Visualizing Implicit Queries For Information Management and Retrieval

Mary Czerwinski, Susan Dumais, George Robertson, Susan Dziadosz, Scott Tiernan and Maarten van Dantzich
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
+1 425-703-4882
marycz@microsoft.com

ABSTRACT
In this paper, we describe the use of similarity metrics in a novel visual environment for storing and retrieving favorite web pages. The similarity metrics, called *Implicit Queries*, are used to automatically highlight stored web pages that are related to the currently selected web page. Two experiments explored how users manage their personal web information space with and without the Implicit Query highlighting and later retrieve their stored web pages. When storing and organizing web pages, users with Implicit Query highlighting generated slightly more categories. Implicit Queries also led to faster web page retrieval time, although the results were not statistically significant.

Keywords
Information management, information retrieval, 3D, similarity, categorization, information visualization, classification

INTRODUCTION
The digital revolution has brought with it the problem of information overload. Even the simplest user query for information is accompanied by a results list that can be overwhelming. In addition, very little support is provided to help users in collecting, organizing and determining relevancy of retrieved items [3, 8, 16]. A careful examination of the graphical user interface and of similarity analysis methods is needed in order to address these sensemaking hurdles.

Usually query results are presented textually, as lists of e-mail, Internet documents or news reports. However, today nearly all web pages include some form of distinguishable graphics (e.g., a company logo) that users might associate in memory with that page. To take advantage of this, we present a visualization that allows users to manually create a spatial layout of the thumbnails of their documents in a 3D environment.

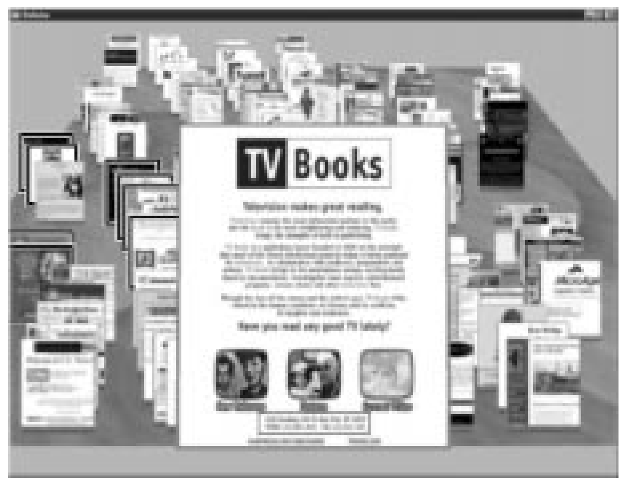

Figure 1. Data Mountain with Implicit Query results shown (highlighted pages to left of selected page).

As an organizational aid to the user, we use document similarity metrics and visual highlighting cues to indicate that web pages are semantically related in this personal information space. This paper will compare two such metrics, one user-driven and one content-driven, used to determine web page similarity relations during sensemaking tasks.

Current web browsers try to alleviate the sensemaking problems raised above through the use of bookmarks or favorites mechanisms, wherein users store the URLs of interesting web pages in order to build a personalized information space. Despite these user interface mechanisms, a 1998 survey of over 10,000 web users revealed that one of the most common problems users have with the web is organizing the information that they gather there [7]. In related research, Abrams, et al. [1] studied the bookmark archives and personal Web habits of users and made recommendations for improving the design of existing favorites management systems. Abrams surveyed 322 Web users, and analyzed the bookmarks of 50 Web users in detail. He found that bookmarks were used to reduce the cognitive load of managing URL addresses (by aiding memory and keeping history), to facilitate access, and to create information spaces for personal and group use. Bookmarks were often added sporadically—perhaps not surprisingly when too many favorite pages were piling up in a user's list. Almost 40% of those studied used no organization and simply left web pages in the order they were added to the favorites list; 50% used a hierarchy of one (30%) or more (20%) levels. Most users organized at the time they created a bookmark and cleaned up only occasionally. The initial use of folders began after a user had about 35 bookmarks. Abrams also found that 50% of the bookmarks had been visited in the last 3 months; 67% in the last 6 months; and 97% in the last year. Some ease of use recommendations provided by Abrams included providing aids in the browser for semi-automatic filing, time- or usage-based orderings, and much better tools for helping users in their organizing task. These findings provided the primary motivation for the research described in this paper.

We describe a new interaction that helps users quickly recognize and use the categorical structure they need to organize their favorite web pages. The interaction technique includes the Data Mountain [13], a novel visual environment for laying out personal web pages in a 3D space (described below), and an Implicit Query technique which shows the user which items are related to their current interest. Our Implicit Query algorithms determine similarities among web pages, and present the results in a visual format that has been observed to be useful and usable. This approach allows users to focus on relevant items instead of searching through large numbers of pages in the space. We have initially applied this idea to interaction with documents on the Web, although the interaction technique could be applied to any electronic document management task.

DATA MOUNTAIN WITH IMPLICIT QUERY

The Data Mountain is a (3D) document management system (see Figure 2). The design and implementation of the Data Mountain is described in [13], so only a short overview of the environment will be provided here. Currently the Data Mountain is being used as an alternative to a web browser's favorites or bookmark mechanism. It should be understood that other forms of documents should work equally well in the new environment.

The Data Mountain uses a planar surface (tilted at 65 degrees; see Figure 2), on which documents are dragged. A document being dragged remains visible so that the user is always aware of the surrounding pages. The user can place her web pages (or documents) anywhere on the surface. In practice, the user creates meaning by organizing the space in which she lays out these documents. In our study, each user was allowed to freely choose an organizational method and adjust it at any point throughout the study.

Figure 2: Data Mountain with 100 web pages.

When the user clicks on a page stored on the Data Mountain, the page is animated forward to a preferred viewing position, as shown in Figure 1. When in the preferred viewing position, another click will put the page back on the Data Mountain in its last known location. In practice, a click on a URL would allow the user to follow that link. Also, a stored page can be moved at any time by dragging it with the mouse. Since the page is visible during the move, the user knows where the page will be when the drag is terminated. The movement is continuous and constrained to the surface of the Data Mountain.

When a user moves a page around on the Data Mountain, it is likely to "bump" into other pages. Objects are not allowed to intersect, and the user's dragging action is not constrained. Instead, we handle collisions by displacing previously placed pages, thus continually maintaining a minimum distance between all pages, and transitively propagating displacement to neighbors as necessary. The user dragging the page continually sees what state will result when the drag is terminated (i.e., there is no animated settling time). This displacement technique also ensures that pages never get fully obscured.

There are a number of cues designed to facilitate spatial cognition. The most obvious are the 3D depth cues of perspective view, accompanying size differences, and occlusion, particularly when pages are being moved. Simple, circular landmarks on the surface of the Data Mountain also offer spatial cues, which may or may not be utilized during page placement or retrieval. Less obvious, but also quite important, are the shadows cast by web pages.

Subtle but pervasive spatial audio cues accompany all animations and user actions to reinforce the visual cues. The sound effects are highly dynamic. For example while dragging a page, the user hears a humming sound that changes pitch based on the speed at which the page is moved. With careful use of the timbre, stereo and attenuation of the humming sound, users are thus provided an additional spatial location indicator. Finally, as the user moves a page, other pages that move out of the way as needed are accompanied by yet another distinctive sound.

For identification purposes, we provided a pop-up labels similar to tool-tips to display web page titles. The title appears as soon as the mouse moves over a page.

In order to support users' information management and retrieval tasks on the World Wide Web (WWW), we use what we call *Implicit Queries*. Implicit Queries are generated automatically, based on the current focus of the user's attention. In our system, the web page that the user has currently selected is the basis for queries launched implicitly in the background. Much richer models of users' interests are easy to incorporate as well [6], but were not needed for this experiment. Web pages that are similar to the currently selected page are then highlighted to aid users' organization and use of their web information space. Figure 1 shows an example where the selected page has something to do with entertainment. For this particular user, most of the entertainment-related pages are on the left side of the layout and are highlighted so that the pattern is easily seen, yet not distracting. In color, this highlighting is very obvious—it is much less obvious in a black and white reproduction, so Figure 1 has been retouched to make clear where the highlighting is.

Prior work on similarity metrics in the user interface

A number of previous systems (for example, Spire [18] and Galaxy of News [11]) have used document similarity metrics as a way to organize (or lay out) documents, bringing similar documents closer together. Information access in these systems is achieved primarily by navigation. The Data Mountain differs fundamentally by allowing the user to determine the layout of the documents, and the Implicit Query mechanism highlights related documents in the context of the user-defined organization. By maintaining spatial consistency with the user's organization, our system is able to more effectively leverage human spatial memory [13].

Many systems show the results of *explicit* queries from users, usually as ranked lists. Some systems post the results of an explicit search against a known structure [4]. These structures may be expert-generated hierarchies or automatically derived (e.g., from the link structure of a Web site or by a statistical analysis of inter-item similarities). The resulting structures are shown using a variety of representations -- clusters, 2D and 3D spaces, tree-maps. All of these systems take a known structure and show how search results fit into that organization. While these approaches are similar to ours in that respect, it is important to note that the organizing structures used by these systems are generated by domain experts or statistical methods and not by the user.

In contrast to these systems that use explicit queries, we have developed an interaction technique which relieves the user from having to generate an explicit query and so reduces the amount of cognitive processing required to retrieve related information. Much less work has been done on Implicit Queries and interfaces to them. The Remberance Agent by Rhodes and Starner from MIT [12] described an approach similar to our general Implicit Query idea. They implemented a continuously running process which monitored what a user was reading and automatically sent this information to a backend server which indexed personal information such as email, notes and papers. The query consisted of whatever information was being read or typed into the current emacs buffer. The results were presented as a ranked list of titles in a separate window, and were thus very limited compared to our Data Mountain.

Schilit et al. [15] developed the XLibris system to support what they call "active reading". Reading was enhanced by computation using a kind of Implicit Query mechanism. Highlights made by the user via pen markings were used to issue an automated query for related materials. It is not clear that the use of highlighting to drive search is optimal, since people highlight and annotate for many different reasons [9]. The results are shown as thumbnails in the margin, which provides a local context, but no global inter-item similarities are shown as in the Data Mountain.

Our work on visualizing the results of Implicit Queries combines two ideas in a novel and a powerful way. We use a rich visual representation of objects and their inter-relations on the user-determined spatial layout of the Data Mountain along with Implicit Queries for information management. In addition, we empirically evaluate the usefulness of our system to support users in information organization and retrieval tasks.

Previous research on highlighting techniques

Much research has been performed on the usefulness of highlighting in attracting attention to relevant information in a display [17]. For instance, it is well known that the techniques of reverse video, color, boldness (or brightness), underlining, and flashing are all effective highlighting techniques, although some of them can actually be disruptive if applied inappropriately. For our Implicit Query highlighting, we examined over 20 highlighting techniques, some inspired by previously reported research [5, 10]. Based on our informal observations of which techniques were more most effective without causing undue disruption, we chose to implement a simple green outline to display which web pages were related to the currently selected web page.

Finally, it should be noted that we have chosen a binary approach to highlighting similarities in the web pages stored on the Data Mountain, as opposed to the continuous approaches we considered. We chose a binary approach purposefully, as we were concerned about the possibly distracting effect that Implicit Query highlighting would have on the user's primary task. We also assumed that a binary approach would lessen the decision-making burden on users. However, we have prototyped many designs that show the relevance score along a continuum in the visualization.

USER STUDIES OF DATA MOUNTAIN IMPLICIT QUERY

We tested our visualization and Implicit Query user interface ideas in two studies. The first study examined how users managed and stored 100 favorite web pages with and without Implicit Queries during their web page interactions. It was our hypothesis that the Implicit Query algorithms and highlighting would provide a useful guide to users during their web page organization. Although users were not required to follow the query recommendations in any way, we further hypothesized that users would indeed group together the highlighted pages. In addition, it was hypothesized that Implicit Query highlighting would improve memory of where pages were stored spatially on the data mountain, due to subjects' having spent more time attending to related pages and considering the suggestions proffered by the system. This hypothesis was based on the theoretical notion of "levels of processing" [2]. According to this theory, information can be processed more or less deeply, ranging from a shallow analysis (attention to surface features) to a deeper, semantic analysis. Information that is processed more deeply has been shown to be more likely to be remembered over time.

In these studies, we also explored two similarity metrics to drive the Implicit Queries. Looking at the content (word) overlap of pages generated one metric, while the other was derived from a group of subjects' previous organizations of the same web pages. We expected that the Implicit Query suggestions coming from subjects' previous organizational strategies might be the "best case" for a similarity metric, and so hypothesized a performance advantage for this algorithm.

In the second study, we examined whether or not this system guidance during web page organization would actually benefit subsequent retrieval of those previously stored favorite web pages. Any retrieval time advantages will be realized over and over again as subjects repeatedly revisit the page later.

EXPERIMENT 1

Methods

Subjects

Thirty-five subjects of intermediate web ability and who were experienced Microsoft Internet Explorer™ 4.0 (IE4) users at work or home participated in the experiment. All users were required to successfully answer a series of screening questions pertaining to web browser and Internet knowledge in order to qualify for participation. The number of females and males was balanced. 15 subjects organized their web pages with no Implicit Query mechanism. 20 subjects were aided by one of two Implicit Query algorithms (9 subjects used algorithm 1 or IQ1; 11 used algorithm 2 or IQ2). The experimental sessions involved two studies, an organizational phase and a retrieval phase. For clarity, methods and results of the two phases will be described separately as Experiment 1 and Experiment 2. The Methods for Experiment 1, the organizational phase, are described below.

Material

One hundred web pages were used in this study; 50 of the pages came from PC Magazine's list of top web sites (and so were likely to have been seen by at least some of the participants) and 50 pages were selected randomly from the Yahoo!™ database. The web pages were downloaded onto a web server located on the computer the subject worked at.

We used two algorithms to generate a set of matching pages for each web page in the study, a co-occurrence algorithm and a content-based algorithm. *IQ1 – co-occurrence similarity*. The first similarity measure was derived from a page-page co-occurrence matrix based on seven previous subjects' categorizations. Only subjects whose categorizations were relatively clear and discrete spatially were used for this algorithm's derivation. We counted the number of times that a pair of pages co-occurred in the same cluster – this number varied between 0 and 7. This algorithm essentially tells the user, "Other people thought these pages were related." *IQ2 – content-based similarity*. For the content-based similarity computations, the popular vector space model from information retrieval was used [14]. Documents were pre-processed to remove the HTML markup. Words on a standard stop list of common words along with 10 web-specific words were omitted, and white space or punctuation was used to delimit words. Each document was represented as a vector of words with entries representing the frequency of occurrence of a word in that document. The similarity between documents was measured by taking the dot product of the document vectors divided by the lengths of the vectors.

Once we generated these two measures of similarity, we set a threshold for each algorithm. Only web pages that matched the target web page at a level of similarity above the threshold were recommended as "related" to the user. We wanted the two algorithms to recommend roughly the same average number of matches per page. The thresholds we chose generated, on average, 4.2 and 4.3 matches per page respectively. In the co-occurrence algorithm (IQ1), this threshold produced 39 pages that had no match above the threshold. There were 28 pages that had no match above the threshold for the content-based algorithm (IQ2).

To indicate which pages were identified as matching the page being viewed, we highlight the related pages with a bright green frame as shown in Figure 1. Highlighting automatically occurs when subjects are presented with a new page for storage during the first phase of the study. Selecting any page on the Data Mountain causes its related pages to be highlighted.

Procedure

Subjects were shown 100 web pages sequentially (order was randomized for each subject) and asked to store them on the Data Mountain. They were allowed to create any organization they wanted and were encouraged to create a personally meaningful structure that mimicked how they stored favorite web pages at home. Subjects were told that they would have to use their organization for a retrieval task in the second half of the test.

For the subjects who were in the Implicit Query conditions, related pages were highlighted according to the IQ1 or IQ2 algorithms. We briefly interrupted each subject in the Implicit Query conditions to discuss the green highlighting after 10 minutes of their organization phase had elapsed. We asked them if they had noticed the highlighting, and what they thought it was for. We then explained its purpose and informed participants that they were free to use or ignore the suggestions. Some subjects noticed the highlighting, figured out its purpose and requested validation of their assumption before 10 minutes had passed, which motivated an early intervention. When this occurred, we discussed it with the subjects at the time of their request. The discussion time, on average, took less than 1 minute.

After all 100 pages had been saved on the Data Mountain, the subjects were given time to fine-tune their organization until they were personally satisfied with it.

The main independent variable of interest in Experiment 1 was the between subjects variable of which Implicit Query matching algorithm was used (no Implicit Query, IQ1, or IQ2). The number and type of categories, organization time, and subjective satisfaction ratings were the dependent measures of interest.

Results

Influences on information management behavior

Most subjects adopted an organization based on semantic categories. Some of these subjects augmented this with temporal or alphabetic cues as well. Five users stored files alphabetically and one used no apparent organization. Table 1 shows the number of subjects following each of the observed organizational strategies. For subjects who categorized using a strategy other than alphabetic, we asked them to circle and label their categories.

Number of categories

The average number of categories for each group is shown in Table 2. Subjects in the Implicit Query conditions created slightly more categories than did subjects without the Implicit Query highlighting during the organization phase of the study, although this difference was not statistically significant, $F(2,27)=1.31$, $p=.28$. Because we had to remove subjects who alphabetized from this analysis, the test was not very sensitive. If we pool these results with those of another condition (not reported here) that used the same organizational procedure and stimuli, but different retrieval cues, there were reliable differences in the number of categories, $F(2,38) = 3.74$, $p=.03$.

Implicit Query Condition	*Semantic*	*Alphabetic*	*Unknown*
IQ0: No Implicit Query	11	3	1
IQ1-: Co-occurrence based	8	1	0
IQ2: Content based	10	1	0

Table 1. Number of participants that used a particular organizational strategy while storing web pages.

Implicit Query Condition	*Avg. # of Categories (Standard Deviations in parentheses)*
IQ0: No Implicit Query	12.0 (3.6)
IQ1: Co-occurrence based	15.8 (5.8)
IQ2: Content based	13.6 (5.9)

Table 2. Averages and standard deviations for the number of categories observed in each Implicit Query condition.

Overlap of category concepts

Subjects' organizations were analyzed for their amount of overlap with each other. In order to do this, we used the layouts that subjects had circled and labeled for us. We identified 20 categories that subjects used very frequently, and reduced their category structures into these 20 categories. In order to check our data reduction procedure, we compared two authors' classification efforts on the same layouts for inter-rater reliability. On average, we obtained 90% agreement across Implicit Query conditions for two independent observers. One subject was primarily responsible for the inter-rater disagreement, and for this subject the 2 raters only agreed 50% of the time.

Once it was determined that the categorization scheme was a reliable one, subjects' clusters were analyzed for consistency. We found no reliable differences in how often subjects agreed with each other in terms of categories maintained during the storage phase of the study $F(2,26)=.164$, $p=.85$. In other words, having an Implicit Query mechanism did not result in more or less agreement between subjects and their organizational schemes.

Organizing Time

Although subjects were not encouraged to be efficient during the organization phase of the experiment, we did record organization times across Implicit Query conditions. Organization time includes both the time used to initially place and organize web pages and the time to reorganize pages after the initial storage. The average organization times were 52 minutes for IQ0, 71 minutes for IQ1 and 81 minutes for IQ2. The effect is marginally reliable $F(2,32) = 2.88$, $p=.07$. This is not surprising, given that Implicit Query users often considered whether or not to follow the system recommendation for where to store a web page, and they created somewhat more categories, both of which could have taken extra time.

Most of the total organization time was spent during initial placement (an average of 53 minutes across all conditions) and substantially less time during reorganization (an average of 15 minutes across all conditions). The variance in reorganization time was quite large, due primarily to two subjects. The subjects who alphabetized in IQ1 and IQ2 each took over an hour to reorganize their alphabetized layouts, which is an interesting finding in itself and warrants further research. One subject moved thumbnail images more than 1000 times! In our prototype it was difficult to reorganize because no grouping of objects was allowed. If subjects who alphabetized are removed, the organization time differences between conditions decrease. (The alphabetizers in IQ0 did not take exceptionally long during reorganization.)

Questionnaire measures

Participants provided a variety of subjective ratings about the user interface and the interaction techniques at the end of the study session. One subject is not included in the analysis of the Implicit Query condition 2 due to her not filling out any answers on the questionnaire. A multivariate ANOVA (using Implicit Query condition as a between subjects factor and each questionnaire item as a multivariate response) revealed a reliable interaction between Implicit Query condition and only one questionnaire item, $F(2,31)=7.09$, $p=0.003$. The questionnaire item that drove this effect, "I was satisfied with my organizational scheme; 1=Disagree, 5=Agree", accounted for over 31% of the variance in the data. Scheffe post-hoc analyses showed that subjects in the no Implicit Query (IQ0) and the co-occurrence-based algorithm (IQ1) groups were not different from each other on this questionnaire item (average ratings = 3.6, SD=0.22 and 4.0, SD=0.28, respectively). Subjects in the content-based algorithm (IQ2) condition, however, were less satisfied than the other two groups (average rating = 2.6, SD=0.25).

A few questionnaire items pertained only to the Implicit Query visualization, and so only groups IQ1 and IQ2 responded to these questions. Analyses of these results revealed that the co-occurrence algorithm built from previous subjects' organizations of the 100 web pages (IQ1), was rated as significantly less distracting than the content-based algorithm (IQ2), $t(18)=-2.04$, $p=.01$, two-tailed. Subjects' ratings of the IQ1 highlighting as more useful than IQ2 reached borderline significance, $t(18)=1.8$, $p=.09$, two-tailed. No other significant effects emerged from analysis of the questionnaire data.

Discussion

The results of the organization phase of this study were mixed, and suggest that good Implicit Queries in the user interface in the storage of information during web interaction might lead to slightly more detailed categorization at the cost of significantly longer storage times. According to Levels of Processing Theory [2], the increased time spent in information management could result in a deeper encoding of subjects' web page organizations on the Data Mountain and might therefore facilitate subsequent web page retrieval in both speed and accuracy. Questionnaire data suggested that subjects were equally satisfied with their organizations with no Implicit Query or with co-occurrence based Implicit Query, but were significantly less satisfied with content-based Implicit Query.

EXPERIMENT 2

In Experiment 2, we assessed whether or not the presence of the Implicit Query highlighting during web page storage improved web page retrieval performance. We compared the IQ0, IQ1 and IQ2 groups on average retrieval times, the number of incorrect pages retrieved, and the number of failed retrieval attempts for the 100 web pages. It is extremely important to note here that the Implicit Query highlighting was disabled for this phase of the experiment. In other words, a subject's target web page did not highlight during its retrieval trial, nor were web pages that might be related to the target highlighted. While such highlighting would be desirable in practice, we felt it would be too beneficial for experimental purposes.

Methods

Subjects, Materials and Procedure

The same thirty-five subjects participated in the second study. After a short break following the organizational study, the subjects started retrieval. For the second study, participants were shown a retrieval cue consisting of the textual title of the web page, and asked to find the corresponding page on the Data Mountain. The retrieval cue was presented in a small, rectangular window below the display window of the Data Mountain. The cues were presented in a random order for each subject. If a subject could not find the target page within two minutes, the subject was instructed to proceed to the next retrieval trial. Since the Implicit Query highlighting was not enabled during this phase of the experiment, there are no visible differences in the user interface among the experimental conditions in this phase.

The three primary dependent measures used in Experiment 2 were web page retrieval time, the number of incorrect pages selected, and the number of failed attempts to retrieve a web page. Retrieval time (or reaction time) was defined as the time to select the correct item. Incorrect selections referred to the total number of pages selected that were not the target page, not including failed retrieval trials. Failed retrievals occurred when the subject either took longer than two minutes for retrieval or chose to stop searching for an item.

Results

Retrieval Time

Only trials in which subjects found the correct web page within the allotted two-minute timeframe are included in the reaction time analysis for both sessions. Figure 3 shows the retrieval time results. For each subject we computed a median response time. The average retrieval time across subjects was 9.5 seconds (SD=8.5) for the subjects with no Implicit Query algorithm, 6.8 seconds (SD=2.01) for the co-occurrence algorithm, and 7.3 (SD=2.14) seconds for the content-based algorithm. The difference between the three conditions was not significant, $F(2,32)=0.73$, $p>.05$. The high variance in IQ0 was primarily responsible for the failure of this effect to be reliable.

Item Effects

As expected, there is large variability in the time required to retrieve individual web pages. The average RT for individual web pages ranges from 4.9 seconds (for the NASA Home Page) to 24.3 seconds (for the Mercury Center Home Page). In general, the pages that were fastest to retrieve were either very distinctive semantically or visually. Popularity of a web page may have had additional effects, but the effect was not large (average retrieval time for 50 popular web pages =12.7, randomly chosen web pages average=12.9).

Incorrect Selections

On average there were very few visits to incorrect pages. The average number of incorrect retrievals in the no Implicit Query algorithm condition = 3.5 (SD=3.6), the IQ1 average = 3.0 (SD=1.6) and for the IQ2 algorithm the average number of incorrect retrievals = 4.7 (SD=3.6). There were no significant differences in the average number of incorrect pages retrieved across the 3 conditions, $F(2,32)=.78$, $p=.47$.

Failed Retrievals

There were an average of 4.8 (SD=3.1) trial failures in the no Implicit Query condition, compared to an average of 2.0 (SD=2.4), on average, in the IQ1 algorithm and 5.0 (SD=6.1) in the IQ2 algorithm, and this difference was once again not statistically significant, $F(2,32)=1.6$, $p=.2$.

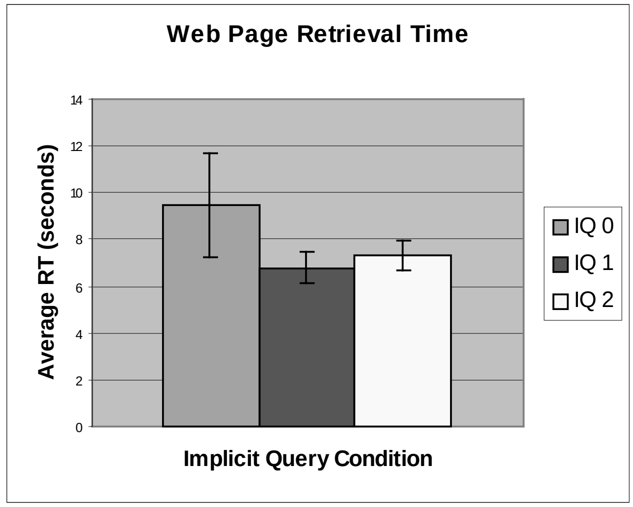

Figure 3. Average retrieval times (including standard error of the mean) for the 3 IQ conditions.

Discussion

It is important to reiterate that no Implicit Query highlighting was used during retrieval in Experiment 2. (We suspect retrieval times in IQ1 & 2 would be reduced even further if highlighting were used during retrieval). Results obtained in Experiment 2 suggest that Implicit Query highlighting during storage can facilitate subsequent retrieval, but the effects were not reliable. Retrieval times were variable both across items and subjects, and additional experimental power will be necessary to fully understand this tendency. Retrieval time differences are important because typical uses of any one piece of information will have one storage incident and many subsequent retrievals. By reducing retrieval time at the expense of storage time, we can shift and reduce overall cognitive load for the user when finding previously viewed information for the task at hand.

A concern about the generality of these results is that we tested subjects' retrievals on the same day that they stored the web pages, an unlikely scenario in everyday web page access. In subsequent studies in our laboratory, we brought subjects back 6 weeks and 6 months after they stored their web pages on the Data Mountain. There was no decrement in performance when subjects returned and were asked to retrieve their web pages after either a 6-week period or a 6-month period of disuse. Details of these further studies are forthcoming.

CONCLUSION

This paper reported two studies that examined users' web page organizations and later retrievals using a combination of two interaction techniques, the Data Mountain and Implicit Queries. Previous research [13] had already demonstrated enhanced performance for web page retrieval in the Data Mountain, when compared to current browser favorites mechanisms (one-dimensional visual text lists). The use of a passive and subtle Implicit Query aid influenced subjects at organization time. They took longer and created somewhat more categories. There are indications of influence at retrieval time, even when no Implicit Query cue is present.

Future work will focus on user interface visualizations for queries in much larger information spaces to see if more powerful results can be obtained in those domains. We will also continue to investigate the influence that Implicit Queries have on individual users' organizational styles. For instance, there was some indication that users without Implicit Queries tended to use an alphabetic organizational strategy more often. We will examine the use of alternative highlighting mechanisms, and explore the use of Implicit Queries at the time of retrieval as well as storage. This future work will be carried out using alternative 3D visual metaphors for visualizing the query results.

ACKNOWLEDGMENTS

We thank the User Interface Research Group at Microsoft and Kevin Larson for their help and valuable comments on this paper. We thank Paul Jaye for his help in data analysis.

REFERENCES

1. Abrams, D., Baecker, R.& Chignell, M. (1998). Information archiving with bookmarks: Personal web space construction and organization, in *Proceedings of CHI '98* (Los Angeles CA, May, 1998), ACM Press, 41-48.
2. Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour,* 11, 671-684.
3. Dumais, S.T. (1988). Textual information retrieval. In M. Helander (Ed.) Handbook of Human-Computer Interaction. Elsevier Science Publishers (North Holland). .
4. Hearst, M. and Karadi, C. Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* July 1997, 246-255.
5. Hirtle, S., Sorrows, M.E. & Cai, G. (1998). Clusters on the World Wide Web: Creating neighborhoods of make-believe, in *Proceedings of Hypertext and Hypermedia '98* (Pittsburgh PA, 1998), ACM Press, 289-290.
6. Horvitz, E., Breese, J., Heckerman, D., Hovel. D. & Rommelse, K. (1998). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.*
7. Kehoe, C., Pitkow, J. & Rogers, J. (1998). GVU's 9th WWW User Survey, http://www.gvu.gatech.edu/user_surveys
8. Marchionini, Gary (1995). *Information Seeking in Electronic Environments.* Cambridge University Press.
9. Marshall, C.C. (1998). Toward an ecology of hypertext annotation, in *Proceedings of Hypertext and Hypermedia '98* (Pittsburgh PA, June, 1998), ACM Press, 40-49.
10. Olsen, D.R. Jr., Boyarski, D., Verratti, T., Phelps, M., Moffett, J.L. & Lo, E.L. (1998). Generalized pointing: Enabling multiagent interaction, in *Proceedings of CHI '98* (Los Angeles CA, May, 1998), ACM Press, 526-533.
11. Rennison, E. (1994). Galaxy of news: An approach to visualizing and understanding expansive news landscapes. In *Proceedings of ACM UIST '94 Symposium on User Interface Software & Technology*, Marina del Ray, CA: ACM, 3-12.
12. Rhodes, B. and Starner, T. A continuously running automated information retrieval system. In *Proceedings of The First International Conference on The Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96)*, London, UK, April 1996, pp. 487-495.
13. Robertson, G., Czerwinski, M., Larson, K., Robbins, D., Thiel, D. & van Dantzich, M. (1998). Data Mountain: Using Spatial Memory for Document Management, Paper to appear in *Proceedings of ACM UIST '98 Symposium on User Interface Software & Technology*, November, San Francisco, CA.
14. Salton, G.McGill, M. (1983). *Introduction to Modern Information Retrieval.* McGraw Hill.
15. Schilit, B.N., Golovchinsky, G. & Price, M.N. (1998). Beyond paper: Supporting active reading with free form digital ink annotations, in *Proceedings of CHI '98* (Los Angeles CA, May, 1998), ACM Press, 249-256.
16. Shneiderman, B., Byrd, D. & Croft, B. (1998). Sorting out searching: A user-interface framework for text searches. *Communications of the ACM, 41(4),* 1998, 95-98.
17. Tullis, T.S. (1997). Screen Design. In (Eds.) Helander, M., Landauer, T.K. & Prabhu, P.'s, *Handbook of human-computer interaction*, 2nd Edition, Elsevier Science, B.V., 503-531.
18. Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Shur, A., and Crow, V. (1995). Visualizing the Non-Visual: Spatial analysis and interaction with information from text documents. In *Proceedings of Information Visualization 1995*, IEEE Computer Society Press, 51-58.