photograph or e-mail message) to receive multiple descriptive labels, each of which may be completely independent from each other in the sense that they labels may not be hierarchical. For example, an email message from Alice may be tagged with the words, “amusing”, “Alice”, and “read later”. The labels, while potentially relevant to the content of the message, cannot easily be organized in a hierarchical fashion. The benefit of the e-mail message being tagged with the labels is that at retrieval, the user can select any of the above tags and be taken to the appropriate e-mail message.

Compare this to taxonomical systems, where information can only exist in a single “location” (e.g., Manuscripts/CogTech/Study2)—that is, the desired information can only be found if the single correct category/sub-category is identified. In the taxonomical hierarchy, the concept of “Manuscripts” must be known beforehand, then the sub-category of “CogTech” and finally the last sub-category of “Study2”. If the user does not happen to know the initial folder or category, they may be unsuccessful at retrieving the desired e-mail message. Thus, tagging systems seem to have more independent “routes” to the desired information than taxonomical systems (which require the knowledge of the top-level category first).

The distinction between tagging and taxonomical organizational systems could be seen to exist on a continuum from structured organization (taxonomy) to unstructured organization (tagging). This distinction is similar to the strategies of personal information management observed by Boardman and Sasse (2004) between “frequent filers” (those who file each piece of information) and “no-filers” (those who do not use an organizational scheme).

To summarize, tagging systems have been embraced by different commercial products and there is some theoretical reason to hypothesize that tagging may help people organize and retrieve information. However, there is still little evidence to support the idea that tagging enhances information organization, and more important, improves retrieval (i.e., the ability of users to get the information they need) compared to more common hierarchical methods especially for information retrieval.

Cognitive Processes Involved in Taxonomy and Tagging Systems

In an attempt to delineate the differences between taxonomical and tagging systems, we first task analyzed the organization and retrieval of information under each type of system. Figure 1a and 2b (adapted from Sinha, 2005) illustrate the generic presumed cognitive processes involved in *organizing* information into taxonomical and tagging systems. The processes described in the diagrams may be generalizeable to different types of information (e.g., individual emails, internet bookmarks, and photographs). However, for clarity, the following example will use the organization of photographs into a taxonomical system. Figure 1a shows the processes involved in organizing taxonomically.

First (Step 1.0), when the individual is presented with a to-be-organized photograph many different concepts may come to mind (e.g., birthday party, vacation, the year the photograph was taken). In order to place the photograph in the appropriate category, the user must first identify the optimal category for the photograph given its main themes (Step 2.0; e.g., summer vacation). This step involves examining and selecting the most appropriate category. If an appropriate category does not exist, the user will generate a new category or new sub-category to place the photograph (Step 2.2; creating a new vacation album). If an appropriate category exists (e.g., an album or envelope for vacation photographs), the user will place the photograph into that category (Step 3.0). They would repeat this process for each photograph until all photographs are categorized.

When the user later wants to retrieve a specific photograph (e.g., the vacation photograph at the beach), he or she may go through the steps outlined in Figure 2a. In the taxonomic system (Figure 2a), the user would start by having a general conception of the photograph they want to retrieve (Step 1.0). First, the user will have to identify the appropriate top-level category (e.g., the main album containing vacation photographs; Step 2.0). Once the main category has been identified, the user will then subsequently "drill down" to more specific categories contained within the main category (e.g., summer vacation section within the vacation album; Step 2.1) until they reach the items within the category. Finally, they can then select the appropriate photograph (Step 3.0). A similar process for organization and retrieval in a tagging system can be followed in Figures 1b and 2b.

Based on our processing model, there is the potential for performance and perceived workload differences between the two organizational systems due to differences in the level of decision making, working memory, attention, or spatial ability requirements. Compare step 2 in the tagging and taxonomical diagrams (Figs. 1a, and 1b). In the taxonomical condition, step 2 involves analyzing the existing categories and selecting the best-fit category for the to-be-organized item. The user would have to decide if existing categories were sufficient to contain the to-be-organized item or whether a new category needs to be created which may be an effortful process.

However, because the tagging system requires the creation of a descriptive label or labels that are unique for the to-be-organized item (step 2), it may require more engagement with the material and thus incur higher workload than the taxonomical method which only requires a general category or sub-category. The critical task difference, we argue, may be that taxonomical systems only require a generic categorization (which may not need deep encoding) while tagging systems require generation of specific descriptor labels (which may require deeper encoding). This task difference should result in performance and workload differences at organization or retrieval.

The goal of the Experiment 1 was to examine potential performance and workload differences between two ways of organizing information. Currently, there exists no research that compares the costs and benefits of these two kinds of information organizational systems. The hypothesized increased engagement of material and subsequent self-generation of a descriptive tag in tagging could lead to a more durable memory trace which may enhance later ability to retrieve the photograph (e.g., Slamecka & Graf, 1978). While both groups of users must generate terms for categories or tags, the tagging group is expected to encounter a higher degree of self-generation because each item must be specifically described compared to a generic classification in taxonomy.

Thus, the question remains, how do the two systems vary in terms of performance and workload at the initial information organization stage? As noted, the vast majority of previous studies examining computer-based information search tasks have utilized a hierarchically organized system. Additionally, there is little empirical evidence to support the notion that tagging systems are more usable (i.e., enjoyable, intuitive, leads to better performance).

Experiment 1: Comparison of tagging and taxonomy for information organization

In this first study, we were primarily interested in potential performance differences at the *organization stage* (i.e., when users are filing information away). Based on our analysis of the cognitive processes during tagging or taxonomical organization, we expected that for information organization, tagging should result in higher perceived workload and increased time to organize a set of information compared to the taxonomical condition. These hypotheses come from the assumption that organizing information within a tagging system requires deeper engagement and elaboration with each photograph compared to organizing using a taxonomy.

Participants organized paper photographs using a simulated taxonomical or tagging system. The cognitive operations involved in our experiment were meant to closely replicate those hypothesized in our decision/action diagram. While most information organization and retrieval tasks may involve interacting with a computer, we felt that because the focus of our initial investigation was a broad assessment of potential differences in performance and workload between systems, a low-fidelity, paper-based study was appropriate.

Method

Participants

Thirty undergraduate students, 17 women and 13 men ranging in age from 18 to 25 years of age, completed the study; the mean age was 20.5 (SD = 2.0). The participants were recruited from psychology classes at Clemson University and reimbursed $7 for their participation. The participants were randomly assigned to an experimental group (described below). Data from three participants was excluded from further analysis because their time to organize the photographs was greater than two standard deviations above the group mean.

Materials

Photographs. Sixty royalty-free color images were obtained from a stock photography company. The predominant content of the images were of people, places or landscapes, and objects (twenty in each category). The photographs were selected at random within each topical category with the constraint that photographs resembled everyday photographs (compared to those that looked professionally composed). An additional 10 images were randomly selected to use as practice photographs. The images were professionally printed on photographic paper (glossy finish) at 4 inches in height and 6 inches in width.

Subjective workload assessment. Participant"s perceived workload was measured with a computerized version of the National Aeronautics and Space Administration Task Load Index (NASA TLX; Hart & Staveland, 1988). Participants estimated their level of workload along six dimensions of workload (mental, physical, temporal, performance demand, effort, and frustration). They used a mouse to indicate their responses on a visual analog scale.

Ability tests. Three standardized cognitive ability tests were used to characterize the participants in this study and to assure randomization of participants between conditions. The tests were the Digit Symbol Substitution (Wechsler, 1997; a measure of perceptual speed), the Shipley Vocabulary Test (Shipley, 1986; a measure of crystallized intelligence), and the Reverse Digit Span (Wechsler, 1997; a measure of working memory). There were no significant differences in any of the abilities between the two experimental groups.

Design and Procedure

The study was a two-group design with organizational system (tagging or taxonomy) as a between groups variable. The dependent variables were organizational time and subjective workload. Upon arriving at the laboratory, participants read and signed an experimental consent form. The participants were familiarized with one of two methods to be used to organize photographs during the experiment.

Participants assigned to the taxonomic condition were told to organize the photographs into main categories. The participants were instructed to write the name of their categories on Post-it notes. They were not limited in the amount of categories they could use, and the categories could be arranged hierarchically such that categories could exist within other categories (e.g., main category of „vacation" could have a sub-category of „summer"). Hierarchical categories were created by using paper folders.

Participants assigned to the tagging condition were told to organize the photographs by assigning an unlimited number of descriptive keywords, or tags to each photograph. The participants were instructed to write down tags relating to each photograph on a Post-it note and attach it to each photograph. Both groups were told to organize the photographs as quickly as possible.

The participants were given 10 photographs to practice organizing using their assigned method. After completing the practice task, each participant organized 60 photographs using their assigned method. The total time to complete the organization task was recorded by the experimenter. After completing the organization task, the participants completed the TLX workload questionnaire. To ensure that participants understood each workload component, they were given an instructional handout describing each workload component. Finally, participants completed the three ability tests.

Results

Mean organizational time and workload components are presented in Table 1. There were no group differences in cognitive abilities (working memory, vocabulary, and perceptual speed). However, participants in the taxonomy condition took significantly longer (about 775 seconds) to organize the photographs than participants in the tagging condition (about 450 seconds), $F(1,26) = 12.3$. $p < .05$. This finding was surprising because it was expected that the task requirements for the tagging condition would be greater than the taxonomy condition. The increased time required to organize photographs into a taxonomical system may be explained by the informal observation that participants in the taxonomy condition seemed more likely to *re-organize* photographs into different categories than participants in the tagging condition (e.g., moving a photograph from one category into another).

While total workload was not significantly different between groups, level of frustration, a component of workload, did show significant group differences, $F(1,26) = 4.2$, $p < .05$, with the tagging group reporting more frustration than the taxonomy group. For this component of workload, participants were asked, "how insecure, discouraged, irritated, stressed and annoyed" they felt during the task (Hart & Staveland, 1988). No other component of workload showed significant group differences. This increased frustration in the tagging condition may come from the need to create a descriptive tag for each photograph, a task step that may be verbal ability demanding. However, it was surprising that this increased frustration in tagging condition was not associated with a concomitant increase in organizational time compared to taxonomy.

Discussion

The goal of Experiment 1 was an empirical evaluation of performance and workload differences between two systems of organizing information (taxonomical and tagging). Using a paper-based sorting technique, the study focused on the initial organization of information (compared to retrieval of information). The results suggest that while taxonomical methods result in longer organization times, tagging led to a greater sense of subjective frustration. The source of the frustration may come from the need to engage each of the 60 photographs and produce descriptive tags.

A major limitation of Experiment 1, however, was the paper-based implementation. While the sorting of paper photographs was presumed to tap the same cognitive processes involved in organizing electronic information, there were sizeable, unanticipated differences in task strategies between conditions. For example, several participants in the taxonomy condition tended to organize their photographs into piles, or unlabeled categories, and then give the piles categorical labels. They essentially broke the task down into two steps (initial gross organization, then fine-tuning). However, participants in the tagging condition, by the nature of the task, were forced to label each photograph one-by-one. These participants tended not to revisit their previous tags for fine-tuning. While the taxonomic strategy of organizing photographs observed in some of our participants (two-step: gross organization and then fine-tuning) was appropriate for paper photographs, it may not be representative of the way in which electronic information is organized.

Experiment 2 was designed to address some of the methodological limitations of Experiment 1 and to assess the potential organizational *and* retrieval differences between the tagging and taxonomy systems. Would there still be performance (i.e., time) differences at organization and retrieval in a more controlled and representative task?

The results of Study 1 suggest potential hypotheses regarding retrieval differences between systems. Organizing information using a taxonomic system took longer than using a tagging system. The longer time to organize using the taxonomic system could result in faster retrieval of information (mediated by a more durable memory trace) compared to the tagging method. The tagging method which requires more self-generation from the user (compared to self-generated categories in taxonomical systems) may also result in a better memory trace (mediated by deeper engagement of the photographs). However, the increased level of workload (driven by frustration) in the tagging condition could translate to a reduced level of performance at retrieval.

Experiment 2: Comparison of tagging and taxonomy for information retrieval

The goal of this experiment was to further investigate the performance and workload differences between taxonomical and tagging information interfaces. More specifically, we were interested in understanding whether there would be information retrieval differences when users organized information using either a taxonomical or tagging interface. Our hypotheses were generated from our process model of each system (Figures 1 and 2).

Experiment 1 showed that the two conditions varied significantly on the amount of time it took participants to organize information, and in terms of subjective frustration at the time of organization but it could not answer the question of whether one kind of system was better in terms of retrieval performance. Would the extra time needed under taxonomic conditions produce better retrieval performance compared to tagging? What would be the effect on workload at retrieval? Additionally, strategic differences in task procedure between groups that may have contributed to differences in the results in Experiment 1 were controlled in Experiment 2.

Similar to Experiment 1, the primary dependent variables were organizational time (how long it took participants to organize information), measures of workload (at organization and retrieval) and retrieval performance (retrieval time, errors). Different from Experiment 1, this study examined retrieval performance using a computer-based task instead of a low-fidelity paper-based task. In this study, participants organized 60 photographs using a computer interface based on either a taxonomy or tagging. After a delay, participants" retrieval performance of a subset of the previously organized photographs was examined.

Method

Participants

Forty five participants, 28 women and 17 men ranging in age from 18 to 34 participated in the second experiment; mean age was 19.4 (SD = 2.6). Participants were undergraduate students who received either course credit or $7. Twenty one participants were assigned to the taxonomy group (11 women, 10 men) while 24 participants were assigned to the tagging group (17 women, 7 men). All participants were intermediate computer users; when asked about the frequency of computer usage (from a scale of 1 indicating “once every few months”, to 7 indicating “daily, most of the day”) the mean response was 5.9 (SD = .71).

Four participants were excluded from further analysis because their organization or retrieval times were two standard deviations greater than the condition means. There were no significant group differences in working memory, verbal ability, or computer experience. However, there was a significant difference in perceptual speed (digit-symbol substitution) between groups, $F(1,40) = 4.0$, $p < .05$. Participants in the tagging condition had higher perceptual speed than the participants in the taxonomy condition. Thus, perceptual speed was statistically controlled in subsequent analyses.

Materials

Photographs. The same set of 60 photographs used in Experiment 1 was used in this study. The photographs were all sized to 400 pixels in height and 700 pixels in width. Throughout the study, when thumbnail images were necessary (i.e., when listing the contents of a particular category) the images were 80 pixels in height and 140 pixels in width.

Equipment. IBM-compatible computers running at 3.2. GHz with 2 GB of RAM were used for the current study. The monitor size was 19 inches diagonally with the resolution set at 1024 pixels in width by 768 pixels in height and a color depth of 32 bits. Participants were seated approximately 18 inches from the monitor screen. The mouse was placed on whichever side the participants chose based on the self report of their dominant hand.

Photograph organization and retrieval task. A web application was created to allow participants to organize and retrieve photographs into hierarchical categories (taxonomy condition) or by applying descriptive labels (tagging condition) using the mouse and keyboard. The application displayed task instructions and collected user responses. The application was presented in the Firefox web browser with all toolbars and status bars hidden from the user.

The web interface consisted of three main task areas (Figure 3a). In the top row, thumbnails of the to-be-organized photographs were displayed horizontally in random order. When the participant selected a thumbnail from the list, the row was replaced with the thumbnail of the selected photograph with an input text box (to enter categories or labels depending on condition) and an enlarged photograph in the center portion of the screen (Figure 3b). On the left side of the screen, participants saw their currently used categories or labels. The application recorded our performance-related dependent variables (e.g., times, number of mouse clicks) at organization and retrieval.

Workload, abilities, and computer experience assessment. The workload and ability measures were identical to Experiment 1. In addition, participants completed a technology and computer experience questionnaire (Czaja, Charness, Fisk, Rogers, & Sharit, 2001).

Design and Procedure

The experiment was a two-group design with organizational system as a between groups variable (tagging or taxonomy). Unlike the first experiment (which only assessed the initial organizational time and workload using a paper-based implementation) Experiment 2 examined both workload and performance at the *organization* and *retrieval* stages using a computerized task. The use of a computerized task facilitated recording of the dependent variables at the retrieval stage (which was impossible with the paper-based implementation of Experiment 1).

Upon arriving at the laboratory, participants (up to three at a time) read and signed an experimental consent form. Participants then started the web-based portion of the study. Participants assigned to the taxonomic condition were told to organize the photographs into categories which could be arranged hierarchically such that categories could exist within other categories (e.g., main category of „vacation" could have a sub-category of „summer"). Participants, however, were limited to three levels-deep of categories (e.g., vacation / summer / family). The decision to use three levels as a maximum was informed by pilot testing which showed that most users (across conditions) rarely used more than 3 levels to categorize a photograph.

Participants in the tagging condition also organized the photographs but were told to assign descriptive labels to each photograph (e.g., vacation, summer, family). In both groups, as participants assigned categories or labels, the hierarchical structure of categories (Figure 3b) or an alphabetized list of labels (Figure 4b) was updated on the left side of the screen. Participants in the tagging condition saw an alphabetically organized list of labels currently used while taxonomy participants saw hierarchically arranged categories. If participants made a spelling error or mis-categorization, they could click on the left-side label or category to view the contents of that category or label and reassign the photographs.

Before starting the organizational task, participants in both conditions were given a guided practice session with one photograph. Both groups of participants were given verbal instructions to work as quickly but accurately as possible. After answering any questions about the procedure, participants organized the 60 experimental photographs using the mouse and keyboard. After the photograph organization task, participants immediately completed the NASA TLX to assess perceived workload. After completing the TLX, participants completed three ability tests and a technology and computer experience questionnaire. The abilities test portion was experimenter-guided in group format and took approximately 25 to 30 minutes. In addition to collecting information about abilities and previous computer experience, this portion of the study was used as a distractor period between the organization and retrieval of photographs.

Once the ability tests and computer experience questionnaire were completed, participants returned to their computers and completed the photograph retrieval portion of the study. For the retrieval portion, 40 random photographs, from the 60 organized earlier, were selected for retrieval. The 40 retrieval photographs were different for each participant and were selected by the program at run-time. Participants saw a randomly selected thumbnail photograph in the top portion of the screen with the words, “locate this image” adjacent. This image remained at the top of the screen until participants correctly located the image among their categories or tags. Participants were told to search for the photograph using their previously created organizational scheme (Figures 5a, 6a) as quickly and accurately as possible using the mouse. Searching for the photograph involved clicking on the links (which represented categories or labels) on the left side of the screen.

For example, in the taxonomic condition, on the left side of the screen, participants saw their earlier categories of images. If participants clicked on the “urban” category, they saw thumbnails of all of the photographs that they had earlier categorized as “urban” in the content viewing portion of the screen (center area). If there were sub-categories under “urban”, they saw a folder icon among the thumbnails indicating further sub-categories of images. Clicking on the folder icon displayed the images within that sub-category.

In the tagging condition, participants saw, on the left side of the screen, an alphabetized list of labels they had used in the organizational portion of the study. When participants clicked a label on the left side of the screen, they saw all of the photographs possessing that label (e.g., all photographs labeled "nature"). If an image was categorized as "nature" and "fall" clicking either label would bring up that image in the content viewing portion of the screen.

As participants searched through their categories or labels, their time to find the photograph was recorded, as well as errors (clicking on the wrong photograph) and number of total mouse clicks. If participants clicked on the wrong photograph, they saw a message indicating that they had clicked the wrong photograph and to keep searching. Once participants found the to-be retrieved photograph in the central content area, they clicked the image with the mouse ending the retrieval trial. After receiving feedback that the correct image was located, the next photograph was presented. After successfully retrieving all 40 photographs, participants immediately completed the TLX to assess their perceived level of workload for the retrieval portion of the study.

Results

Mean performance and workload components are presented in Table 2. Organizational time was defined as the total amount of time participants spent on organizing the 60 photographs (either into categories or by applying labels). Retrieval time was measured as the amount of time participants spent retrieving the 40 previously viewed images by navigating their categorical scheme or list of labels and clicking the correct photograph. For the measurement of organizational and retrieval time, inter-trial time was excluded (i.e., time spent between trials).

Another dependent measure was the total number of clicks of the mouse involved in either phase (organization or retrieval). A data point was recorded every time participants clicked on any active area of the interface (e.g., a thumbnail image, a category name or label) during a trial. Finally, our measure of user error was the number of times in which participants clicked the wrong thumbnail image at retrieval. Error rate was extremely low (less than 1 error per participant over the 40 retrieval trials) and was not further analyzed.

Because of the significant perceptual speed difference between our two conditions, score on the digit-symbol substitution test was used as a covariate in these analyses (Table 2 illustrates the unadjusted means). Task completion times and workload components were analyzed with a 2 (condition: tagging, taxonomy) × 2 (phase: organization, retrieval) repeated measures analysis of variance (ANOVA) with condition as a between-groups variable and phase as a within-subjects variable. There was a significant interaction between condition and phase meaning that condition had different effects at the information organization stage and the retrieval phase ($p < .05$). The source of this interaction was a significant difference in *organizational* time between conditions but no significant difference in *retrieval* time (see Figure 7) [1]. Participants in the taxonomy condition were faster at organizing the photographs (973 s) compared to participants in the tagging condition (1132 s). This translates to a 3 second advantage for taxonomy participants in terms of organization time per photograph. The taxonomy advantage at organization is opposite of the results from Experiment 1 (which found a time advantage for tagging).

A longer organizational time for the tagging group in Experiment 2 could be explained by arguing that the tagging group used more labels to describe each photograph than the taxonomy group‟s categories. However, there were no significant differences in the number of labels or levels of categories used to describe each picture by condition (1.8 labels used in tagging, 1.7 categories used in taxonomy). In addition there were no significant differences in the total number of characters used in the tags compared to categories ($p > .05$).

For the number of mouse clicks required to complete the task, there was no significant difference at organization. However, at retrieval the participants in the taxonomy condition required significantly more mouse clicks (91) to complete the task than participants in the tagging condition (84). At retrieval, participants had to search through their organizational scheme to find a specified photograph (40 total). In this task, the smallest possible number of clicks required to retrieve an image was 80 clicks (one click to select the tag or category and another click to identify the image). The small but significant difference in mouse clicks at retrieval favoring the tagging system seems to indicate that searching through the taxonomy was potentially more error prone (i.e., lead to more clicks on the wrong category). From the

[1] When the perceptual speed measure was removed as a covariate, the significant difference between conditions on organizational time was not significant. All other results were unchanged.

perspective of depth versus breadth in menu design, these results support previous research showing an advantage of breadth over depth in terms of performance (e.g., Parkinson, Hill, Sisson, & Viera, 1988).

The only significant workload difference between conditions was in the temporal demand component. For temporal demand, participants rated the extent to which they felt any time pressure to complete the task. Participants in the tagging group reported significantly higher temporal demand (9.0) compared to participants in the taxonomy group (5.6) but only at the organization phase. This difference is puzzling because, except for the specific tagging/taxonomy difference, the two conditions were nearly identical in terms of task procedure (cf. Figure 3b, 4b). One explanation of the increased temporal demand for tagging may be that participants in the tagging condition felt more time pressure because they were required to generate *specific* tags that were descriptive of the photograph. Participants in the taxonomy condition instead only had to create general categories for each photograph. This difference in presumed processes might be expressed as increased temporal demand (as participants try to think of specific tags).

General Discussion

The goal of Experiment 2 was to more closely examine how tagging and taxonomical systems might differ in terms of performance and workload. The results of Experiment 2 showed that tagging resulted in longer photograph organization time than taxonomy. Interpreted within our model of the processes involved in tagging (Fig. 2), it could be assumed that the source of the increased tagging time at organization was due to the difficulty of creating specific, descriptive labels (Fig. 1b, step 2) for each photograph. The other steps in the process of organizing using a tagging or taxonomic system are shared between the two organizational schemes (e.g., step 1, and step 3) and can be assumed to be the same (i.e., require similar times).

Temporal demand differences at organization are somewhat surprising given that both groups were given the exact same instructions. Both groups were told to emphasize speed and accuracy. It may have been the case that tagging participants engaged in a speed/accuracy tradeoff wherein they sacrificed accuracy for speed. However, the equivalence of the number of mouse clicks at retrieval between groups suggests that this was not the case (i.e., more mouse clicks could indicate more error). While the frustration component of workload was not significantly different in Experiment 2 (like Experiment 1), there was a non-significant trend for higher frustration ratings in the tagging group compared to the taxonomy group at organization and retrieval (Table 2).

At retrieval, participants in the taxonomy condition required more mouse clicks to properly locate and identify a photograph. This effect could have two possible sources. The first is that at the organization stage, each method resulted in a different level of memory encoding (i.e., tagging resulted in better memory which resulted in fewer clicks required at retrieval). Another possibility is that the differences in clicks required at retrieval were due to the differences in the retrieval interfaces. This possibility is not surprising given previous research that has shown performance differences in hierarchical menu search. Recall that one major difference in tagging and taxonomic information retrieval interfaces is that taxonomic interfaces show hierarchical structure (i.e., have depth) while tagging interfaces have no hierarchical structure (i.e., are flat). The need to navigate the depth of the hierarchical system may have contributed to the need to have more mouse clicks.

One limitation of the current study was the relatively short distractor period between organization and retrieval (25-30 minutes). This short interval may not have allowed enough time for the presumed benefits of elaborative encoding engendered by tagging to manifest itself.

Another limitation of the current studies was the use of photographic stimuli. There has been some research that has shown that memory for images, particularly detailed photographic material is much better than memory for words (e.g., Park, Puglisi, & Sovacool, 1983). This may explain the generally high level of performance (low errors, fast performance) in our retrieval condition of Experiment 2. Because the current study did not manipulate the material to be organized, it is unknown whether the observed effects were specifically due to the photographic stimuli, or whether these effects would vary with textual material (e.g., e-mails, web pages).

Taken together, the results of the studies show some performance and workload differences in using tagging or taxonomical information organization. Discrepant findings between Experiment 1 and 2 may be explained by methodological differences. Experiment 1 was a broad assessment of major differences between organizational methods using a paper-based implementation. It was thought that this type of study would be sufficient to determine time and workload differences. However, it was not anticipated that specific organizational strategies would differ from electronic methods of organization. Paper-based organization induced some participants to first create unnamed piles of photographs which were labeled and refined later. However, electronic methods of organization were more linear, with each photograph being organized in-turn; with little re-organization afterwards. Experiment 2 showed that tagging may lead to longer times to organize information and potentially more workload. The source of the increased time to organize is presumed to come from the need to generate descriptive tags for each to-be-organized item.

Conclusion

It is hoped that the studies just presented shed some much needed light on the usability differences between two ways of organizing information electronically. The results suggest that if performance at organizing information is an important criterion, taxonomical systems are preferred because they lead to faster information organization. However, at retrieval, navigating taxonomical systems requires more mouse clicks than navigating a tagging system. There are also workload differences such that tagging systems impose a higher level of workload (induced by temporal demand) from the need to generate descriptive tags. In addition, many of the predicted trends were present but not significant (e.g., higher frustration and total workload in the tagging condition).

It is important to note that the current study examined individual-level performance (i.e., organizing information for oneself). Under these conditions, our results suggest that tagging may not be an efficient system of information organization. However, when the burden of information organization is shared (i.e., social tagging) the results may show an advantage for tagging. Again, examining our process model, it was presumed that the most demanding stage of organizing information was the stage where users generate descriptive labels. In social tagging situations, this burden is distributed among many users. Alleviating this burden would have obvious implications for perceived workload (e.g., workload would be lower).

However, the extent to which tags chosen by others would be meaningful to oneself and how it would affect performance and workload at retrieval is unknown (however, see Pitman & Payne, 2006). Previous research has shown in a variety of domains, that the probability that two people will choose the same label to describe a common object is low (Furnas, Landauer, Gomez, & Dumais, 1984). Another method of alleviating the burden of generating tags is to have it automated. However, automatic generation of tags might mitigate any enhanced memory benefit (from elaborative encoding) and thus interfere with retrieval.

Currently, we are examining ability/performance relationships between tagging and taxonomies. For example, previous research has shown that individual and age-related differences in spatial abilities are an important component in the ability to navigate a hierarchically organized menu structure (not unlike the taxonomical condition). The relationship between spatial abilities and hierarchical navigation is thought to come from the need to create and manipulate a mental model, or abstract representation of the computer menu system (e.g., Sein & Bostrom, 1989). What would be the spatial ability/performance relationship in a tagging environment? Tagging may not require users to create a spatially organized model of the system.

Similarly, how do individual differences in other abilities contribute to a preference or performance difference between these two systems? For example, according to our model when users are presented with the to-be-organized information, multiple concepts may be activated within the user. Individual differences in vocabulary may limit the quality and quantity of descriptive tags that a user is able to produce. An inability to easily produce descriptive tags may result in frustration, increased workload, and reduced retrieval performance. Verbal ability was measured in the current studies but the use of a college-aged sample (with little variance in vocabulary test scores) precluded an analysis of the influence of verbal ability on performance.

Finally, we are interested in the performance differences between older and younger users when using taxonomical or tagging systems. Specific age-related changes in cognitive abilities and memory could suggest potential hypotheses. For example, in general, it has been shown that fluid abilities (e.g., spatial abilities) show significant age-related declines while crystallized intelligence (e.g., accumulated knowledge) does not (e.g., Schaie, 1996). It may be the case that tagging, because it presumably relies on verbal ability (which, as a type of crystallized intelligence is resistant to age-related decline), may show smaller age-related differences in performance. Additionally, older adults, partly due to reduced spatial abilities, have more difficulty recovering from navigational errors in hierarchical systems (Mead, Batsakes, Fisk, & Mykityshyn, 1999; Lin, 2003; Zaphiris, 2001). The use of a flat tagging interface may mitigate age-related differences in information search performance.

Finally, the social implications of tagging and taxonomical systems, however, cannot be ignored. Having a group of people categorize information together would dramatically reduce the initial workload of having to organize information for oneself. However, there are several limitations that may mitigate the usefulness of social tagging (or taxonomy). First, people may not share similar labels or categories for similar objects. Conversely, different users who work in very different domains may use similar words with very different meanings (e.g., homonyms). These issues can be somewhat controlled through the use of constrained tagging vocabularies (limiting tags to a pre-defined set of words). Second, social tagging is only useful if the information to be organized is not private. Users would be less inclined to have others organize private information (e.g., e-mail messages).

To conclude, the current research attempted to formalize the implicit theories underlying information organizational systems. Why do they work? Is one better than the other? The results of our initial analyses seem to indicate that there are some performance and workload differences between the two systems of information organization and retrieval.

References

Boardman, R., & Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": A cross-tool study of personal information management. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems, 583-590*.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human-Computer Interaction, 11,* 125-156.

Czaja, S. J., & Sharit, J. (1998). Ability-performance relationships as a function of age and task experience for a data entry task. *Journal of Experimental Psychology: Applied, 4,* 332-351.

Czaja, S. J., Charness, N., Fisk, A. D., Rogers, W., & Sharit, J. (2001). The center for research and education on aging and technology enhancement: A program for enhancing technology for older adults. *Gerontechnology, 1,* 50-59.

Davies, S. P. (1993). The structure and content of programming knowledge: Disentangling training and language effects in theories of skill development. *International Journal of Human-Computer Interaction, 5*, 325-346.

Echt, K. V., Morrell, R.W., & Park, D. C. (1998). Effects of age and training formats on basic computer skill acquisition in older adults. *Educational Gerontology, 24*, 3-23.

Edwards, D. M., & Hardman, L. (1989). Lost in hyperspace: Cognitive mapping and navigation in a hypertext environment. In R. McAleese (Ed.), *Hypertext: Theory into practice* (pp. 105-125). Oxford, UK: Intellect Books.

Feldman, S. (2004). *The high cost of not finding information*. Retrieved August 3, 2005. from the World Wide Web: http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9534

Fiske, S. T. & Taylor, S. E. (1991). *Social Cognition*, 2nd ed. New York: McGraw-Hill.

Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., and Naaman, M. 2006. Why do tagging systems work? In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM Press, New York, NY.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1984). Statistical semantics: Analysis of the potential performance of keyword information systems. In J. C. Thomas and M. L. Schneider (Eds.), *Human Factors in Computer Systems*. Norwood: Ablex.

Golder, S., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32,* 198-208.

Gordon-Murnane, L. (2006). Social bookmarking, folksonomies, and web 2.0 tools. *Searcher, 14,* 26-38.

Hart, S. G., & Staveland, L. E. (1998). *Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research*. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload. Amsterdam: North-Holland.

Iden, R., Mook, A. C., & Pak, R. (2006). Performance-based evaluation of two photo-organizing interfaces. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 2311-2315.

Jacko, J., & Salvendy, G. (1996). Hierarchical menu design: Breadth, depth, and task complexity. *Perceptual and Motor Skills, 82,* 1187-1201.

Lin, D.-Y. M. (2003). Age differences in the performance of hypertext perusal as a function of text topology. *Behaviour & information technology, 22*, 219-226.

Mead, S. E., Batsakes, P. J., Fisk, A. D., & Mykityshyn, A. (1999). Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development, 23*, 553-573.

Morrell, R. W., & Echt, K. V. (1996). Instructional design for older computer users: The influence of cognitive factors. In W. A. Rogers, A. D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and applications*, (pp. 241-265). Hillsdale: Erlbaum.

Morrow, D. O., Stine-Morrow, E. A. L, Leirer, V. O., Andrassy, J. M., Kahn, J. (1997). The role of reader age and focus of attention in creating situation models from narratives. *Journal of Gerontology: Psychological Sciences, 52B*, 73-80.

Morville, P., & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web*. Sebastopol: O'Reilly.

Pak, R., Rogers, W. A., & Fisk, A. D. (2006). Spatial ability sub-factors and their influence on an information search task. *Human Factors, 48*, 154-165.

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial locations in older adults: evidence for a pictorial superiority. *Journal of Gerontology, 38*, 582-588.

Parkinson, S. R., Hill, M. D., Sisson, N, & Viera, C. (1988). Effects of breadth, depth and number of responses on computer menu search performance. *International Journal of Man-Machine Studies, 28*, 683-692.

Pew Internet and American Life Project. (2006). Internet penetration and impact. Retrieved September 22, 2006, from the World Wide Web: http://www.pewinternet.org/PPF/r/182/report_display.asp

Pitman, J. A., & Payne, S. J. (2006). Creating names for retrieval by self and others. *Behaviour & information technology, 25*, 489-496.

Schaie, K. W. (1996). Intellectual development in adulthood: *The Seattle longitudinal study*. New York: Cambridge University Press.

Seagull, F. J., & Walker, N. (1992). The effects of hierarchical structure and visualization ability on computerized information retrieval. *International Journal of Human-Computer Interaction, 4*, 369-385.

Sein, M. K., & Bostrom, R. P. (1989). Individual differences and conceptual models in training novice users. *Human-Computer Interaction, 4*, 197-229.

Sharit, J., Czaja, S. J., Nair, S., & Lee, C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors, 45*, 234-251.

Shipley, W. (1986). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Press.

Sinha, R. (2005). A cognitive analysis of tagging. Retrieved August 5, 2007, from the World Wide Web: http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of experimental psychology: Human learning and memory, 4*, 592-604.

Stanney, K. M., & Salvendy, G. (1995). Information visualization; assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics, 38,* 1184-1198.

Stronge, A. J., Rogers, W. A., & Fisk, A. D. (2006). Web-based information search and retrieval: Effects of strategy use and age on search success. Human *Factors, 48*, 434-446.

Taylor, S. E. & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, M. P. Zanna (Eds.), *Social Cognition: The Ontario Symposium,* (pp. 89-134). Hillsdale: Erlbaum.

Vicente, K. J., & Williges, R. C. (1988). Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies, 29,* 647-668.

Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical files system. *Human Factors, 29*, 349-359.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd Ed.). San Antonio, TX: The Psychological Corporation.

Zaphiris, P. (2001). Age Differences and the Depth-Breadth Tradeoff in Hierarchical Online Information Systems. In C. Stephanidis (Ed.), *Universal Access in HCI*, pp. 540-544. Hillsdale, NJ: Erlbaum.

Author Note

Richard Pak & Steven Pautz, Department of Psychology, Clemson University; Rebecca Iden, Department of Industrial Engineering, Clemson University.

Experiment 1 was partially supported by the National Science Foundation and the Department of Defense ASSURE Program under Grant No. SES-0353698. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the Department of Defense. Special thanks to Abigail C. Mook for her help in data collection in Experiment 1. Thanks to Chuck Robertson and Lee Gugerty for helpful discussions concerning the study design.

Portions of Experiment 1 were presented at the Human Factors and Ergonomics Society 50th Annual Meeting in San Francisco, CA (October 2006) with proceedings (Iden, Mook, & Pak, 2006).

Correspondence concerning this article should be addressed to Richard Pak, Department of Psychology, Clemson University, Clemson, SC 29634. E-mail: richpak@clemson.edu

Table 1.

Summary of performance and workload measures from Experiment 1 by condition (taxonomy, tagging).

	Taxonomy (N=13)		Tagging (N=14)		
	Mean	SD	Mean	SD	*F*
	Abilities				
Perceptual speed[a1]	71.0	9.1	71.7	12.6	0.0
Working memory[b1]	10.5	2.4	9.4	2.4	1.4
Verbal ability[c1]	30.8	3.3	30.9	4.9	0.0
	Organization of Photographs				
Performance					
Time (s)	775.6	70.6	450.5	60.7	12.3*
Workload					
Mental demand	3.9	0.6	7.5	2.0	2.7
Physical demand	0.7	0.6	0.0	0.0	1.6
Temporal demand	13.0	2.6	20.5	8.2	0.7
Performance	5.3	1.2	7.2	2.1	0.6
Effort	5.3	1.4	4.9	1.3	0.1
Frustration	1.6	0.5	6.1	2.1	4.2*
Total workload[d]	29.9	3.5	36.6	4.6	1.3

Note. * indicates significant between-group difference, $p < .05$. [a]Digit symbol substitution (number correct; Wechsler, 1997), [b]reverse digit span (Wechsler, 1997), [c]Shipley vocabulary test (Shipley, 1986), [d]The total workload composite was the sum of all six workload sub-scales. The range of possible total workload was 0 to 100 with higher values indicating more subjective workload. [1]Higher is better

Table 2.

Summary of performance and workload measures from Experiment 2 by condition (taxonomy, tagging) and phase (organization, retrieval).

	Taxonomy (N=20)		Tagging (N=21)		
	Mean	SD	Mean	SD	*F*
Abilities and Technology Experience					
Perceptual speed[a1]	63.3	12.5	70.4	10.5	4.0*
Working memory[b1]	9.1	2.6	9.2	2.5	0.2
Verbal ability[c1]	27.0	5.9	29.6	3.8	3.0
Technology experience[d1]	5.9	0.9	6.1	0.4	0.9
Organization Phase					
Performance measures					
Time to organize (s)	973.2	379.5	1132.4	281.7	5.9*
Mouse clicks	138.7	34.2	143.9	37.8	0.0
Number of labels or levels used	1.8	0.9	1.7	0.6	0.5
Number of characters typed	509.7	493.8	638.8	405.5	0.8
Workload measures					
Mental demand	14.6	1.6	15.4	2.0	0.3
Physical demand	2.5	0.7	0.9	0.4	2.4
Temporal demand	5.6	1.1	9.0	1.4	5.3*
Performance	10.0	1.8	8.6	1.1	0.6
Effort	10.2	1.5	10.4	1.1	0.0
Frustration	2.5	1.2	6.4	1.7	1.4
Total workload[d]	45.3	3.5	50.6	3.1	1.0
Retrieval Phase					
Performance measures					
Time to retrieve (s)	423.8	169.7	378.0	122.3	0.0
Mouse clicks	91.4	11.9	84.1	4.9	6.5*
Workload measures					
Mental demand	14.1	1.8	13.4	2.0	0.1
Physical demand	3.5	1.2	2.2	0.7	0.4
Temporal demand	9.2	2.1	8.2	1.7	0.3
Performance	6.9	1.7	10.0	2.0	0.7
Effort	9.0	1.5	8.5	1.4	0.1
Frustration	1.5	0.8	3.4	1.2	1.0
Total workload[d]	44.1	4.5	45.7	3.7	0.0

Note. * indicates significant between-group difference, $p < .05$. [a]Digit symbol substitution (number correct; Wechsler, 1997), [b]reverse digit span (Wechsler, 1997), [c]Shipley vocabulary test (Shipley, 1986), [d]Technology and computer experience, question 6 (Czaja, Charness, Fisk, Rogers, & Sharit, 2001). [d]The total workload composite was the sum of all six workload sub-scales. The range of possible total workload was 0 to 100 with higher values indicating more subjective workload. [1]Higher is better

Figure 1a. Processes involved in *organizing* information in a taxonomic system (adapted from Sinha, 2005).

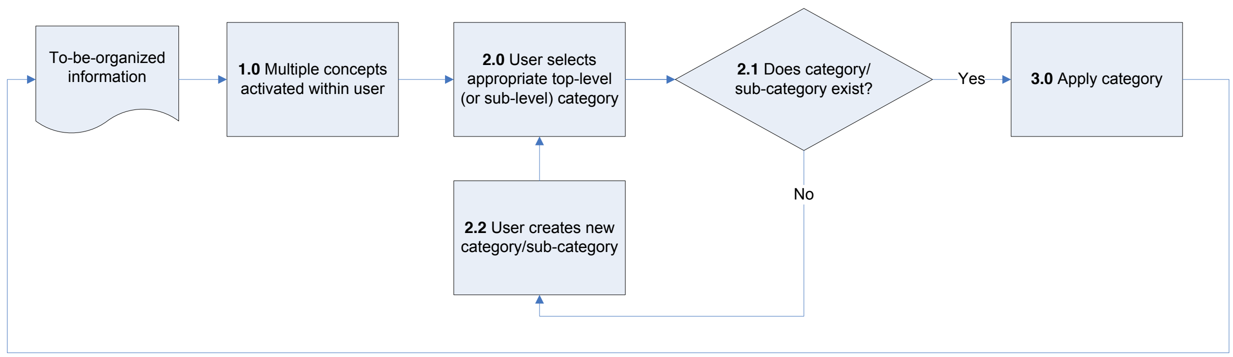

Figure 1b. Processes involved in *organizing* information in a tagging system (adapted from Sinha, 2005).

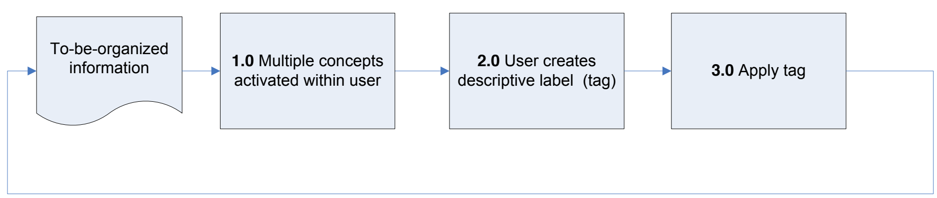

Figure 2a. Processes involved in *retrieving* information from a taxonomic system.

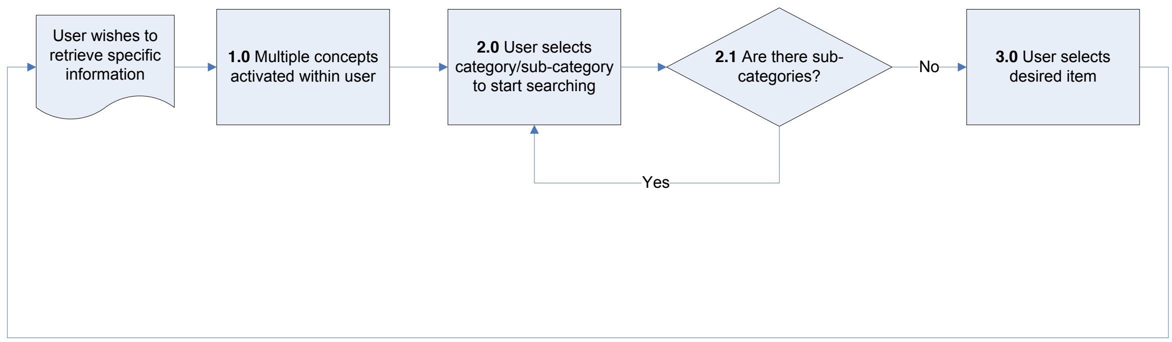

Figure 2b. Processes involved in *retrieving* information from a tagging system.

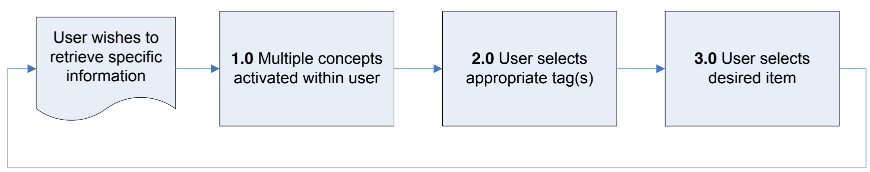

Figure 3a. Screen capture of the taxonomy condition start page (organization phase). When the participant clicked on a thumbnail image, they saw the categorization interface (Fig. 3b).

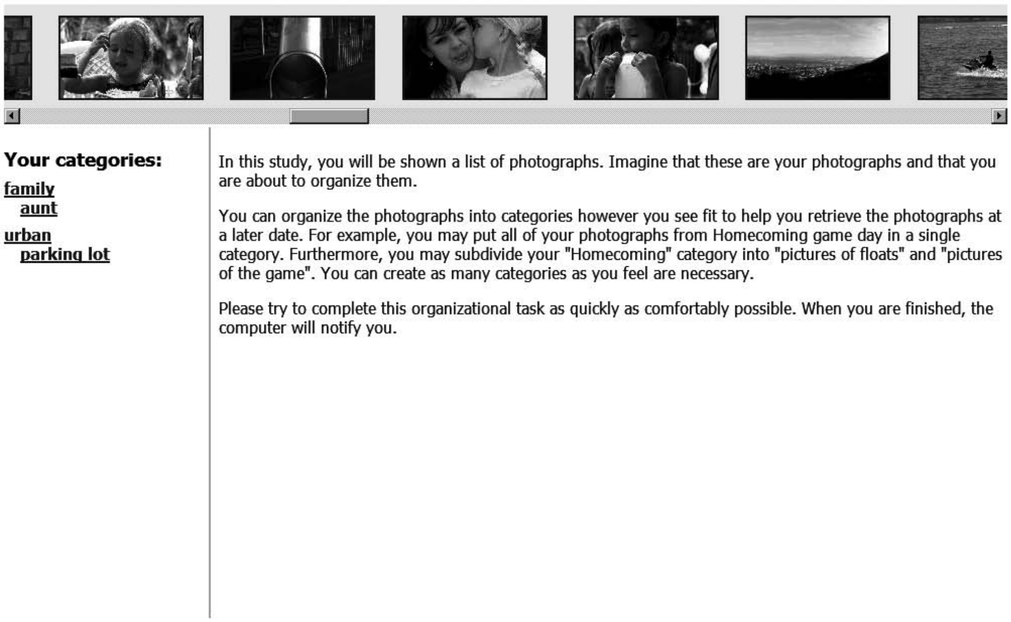

Figure 3b. Screen capture of applying category and subcategories to photograph in taxonomy condition (organization phase). Participants used categories appeared on the left.

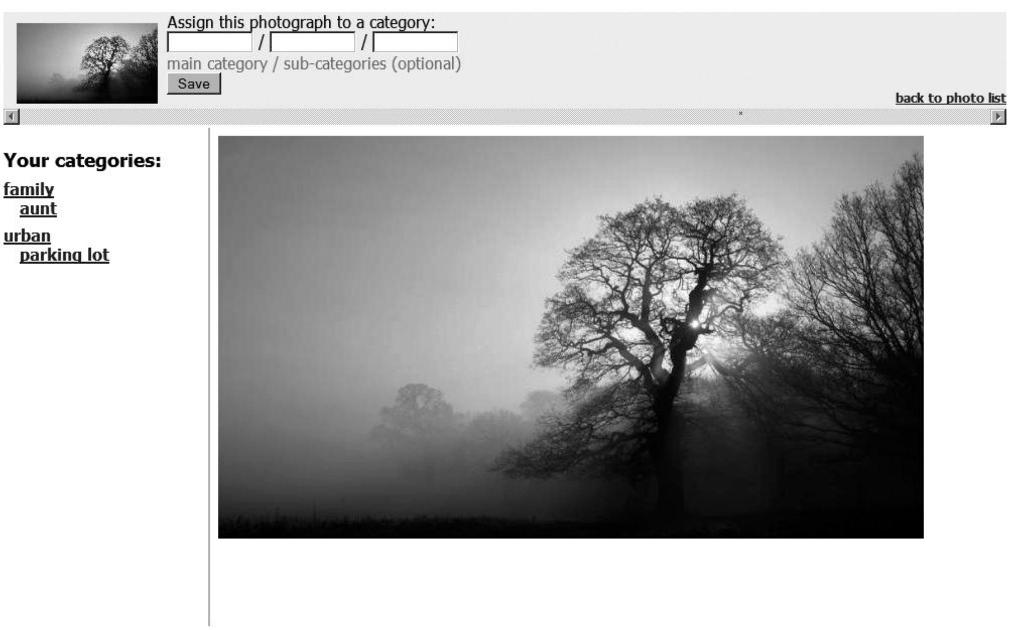

Figure 4a. Screen capture of the tagging condition start page (organization phase). When the participant clicked on a thumbnail image, they saw the labeling interface (Fig. 4b).

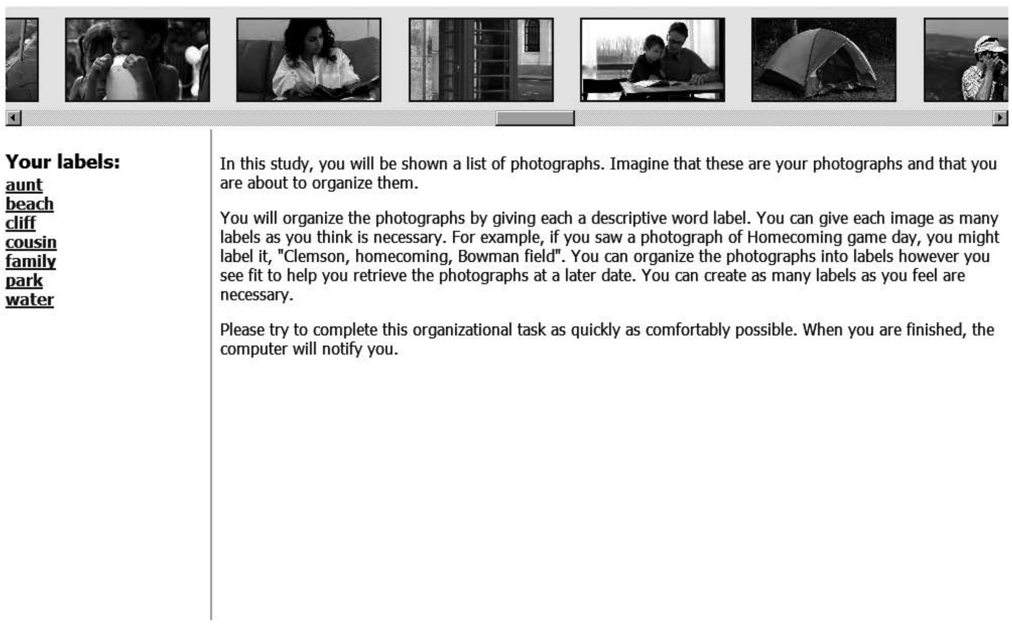

Figure 4b. Screen capture of applying labels or tags to photograph in tagging condition (organization phase). Participants currently used labels appeared on the left.

Figure 5a. Screen capture of the taxonomic condition start page (retrieval phase)

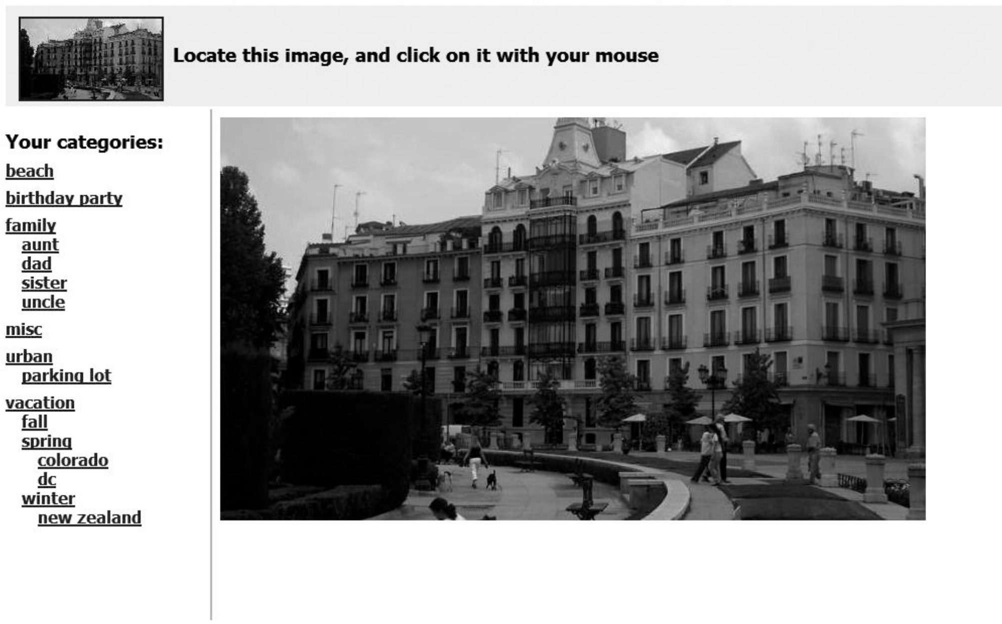

Figure 5b. Screen capture of viewing contents of a the “urban” category in the taxonomy condition (retrieval phase).

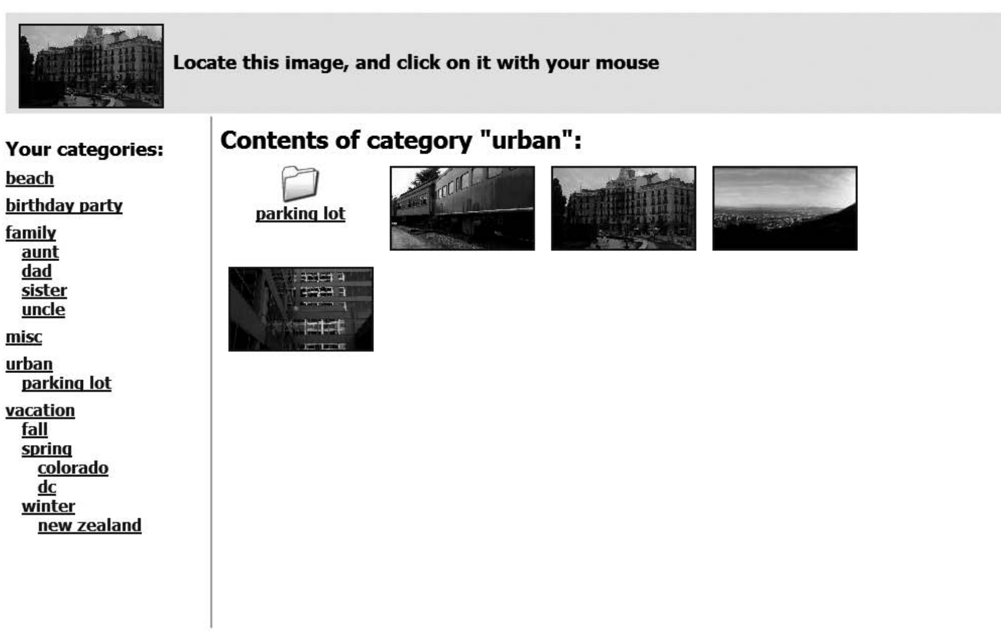

Figure 6a. Screen capture of the tagging condition start page (retrieval phase)

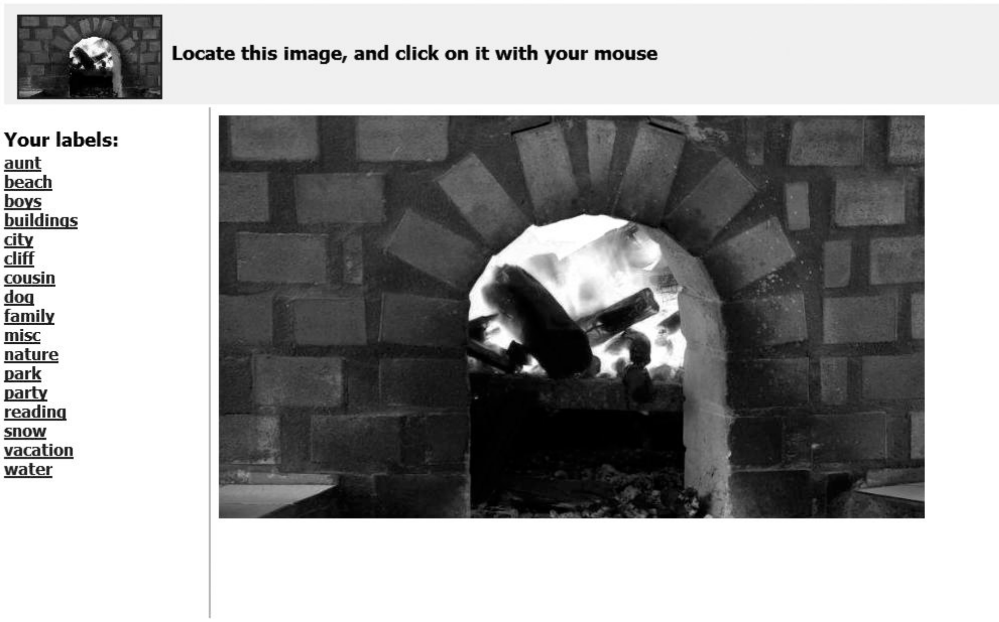

Figure 6b. Screen capture of viewing the photographs that have the nature label in the tagging system (retrieval phase)

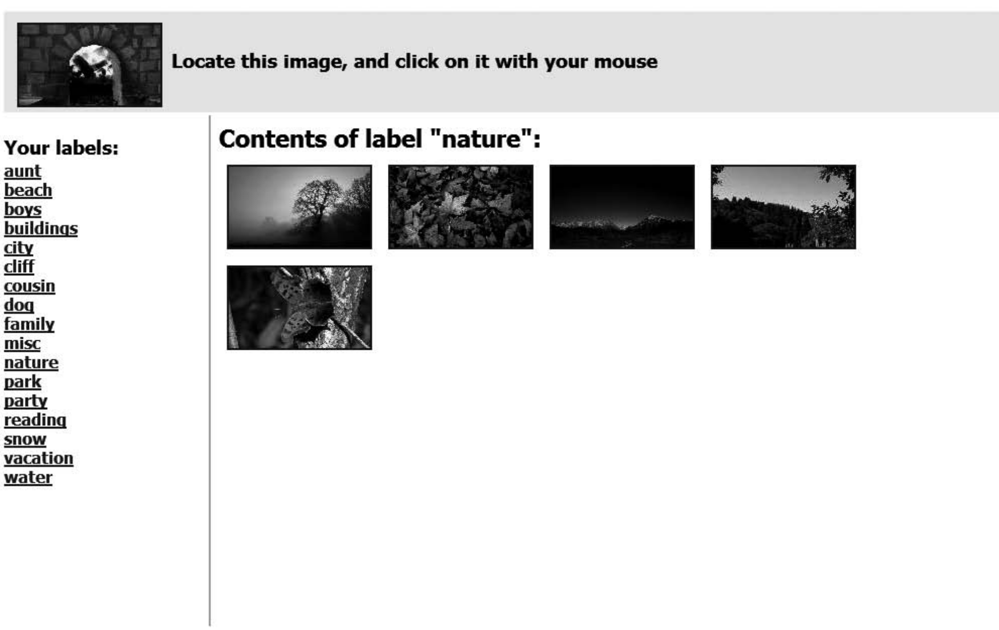

Figure 7. Mean times to organize and retrieve photographs in the tagging and taxonomy conditions.