Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces

Marc M. Sebrechts,
Joanna Vasilakis, Michael S. Miller
Department of Psychology
The Catholic University of America
Washington, DC 20064 USA
(202) 319-5750
sebrechts@cua.edu

John V. Cugini,
Sharon J. Laskowski
Information Technology Laboratory
National Institute of Standards & Technology (NIST)
Gaithersburg, MD 20899-8940 USA
(301) 975-3265
cuz@nist.gov

ABSTRACT

Although there have been many prototypes of visualization in support of information retrieval, there has been little systematic evaluation that distinguishes the benefits of the visualization per se from that of various accompanying features. The current study focuses on such an evaluation of NIRVE, a tool that supports visualization of search results. Insofar as possible, functionally equivalent 3D, 2D, and text versions of NIRVE were implemented. Nine novices and six professional users completed a series of information-seeking tasks on a set of retrieved documents. There were high interface costs for the 3D visualization, although those costs decreased substantially with experience. Performance was best when the tool's properties matched task demands; only under the right combination of task, user, and interface did 3D visualization result in performance comparable to functionally matched 2D and textual tools.

Keywords

Information Visualization, 3D, 2D, text, Evaluation, Interface Design

1. INTRODUCTION

As information spaces have become more complicated, substantial attention has been devoted to considering how spatial dimensions can be used to decrease the task demands. Although there are numerous prototypes for visualizing large document spaces (e.g.: [1], [2], [5], [6], [12], [17]), there is little in the way of systematic comparison of the value of these approaches.

The current study focuses on trying to provide a controlled comparison of text, 2D, and 3D approaches to a set of fairly typical information seeking tasks on a small collection of 100 top-ranked documents that have been retrieved from a much larger set. Insofar as possible, the three interfaces are designed to provide common basic functionality in order to be able to isolate differences in the visualization rather than in the underlying algorithms or ancillary tools that are added to the visualization.

A few studies have conducted evaluations of the value of 2D displays. Nowell, France, Hix, Heath, and Fox [9] described the Envision system, in which the user can control how document attributes, such as author, date, or relevance score, are mapped into a 2D display using graphical attributes, such as size, shape, and color. A satisfaction survey showed that users liked the system. User performance was judged successful, but it was measured against the performance of the interface designer using the same system, not against the use of another (presumably less graphical) system.

Another 2D evaluation was conducted by Veerasamy and Belkin [15]. A simple 2D-grid system was used with search keywords along the y-axis, and document IDs along the x-axis. Each cell of the grid showed the frequency of the corresponding keyword within that document. A user study showed a weak effect in favor of the performance of the visualization over a text-based system. The authors note that traditional measures of information retrieval (IR) effectiveness, such as precision and recall, may not be appropriate for capturing the advantage of visualization-based systems over text-based ones. In a subsequent paper, Veerasamy and Ileikes [16] addressed these limitations by using a constrained relevance judgment task and demonstrated that the graphical tool increased speed and accuracy over textual displays.

There have also been a few studies of 3D-based visualization tools. The hyperbolic browser [8] simulated changes in appearance of documents spread over a 3D spherical surface. That tool used a focus+context fisheye approach to visualize and manipulate large hierarchies. A laboratory experiment contrasted using the hyperbolic browser against a conventional 2D scrolling browser with a horizontal tree layout. Although users preferred the hyperbolic visualization, there was no performance advantage on the task of finding specific node locations.

More recently, Swan and Allan [13] performed a controlled study comparing 1) a text-based system [11], 2) a GUI oriented system, and 3) the latter enhanced with 3D visualization of document clusters. They wanted to improve so-called "aspect-oriented" IR, which emphasizes finding some specified information, not documents per se. In that context, using recall as a measure, there was a small advantage for the 3D system over text-based, and for the text-based over the plain GUI. However, there was no evidence for the overall effectiveness of the use of 3D; the system's utility depended on the tasks and the users. Experienced users preferred the text-based system, while novices liked the GUI systems. Some users thought the 3D approach was "worthless", others thought it natural and intuitive.

In sum, there have been very few studies directly comparing the effectiveness of visualization against functionally equivalent traditional interfaces for IR applications. Previous data shows only modest benefits of visualization in this context.

1.1 NIRVE Prototype

Any assessment is of course likely to depend on the particulars of the visualization that is used and its specific objectives. For our purposes, we have focused on the issue of using visualization to analyze a set of documents that have already been retrieved by a search engine. To that end, the Visualization and Virtual Reality Group at the National Institute of Standards and Technology (NIST) has developed several prototypes that give users an overview of the results of a document search [3],[4].

Keyword queries were based on topics from the Text Retrieval Conference [14] that were designed for the database of all the news stories issued by the Associated Press for 1988. The PRISE [11] search engine returned a set of entries from that database, one per document. Each entry contained: document title, a unique document identifier, a relevance score (indicating the search engine's estimate of the "goodness" of the match between the document and the query), document length, and the number of occurrences of each keyword.

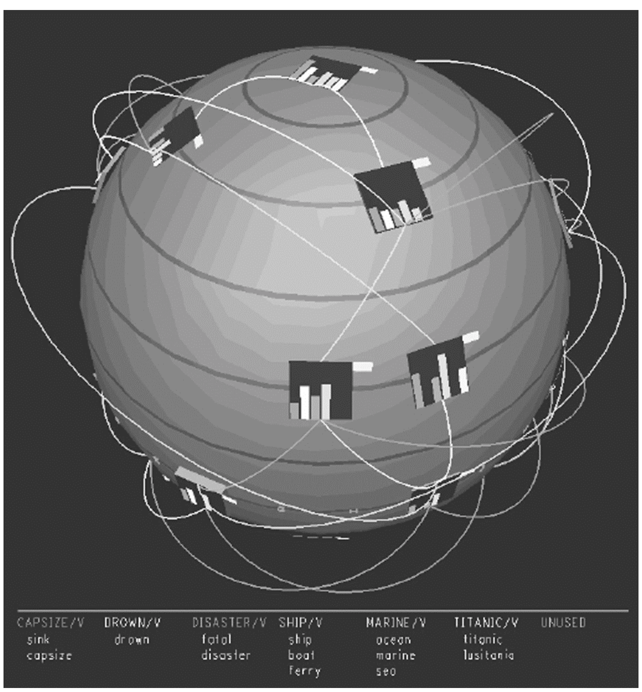

Figure 1. 3D NIRVE document space widow with clusters of documents displayed as boxes on the surface of a sphere. Colored bars within each cluster box indicate concepts. Arcs are similarly color-coded.

The NIRVE display is organized into two windows: a "document space" window in which document titles, clusters, keywords and concepts are depicted, and a smaller control window for operations that would not map naturally into the document space window. A typical example of a 3D NIRVE document space window is shown in Figure 1. The control panel window is shown in Figure 2.

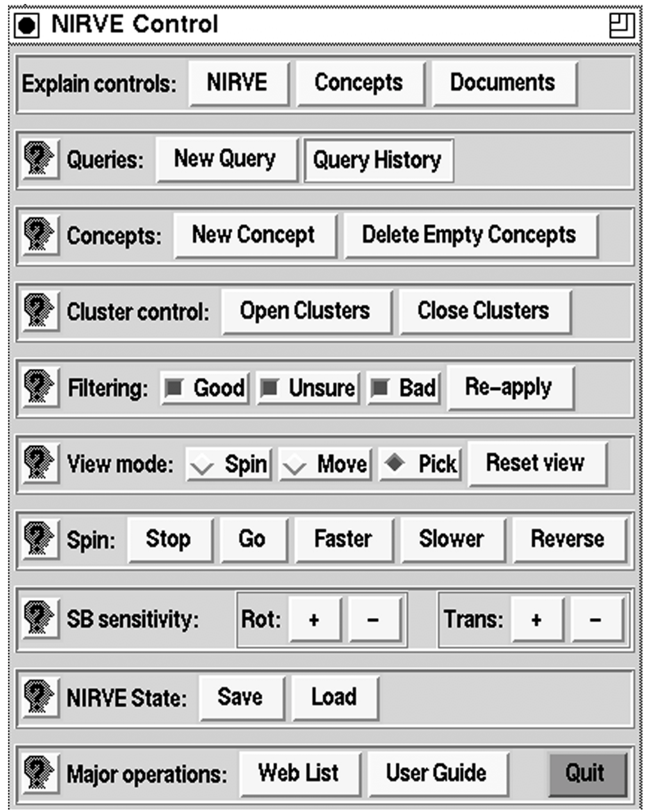

Figure 2. The NIRVE control panel window for 2D and 3D.

The user can dynamically map a subset of keywords of the query into a concept. The concept can be assigned a name and a color. The association of keywords and concepts is displayed as an interactive legend at the bottom of the document space window.

A concept profile is computed for every document, based on the frequency of the keywords therein. A cluster is then defined as a set of documents all of which have the same subset of concepts. Each cluster is represented as a small box on the face of which is a color bar chart indicating the average concept profile of the documents therein. The thickness of the box is proportional to the number of contained documents.

These cluster icons are arranged on the surface of a globe. An icon's latitude is determined by the number of concepts exemplified by the cluster: more concepts cause an icon to be located nearer to the "North Pole" of the globe. Longitude has no intrinsic meaning; icons are arranged so as to try to place clusters with similar concept profiles near each other. Clusters that differ by a single concept are connected by an arc whose color represents the conceptual difference between them. E.g. if cluster

A has the concepts "boat", "sink", and "ocean", and cluster B has "boat", "sink", "ocean", and "storm", then they will be connected by an arc color-coded for "storm".

When the user "opens" a cluster, a 2D-document rectangle containing all the document titles is projected outward from the cluster icon, as shown in Figure 3. These titles are arranged such that similar titles (i.e. those containing some matching words) have nearby horizontal positions. Vertical position is controlled by the relevance score assigned by the search engine. Thus similar titles will appear approximately in the same column, with the better scoring titles towards the top. When the user selects a title, Netscape displays the full document text. Each keyword is highlighted in the color of its containing concept.

Users can "mark" documents or entire clusters with a small colored flag to signify their evaluation of that entity as "good" (green), "bad" (red), or "unsure" (yellow). The initial value is "unsure". The user can then control the display of any subset of these categories (e.g. suppress "bad" documents, show only "good" or "unsure" documents).

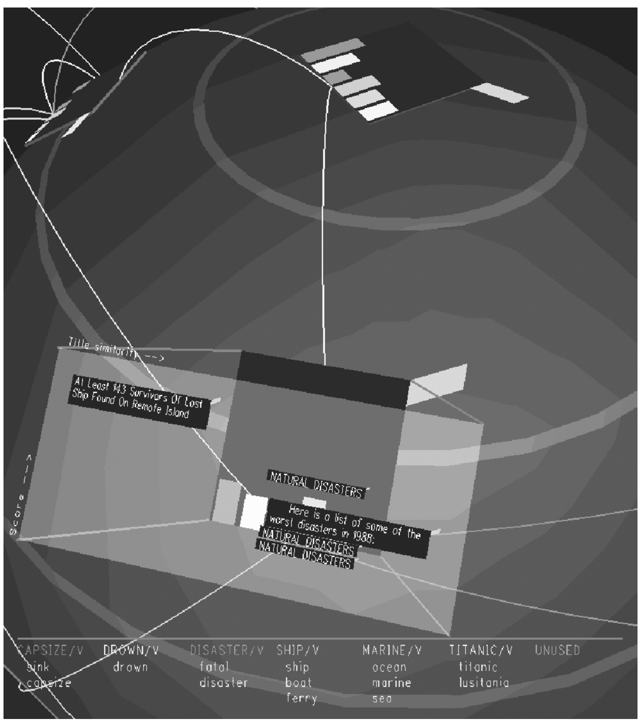

Figure 3. Document rectangle projected outward from the cluster icon, revealing the document titles

2. METHOD

2.1 Participants

A total of fifteen people participated in the experiment. Nine university students without professional computer experience participated as "novice" users. The remaining six participants had professional experience with graphical user interfaces and/or information retrieval systems and participated as "professionals."

2.2 Equipment

A Silicon Graphics Indy 100 MHz workstation, using software-based OpenGL, was used to test novice users. A Silicon Graphics Onyx workstation, using hardware accelerated OpenGL, was used to test visualization professionals.

2.3 Experimental Condition

There were three experimental conditions in which the dimensionality of the basic information display varied while preserving as much common functionality as possible. As described in detail above, in the 3D condition, the document space is presented as set of concept cluster boxes placed on the surface of a sphere (Figure 1).

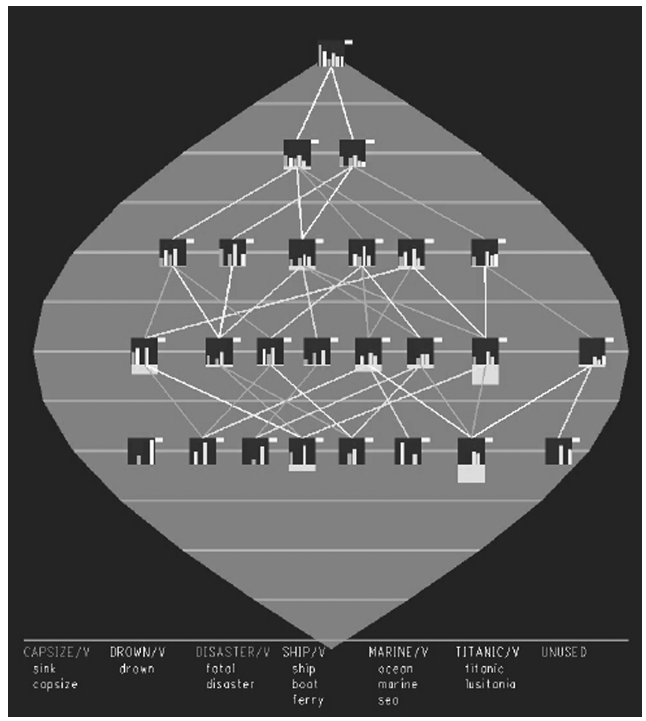

Figure 4. 2D NIRVE screen; the display is flattened version of 3D NIRVE

For the 2D condition, the globe was flattened into a map on which all clusters could be displayed simultaneously, as shown in Figure 4. Since there is no third dimension to convey cluster box density, this information is conveyed as the width of a gray bar located at the bottom of the box. Arcs indicating conceptual similarity are depicted as straight lines, and the field of document titles is simply drawn over the display of cluster icons.

In the text condition, documents were organized as a list grouped by cluster. Clusters were labeled with their colored concept profile; each cluster then was followed by its list of document titles, as shown in Figure 5.

In the 3D and 2D conditions, participants could zoom in or out of a cluster and could mark clusters and individual documents with colored flags. The number of documents within a cluster and the associated concept strengths were displayed at the top of the screen when the cursor was placed over the cluster box.

In the text condition, only individual documents could be marked using a red checkmark and only scrolling (not zooming) was available. The number of documents and associated concepts for a cluster were displayed as part of the scrollable text list at the beginning of each cluster.

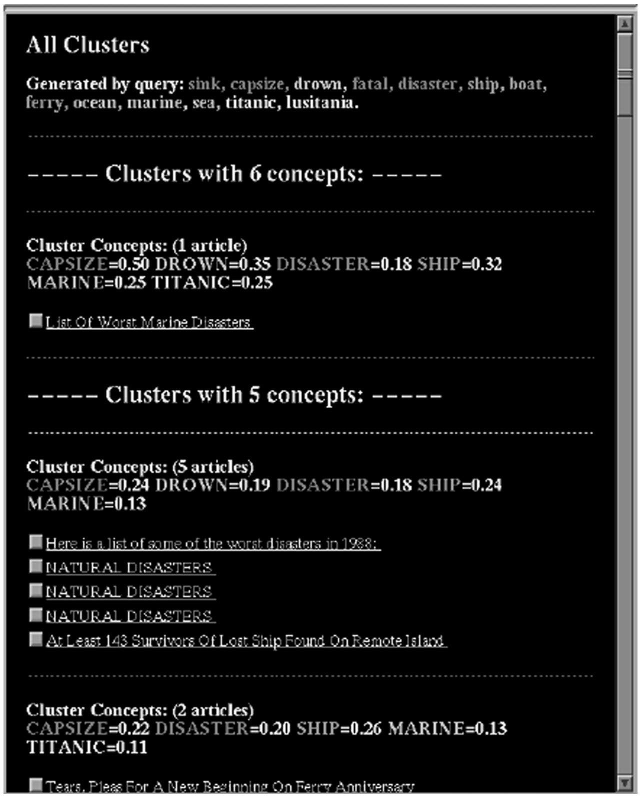

Figure 5. Text NIRVE screen with document titles grouped under cluster headings in a web browser list.

2.4 Topics and Tasks

A variety of topics and tasks were used in order to provide some generality of results. All experimental conditions used the same six topics: smoking bans, gene therapy, acid rain, aviation, gun control, agent orange. For each topic, participants were given a series of sixteen tasks (see Table 1) consisting of basic activities that are often included as part of more complex information analysis and retrieval. Participants had to locate, compare, and describe documents or clusters given specific types of information such as title, key concepts, or content.

2.5 Procedure.

A between-subjects design was used in order to maximize information on performance changes with system use within the given time constraints. Three novices and two professionals were quasi-randomly assigned to each of the three visualization conditions. All participants completed questionnaires providing demographic information and self-ratings of computer skills.

1. Locate and skim the article entitled, "On the science front." Mark as green the article and cluster box so that you can find it later.
4. Locate and skim the article that discusses where Aids came from. Mark as green the article and cluster box so that you can find it later.
8. Display in Netscape the article you marked earlier, "On the science front." Locate another article with a similar title.
13. Locate all clusters that discuss research and cure.
15. List in order the three-five main topics that are most frequently represented in the document space.

Table 1. Sample Tasks, (from a set of sixteen) Used in the Gene Therapy Topic

Participants completed two topics on each of three consecutive days. Each topic session lasted approximately 45 to 75 minutes. Visual and audio recordings were made of all experimental sessions.

Participants were informed that they would be using a computer program to assist in searching through Associated Press newspaper articles from 1988. They were given a brief description of the Document Space and Control windows (see above). The experimenter provided instructions and examples for each of the basic interaction techniques.

For each topic, the experimenter submitted the query and defined the concepts in order to maintain constancy across participants. Each of the sixteen tasks was presented individually. Participants were asked to "think aloud" or verbalize what was going through their minds as they were executing each task. All participant responses were timed. Participants were given three minutes to complete each task. If a task was not completed in the three-minute time period, a hint and an additional minute was provided to assist the participant. After each topic participants were offered a 5-minute break.

The first topic tasks were conducted as a training session guided by the experimenter. Like all other topic sessions, it consisted of retrieval followed by restructuring of keywords into concepts, followed by sixteen tasks. During this tutorial session, participants were allowed to ask questions and receive guidance from the experimenter.

Upon completion of all six sessions, a post-experimental questionnaire asked the participants to list what they liked and disliked about the tools they had used.

3. Results: Quantitative

Both quantitative and qualitative results were collected. The quantitative data consist of speed of task completion. This measure was selected since, given a restricted number of documents and enough time and motivation, any task could be completed successfully. The restricted time frame focuses on the ability to complete a task efficiently and avoids some of the potential problems with self-imposed time limits. Tasks that were not completed by the initial deadline were assigned a time of three minutes for purposes of analysis.

Preliminary analyses indicated that the majority of participants failed to complete a task requesting "all relevant articles" on a topic, so the data from that task were dropped from subsequent analyses. In addition, there were inconsistencies in the way participants dealt with the tasks of identifying the major topics, so these were also dropped from subsequent analyses. The remaining thirteen tasks were grouped into eight types, as shown in Table 2, based on the main component activities.

3.1 Overall visualization effects

Overall, the text condition showed the fastest response times. The reliability of these differences is demonstrated by the standard F-statistic with 2 and 9 degrees of freedom, $F(2,9) = 8.767$, $p = .008$.* The following sections outline how this global result may be due to prior familiarity with the alternative approaches to information access, interface constraints, and task characteristics.

3.2 Experience with the system

The differences among text, 2D, and 3D performance could be, in part, a function of experience. As shown in Figure 6, the greatest improvement in response time over the course of the experiment was for the 3D condition (linear contrast interaction: $F(2,9) = 45.35$, $p < .01$). Some improvement was also found for the 2D condition, with a slight decrement in performance for the text condition. The effect of limited 3D experience is also evident in the contrast between the first tutorial session, and the second session which participants completed without assistance, as shown in Figure 4. Whereas text and 2D performance improved during the second session, 3D performance declined.

A. Locate a cluster given its concepts

B. Locate a document given its concepts & title

C. Locate a document given its concepts

D. Recover a document given its title and follow a link to a new document

E. Recover a document and locate a new document given its title or content

F. Recover and compare contents of documents

G. Determine the concepts for a document and locate it given its title

H. Determine the concepts for a document and locate it given its contents

Table 2. Task Type Groupings & Component Activities

3.3 Testing Groups / Machine Properties

The pattern of responses was similar for the two experimental groups. However, for the "professionals" there was no mean response time difference between the 2D and 3D conditions, whereas there was a difference of almost forty seconds for the "novices", probably due to the substantially slower machine. This large difference between groups is not reliable, $F(1,9) = 0.636$, $p = .446$, in part reflecting the substantial variability in performance.

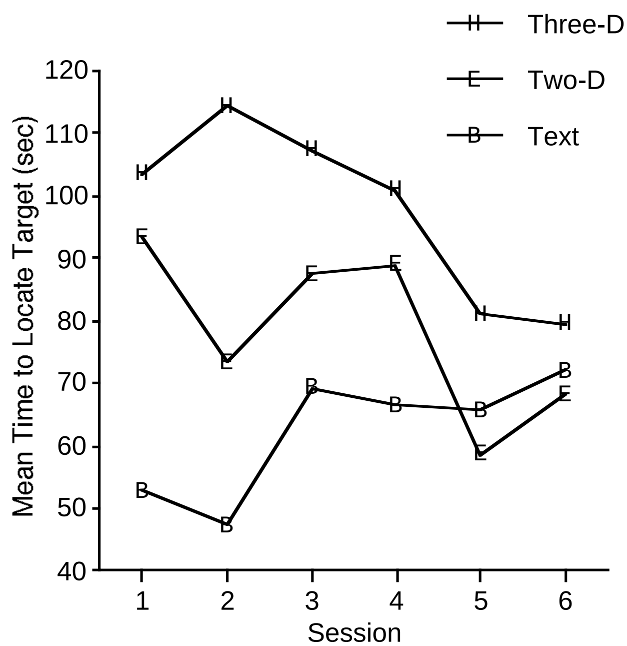

Figure 6. Time to locate target by Session and Visualization Method collapsed across Task Type

3.4 Task Type

There were reliable differences in response time among the different types of task as depicted in Figure 7. Overall response time depended on task demands for all conditions. For example, when a task required locating a cluster based on specified concepts, as in task type A, responses were relatively fast and accurate. However, when the participants had to determine which clusters to access by extracting concepts from a title (G) or a content description (H), response time was substantially longer.

The relative effectiveness of different visualization techniques also depended upon the task type, $F(14,63) = 2.253$, $p = .015$. Overall, this interaction seemed to be due to the match between task characteristics and interface properties. For example, in searching for a title (G), participants in the text condition could just scroll through the entire list of titles, whereas those in the 2D and 3D condition had to first find the appropriate cluster or grouping.

4. Results: Qualitative

In addition to task performance, several qualitative measures were used. First, the experimenter observed and recorded sequences of actions of each user. These were also videotaped and reviewed to augment the description of performance. Second, participants were asked to say aloud what they were thinking. These were combined into a set of strategies used for each task. Finally, at the conclusion of the study, participants were asked to list what they liked and disliked about the visualization tools used in the experiment.

The outcomes of those assessments focused on a few basic issues: grouping of documents into clusters, coding of concepts by color, visibility and legibility of displays, speed and complexity of system response, and the ability to maintain a sense of spatial location. The consequences of each of these factors for performance are described below.

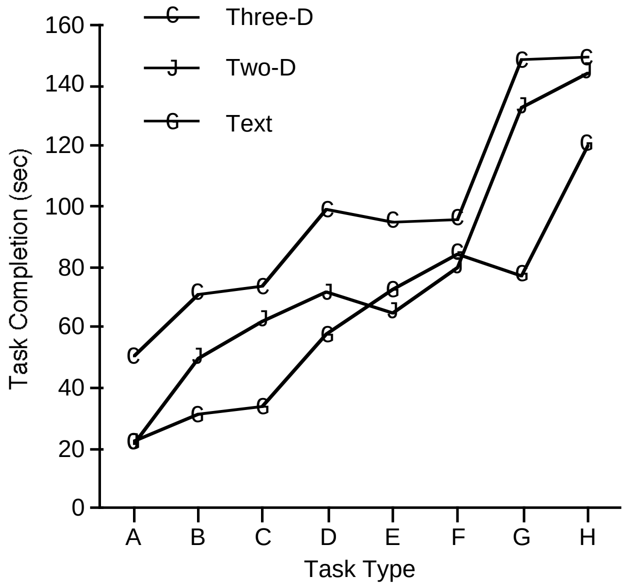

Figure 7. Time to locate target by Task Type (identified in Table 2) and Visualization Method collapsed across Session

4.1 Cluster Grouping

Overall, participants' reports and performance suggested that they understood and generally liked the organizational aspects of NIRVE including clustering of documents and the relational arrangement of clusters.

Participants used the grouping of concepts into clusters to narrow their search for particular documents. If a particular concept was not of interest, the participant knew which set of documents to avoid. The grouping also contributed to the selection of potential documents because it showed concept combinations that might not have otherwise been considered. In a number of cases, participants did not immediately appreciate what it meant to be grouped into a cluster. As a consequence, for example, during the first session or two participants often had difficulty with a task that required locating similar articles. That task was substantially easier once the search could be confined to a single cluster or a few neighboring clusters.

The relational structure of the clusters in the 2D and 3D displays was also used to keep track of preferred clusters. The vertical placement of clusters according to the number of concepts helped users adopt the strategy of linking up or down the space depending on their needs of adding or subtracting concepts. Many 2D and 3D participants would start from one pole of the globe and navigate through various links. In a number of cases, they began with the lowest level containing the minimal number of "potential" concepts required to find a document, and then worked their way up the globe or map until they found a matching document.

4.2 Color Coding

The most frequently used feature of the NIRVE interface was color. Users in all three conditions took advantage of color-concept mapping. The text condition benefited the most from this dimension, making this otherwise tedious list more efficient than anticipated. Instead of skimming or quickly reading the list of concepts at the beginning of each cluster, participants adopted the strategy of scanning the associated colors. This strategy is efficient because visual scanning of color, an automatic process, takes less time and effort than scanning words.

The text users' ability to understand the information "space" was also enhanced by the frequency of appearance of the color-coded concepts. When asked to identify the most frequently represented concepts and their spatial layout, users indicated, through their verbal protocols, that it depended on which color appeared the most in the list. Scrolling quickly up and down the list was sufficient enough to extract this information.

The strategy of using color for quick identification was adopted among the 2D and 3D users as well.

4.3 Visibility / Legibility

A critical feature of each display was the ability to easily identify concepts and read titles. In the text condition, this was rarely a problem since the concept labels and titles remained a fixed size. In the 2D and 3D conditions, in contrast, there were substantial problems with identification. As more concepts were added, the number of clusters increased, and their size on the screen for a given view decreased, making it more difficult for participants to distinguish colors in the cluster profiles. In general, the ability to use color-coding for searches decreased markedly once the number of concepts exceeded about five.

Legibility of concepts at the bottom of the document space window was also a problem. The size of the text changed with the number of items. Beyond roughly eight concepts, it became difficult to read the keywords and concept labels. A similar legibility problem occurred for the document titles. All 2D and 3D participants had difficulty reading the article titles that projected from the cluster boxes. Users dealt with this problem by clicking on the article title to bring it up in Netscape. This clicking required additional time for any task, since participants had to call up and read each title individually in Netscape instead of scanning the title as is intended by the design. There were times when users forgot a previously read article title and would have to click on it again to bring it up in Netscape to access the legible version.

These results point to the more general phenomenon of visual real-estate tradeoffs. As more space is given to visualization, as in the current 2D and 3D display of clusters, less space is available for textual displays. These can be addressed by a new level of tradeoffs in degree of overlap, sequential displays, and a variety of other options, but each such solution requires its own cost-benefit analysis.

4.4 Interface Speed And Complexity

As the demands on the display and interaction increase, the speed of the machine becomes a more dominant factor in a user's performance. The effects of slower machine speed for the novices were noticeable primarily for the 3D interface. There was always a delay or lag between the physical manipulation of the spaceball and the "corresponding" movement of the sphere. In a number of cases, participants exerted too much force on the spaceball (hoping to speed up the process) inadvertently causing it to rotate out of control; this often caused those participants to "get lost in the space." Participants often expressed frustration and adopted strategies for task execution that required less updating of the document space. In a number of instances, participants would reset the view of the document space because they lost track of their current procedure while focusing on their inability to control the sphere.

Delays in system response also impeded performance. When participants clicked to zoom a cluster, the motion was not always apparent. Attempting the zoom a second time caused the motion to stop, requiring the participant to start the process over again. In other instances, participants did not press a mouse button hard enough, but interpreted the lack of response as the typical system delay.

The number and complexity of actions was also an issue. Scrolling and mouse-clicking are familiar text-based interaction techniques. In contrast, the 3D interaction techniques - using a spaceball for navigation and using the mouse for opening, zooming, centering, and marking documents and clusters - were more demanding, especially among novice users. Most participants had no problems grasping the conceptual aspect or layout of the document space, but their ability to work in that space was hindered by the difficulty associated with maneuvering through the document space.

4.5 Maintaining Spatial Location

One of the functions of 3D is to provide a means of accessing larger sets of information that cannot readily be captured in a single screen. All participants seem to grasp this functionality. However, once rotation of the sphere was initiated for access to specific information, some participants became confused and unable to efficiently navigate the space. Remembering the spatial relationships that were no longer in view was one difficulty. The other was tracking relative location on the sphere. Pilot testing had suggested that this might be a problem, so we placed an alphabet sequence around the globe's equator. However, most participants tended to ignore this potentially useful set of "landmarks" in their task performance.

5. CONCLUSIONS AND FUTURE WORK

The utility of visualization techniques derives in large part from their ability to reduce mental workload. The results of this study suggest that such reductions are dependent upon an appropriate mapping among the interface, the task, and the user. In the case of the interface, for example, pilot studies had demonstrated that a mouse provided reasonably low mental workload for navigation in text or 2D modes, but produced a high load for 3D navigation. A spaceball dramatically reduced the cognitive load for the 3D condition because of its natural mapping to our 3D sphere. However, when the spaceball was used with the slower machine, delays in responses resulted in greater numbers of participant errors and more frustration.

User experience also influenced performance. The 3D condition showed the greatest decrease in response time during the experiment, and after six sessions, 3D response time was comparable to 2D and text conditions when machine speed was not an impediment. In order to evaluate the effect of user experience more broadly, we reanalyzed the data, splitting participants into two groups based on whether they were above or below the mean in self-reported computer experience (Figure 8). Computer skills were relatively unimportant for text, they mattered somewhat for 2D, and they mattered a great deal for 3D. In fact, for those participants with greater computer skills, the 3D condition resulted in more rapid response times than the 2D condition. The combination of these results suggests that 3D visualization cannot be adequately evaluated using only short-term studies of novice users.

Participants also tended to structure the tasks in ways that would decrease cognitive load. In our studies, color-coding turned out to be an especially useful tool across conditions and tasks. It provided a quick means to access clusters of documents. In contrast, zooming in to read a document title tended to be time-consuming on the slower machine. So, instead of using zooming, participants would simply open the entire article in a separate window just to read the title.

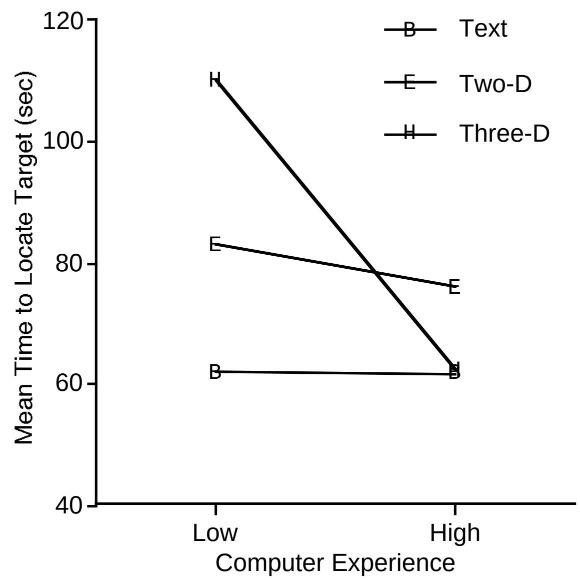

Figure 8. Response Time for the three interface conditions split by below (low) and above average (high) self-rated computer skills.

The perceived load of the interface was dependent not only on the functionality, but also on the extent to which those functions matched users' expectations, or what we might call interface "fidelity." In the display of documents in the 2D and 3D conditions, for example, columnar alignment was used to indicate title similarity. A number of participants initially tried to use columnar matching to identify similar titles. However, in several instances, although the most closely aligned titles did show the greatest "relative similarity," they did not match participants' expectation of similar content. As a consequence, those participants tended to abandon the intended aid and scan titles for meaning instead.

Overall the available features and the required tasks seemed to matter more than the dimensionality of the visualization. Color-coding of concepts helped to provide grouping regardless of the visualization tool. Searching for a title through a text list was often easier than having to locate a title by first identifying a cluster. In contrast, if documents were linked in "neighboring" clusters, the 2D and 3D tools were easier than a text-based search. It is also important to remember that the current evaluation was limited to 100 retrieved documents. With larger sets of 500-1000 documents scrolling and scanning through text becomes impractical. In that case, 2D and 3D visualization may produce a larger relative benefit.

Finally, this study, as others, is in many ways constrained by specific visualization tools. In order to generalize the results to other information retrieval applications, a more principled approach to design and evaluation is needed. In the present study, we focused on controlling for effects of functionality across alternative visualization tools; the results suggested that in at least several cases, the functionality was more consequential than the visualization. In addition, we need to isolate the mental workload imposed by our visualization tools for each of our tasks; preliminary task analyses of this sort, matching models and performance, are promising [10]. In order to build on this approach, tasks need to be constructed that more clearly isolate perceptual, cognitive, and motor components, and additional data needs to be collected demonstrating the relationship of those components to the effectiveness of visualization tools.

6. ACKNOWLEDGMENTS

This research was supported in part by NIST contract 60NANB6D0126 and is not subject to copyright.

7. REFERENCES

[1] Allen, R.B., Obry, P. Littman, M. An Interface for Navigating Clustered Document Sets Returned by Queries. Proceedings of SIGOIS 203-208, Milpitas CA, June 1993.

[2] Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I., Brown, C. VR-VIBE: A Virtual Environment for Co-operative Information Retrieval. Eurographics '95, Maastricht, The Netherlands, August 1995, 349-360.

[3] Cugini, J., Piatko, C., Laskowski, S. Interactive 3D Visualization for Document Retrieval. Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation, ACM Conference on Information and Knowledge Management (CIKM '96), November 1996. See http://zing.ncsl.nist.gov/~cugini/uicd/viz.html.

[4] Cugini, J., Laskowski, S., Piatko, C. Document Clustering in Concept Space: The NIST Information Retrieval Visualization Engine (NIRVE). CODATA Euro-American Workshop on Visualization of Information and Data, Paris France, June 1997. See http://zing.ncsl.nist.gov/~cugini/uicd/concept-clusters.html and http://www.nist.gov/itl/div894/vvrg/VACD.html.

[5] Fowler, R., Wilson, B., Fowler, W. Information Navigator: An information system using associative networks for display and retrieval, Report NAG9-551, No. 92-1, Univ. of Texas - Pan American, Edinburg, TX. See http://bahia.cs.panam.edu/info_vis/inf_nav/info_nav_tr_92.html.

[6] Hearst, M., Karadi, C. Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy. Proceedings 20th Annual International ACM/SIGIR Conference, Phil PA, July 1997. See: http://www.parc.xerox.com/istl/projects/ia/papers/cac-sigir97/sigir97.html.

[7] Kominek, J., Kazman, R. Accessing Multimedia through Concept Clustering, CHI'97 Conference Proceedings, Atlanta GA, March 1997.

[8] Lamping, J., Rao, R., Pirolli, P. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. CHI '95 Proceedings: Conference on Human Factors in Computing Systems: Mosaic of Creativity, Denver, CO, May, 1995.

[9] Nowell, L.T., France, R.K., Hix, D., Heath, L.S., Fox, E.A. Visualizing Search Results: Some Alternatives to Query Document Similarity. Proceedings of SIGIR'96, Zurich Switzerland, August 1996.

[10] Miller, M., Sebrechts, M., Vasilakis, J., Laskowski, S., Cugini, J. Preliminary Task Analysis of a 3D Information Visualization Tool: Implications for Interface Design. Unpublished manuscript, 1999.

[11] NIST PRISE Search Engine: http://www.itl.nist.gov/div894/894.02/works/papers/zp2/zp2.html

[12] Rennison, E. Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes. Proceedings of UIST'94.

[13] Swan, R.C., Allan, J. Aspect Windows, 3D Visualizations, and Indirect Comparisons of Information Retrieval Systems. Proceedings of the 21st Annual International ACM/SIGIR Conference, Melbourne Australia, August 1998.

[14] Text Retrieval Conference http://trec.nist.gov

[15] Veerasamy, A., Belkin, N. Evaluation of a Tool for Visualization of Information Retrieval Results. Proceedings of SIGIR'96, Zurich Switzerland, August 1996.

[16] Veerasamy, A., and Ileikes, R. Effectiveness of a graphical display of retrieval results. Proceedings of SIGIR'97, Philadelphia, PA, 1997, 236-243.

[17] Wise, J. The Ecological Approach to Text Visualization. JASIS, Sept. 1998.

*Footnote: All statistical analyses are based on standard Analysis of Variance techniques. This approach determines the ratio of variances based on the effects of interest to variance associated with error. The format is F(degrees of freedom for effect, degrees of freedom for error) = value of F for the data, p = probability that a value as large or larger than the obtained F could occur by chance alone. A probability value of less than .05 is normally taken as indicating that the effect is reliable or not due to solely to chance.