



Clustering-Based Visual Interfaces for Presentation of Web Search Results: An Empirical Investigation

Ozgur Turetken*

*Fox School of Business and Management, Temple University,
1810 N. 13th Street, Philadelphia, PA 19122
E-mail: turetken@temple.edu*

Ramesh Sharda

*College of Business Administration, Oklahoma State University,
Stillwater, OK 74078
E-mail: sharda@okstate.edu*

Abstract. *The result of a typical web search is often overwhelming. It is very difficult to explore the textual listing of the resulting documents, which may be in the thousands. In order to improve the utility of the search experience, we explore presenting search results through clustering and a zoomable two-dimensional map (zoomable treemap). Furthermore, we apply the fisheye view technique to this map of web search clusters to provide details in context. In this study, we report on our evaluation of these presentation features. The particular interfaces evaluated were: (1) a textual list, (2) a zoomable two-dimensional map of the clustered results, and (3) a fisheye version of the zoomable two-dimensional map where the results were clustered. We found that subjects completed search tasks faster with the visual interfaces than with the textual interface, and faster with the fisheye interface than just the zoomable interface. Based on the findings, we conclude that there is promise in the use of clustering and visualization with a fisheye zooming capability in the exploration of web search results.*

Key Words. *information overload, information visualization, human-computer interaction, user study, web search*

1. Introduction

The World Wide Web is the largest information repository, and is continuously growing in size and prominence. There is no strict structure in the content or format of information on the web, resulting in various forms of information overload. Information overload is the exposure of an information user to more information than she needs, and more importantly, is able to process. The overload problem is important due to its significantly adverse impacts on the use of information. Beyond a relatively low quantity of information, peo-

ple will begin to filter out information that they use to make decisions, increasing the probability of bypassing important information (Jacoby, 1984).

In this study, we address a specific web-related information overload problem, one that occurs while performing a typical search on the web using a search engine. Typically, a search engine presents its results as a ranked-list based on a relevance score for each document. For broadly formulated search queries, such a list may contain thousands of documents. Previous research has suggested that search engine users are not likely to go beyond the top 20 to 30 documents on these lists before getting bored or frustrated, and subsequently quitting the search (Allan et al., 2001; Amento, Terveen, and Hill, 2000a; Amento et al., 2000b; Bharat and Henzinger, 1998; Chen and Dumais, 2000; Mukherjea, Hirata, and Hara, 1998; Pratt and Fagan, 2000; Pratt, Hearst, and Fagan, 1999; Roussinov and Chen, 2001). Given that more and more people are using search engines for a variety of information needs, there is a need for research to enhance the ranked-list presentation of search results.

This paper investigates the combined use of clustering and visualization in the presentation of search results as a potential remedy to the overload problem inherent in the ranked-list presentation. We introduce a system that groups search results, and then presents these groups by means of a two-dimensional zoomable map. The idea of using clustering and visualization

*Author to whom correspondence should be addressed.

together for exploring web search results is not new, (for example Roussinov and Chen, 2001), but there is a dearth of rigorous empirical studies for the identification of desirable features in such visual presentations of web search results (Dumais, Cutrell, and Chen, 2001). This opinion is shared by practitioners who focus on usability testing of recent and popular visualization systems, such as Inxight's Star Tree (Hane, 2000). We believe that different implementations of the concept combined with rigorous user studies with varying groups of users and with different task settings are necessary. This leads us to address the following research question through an empirical study:

Can a clustering-based zoomable map of web search results improve search success over the textual ranked-list presentation?

To address this question, we developed a prototype system which presents an overview of search results that are hierarchically clustered based on document contents. This overview summarizes the document collection, and allows its viewers to quickly recognize certain patterns such as the number and size of the main groups that emerge (Greene et al., 2000). Based on this recognition, the searcher can determine what groups may be of higher interest, and focus on those. We propose that this approach, where the searcher can make a quick decision as to which documents to eliminate from further exploration, would reduce overload and hence provide better information access than a ranked-list.

The result of clustering is a hierarchical organization. It is known that when people explore hierarchies such as the contents of a web site (Bederson et al., 1998), switching between different levels of the hierarchy may be disorienting and confusing. This feeling of "being lost in the information space" could be eliminated if the interface provided an idea of how the immediate details fit in with the overall context of information. This argument is theoretically supported by the gestalt theory in psychology, for example (Wertheimer, 1924), which emphasizes that the whole is larger than the sum of its parts, and parts (details) are more meaningful when viewed within the whole (the context). In visualization of search result clusters, where the boundaries between different parts of the visualization are imposed rather than natural, it is desirable to examine local detail in its global context (Dumais, Cutrell, and Chen, 2001).

We believe that a fisheye view (Furnas, 1986) that shows immediate (i.e. zoomed-in) areas of interest in

full detail while providing a summary of the out-of-zoom information is a good way to present details in context. A fisheye view is helpful in elimination of the disorientation caused from switching between different views of an information collection, because it provides a continuous view of the context. Meanwhile, it still controls the information overload by gradually deemphasizing further details of the context. Therefore, the second research question we address is:

Can a clustering-based zoomable map of web search results with fisheye zooming improve search success over the zoomable map without fisheye zooming?

To summarize, the objectives of our research are:

1. To develop an interface that applies a clustering-based visual presentation (with or without fisheye zooming) for displaying web search results,
2. To determine if a graphical display (zoomable map) of clustered web search results is better than the ranked list, and
3. To determine if the fisheye zooming can enhance search success over the zoomable map with no fisheye zooming.

This paper focuses primarily on the last two objectives. The scarcity (Roussinov and Chen, 2001; Zamir and Etzioni, 1999), and the practical difficulty (Allan et al., 2001) of empirical studies on the usefulness of visual presentation of information, especially in the web domain, reinforce the contribution of the paper. The rest of the paper is organized as follows. In Section 2, we review previous work in the study of the interface features we are applying in our study. Section 3 introduces the interfaces that are tested. Section 4 presents the research hypotheses based upon previous theoretical work in user interface success. Section 5 explains our methodology in testing a proof of concept system for our proposed interface features. Discussion of the statistical analyses can be found in Section 6. The paper concludes with a discussion of the results and directions for future research.

2. Related Work

The focus of this study is the exploration of web search results. Arguably, once search results are retrieved, the exploration of this collection is a browsing task. Therefore, the result of the search process can be improved through superior browsing of the results.

Though there have been considerable improvements in presentation of a list of results obtained through a search (ranking, color coding, text size, etc.), our focus is on investigating the value of a clustering-based visualization of these results. The following three subsections review related work in three areas: clustering of search results, visualization of document collections with an emphasis on search results, and the concept of fisheye views for the visualization of clustered information. Table 1 provides a summary of the representative work in these domains.

2.1. Clustering based presentation of search results

There is very little structure in the list display of web search results, making it difficult to browse them (Allan et al., 2001; Amento, Terveen, and Hill, 2000a; Amento et al., 2000b; Bharat and Henzinger, 1998; Chen and Dumais, 2000; Mukherjea, Hirata, and Hara 1998; Pratt and Fagan, 2000; Pratt, Hearst, and Fagan 1999; Roussinov and Chen, 2001). According to the cluster hypothesis, mutually similar documents will tend to be relevant to similar information needs (Van Rijsbergen, 1979). Hence, clustering would impose a structure on the collection of search results, increasing browsing efficiency.

Cutting et al. (1992), and Hearst (1995) were some of the early adopters of this idea for developing a clustering-based browsing method called "Scatter/Gather". This method is directed towards a focus set of documents that are potentially interesting to the user. The focus set is clustered into smaller subsets and summarized to form an outline from which the user can select a smaller focus set. The indicated subcollection becomes the focus set, and the process repeats. Pirolli et al. (1996), and Hearst and Pederson (1996) tested the Scatter/Gather interface by means of an empirical study, and concluded that successful use of clustering can increase the effectiveness and efficiency of information search.

In another study, Allen, Obry, and Littman (1993) applied Ward's hierarchical clustering algorithm (Everitt, 1977) to search results. Through informal experiments, they concluded that clustering is useful for retrieval when queries have multiple parts or include terms with multiple meanings. On the other hand, their clustering-based presentation did not lead to improvements when the quality of the result set is not good enough, or is too good to require further processing.

More recently, researchers have worked on other aspects of applying clustering to search results. For exam-

ple, Mukherjea, Hirata, and Hara (1998) used clustering in processing the results of a WWW image search engine. Their clustering approach was based on multiple attributes of the image files such as their URLs, key words associated with the image files, and the actual images. The authors do not report on any usability study associated with this system. Zamir and Etzioni (1999) proposed a method, Suffix Tree Clustering (STC), for improving the speed of clustering algorithms for on-line systems by using snippets instead of full web documents. Based on the analysis of the server logs, the system (Grouper) that was built based on STC led to more efficient searches in terms of speed and the number of documents examined. Roussinov and Chen (2001) implemented a method that takes advantage of particular mathematical representations of documents for faster clustering. Due to this increased speed in clustering, this system could be highly interactive, and lets its users refine their searches multiple times. Through a controlled user study, the authors showed that this approach (Adaptive Search) leads to speed improvements, and is preferred by its users.

Other researchers have directed efforts to base clustering on the link structures between documents (Amento, Terveen, and Hill, 2000a; Bharat and Henzinger, 1998; Chen et al., 1999). The Cha-Cha system described in Chen et al. (1999) creates a "table of contents" structure for the collection of search results. The authors report that Cha-Cha leads to slight improvements in understanding of the collection, search speed, and user satisfaction. It was also preferred by its users over the ranked-list presentation. The (Bharat and Henzinger, 1998) study reports that their approach resulted in efficiency (precision) improvements, and the study found that their system's organization of web pages agreed with human experts. There have also been approaches to support information seekers by facilitating manual grouping of web documents (Amento et al., 2000b).

The usability tests performed in these studies verify that summarization of search results by clustering similar pages together could help in reducing overload due to the imposed structure on the document collection. The value of clustering in presenting and examining a document collection has also been recognized by commercial search engines such as NorthernLight (<http://www.Northernlight.com>), and vivisimo (<http://vivisimo.com>). These search engines present search results through an automatically generated

Table 1. Summary of representative work

System	Reference	Basic approach	Empirical testing
Adaptive search	(Roussinov and Chen, 2001)	SOM based clustering and visualization of search results	User study (36 subjects) found speed improvements and user preference over ranked list
CategoryMap	(Chen et al., 1998)	SOM based clustering and visualization of documents	User study (34 subjects) found method was inferior to Yahoo
Cha-Cha	(Chen et al., 1999)	Table of contents structure based on links between search results.	User study (18 subjects) found slight improvements in understanding of the collection, search speed, and user satisfaction.
CZWeb	(Collaud et al., 1995)	Hierarchy to display some web pages in great detail and others in less detail or no detail at all.	Not reported
Fisheye Menus	(Bederson, 2000)	Application of a fisheye distortion to computer menus	Pilot study (10 subjects) found users preferred fisheye menus for browsing tasks, hierarchical menus for goal-directed tasks
Generalized Fisheye Views	(Furnas, 1986)	"Details in context" view of various structures	User study (20 subjects) found that fisheye views were superior to "flat files" for navigation of a hierarchy
Graphical Fisheye Views	(Sarkar and Brown, 1992)	Formalized rules for fisheye views of a generic graph	Not reported
Grouper	(Zamir and Etzioni, 1999)	Clustering of search results based on the contents of their "snippets"	Analysis of server logs showed clustering led to improvement in the number of documents followed, amount of searcher time and effort
Hyperbolic browser	(Lamping and Rao, 1996)	Laying out the hierarchy uniformly on the hyperbolic plane and map this plane onto a circular display region	Not reported
Lighthouse	(Allan et al., 2001)	Visualization of inter-document similarities by means of a spring embedding algorithm	User study (20 subjects) showed that users were more successful with the visualization than they would be by following the ranked list
Motif X	(Allen, Obry, and Littman, 1993)	Content-based hierarchical clustering of search results, tree-visualization	Informal experiments found clustering useful when queries have multiple parts or include terms with multiple meanings, not useful when quality of results is not good enough, or too good
PDQ Tree-browser	(Kumar et al., 1995)	Separate presentation of overview and details, which are obtained by user queries and filtering.	Controlled experiment (24 subjects) showed that pruning significantly improved performance speed and subjective user satisfaction.
QRVE	(Mukherjea, Hirata, and Hara, 1998)	Clustering based on URLs, key words, and images and presentation of clusters by graphical elements (glyphs).	Not reported
Scatter/Gather	(Cutting et al., 1992; Hearst, 1995; Hearst and Pedersen, 1996; Pirolli et al., 1996)	Clustering and reclustering based on document contents	User study (16 subjects) found that Scatter/Gather alone was not more effective than ranked-list presentation, but would successfully complement it
Topic distillation	(Bharat and Henzinger, 1998)	Combination of content and connectivity analyses for finding "quality" documents in a search	No user study, test data showed improvement over basic connectivity analysis
TopicShop	(Amento et al., 2000b)	Visual representations of documents to help manual organization	User study (40 subjects) showed that users found it easier and faster to select better sites, create groups, and integrate these two activities
Variable zoom	(Schaffer et al., 1993, 1996)	Modification of the basic fisheye view approach for a two-dimensional hierarchical network to show detail while in the rest of the (uniformly scaled down) network	User study (20 subjects) showed that variable zooming led to significant speed improvement and effort savings.
Visual Site map	(Lin, 1997).	SOM based clustering and visualization of search results or a static set of documents	User study (sample size not reported) found speed improvement over traditional presentation of documents

(Continued on next page.)

Table 1. (Continued).

System	Reference	Basic approach	Empirical testing
VITESSE	(Nigay and Vernier, 1998)	Fisheye presentation of a collection of search results without any grouping	User study (17 subjects) found users preferred using the non-distorted view over the fisheye
WebBook and Web forager	(Card, Robertson, and W Y, 1996)	Manual organization of web documents into books and into collection of books	Not reported
WebSOM	(Honkela et al., 1997; Lagus et al., 1996)	SOM based clustering and visualization of a collection of web pages	Not reported
WebTOC	(Nation, 1998)	Creating a table of contents structure based on links and directory structure and visualizing this structure (overview) and individual web pages (details) simultaneously, but separately	Not reported
Zooming Web Browser	(Bederson et al., 1998)	Fisheye views of multiple web pages and the links between them instead of a view of one page at a time.	User study (30 subjects) found speed improvements over traditional display method of Netscape in several different information seeking scenarios
Fisheye views of Topological Networks	(Hollands et al., 1989)	Application of fisheye views to a fictitious subway network	User study (48 subjects) found that fisheye views led to slight speed improvements for large networks

cluster structure using proprietary implementations of text clustering algorithms (Salton, 1989).

2.2. Visual presentation of document collections

People usually have an easier time in understanding information when it is visually presented (Card, Moran, and Newell, 1986; Shneiderman, 1996; Tufte, 2001). As argued by Shneiderman (1996):

“The bandwidth of information presentation is potentially higher in the visual domain than for media reaching any of the other senses. Humans have remarkable perceptual abilities that are greatly under-utilized in current designs. Users can scan, recognize, and recall images rapidly, and can detect changes in size, color, shape, movement, or texture. User interfaces have been largely text-oriented, so as visual approaches are explored, appealing new opportunities are emerging.”

The basic cognitive tasks of information exploration such as scanning, sorting, and selection for which information visualization is proven useful are also performed in browsing a collection of (web) documents. Therefore, various researchers have explored visual presentations of a document collection. Most of the clustering-based systems discussed in the previous section have a visual presentation component. For example, the (Roussinov and Chen, 2001) approach is based on Kohonen’s self-organizing maps (SOM) (Kohonen,

1998), which creates two dimensional maps of the clustering structure simultaneously with the clusters themselves. On such a map, a large area usually represents a large or an important collection of information, and physical proximity usually implies semantic proximity. SOM based visualizations are not limited to web search results. Other similar systems for the web are WebSOM (Honkela et al., 1997; Lagus et al., 1996), the CategoryMap (Chen et al., 1998), and the Visual Site map (Lin, 1997).

The Query Visualization environment (QRVE) (Mukherjea, Hirata, and Hara, 1998) discussed in the previous section does not only provide clusters of search results, but also presents these clusters by means of glyphs (graphical elements). These glyphs represent both document clusters that are formed automatically and web sites that are organized by people. Similarly, the TopicShop system (Amento et al., 2000b) provides visual representations of each document in the collection to help its users with the manual organization of those documents. It also provides other visual support such as shading to make it easier to differentiate between groups.

A tree structure is a well-known way to represent a hierarchy. The hierarchy that results from document clustering as described in Allen, Obry, and Littman (1993) is visualized by means of a tree. Their Motif X-windows interface has a query window, a subtree document list, and a text window component in addition to

the interactive dendrogram (tree). In this interface, the overall hierarchy is static and details are shown in these separate windows using color clues. The PDQ Tree-browser (Kumar, Plaisant, and Shneiderman, 1995) is a similar system that was designed to help information in browsing a web. PDQ (Pruning with Dynamic Queries) is based on the following requirements: browse the entire tree and view at different levels, query nodes at all levels on the basis of attribute values, and hide uninteresting nodes and branches rapidly, and thus reduce the data set progressively. Both of these systems were designed to provide summarized context and detailed information on two separate graphs on the same screen. The main disadvantage of the tree-based systems is that the visual cues to smoothly connect the details and context are not very strong.

Similar to the Cha-Cha system discussed before, the WebTOC (Nation, 1998) system summarizes the contents of a web site by means of a table of contents structure. WebTOC uses two different strategies: following existing links, or using the underlying directory structure. The automatically generated table of contents provides graphical information indicating the number of branches as well as individual and cumulative sizes of these branches. WebTOC presents context (the web site) and details (the web page of interest) simultaneously, but separately. The problem with this separation is the difficulty to mentally connect the two graphs.

The Lighthouse system described in Allan et al. (2001) presents the results of a search by visualizing the inter-document similarities. The visualization is based on a spring embedding algorithm where the similarities between individual documents create a pseudo physical force, which in turn, determines the physical location of the documents. The Lighthouse system was tested through a usability study. The results were that the users were significantly more successful (in terms of precision) by the visualization than they would be by following the ranked list. A similar approach was followed by the KartOO search engine (<http://www.kartoo.com>), which uses meta-information (i.e. existing links between pages) to display how the results of a web search are related to each other. It also displays phrases that are extracted from the collection of search results to present other commonalities among the results.

Card et al., (1996) propose two moves from the traditional “one page at a time” display of web pages: a move from the single web page as the unit of interaction to a higher aggregate entity (the WebBook), and

a move from a work environment containing a single element to a workspace in which the page is contained with other entities, including WebBooks (the Web Forager). This is an interesting application of the idea of organizing web documents into groups and presenting a visual overview of these groups. However, the organization of pages in the Web Books and Web Foragers are done manually, and hence the scalability of the application for a very large collection of documents is questionable.

The value of visualization in enhancing one’s understanding of an information collection is clear, but understanding visual presentations of information is not without challenges. One of the most important challenges in the visual presentation of information is the trade-off between overload and disorientation (Chen et al., 1998). A fisheye view is a possible way to address this trade-off.

2.3. *Fisheye views*

As explained in Leung and Apperly (1994), there are a number of ways a fisheye view can be created. In his original work, Furnas (1986) demonstrated an application for tree structures and a specific example for tree structured text files. The results of his usability study showed that fisheye views are more effective than flat views in navigating hierarchically organized text. Sarkar and Brown (1992) applied the fisheye view technique for viewing and browsing computer graphs. They built a framework to incorporate arbitrary structures by redefining the concept of “distance” between graphical elements.

Lamping and Rao (1996) describe an implementation for presenting a two dimensional graph through a fisheye zoom. This application, named “hyperbolic browser” provides a smoothly varying “focus plus context” view where the display space allocated to a node decreases continuously with the distance from the focus, yet does not disappear abruptly. This application of fisheye views is one of the best known and successfully commercialized (Hyperbolic tree form Inxight[®], (Hane, 2000)). The integration of details within context by means of a spherical distortion is promising for looking at those details without losing awareness of the overall picture.

The CZWeb tool (Collaud et al., 1995) graphically displays a network in a rectangular two dimensional display space. A hierarchy is used to display some web pages in great detail and the others in less detail or no detail at all. This study is important in that it was one

of the first attempts to apply the fisheye view technique to the web to remedy the “lost in hyperspace” problem, which is important for web users.

Bederson et al. (1998) report on the development of a web-browsing prototype in their multiscale graphical environment, Pad++ (Bederson et al., 1996). This prototype displays multiple web pages and the links between them instead of showing one page at a time. A fisheye view approach is used where the page in focus is clearly readable whereas the others are shown in smaller scale to provide context. The authors compared their system to the traditional display method of Netscape in several different scenarios, and found that subjects using their system were 23% faster than those using Netscape in performing the same information seeking tasks. In another study, Bederson (2000) applied the fisheye view technique to menus in graphical user interfaces, and found that many of the users preferred such views to more traditional displays of menus.

Schaffer et al. (1993, 1996) describe a fisheye view approach (the variable zoom method) for the visualization of two-dimensional networks. This method magnifies lower levels of the network hierarchy to show detail while embedding this detail in the remaining, uniformly scaled down, network. The authors applied this method to a simulated telephone network, and compared its success to that of a traditional view of the same network. The results showed that the users of the fisheye-based display spent less time and effort in navigating and rerouting the links of the network. A similar study based on a fictitious subway network reported in Hollands et al. (1989) also found that fisheye views affect the speed of locating elements in the visual display, especially when the target element is not in the initial display.

To our knowledge, the only system that has attempted fisheye view visualizations of web search results is VITESSE (Nigay and Vernier, 1998). This system presents a visual display of search results without any link-based structure or any content-wise organization such as clustering as in the systems discussed in Section 2.1, or categorization into predetermined groups as discussed in Pratt and Fagan (2000) and Pratt, Hearst, and Fagan (1999). Thus, its benefit in reducing information overload is limited.

As evident from the discussion in this section, fisheye view systems have found applicability in the visualizations of complex information collections. The smooth integration of context and details is promis-

ing, but also challenging, because each implementation needs to be modified based on the specific application.

Table 1 presents a summary of our review of related studies. An important observation in this review shared by other comprehensive reviews (for example, (Leung and Apperly, 1994)) is that most fisheye view systems developed so far have not been tested in rigorous usability studies. Many of the studies that report on usability testing utilized very small sample sizes, or only did pilot testing. Accordingly, not much is known about their actual usefulness in spite of their conceptual appeal. This observation also applies to the more generic web visualization systems discussed in Section 2.2. This study is an attempt to address this issue by empirically testing the usability of our visual interfaces. For this purpose, we derive testable hypotheses from the two research questions that are addressed in this study in Section 4. In order to clarify the comparisons in these hypotheses, we next describe the different interfaces we compare.

3. The Interfaces

Our experimental study compares three different interfaces. The first provides the traditional ranked list presentation of search results. Fig. 1 displays an example ranked-list from the AltaVista search engine. The “results page” from AltaVista is our “textual” experimental condition (The earlier versions of our visual prototype used AltaVista, therefore we used the results page from AltaVista as our textual condition to eliminate any effect that is attributable to the source of web pages. Our most recent implementation uses Google). The results are presented in their original presentation format. The literature (Allan et al., 2001; Amento, Terveen, and Hill, 2000a; Amento et al., 2000b; Bharat and Henzinger, 1998; Chen and Dumais, 2000; Mukherjea, Hirata, and Hara, 1998; Pratt and Fagan, 2000; Pratt, Hearst, and Fagan, 1999; Roussinov and Chen, 2001) suggests that a typical searcher rarely looks beyond the first few pages of the results. Therefore, the visualizations can work with a certain predetermined number of documents that have the highest ranking. We follow this approach by displaying the top 100 search results while also making sure that all experimental conditions present the same exact collection of documents.

We explore visual presentation of search result clusters instead of presenting each individual result. The



Fig. 1. Textual presentation of search results (AltaVista search engine) to the query "Jaguar".

clustering is based on the similarities between web documents where documents are represented as vectors of term frequencies for terms (words or phrases) that are obtained from the overall collection of documents resulting from a specific search. Once there is a vector representation for each document, clustering these documents boils down to the clustering of the corresponding vectors for which there are very well known approaches (Everitt, 1977). Modeling each document as a vector based on its content is a very common approach in document clustering, details of which can be found in Salton (1989). More examples were detailed in Section 2.

The clusters can be visualized in many different metaphors (Section 2). Our implementation presents an overview of clusters as displayed in Fig. 2. This overview is obtained by slicing the available map space vertically for the first level clusters where the width of each rectangle is proportional to the size (number of documents included) of the corresponding cluster. Coloring is used to clearly differentiate different regions from each other. The clustering algorithm labels each cluster by the most frequently occurring terms in that cluster. For better readability, the label for the cluster over which the user moves his/her mouse is displayed in the text window above the map.

There are two different versions of the interface that present further details of the clusters identified in the overview. One, the "full zoom" approach, presents the details of a cluster without any context while the other, the "fisheye zoom" approach, presents the same details within the context of the other clusters. Figs. 3 and 4 are the maps a user reaches when she clicks on the "Club, Jaguar" cluster in the visual overview of Fig. 2. Fig. 3 shows the full-zoom visualization. As seen in the figure, the only visible information in this visualization is the contents of the cluster in zoom without any context information. This view is basically the overview of the selected group of documents, so all the visual properties of the map in Fig. 2 apply to this map.

Alternatively, the fisheye visualization of Fig. 4 keeps a general view of the other clusters while showing the cluster of interest. Regardless of the region of the map that is currently in zoom, the user of the interface is constantly aware of how the details (lower-level clusters or individual documents) fit in with the overall collection of information. Here, the zoomed-in regions are magnified only to a level that they pass a predefined visibility threshold, and the out-of-zoom regions shrink accordingly. Therefore the relative size of a region still has a meaning among the in-zoom or out-of-zoom regions. However, the sizes are not

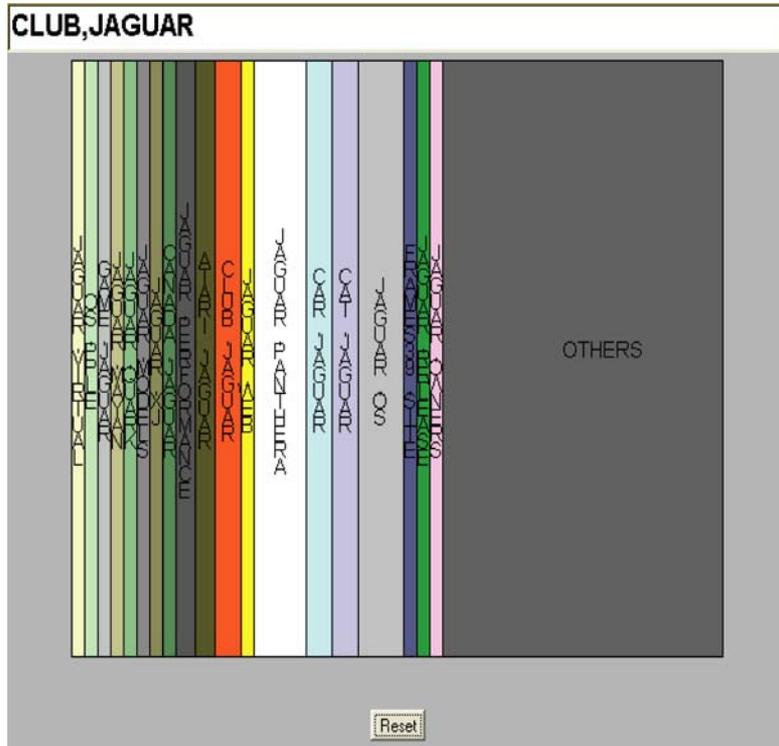


Fig. 2. Visual overview of the first level clusters for the query "Jaguar".



Fig. 3. "Full" zooming of the "Club, Jaguar" cluster.

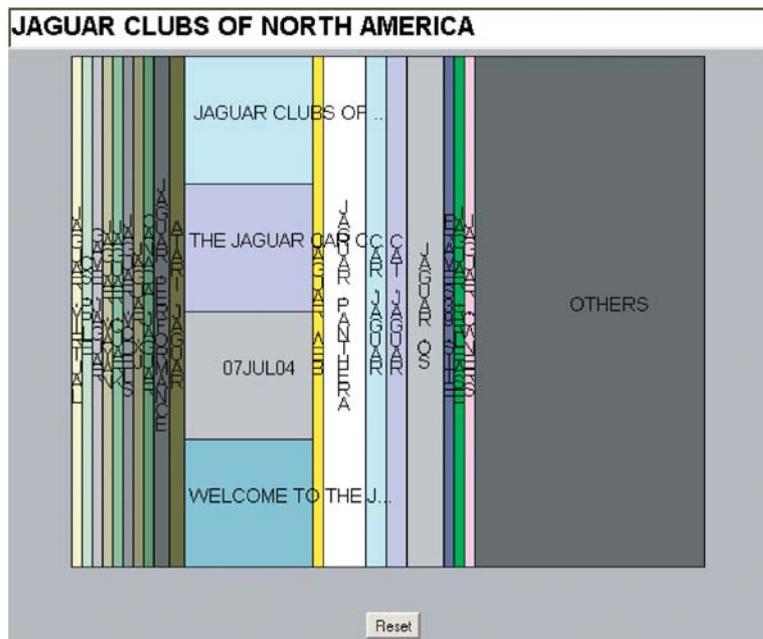


Fig. 4. "Fisheye" zooming of the "Club, Jaguar" cluster.

meaningfully comparable between in-zoom and out-of-zoom clusters. This sort of visual distortion is the basis of the fisheye technique.

Fig. 5 displays what the user would see when she clicks on the region labeled "Welcome to Jaguar Drivers' Club" on the map of Fig. 4. As seen in the figure, when the zoomed-in region is an individual web page, the page is displayed in a separate window. In case the document is irrelevant to the user's information needs, she simply closes the window. Likewise, to quit looking into an irrelevant cluster, the user can simply return to the overview diagram by clicking on the "reset" button, or simply click on a different region to see the contents of another cluster.

These visualizations are based on the "Zoomable Treemap" algorithm (Turetken and Sharda, 2004), which is our modification of the Treemap algorithm by Johnson and Shneiderman (1991).¹ The mentioned work presents other details of this technique.

The implemented prototype has a main module created by server-side scripting (Microsoft Active Server Pages), which prompts users for a query, and then sends that query to the search engine. Once the search results are retrieved, the main module parses the results, in-

dexes them, and saves them locally. Then it uses the hierarchical clustering routine of IBM's IntelligentMiner tool to cluster the documents. The hierarchical clustering in IntelligentMiner works on the document clustering principles discussed earlier in this section. To be specific, first the documents are represented as vectors using frequencies of terms in the collection as an index. Then the similarity between every document pair is determined by the cosine measure. Based on inter-document similarities, the most similar documents are combined into a cluster and the cluster is represented by the midpoint of the two vectors that represent each of the combined documents. The similarity calculations repeat with the new collection, and the pair with the highest similarity score is combined, and this process repeats itself until every document is assigned to a cluster.

The result of document clustering is a hierarchy structure, which is visualized through our system as described above. The algorithm was implemented through a Java applet that takes the index of documents and the hierarchical structure as inputs, and creates the interactive maps. Other details of this implementation along with the algorithms and design ideas can be found in Turetken and Sharda (2004).

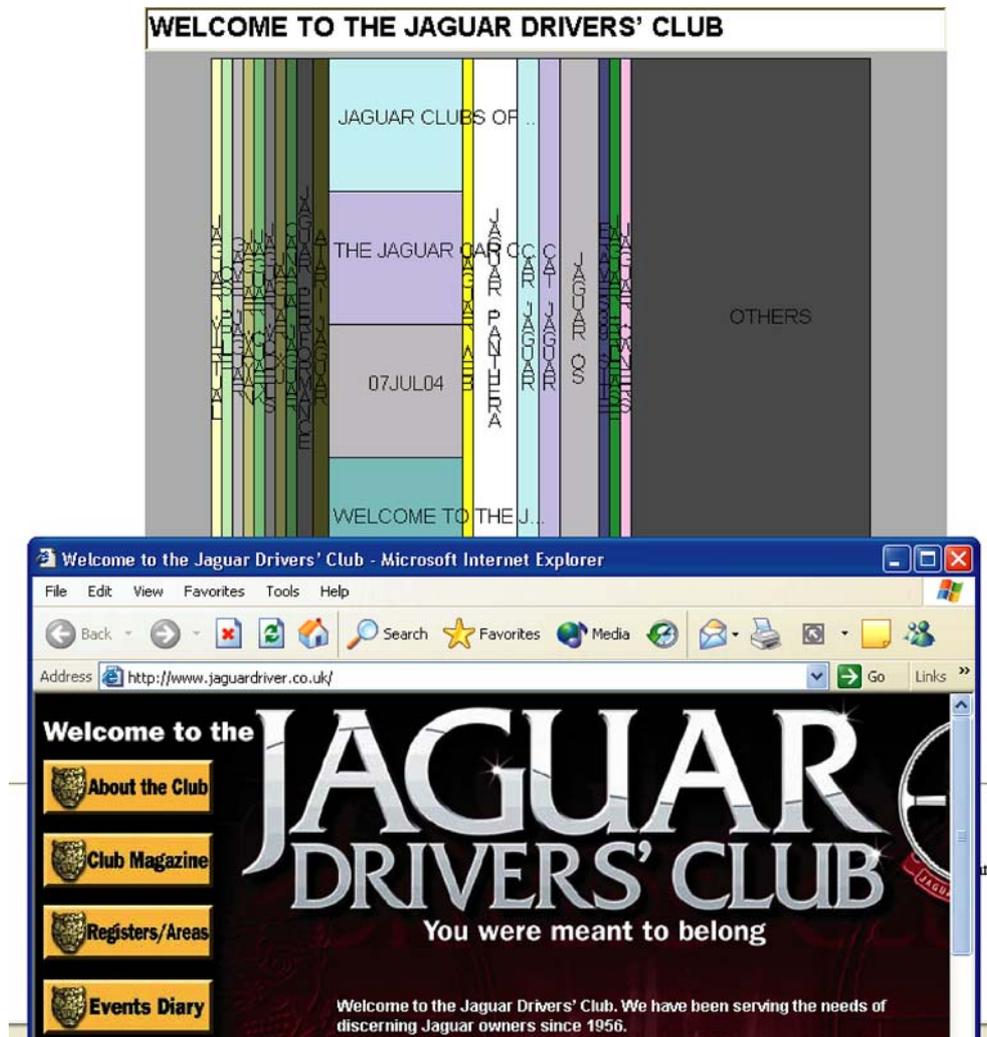


Fig. 5. Looking into the contents of the Web page titled "Welcome to Jaguar Drivers' Club".

4. Hypotheses

The independent variable of interest in this study is the interface for presentation of web search results, which can take three levels: textual, full zoom, and fisheye zoom. Previous work in information systems has identified a number of other factors that may affect the general success of the interaction between a user and the user interface. Among these are the amount of training on the interface (Suh and Jenkins, 1992), the characteristics of the task to be performed using the interface (for example its level of difficulty) (Suh and Jenkins, 1992), contextual variables (for example indi-

vidual characteristics and experience) (Santhanam and Sein, 1994), and interactions between some of these factors (for example "cognitive fit," (Vessey, 1991) and "task and interface match" (Tan and Benbasat, 1990, 1993)). In this study, these variables are either held constant (amount of training and task) or controlled for (contextual variables).

The success of the interaction between the user and the interface manifests itself in the performance of the user and her satisfaction with the system. There are two typical measures of end-user performance: effectiveness and efficiency. Effectiveness is a measure of how desirable the user's outcomes are, whereas efficiency

refers to how well she uses the available inputs (physical resources or time) in producing those outcomes.

We operationalize the dependent variables of effectiveness and efficiency as suggested by Roussinov and Chen (2001), Lim, Benbasat, and Todd (1996), Hertzum and Frokjaer (1996), Tan and Benbasat (1993), and Pratt and Fagan (2000) among others. The number of correct answers given to a set of objective questions having answers within the search results is our surrogate measure for effectiveness. The surrogate measure for efficiency is the time it takes to complete the task of answering these questions.

Following this discussion, we convert our research questions to the following hypotheses:

H1a: The visual interfaces will result in higher number of correct answers than the textual interface.

H1b: The fisheye-zoom interface will result in higher number of correct answers than the full-zoom interface.

H2a: The visual interfaces will result in higher speed in finishing the experimental tasks than the textual interface.

H2b: The fisheye-zoom interface will result in higher speed in finishing the experimental tasks than the full-zoom interface.

As a surrogate for end-user satisfaction, we use the scores obtained from a satisfaction survey administered to the subjects regarding each presentation. Subsequently, the following additional hypotheses are formulated:

H3a: The visual interfaces will result in higher user satisfaction (score) than the textual interface.

H3b: The fisheye-zoom interface will result in higher user satisfaction (score) than the full-zoom interface.

5. Experimental Design

5.1. Tasks

The motivation of this study is to test the ability of the developed interface in browsing the results of a search query that retrieves a large number of pages (hits). Therefore, our first criterion in designing the experimental tasks was to receive a large number of hits. At the same time, we did not want the personal traits

or backgrounds of the subjects to influence the results. Thus the search tasks had to be on general topics so that one group of subjects would not be more knowledgeable than another.

Shneiderman (1997) classified information search objectives into the following general categories:

1. Specific fact-finding (Searching *directly* for a readily identifiable outcome)
2. Extended fact-finding (Searching *indirectly* for relatively uncertain but replicable outcomes)
3. Open-ended browsing (Gaining an understanding of a general subject area)
4. Exploration of availability (Self explanatory)

In the web domain, the availability of material is subject to continuous change. Therefore, it is difficult to determine exploration of availability. In real-life information search activities, open-ended browsing is essential and inevitable. Conversely, in a controlled experiment, it is very difficult to measure the outcomes of a loosely defined objective such as “finding new work on voice recognition in Japan,” or “possible relationships between carbon monoxide levels and decertification.” This left us with fact-finding questions for testing the success of the interface. The answers to such questions can be found within the collection of search results. Yet, finding those answers, especially in a fast manner, requires the ability to effectively overview the document collection, and to focus on a specific part when needed. We believe that the visual interface, particularly with the fisheye zooming, will better facilitate the tasks of focusing and refocusing for answering fact-finding questions. Subsequently, the search tasks in this study were on specific or extended fact finding.

We designed three similar tasks: Task A, Task B, and Task C as displayed in Table 2.² This similarity allows us to control the “task” variable. These tasks are on general topics, and result in large number of search results. Figs. 6 through 8 show the three alternative presentations for the results of the second query of Task A.

5.2. Measurements

In order to eliminate the effect of the subjects’ ability to formulate effective search queries and the speed effect of network traffic at the specific times that the experiments were being conducted, the search queries were formulated and executed in advance (by using each of the experimental tasks as displayed in Table 2 in their entirety). Each experimental condition used the

Table 2. *The experimental tasks*

Task	Questions
A	1. Where can I get good filet mignon in Madison, WI?
	2. What was the population of Hong Kong in 1998?
	3. Find two other books by the author of Jurassic Park.
	4. Name three shows that took stage in Broadway in 1989
B	1. Find the names of two hotels in Kyoto (Japan) that are near the train station.
	2. How long does it take to get by train from Paris to Munich?
	3. I need a map of Kusadasi, Turkey.
	4. Where did William Shakespeare die?
C	1. I'm looking for the names of campgrounds around Lake Tenkiller (Oklahoma) that have showers.
	2. How many track medals did the country that had won the most gold medals in track in the 1972 Olympics win in the 1996 Olympics?
	3. What are the two most recent movies from the director of "Full Metal Jacket"?
	4. Which countries are the neighbors of Ukraine?

first 100 of the results that were obtained from these queries. None of the interfaces (text-based or visual) allowed the subjects to reformulate search queries. This way, we were able to attribute any differences in search success to the presentation style of the interface since the collection of results that was generated was the same for each interface. The speed of the search process depended only on the local computers, each of which had identical hardware and software configurations. This provided additional control in assessing the users' performance in the experiments.

Similarly, the effect of the subjects' past experience and knowledge on experimental results were reduced by asking them to provide us with the URL of the site in which they found the answers to the search questions rather than giving factual answers. After the experiments, these URLs were compiled and the duplicates were eliminated. There were a few potential sites that had relevant information for each question. All such sites with relevant information were considered correct answers, and others were considered incorrect. In other words, the accuracy score assigned to a web site was either 0 or 1. The score of each subject under each treatment condition was then a simple sum of the correct answers.

We adopted a multi-item scale from Stasko et al. (2000) for measuring satisfaction. The measurement of time was done by recording the beginning of each experimental phase, which was com-

mon for all subjects, and by automatically marking the time when a user submitted her answer to each question.

To measure web search experience we adopted a two-item scale (Wang, Hawk, and Tenopir, 2000) composed of the frequency and duration of search engine use. We used the "Group Embedded Figure Test" (GEFT) (Witkin et al., 1971) for determining cognitive style. This test is known to measure a very salient dimension of cognitive style, i.e. field dependence/independence (Witkin et al., 1971). Subjects of the GEFT are assigned a score between 0 and 18 depending on how many simple figures embedded in more complex ones they can identify. A higher score on the GEFT means that a subject is more "field-independent", i.e. he or she can better locate details in a large context.

5.3. Experimental procedure

We conducted controlled experiments with several groups of business students at a major university. The students were offered course credit for the completion of the whole experiment. We had a total sample of 78 subjects, of which 52.6% were male, and 47.4% were female. 61.4% of the subjects reported that they were between age 20 and 22, 19.3% were between age 23 and 25, and 19.3% were 26 or above.

Each experimental session started with the subjects filling out a questionnaire on demographic data such as age, gender, and native language. We also collected data on the subjects' web search experience. In addition, the subjects took the cognitive style test (GEFT).

The subjects were randomly assigned to one of three groups. Each subject underwent a training session resembling the experimental phases. For this purpose, we used a task similar to the ones that the subjects would actually be working on during the experiment. These sessions familiarized the subjects with the experimental procedure and the web-based forms through which they would submit their answers. After the training, each group of subjects underwent three phases of experimentation. The tasks involved browsing the search results to find a URL that would potentially answer the experimental questions. In each of the three phases, subjects were given tasks A, B, and C in the same order. The difference was in the presentation of search results. For example, for the first group, Task A was supported by the textual listing while Task B was supported by the full zoom system, and Task C was supported by the fisheye zoom system. The other

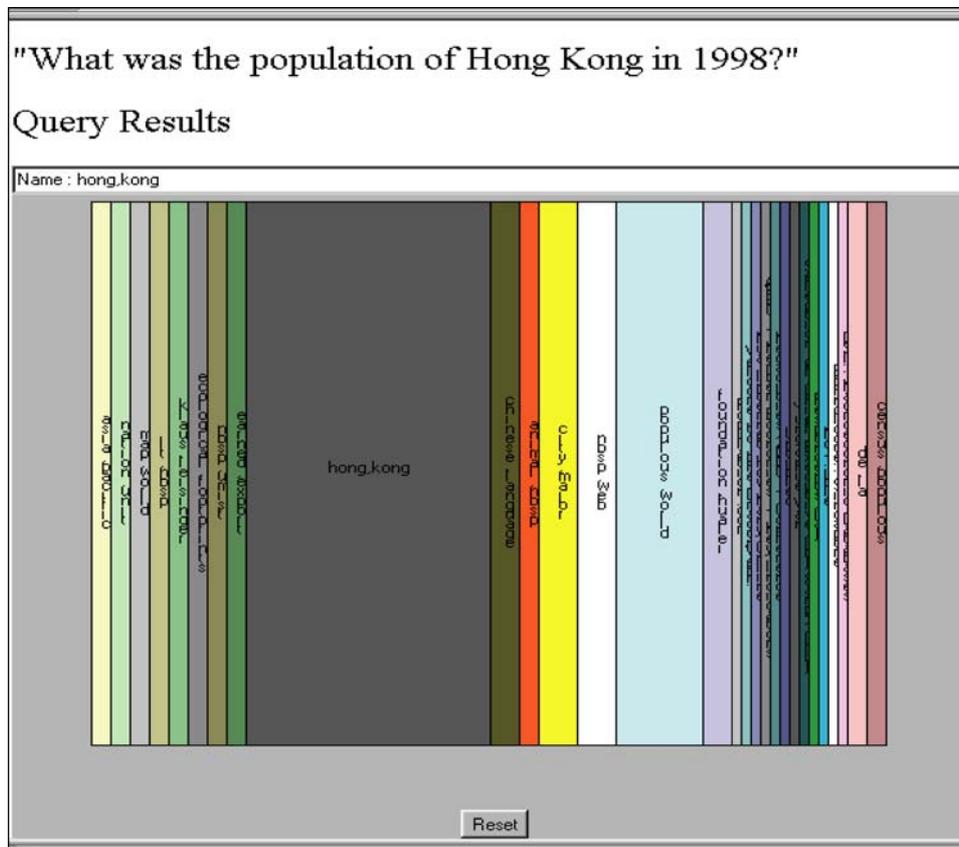


Fig. 6. Visual overview of the search results to query 2.

task-presentation method combinations were as listed in Table 3.

This experimental design facilitates the study of all three modes of the interface, i.e. textual, full zoom, and fisheye zoom. If the order the subjects are exposed to the interfaces has no significant effect (i.e. no phase-task combination and learning effect), the performance (number of correct answers and the time to finish the task) or satisfaction of the subjects using a specific interface can be found by aggregating the respective measures from the three different groups. Subsequently, the data collected by this design can be analyzed as if they were collected through a repeated measures design. We test the validity of this assumption in Section 6. Table 3 displays our experimental design.

We chose this design over a regular repeated measures design, in which all subjects would go through the same task and interface combination in the same order, to control the effect of the sequence in which the subjects were exposed to differing presentation methods.

If significant differences between presentation methods were found with a regular repeated measures design, it would be impossible to estimate whether this was a learning effect and the difference was caused because of this learning, i.e. the different sequence that the subjects were exposed to the interfaces. Hence, the internal validity of a regular repeated measures design for this experiment would be questionable.

Table 3. The experimental design

Phase	Task/Support		
	Group 1	Group 2	Group 3
Phase 1	1 Task A No visualization	2 Task A Full zoom	3 Task A Fisheye zoom
Phase 2	4 Task B Fisheye zoom	5 Task B No visualization	6 Task B Full zoom
Phase 3	7 Task C Full zoom	8 Task C Fisheye zoom	9 Task C No visualization

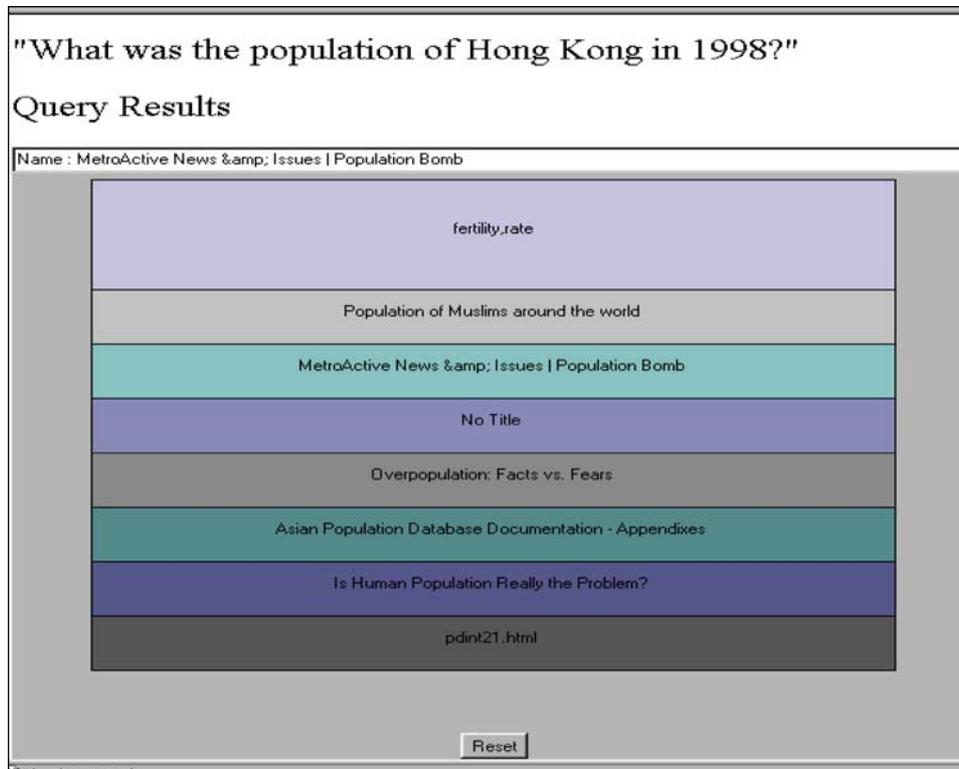


Fig. 7. "Full" zooming of the "populous world" section of Fig. 6.

During an experimental session, the subjects were given 12 minutes (an average of 3 minutes per question) for the completion of each experimental phase. The time limit was a decision based on our pilot tests where we observed that three minutes were sufficient to find answers to our questions. As has been done in previous studies (Dumais, Cutrell, and Chen, 2001), the subjects were reminded of the time left for the specific phase every three minutes so that they knew the expected completion time for each question, and could adjust their pace accordingly. However, they were free to move on to the next question whenever they wanted to within a given phase. At the end of each phase, the subjects were asked to evaluate the mode of presentation that they had just experienced. A total of six sessions of the experiment were conducted. Each session involved a different sample. The first session was used for testing the experimental procedure, and the second one did not yield usable data due to a technical problem with the connection of the web forms to the central database. Thus a total of 78 usable observations were collected.

6. Data Analyses

6.1. Test of hypotheses

We conducted a Multivariate Analysis of Variance (MANOVA) with a within-subjects factor of "interface," and between-subjects factors of "sex," "native language," "age," "web experience," and "cognitive style" (field independence). We included multiple paired contrasts to compare the "visual" interfaces to the "textual," and the "fisheye zoom" interface to the "full zoom" without the need for separate tests. The results of the comparative tests show that at the 0.05 level, the effect of the interface was significant on the "time" variable only (see Table 4). The significance exists for both the comparison between the visual and textual interfaces, $F(1, 46) = 12.78, p = 0.001$, and the comparison between the full zoom and fisheye interfaces, $F(1, 46) = 4.47, p = 0.040$. In addition, as seen in Table 5, the average time to complete an experimental task using each different type of interface was ranked as expected, i.e. $time_{fisheye} < time_{full\ zoom} < time_{no\ visualization}$.

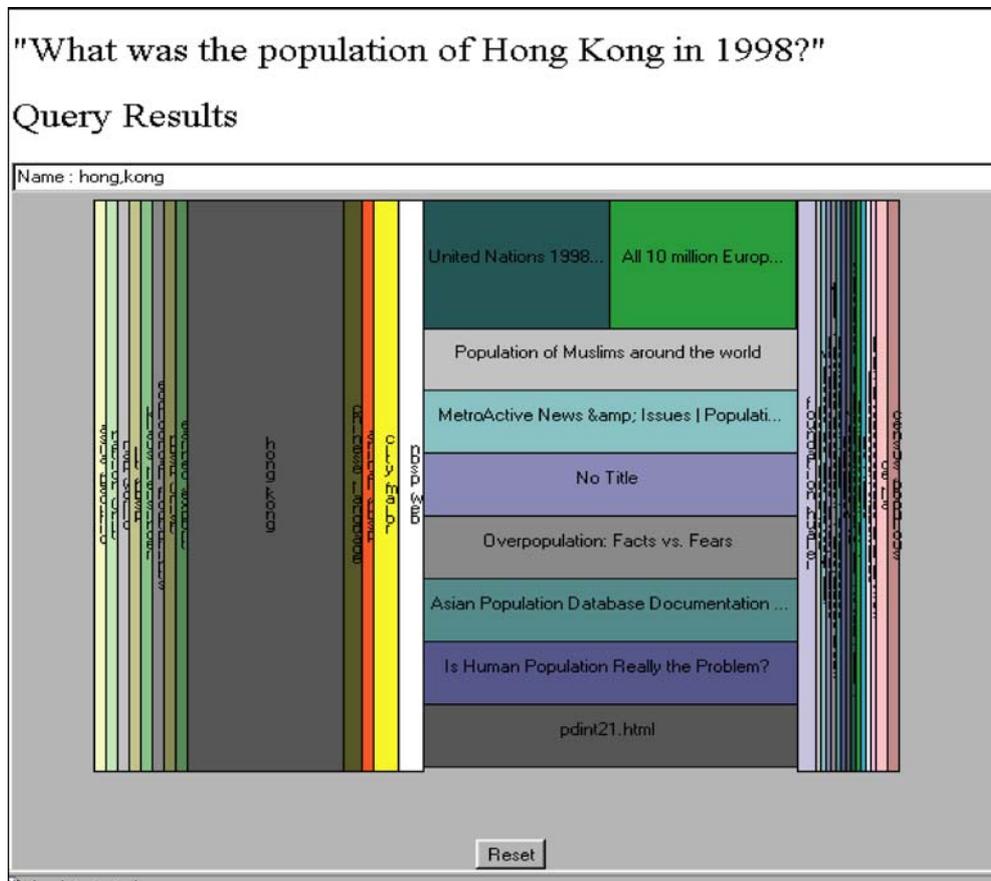


Fig. 8. "Fisheye" zooming of the "populous world" section of Fig. 6.

These findings suggest partial support for our hypotheses as summarized in Table 6. We next discuss the details of our data analyses.

6.2. Analyses details

6.2.1. Preliminary analyses. The original data set contained missing variable values for a total of 12 cases. Since there were only a few missing values for each of

these cases, we chose to replace these values with the average of the same variable from the other cases. Also, during data collection, it was observed that some subjects finished the tasks in an unreasonably short period of time. An examination of the data set showed that some subjects submitted totally irrelevant web sites as the source of their answers for three or four questions in each phase. This is a clear indication that these subjects

Table 4. Paired comparisons

Independent variable	Dependent variable	Comparison	F	Sig.
Interface	SCORE	Fisheye zoom vs. Full zoom	2.455	.124
		Visual vs. Textual	.505	.481
	TIME	Fisheye zoom vs. Full zoom	4.470	.040
		Visual vs. Textual	12.776	.001
	SATISFACTION	Fisheye zoom vs. Full zoom	.123	.727
		visual vs. Textual	.250	.620

Table 5. Descriptive statistics ($N = 57$)

Dependent variable		Mean
Time (sec)	Full zoom	507.2105
	Fisheye zoom	479.4035
	Textual	580.1053

Table 6. Summary of the Hypotheses testing results

Hypothesis	Result
H1a	No support
H1b	No support
H2a	Support in hypothesized direction
H2b	Support in hypothesized direction
H3a	No support
H3b	No support

did not even look at the search results for the answer of a question. The reason we could make this judgment was that the web site collection within which the answer to each question was to be found was available in advance. Therefore, we could easily decide whether the subjects were actually working on the task as instructed, i.e. finding the answer by browsing the available collection, or were giving random answers because they chose not to work on the tasks. We eliminated data from such subjects to improve the quality of the data and the following analyses. This reduced the size of our analysis sample to 57.

The next thing to do before hypotheses testing was to convert the measurements of some of the variables to usable categories or quantities. The measurement scale used for “satisfaction” is a 7-point Likert scale with five items. In order to decide whether this scale is reliable (i.e. repeatable) and whether these different items can be combined into a single satisfaction score, we performed two different analyses. First, to measure inter-item reliability, we calculated coefficient α (Nunnally, 1978), and decided that the scale had high reliability ($\alpha = 0.89$). Next, to understand whether the scale measures a single concept (i.e. the assumed satisfaction construct) or multiple concepts, we performed a confirmatory factor analysis. The results of this analysis showed that a single factor explains most (73.5%) of the variance for the five items on the scale, and the scree plot supports this observation. This led us to conclude that the scores for the items in this scale can be combined into a single “satisfaction score.” To decide

on the weight distribution of these items, i.e. how much of the satisfaction each item actually measures, we examined the component score matrix, and concluded that the weight of each item in explaining satisfaction is close enough to be considered practically equal to each other. With this observation, we transformed the measurements from the satisfaction survey to a single “satisfaction score” by averaging the score of each item on the scale.

The examination of the answers that were given to the two “web search experience” questions (duration and frequency of web search) revealed that there were four distinct categories of subjects in our sample: the subjects with “high” experience reported frequent web search engine use or long period of familiarity with web search engines. The subjects with “very low” experience indicated that they had not used a search engine before, and those with “low” experience indicated that they had a relatively short exposure to web search engines (less than 2 years), and that they very seldom use them (once a month or less frequently). We classified all the other subjects into the “moderate” experience category. Likewise, the exploration of the distribution of scores on the cognitive style tests led us to categorize our test sample into four different cognitive style groups, where the first group contains the most field independent individuals, and the fourth group contains the most field dependent ones.

6.2.2. Test of assumptions. As explained in Section 5.3, the data collected in our experiments could be analyzed using a repeated measures model if we could assume that the order of exposure to different presentation methods had no significant effect on the subjects’ performance and satisfaction. Accordingly, we tested the validity of this assumption by revisiting our experimental design (See Table 3 with the cell numbers from 1 to 9). Three factors differentiate these cells from one another: interface, task, and phase. To isolate the effect of the task and phase combination from the effect of the interface, we performed a test to compare the levels of the three dependent variables between cells 1, 5, and 9, between cells 2, 6, and 7, and between cells 3, 4, and 8 separately. This way, we only compared the effect of the task and phase combination since the user interface in each of these comparisons is constant.

We performed a separate MANOVA to test the effect of the phase-task combination on the dependent variables for each interface (after testing the basic MANOVA assumptions of normality and correlation

between dependent variables). The test results showed no significant difference between the three groups in terms of performance and satisfaction (*textual*, $F(6,78) = 1.478, p = 0.191$, *full zoom*, $F(6, 178) = 1.478, p = 0.185$, *fisheye zoom*, $F(6,78) = 1.478, p = 0.899$). Thus, we conclude that the phase-task combination had no significant affect on the performance or satisfaction of the experimental subjects. This allows the dependent variable measures to be treated as the repeated measures of the same variables with a within-subjects factor of the user interface. Accordingly, the test of hypotheses could be performed simultaneously by means of repeated measures ANOVA or repeated measures MANOVA, and multiple (paired) comparisons. The normal P-P plots showed that the dependent variables are normally distributed; therefore either ANOVA or MANOVA are appropriate to analyze the data. We preferred repeated-measures MANOVA to three separate repeated-measures ANOVAs due to the significant correlations between the dependent variables.

6.2.3. Stepwise analyses. We conducted the data analyses in a stepwise manner starting at the multivariate level. As seen in Table 7, at the 0.05 level of α , the effect of “interface” is significant, $F(6, 182) = 3.462, p = 0.003$, and those of all interaction terms are not ($F(18,255) = 1.371, p = 0.146$, $F(12,238) = 1.342, p = 0.196$, $F(6,180) = 1.009, p = 0.421$, $F(6,180) =$

Table 7. Multivariate tests

Within subjects effect	Value	F	Sig.
Interface			
Pillai’s Trace	.203	3.426	.003
Wilks’ Lambda	.799	3.563(b)	.002
Interface * Cognitive style			
Pillai’s Trace	.239	1.327	.170
Wilks’ Lambda	.770	1.371	.146
Interface * Age			
Pillai’s Trace	.165	1.337	.197
Wilks’ Lambda	.841	1.342	.196
Interface * Sex			
Pillai’s Trace	.065	1.017	.416
Wilks’ Lambda	.936	1.009(b)	.421
Interface * Language			
Pillai’s Trace	.030	.459	.838
Wilks’ Lambda	.970	.455(b)	.841
Interface * Web experience			
Pillai’s Trace	.221	1.221	.243
Wilks’ Lambda	.791	1.223	.243

$0.455, p = 0.841, F(18,255) = 1.223, p = 0.243$). This implies that there is an overall difference between the success of the three different interfaces, and this difference is independent of the level of the control variables. To decide whether there is a significant difference between the three different interfaces in how they affect the individual outcomes, i.e. “score,” “time,” and “satisfaction,” we next examined the univariate tests for the “interface” effect.

As portrayed in Table 8, the effect of interface was significant for the “time” variable only, $F(1,92) = 8.883, p = 0.005$. This means there is insufficient evidence to support hypotheses H1a, H1b, $F(1,92) = 1.387, p = 0.245$, H3a and H3b, $F(1,92) = 0.214, p = 0.808$. Hypothesis H2a will be supported if there

Table 8. Univariate tests

Independent variable	Dependent variable	F	Sig.	
Interface	Score	Sphericity Assumed	1.387	.255
		Greenhouse-Geisser	1.387	.255
		Huynh-Feldt	1.387	.255
		Lower-bound	1.387	.245
Time	Time	Sphericity Assumed	8.883	.000
		Greenhouse-Geisser	8.883	.000
		Huynh-Feldt	8.883	.000
		Lower-bound	8.883	.005
Satisfaction	Satisfaction	Sphericity Assumed	.214	.808
		Greenhouse-Geisser	.214	.760
		Huynh-Feldt	.214	.808
		Lower-bound	.214	.646

Table 9. Tests of control variables

Independent variable	Dependent variable	F	Sig.
Cognitive style	Score	1.028	.389
	Time	1.867	.148
	Sat	1.180	.328
Age	Score	.028	.972
	Time	.115	.891
	Sat	.615	.545
Sex	Score	2.856	.098
	Time	.406	.527
	Sat	.046	.832
Language	Score	1.377	.247
	Time	2.900	.095
	Sat	.630	.432
Web experience	Score	.715	.548
	Time	.961	.419
	Sat	2.969	.041

Table 10. The effect of web experience on satisfaction

Interface	Web experience			
	Very low	Low	Moderate	High
Textual	5.70 (SD = 0.42)	3.30 (SD = 2.12)	3.35 (SD = 1.77)	3.92 (SD = 1.39)
Full-zoom	6.20 (SD = 0.57)	4.00 (SD = 1.98)	4.05 (SD = 1.28)	3.03 (SD = 1.45)
Fisheye-zoom	6.10 (SD = 0.71)	3.60 (SD = 2.55)	3.80 (SD = 1.71)	3.13 (SD = 1.27)

is a significant “time” difference between the textual system and the visual systems, and H2b will be supported if there is a significant “time” difference between the fisheye-zoom and full-zoom systems. Following the stepwise analysis approach, these two hypotheses were tested by means of a priori defined paired comparisons as discussed in Section 6.1.

6.3. Post Hoc analyses

In addition to the factor of main interest, i.e. the interface, it is worth mentioning the effects of the control variables on the outcomes. As seen in Table 9, only one of these effects is significant at the 0.05 level: the effect of “web experience” on “satisfaction.” The examination of the satisfaction scores for each different type of interface, and for each web experience group revealed that the subjects with the highest experience level were the least satisfied using the visual interfaces. Meanwhile, they were the most satisfied using the textual interface. On the other hand, the least experienced subjects were the most satisfied using both of the visual interfaces, and their satisfaction scores were very similar for each of the full zoom and the fisheye interfaces. These results are summarized in Table 10. Although there were no predefined hypotheses regarding these effects, the results are interesting. The next section includes our discussion of the data analyses results as well as the conclusions drawn from those.

7. Discussion and Conclusions

The World Wide Web is used for a myriad of reasons ranging from financial interactions to entertainment. The way web content is presented is an important factor, especially for computer novices, to have a positive

experience in web use. From that point of view, research aiming to enhance the quality of a presentation method for web content is important. Meanwhile, it is equally important to carefully examine the usefulness and practicality of a given presentation method. The specific purpose of this study has been to accumulate empirical evidence for comparing the traditional ranked-list presentation of search results to the map-based presentation of search clusters with different zooming methods.

Our tests showed that the use of clustering-based visualization for the presentation of web search results resulted in significant speed improvements. Furthermore, the application of the fisheye view idea to the visual interface led to better improvements. On the other hand, the task outcomes or the satisfaction with the different systems did not significantly vary among subjects. The fact that we found significant speed improvement without a significant change in task outcome can be potentially explained by the “effort minimization” concept as discussed by Todd and Benbasat (1992). They argue that decision makers do not necessarily analyze problems in more depth to make better decisions when they have expanded processing capabilities. Rather, many decision makers may use these tools to reduce their effort to achieve a similar level of outcome quality. In their study, experimental subjects behaved as if effort minimization was an important consideration, and did not produce higher quality decisions with better aids. Although the tasks supported by our systems are not decision-making tasks, Todd and Benbasat’s observations are broad enough to encompass situations where an information user is supported by a computerized system, and where there is no specific emphasis on effectiveness, but rather on the natural behavior of users when, in a way, they are left free to choose between effectiveness and efficiency. This was the case

in our studies. On the other hand, a more successful system could still improve effectiveness if users were required or rewarded to spend comparable amounts of time and effort using both systems. This argument is consistent with previous research findings in similar research settings to ours (Allan et al., 2001; Roussinov and Chen, 2001) as well as slightly different ones (Tan and Benbasat, 1993; Vessey, 1991). We believe that it is important for future research to study in detail how the trade-off between effectiveness and efficiency affects system use in different settings.

We believe that our results have practical significance. Our experiments showed that the average time it took to complete an experimental task, which in turn included four individual search tasks, were 479, 506, and 580 seconds for the fisheye, full zoom, and textual interfaces respectively. This corresponds to an average time of 120, 126, and 145 seconds for each search. In itself the number of seconds gained may not seem significant. However, if the efficiency holds for the millions of searches, the time savings facilitated by visual interfaces could lead to considerable efficiency improvements for those individuals that do knowledge intensive work. Furthermore, if the savings in time lead to lower cognitive effort as suggested by Todd and Benbasat (1992), then regardless of its amount, any time saving would improve decision quality in the long run.

In spite of these contributions, there are some issues that limited the validity of our study. As displayed in the previous section, across all differing interfaces, more experienced users were the least satisfied. This could possibly be explained by the complaints that we heard from more experienced web searchers regarding the lack of freedom in the formulation of search queries. The lack of such freedom was primarily due to the experimental controls we wanted in place. The fact that the search engine was not optional and that the tasks had easily identifiable outcomes allowed us to exercise a higher level of control, desirable for the internal validity of the study. We realize that this approach led to a certain level of degradation in the external validity of the research findings. However, such a trade-off between internal and external validity is in the nature of experimental designs, and is not a weakness particular to this study.

Another observation was that subjects who did not “buy into” the experiment had a harder time learning and using the visual systems. This observation points to the existence of an interaction between the interface type and an omitted contextual variable: “sub-

ject involvement.” We had anecdotal evidence showing that some subjects found the systems interesting and believed in their promise, but they thought there was no incentive at this time for them to orient themselves to a new paradigm. In fact, the following comment from a participant in our experiments supports this observation:

“It is sometimes hard to get people out of a paradigm. Anyway, when I went home and thought about it, I realized that I was faster in finding information using your system”

Naturally, this observation was more applicable to the more experienced subjects. As mentioned in Section 6.3, the satisfaction scores for the visual interfaces dropped drastically with increasing levels of web experience. We believe that the reason the interaction between the interface and web experience was not statistically significant is due to the large number of low and moderate experienced users. However, the mentioned comparison between subjects with very low and high experience levels suggests that such an interaction is practically significant. This is a probable cause as to why we did not obtain the expected results in user satisfaction.

This issue would naturally resolve itself in a non-experimental setting. Therefore, a way to address the point is to conduct less controlled field studies where people use the systems for their daily needs for an extended period of time. On the other hand, more reliable results in experiments could be obtained by including “non-student” subjects for whom the web search activity is part of their daily routine and is hence more relevant. For both of these scenarios, the initial resistance to change would be less of an issue, hence the users’ true attitudes towards our interface could be observed.

Meanwhile, despite the strong evidence in the literature for the usefulness of clustering for presenting search results, the clustering algorithms in use are not without problems. The fully automatic methods of generating clusters out of textual documents still have room for improvement especially in how synonyms for terms are identified, how the similarity between clusters are defined, and how good similarity thresholds for combining clusters are found. Another possible direction for improvement is the use of semi-automated methods where clustering is supported by supervision from already existing knowledge such as thesauri or existing categories. Such an approach could especially

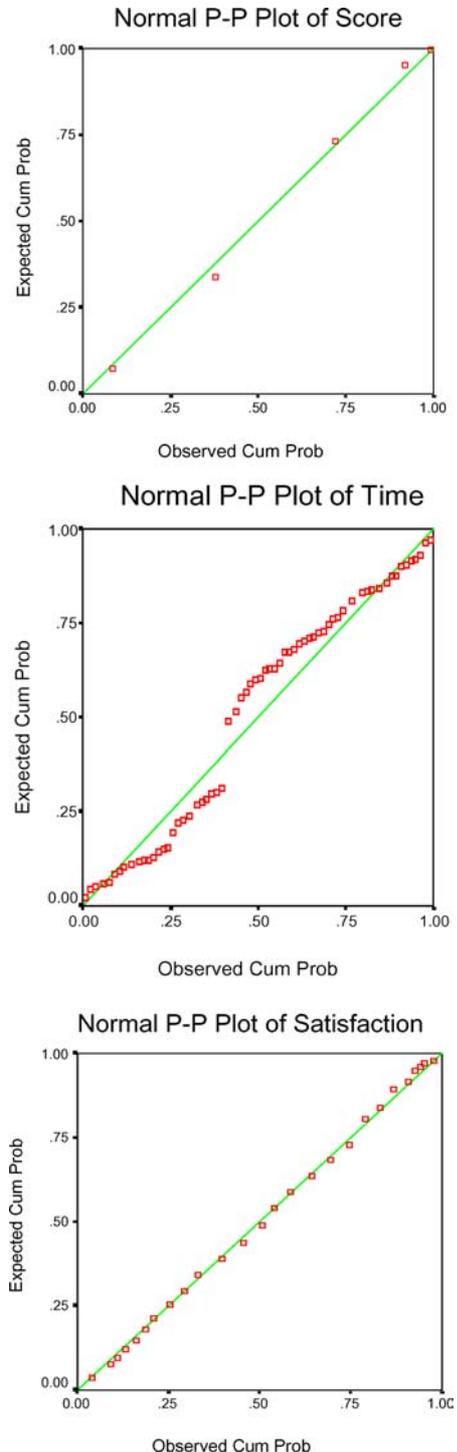
help in the labeling of clusters since cluster labels are not very intuitive with most of the current clustering tools.

A first look at the quality of search result clusters by recent search engines such as vivisimo (vivisimo.com) suggests that a good amount of improvement is already achieved, however these intuitive observations should be verified in more rigorous empirical evaluations. For such evaluations, an alternative to the fact-finding tasks that we used in this study is exploring the success of the interfaces by means of open-ended browsing tasks as discussed in Section 5.1. In the future, it would be interesting to integrate some of these newer clustering algorithms into our systems and test them with complex and open-ended tasks such as collecting information on a well-known philosopher. Such a study would not yield easily quantifiable outcomes. Instead, it would yield results, the quality of which may be evaluated by domain experts. We believe that the value of summarizing search results by means of clustering and being able to see details in context would be even clearer for such tasks. Experimenting with such scenarios would add to our understanding of the usability of our design ideas.

These discussions also suggest that there is a need for more theory-building research on interface design for modern systems and the behavior of today's information system users. Such efforts should go hand in hand with empirical studies that validate the theorized relationships between the causes and indicators of interface success. Nonetheless, the support for our hypotheses suggests that our design ideas are promising, and it is worthwhile to focus on improving the implementation of the presentation system. Processing of search results into clusters and then applying visualization, particularly with a fisheye-based zoom, is likely to give the user significant power in searching the World Wide Web. Organizations that do knowledge work could reevaluate their knowledge management tools in light of these discussions, possibly by adopting visual interfaces as described in this study, and by getting involved in field studies to enhance the understanding on the usability of such systems. The fact that the web will gradually host a larger collection of information, and the purpose in using web-based information will be more diversified, presents many opportunities for the proactive knowledge manager to utilize the wealth of available information without getting overwhelmed by that very abundance.

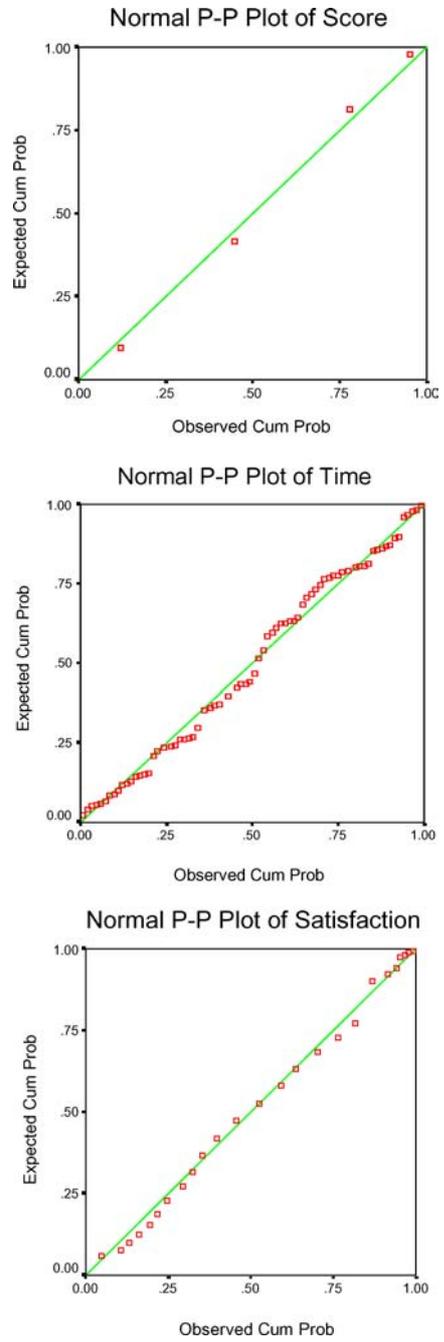
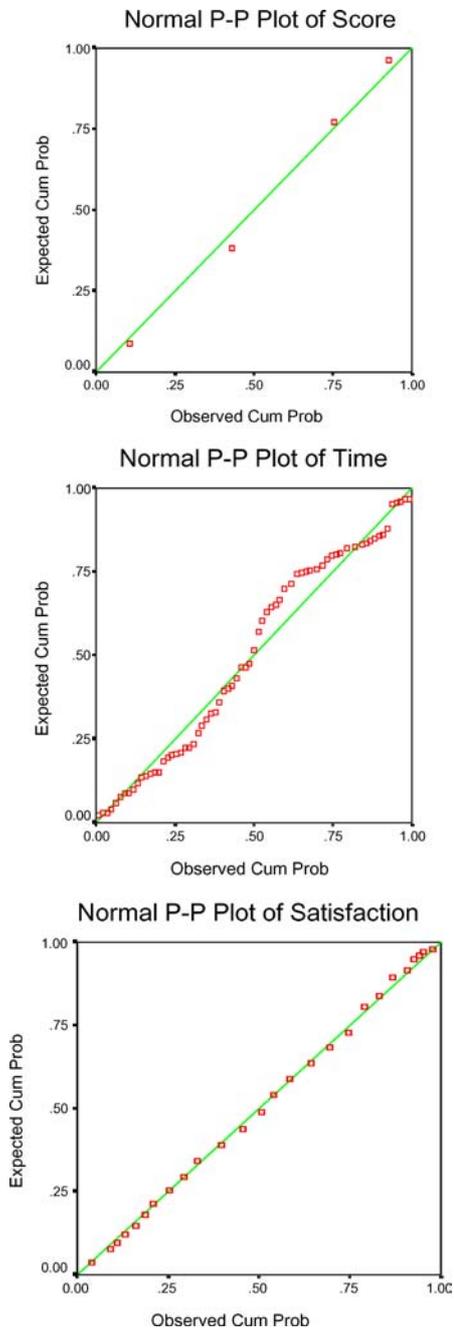
Appendix

Normal PP Plots for "No Visualization"



Normal P-P Plots for the “Full Zoom” Visualization

Normal P-P Plots for the “Fisheye Zoom” Visualization



Notes

1. We are thankful to Ben Shneiderman and the Human Computer Interaction Laboratory of University of Maryland for sharing the original Treemap algorithm. Of course, we are responsible for any errors in adaptation of the Treemap algorithm to include fisheye visualization.
2. Some of these questions were modified from those that were used in a panel on Web Search at the 1998 ACM Conference on Advances in Informational Retrieval.

References

- Allan J, Leuski A, Swan R, Byrd D. Evaluating combination of ranked lists and visualizations of inter-document similarity. *Information Processing and Management* 2001;37:435–458.
- Allen R, Obry P, Littman M. An interface for navigating clustered document sets returned by queries. Paper presented at the the ACM SIGOIS: Conference on Organizational Computing Systems, Milpitas, CA, 1993.
- Amento B, Terveen L, Hill W. Does “Authority” Mean Quality? Predicting Expert Quality Ratings of Web Documents. Paper presented at the Annual ACM Conference on Research and Development in Information Retrieval, Athens, Greece, 2000a, 7/2000.
- Amento B, Terveen L, Hill , Hix D. TopicShop: Enhanced Support for Evaluating and Organizing Collections of Web Sites. Paper presented at the 13th annual ACM symposium on user interface software and technology, San Diego, CA, USA, 2000b.
- Bederson BB. Fisheye menus. Paper presented at the 13th annual ACM symposium on user interface software and technology, San Diego, CA, USA, 2000.
- Bederson BB, Hollan JD, Perlin K, Meyer J, Bacon D, Furnas GW. Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *Journal of Visual Languages and Computing* 1996;7:3–31.
- Bederson BB, Hollan JD, Stewart J, Rogers DAD, Vick D, Ring L, Grose E, Forsythe C. A zooming web browser. In: Forsythe C, Ratner J, Grose E, eds. *Human Factors and Web Development*. New Jersey: Lawrence Erlbaum, 1998:255–266.
- Bharat K, Henzinger MR. Improved algorithms for topic distillation in a hyperlinked environment. Paper presented at the Annual ACM Conference on Research and Development in Information Retrieval, Melbourne, Australia, 1998, August 1998.
- Card SK, Moran TP, Newell A. The Model Human Processor: An Engineering Model of Human Performance. *Handbook of Perception and Human Performance*. New York, New York: John Wiley and Sons, 1986.
- Card SK, Robertson GG, W Y. The WebBook and the Web Forager: An information workspace for the World-Wide Web. Paper presented at the Conference on Human Factors and Computing Systems, Vancouver, British Columbia, 1996, April 13-18, 1996.
- Chen H, Dumais S. Bringing order to the Web: Automatically categorizing search results. Paper presented at the Conference on Human Factors in Computing Systems, The Hague, The Netherlands, 2000, 1–6 April.
- Chen H, Houston AL, Sewell RR, Shatz BR. Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques. *Journal of the American Society of Information Science*, 1998;49(7):582–603.
- Chen M, Hearst M, Hong J, Lin J. Cha-Cha: A system for organizing Intranet search results. Paper presented at the 2nd USENIX Symposium on Internet Technologies and SYSTEMS, 1999, October 11–14.
- Collaud G, Dill J, Jones CV, Tan P. A distorted-view approach to assisting web navigation. Paper presented at the New Paradigms in Information Visualization and Manipulation, (in conjunction with CIKM’95), 1995, December 1, 1995.
- Cutting DR, Karger DR, Pedersen JO, Tukey JW. Scatter/gather: A cluster-based approach to browsing large document collections. Paper presented at the 1992 Conference on Research and Development in Information Retrieval, Copenhagen, Denmark, 1992.
- Dumais S, Cutrell E, Chen H. Optimizing search by showing results in context. Paper presented at the Conference on Human Factors in Computing Systems, Seattle, Washington, United States, 2001, March 31–April 04.
- Everitt, B. *Cluster Analysis*. London, 1977.
- Furnas GW. Generalized Fisheye views. Paper presented at the Conference on Human Factors in Computing Systems, Boston, MA, 1986.
- Greene S, Marchionini G, Plaisant C, Shneiderman B. Previews and overviews in digital libraries: Designing surrogates to support visual information-seeking. *Journal of the American Society of Information Science* 2000;51(3):380–393.
- Hane PJ. LEXIS-NEXIS tests data visualization technology. *Information Today* 2000;17(1).
- Hearst M. TileBars: Visualization of term distribution information in full text information access. Paper presented at the Conference on Human Factors in Computing Systems, Denver, CO, 1995.
- Hearst M, Pedersen P. Reexamining the cluster hypothesis: Scatter/gather on retrieval results. Paper presented at the Conference on Research and Development in Information Retrieval, Zurich, Switzerland, 1996.
- Hertzum M, Frokjaer E. Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on Computer-Human Interaction* 1996;3(2):136–161.
- Hollands JG, Carey TT, Matthews ML, McCann CA. (eds.) *Presenting a Graphical Network: A Comparison of Performance Using Fisheye and Scrolling Views*. Amsterdam, Netherlands: Elsevier Science Publishers, 1989.
- Honkela T, Kaski S, Lagus K, Kohonen T. WEBSOM—self-organizing maps of document collections. Paper presented at the WSOM’97, Workshop on Self-Organizing Maps, Espoo, Finland, 1997, June 4–6, 1997.
- Jacoby J. Perspectives on information overload. *Journal of Consumer Research* 1984;10:432–435.
- Johnson B, Shneiderman B. Treemaps: A space filling approach to the visualization of hierarchical information structures. Paper presented at the the Second International IEEE Visualization Conference, 1991.
- Kohonen T. Self-organization of very large document collections: State of the art. Paper presented at the ICANN98, the 8th International Conference on Artificial Neural Networks, 1998.
- Kumar HP, Plaisant C, Shneiderman B. Browsing Hierarchical Data with Multi-Level Dynamic Queries and Pruning (Technical Report

- CSHCN TR 95-10). College Park, MD: University of Maryland, 1995.
- Lagus K, Kaski S, Honkela T, Kohonen T. Browsing Digital Libraries with the Aid of Self-Organizing Maps. Paper presented at the Fifth International World Wide Web Conference, Paris, France, 1996, May 6–10.
- Lamping J, Rao R. Visualizing *Large Trees Using The Hyperbolic Browser*. Paper presented at the ACM Conference Companion on Human Factors and Computing Systems, Vancouver, British Columbia, 1996, April 14–18, 1996.
- Leung KY, Apperly MD. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction* 1994;1(2):126–160.
- Lim KH, Benbasat I, Todd PA. An experimental investigation of the interactive effects of interface style, instructions, and task familiarity on user performance. *ACM Transactions on Computer-Human Interaction* 1996;3(1):1–37.
- Lin X. Map displays of information retrieval. *Journal of the American Society of Information Science* 1997;48(1):40–54.
- Mukherjea S, Hirata K, Hara Y. Using Clustering and Visualization for Refining the Results of a WWW Image Search Engine. Paper presented at the New paradigms in information visualization and manipulation, Bethesda, MD, 1998.
- Nation D. WebTOC: A tool to visualize and quantify Web sites using a hierarchical table of contents browser. Paper presented at the Conference on Human Factors and Computing Systems, Los Angeles, CA, 1998, April 1998.
- Nigay L, Vernier F. Navigational interaction techniques in the search results space. Paper presented at the Workshop on Innovation and Evaluation in Information Exploration Interfaces, Berkeley, CA, 1998.
- Nunnally JC. *Psychometric Theory* (2nd ed.). New York, New York: McGraw-Hill Book Company, 1978.
- Pirolli P, Schank P, Hearst M, Diehl C. Scatter/gather browsing communicates the topic structure of a very large text collection. Paper presented at the Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada, 1996, April 13–18.
- Pratt W, Fagan L. The Usefulness of Dynamically Categorizing Search Results. *Journal of the American Medical Informatics Association* 2000;7(6):605–617.
- Pratt W, Hearst M, Fagan L. A Knowledge-Based Approach to Organizing Retrieved Documents. Paper presented at the Sixteenth National Conference on Artificial Intelligence, 1999.
- Roussinov D, Chen H. Information navigation on the web by clustering and summarizing query results. *Information Processing and Management* 2001;37(6):789–816.
- Salton G. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley, 1989.
- Santhanam R, Sein MK. Improving end-user proficiency: Effects of conceptual training and nature of interaction. *Information Systems Research* 1994;5(4):378–399.
- Sarkar M, Brown MH. Graphical Fisheye Views of Graphs. Paper presented at the Conference on Human Factors in Computing Systems, 1992.
- Schaffer D, Zuo Z, Greenberg S, Bartram L, Dill J, Dubs S, Roseman M. Comparing Fisheye and Full-Zoom Techniques for Navigating Hierarchically Clustered Networks. Paper presented at the Graphics Interface, 1993.
- Schaffer D, Zuo Z, Greenberg S, Bartram L, Dill J, Dubs S, Roseman M. Navigating hierarchically clustered networks through fisheye and full-zoom methods. *ACM Transactions on Computer-Human Interaction* 1996;13(2):162–188.
- Shneiderman B. The eyes have it: a task by data type taxonomy of information visualizations. Paper presented at the IEEE Symposium on Visual Languages, 1996.
- Shneiderman B. *Designing The User Interface, Strategies for Effective Human-Computer Interaction*. Reading, MA: Addison Wesley, 1997.
- Stasko J, Catrambone R, Guzdial M, McDonald K. An Evaluation of Space-Filling Information Visualizations For Depicting Hierarchical Structures (Technical GIT-GVU-00-03). Atlanta, GA: Georgia Institute of Technology, Graphics, Visualization and Usability Center, 2000.
- Suh KS, Jenkins AM. A comparison of linear keyword and restricted natural language database interfaces for novice users. *Information Systems Research* 1992;3(3):252–272.
- Tan JKH, Benbasat I. Processing of graphical information: A decomposition taxonomy to match data extraction tasks and graphical representations. *Information Systems Research* 1990;1(4):416–439.
- Tan, JKH, Benbasat I. The Effectiveness of Graphical Presentation for Information Extraction: A Cumulative Experimental Approach. *Decision Sciences* 1993;24(1):167–191.
- Todd P, Benbasat I. The use of information in decision making: An experimental investigation of the impact of computer based decision aids. *Management Information Systems Quarterly* 1992;16(3):373–393.
- Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 2001.
- Turetken O, Sharda R. Development of a fisheye-based information search processing aid (FISPA) for managing information overload in the web environment. *Decision Support Systems* 2004;37(3):415–434.
- Van Rijsbergen CJ. *Information Retrieval*. London, UK: Butterworths, 1979.
- Vessey I. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences* 1991;6(1):219–240.
- Wang P, Hawk WB, Tenopir C. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing and Management* 2000;36:229–251.
- Wertheimer, M. Gestalt theory. *Social Research*, 11. Berlin, Germany, 1924.
- Witkin HA, Oltman PK, Raskin E, Karp SA. *A Manual For The Embedded Figures Test* (Consulting Psychologists Press ed.). Palo Alto, CA, 1971.
- Zamir O, Etzioni O. Grouper: A Dynamic Clustering Interface to Web Search Results. Paper presented at the 8th International World Wide Web Conference, 1999.

Ozgur Turetken is Assistant Professor of Management Information Systems in the Fox School of Business and Management at Temple University. He received a B.S. in Electrical Engineering and an M.B.A. from Middle East Technical University in Ankara, Turkey, and a Ph.D. from Oklahoma State University. Dr. Turetken's

research interests are in information visualization, decision support systems, and distributed work arrangements. His prior research appeared in *Communications of the ACM*, *Decision Support Systems*, *Information Systems Frontiers*, and many international conferences.

Ramesh Sharda is ConocoPhillips Chair of Management of Technology and a Regents Professor of Management Science and Information Systems in the College of Business Administration at Oklahoma State University. He received his B.Eng. degree from University of Udaipur, M.S. from The Ohio State University and an MBA and Ph.D. from the University of Wisconsin-Madison. His research has been published in major journals in management science and information systems including *Management Science*, *Information Systems Research*, *Decision Support Systems*, *Interfaces*, *INFORMS Journal on Computing*, *Com-*

puters and Operations Research, and many others. He served as the Founding Editor of the *Interactive Transactions of ORMS* and on the editorial boards of other journals such as the *INFORMS Journal on Computing*, *Information Systems Frontiers*, *Journal of End User Computing*, and *OR/MS Today*. One of his major activities in the last few years was to start the MS in Telecommunications Management Program at Oklahoma State. Now he is establishing a major interdisciplinary Institute for Research in Information Systems (IRIS) at OSU. His research interests are in optimization applications on desktop computers, information systems support collaborative applications, neural networks, business uses of the Internet, and knowledge networks. Defense Logistics Agency, NSF, Marketing Science Institute, and other organizations have funded his research. Ramesh is also a cofounder of a company that produces virtual trade fairs, iTradeFair.com.