My Exploration in Science Research

My initial inspiration to perform scientific research in the field of computer science came from my parents, whom I admire for their endless interest and successful efforts in their field of work. As a young child, I had visited their offices and saw some of their work. Although at the time I didn't completely understand their jobs, I was amazed that the computer had the power to create such an impact on people's lives. When I started my science research in high school I knew it would be the perfect opportunity to experience first hand what I had witnessed as a child. However, I also had personal aspirations. As a long time art enthusiast and visual learner, I wanted to find a field that combined my passion for art with my interests in computers, research that would affect me personally. My science research teachers in high school have always encouraged me to pursue a field that I am interested in. My mentor has also helped me to select and focus on a valuable research area. Ultimately, I found the field of Information Visualization, which deals with human computer interaction through visualized information.

My research consisted of the study of user preferences in the web search process, popular search engines, existing visualization techniques, the design of novel visualizations based on the user preferences, and the verification of the effectiveness of my new designs. To better characterize the user preferences and the attractive features in the popular search engines, I conducted a series of user surveys. Based on the survey results I quantitatively ranked the different user preferences from and the effectives of each search engine feature through the use of statistics and charts. I also tried different approaches to compare the existing visualization techniques and finally used a matrix comparison to depict the commonalties and differences among these techniques. These statistical results steered my research in the right direction. This was the first time I found that the mathematics I learned in school was truly applicable in real

research. In some cases, I had to learn more about specific statistical tests from other textbooks and references to better understand and correctly use them. After I came up the preliminary design and final designs of the novel visualizations, which graphically categorized the user preferred features, all subjects in the user study responded very well to the new designs. To determine the significance of these results, I searched and compared several possible statistical tests. I performed a matched pairs T-test, which compares the deviation in results between two groups under different conditions and decides whether this deviation in results is statistically significant. These two groups contained the search engine effectiveness results, in terms of the average amount of time needed to complete the search tasks. The two different conditions were my novel visualization designs versus the current Google layout. The results of T-test indicate that the novel visualization design is statistically more effective than the Google layout. This was an essential component of my research because it strengthened the validity of my results, creating on overall more influential piece of work.

After completing my research, I have realized that science and mathematics are truly the building blocks of knowledge and the foundation of our life. I have witnessed the power of these two unique yet interrelated fields of study as continuing research gradually changes the world and our body of knowledge.

Studying and Designing Effective Information Visualizations For Graphics-based Search Engines

Review of Literature

The introduction of the World Wide Web in the early 1990's transformed personal computers, originally used to organize information, into reliable sources for information [2]. The Web's increased value is due to search engines, computer programs designed to locate desired information on the Web [5]. Since their creation, search engines have become indispensable tools for information retrieval, consistently growing in popularity [12].

The advances made in text-based search technology keep the search engines' information up-to-date and reliable [14]. However, a majority of the technological advances have focused mostly on obtaining more relevant results and techniques to rank these results. Web pages are also being developed at a very rapid rate. This unprecedented increase in information requires users to take extra time in processing and evaluating the information, creating an overall ineffective search process [1].

Therefore, the *presentation of results* has become a more critical factor in the effectiveness of a search process. Web search engines only present users with a textually organized list of retrieved documents, such as

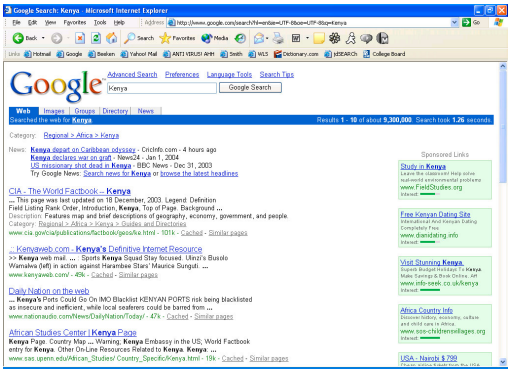

Figure 1 Text-based Search Engine Results

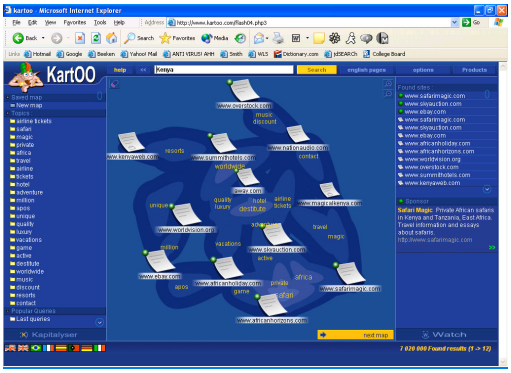

Figure 2 Graphics-based Search Engine Results by Kartoo

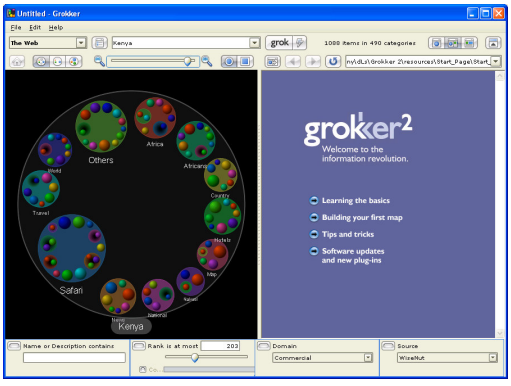

Figure 3 Graphics-based Search Engine Results by Grokker

www.Google.com shown in *Figure 1*. To accommodate the increasing amount of information on the Web, search result visualization must require minimal cognitive effort to interpret the content and relevancy of results [15].

Aiming to take advantage of human visual comprehension capabilities, researchers have developed graphics-based search engines [11]. For example, given a search keyword, Kartoo (www.kartoo.com) will return a visual representation of the search results (Figure 2). In this case, each document is represented as a rectangular icon and placed in a topographic map depending on its level of relevancy. Grokker (www.grokker.com) organizes its results into separate color palettes, each representing a category (Figure 3). Graphics-based search engines promise to help users perceive hidden relationships among retrieved results, such as content similarities, to create a faster and better search process [8].

However, graphics-based search engines have not become mainstream search engines. An analysis on search engine ratings shows that none of the top 15 search destinations were graphics-based [13]. Therefore, there is a need to enhance graphics-based search engines to decrease the amount of time users take to locate desired information.

Objectives

To identify: (a) the characteristics of existing graphics-based search engines and (b) the key properties of an effective graphics-based search engine, this research addressed the following objectives:

1. Characterize users' search preferences.
2. Analyze the content properties of text-based and graphics-based search engines.
3. Determine how existing visualization techniques may improve the average time needed to obtain a quality result using graphics-based search engines.

4. Create novel visualization techniques that are more effective than the currently most preferred text based search results organization.

Objective 1: Methods, Results and Discussion

A. Characterize User Search Preferences

An Institutional Review Board approval was obtained to conduct all user studies. To accomplish the first objective, a user study was designed and conducted to compare the effectiveness of text-based search engines to graphics-based search engines and solicit user feedback. There were a total of 24 participants in this study, including 12 males and 12 females ranging from ages 14 to 50 of different professions.

Table 1 Part I User Profile Questionnaire

1	What gender are you?
2	What age group do you fall in?
3	About how often do you use the Internet?
4	About how often do you user Google.com?
5	About how often do you use Kartoo.com?
6	Have you planned a trip/traveled to Kartoo/Switzerland?

This user study was divided into three parts: user profile (Part I), completion of search tasks (Part II) and user feedback interview (Part III). Part I contained a six multiple-choice questionnaire to profile the subjects (*Table 1*). In Part II, subjects completed two search tasks, A through D, one using Kartoo and the other using Google™ (*Table 2*). Users were allowed to put the keywords in any order, giving them searching freedom without affecting the results. Subjects were asked to record their answer along with the URL of the relevant documents they found.

Table 2 Assigned Tasks

	Task 1	Task 2
A	Search: Capital of Kenya	Search: Capital of Switzerland
	Keywords: Capital, Kenya	Keywords: Capital, Switzerland
B	Search: Airline from NYC to Kenya Capital	Search: Airline from NYC to Switzerland Capital
	Keywords: Airline, Capital, Kenya	Keywords: Airline, Capital, Switzerland
C	Search: Hotel in Kenya Capital	Search: Hotel in Switzerland Capital
	Keywords: Airline, Capital, Hotel, Kenya	Keywords: Airline, Capital, Hotel, Switzerland
D	Search: Attraction in Kenya Capital	Search: Attraction in Switzerland Capital
	Keywords: Airline, Attraction, Capital, Hotel, Kenya	Keywords: Airline, Attraction, Capital, Hotel, Switzerland

Table 3 User Feedback Questions

	Questions
1	Which site do you prefer: Google or Kartoo?
2	Why do you prefer Kartoo.com?
3	Why do you prefer Google.com?
4	What factor do you consider first, second and third when selecting one document over another?
5	If you were to create an ideal search engine results page, which elements would exist on your page?

Results were recorded as: a) whether or not the task was completed, b) the keyword combinations used, c) the number of clicks needed to fulfill the tasks, and d) the total of amount of time needed for each individual to complete each sub-task. In Part III, an interview was conducted to solicit user feedback. This questionnaire included five questions to understand users' preferences (*Table 3*).

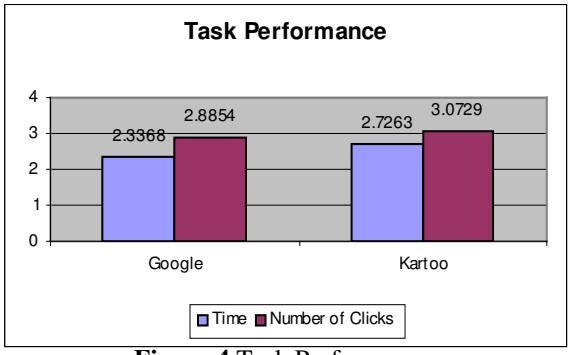

Figure 4 Task Performance

B. Result Analysis of User Search Preferences

The results from the user study Part II, is found in *Figure 4*. This data indicates that the average task completion time and number of clicks were less using Google than using Kartoo. Therefore, it supports that Goggle's text-based results were more effective than Kato's graphics-based results.

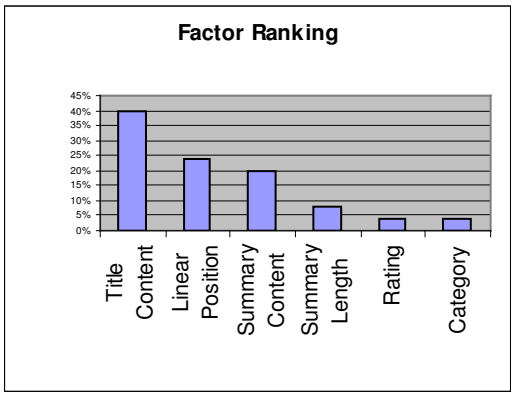

Figure 5 Top Factors in Search Engines

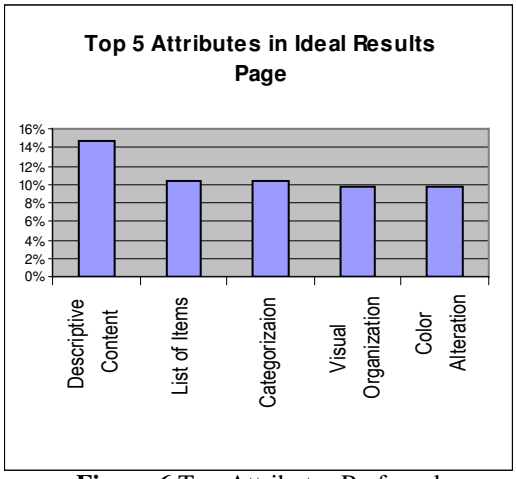

Figure 6 Top Attributes Preferred

Part III results (*Table 3*) were used to determine user preferences. Among the 24 subjects, 22 preferred Google while two preferred Kartoo. Most subjects found Google more useful because of its detailed linear organization, informative title, summary, and categorization, thus characterizing user search preferences. As shown in *Figure 5*, 40% of users ranked Title-Content as the first factor they consider when

choosing one Web page over another. When combining the Title-Content results (40%) with Summary-Content results (22.5%), 62.5% of subjects selected content as a top factor while searching (*Figure 5*). In addition, users chose Descriptive-Content, which includes the document's title and description, as the number one attribute that would appear on their ideal results page (*Figure 6*). Users also identified a visual organization of results as a top results page attribute (*Figure 6*). Therefore*, it is evident that users prefer a visually organized, direct display of informative content to help them quickly determine the relevancy of retrieved documents*, addressing Objective 1.

Objective 2: Methods, Results and Discussion

Content Analysis of Text vs. Graphics

This analysis compared the content in three major text-based search engines, Google™, Yahoo!®, and Vivisimo®, to that in two graphics-based search engines, Kartoo and Grokker. Search results normally consist of two types of information: *descriptive information* and *meta-information*. Descriptive information provides information of a retrieved document, including the title and a short summary of the document. Meta-information contains properties of a retrieved document, including its categories, document size, and relevant sponsored links.

Figure 7 Google Results for "kenya" Search

Figure 8 Yahoo Results for "kenya" Search

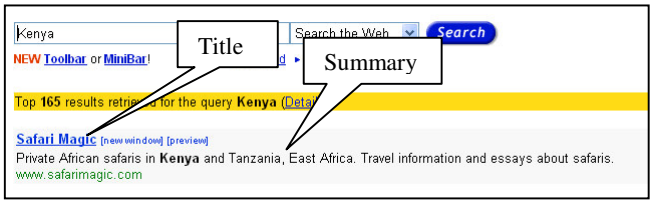

Figure 9 Vivisimo Results for "kenya" Search

Both text and graphics-based search engines organize the retrieved documents by categories to help users locate desired information. *Figures 7, 8 and 9* present images of text-based search engines, which usually use a list-based organization to categorize results. In contrast, *Figures 10 and 11* present images of two graphics-based search engines, which often use linked icons or graphical representations to depict the relevant categories.

Text-based search engines provide a meaningful title and a short summary of retrieved documents to show document relevancy (*Figures 7, 8, 9).*

To aid users further in their search process, text-based search engines also provide meta-information about the retrieved documents. The most common meta-information provided is the size of a retrieved document in kilobytes [3]. This information allows users to determine the amount of time needed to download the document. Cached pages, similar or related pages are also provided for users to retrieve in case the current URL is no longer accessible (*Figures 7 and 8*).

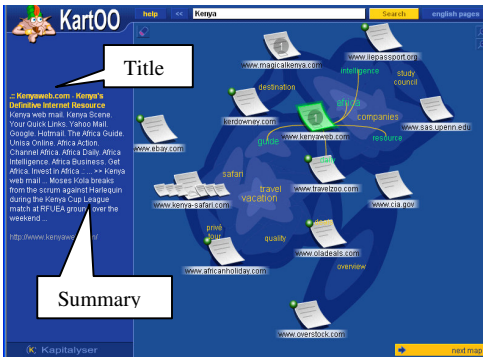

Figure 10 Kartoo Results for "kenya" Search

In contrast, the graphics-based search engines use graphical representations to display the descriptive information of search results. Kartoo and Grokker often directly present the title and a descriptive excerpt from a source document with little processing [4][9]. To view descriptive information in a flyover display users need to place their cursor over the icon (*Figure 10, 11*). Users are then required to remember the descriptive

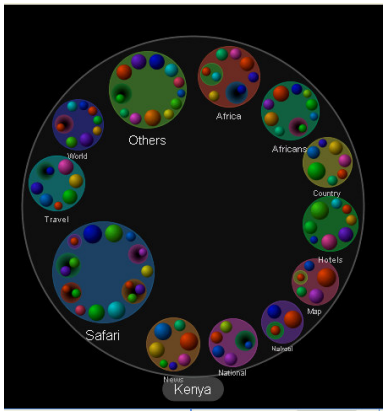

Figure 11 Grokker Results for "kenya" Search

information regarding each retrieved document in order to ultimately compare the results. This dependency on users' actions creates obstacles that prevent the graphics-based search engines from being more effective. In terms of meta-information, graphics-based search engines only provide users with viewing options, such as preview page and popup window.

Table 4 Comparison of Content From Two Types of Search Engines

Content Type	Feature	Function	Text Based	Graphic Based
Descriptive Content	Title	Gives search engine users an one phrase overview of the result document's content	✍	✍
	Short Summary	Enables search engine users to view the context in which the result document appears on the results page and presents a concise summarization of the result document	✍	✍
Meta-Info	Type of Web Page	Allows users to view the web page is a different format	✍	👂
	Document Size	Allows non-Broadband search engine users to determine the amount of time needed to download the result document	✍	👂
	Viewing Option	Allows search engine users to customize their viewing windows [Preview Page, Open Window, Open Frame]	✍	✍
	Related Information	Presents search engine users with additional information regarding their keyword query	✍	👂

In summary, as shown in *Table 4*, contrary to text-based search engines, which provide a useful overview of retrieved documents through their informative title and short summary, graphics-based search engines require users' clicks to reveal the content, increasing the amount of time users need to locate desired information. *This analysis reveals that text-based search engines are presently more effective than graphics-based search engines in terms of related content to support the search query*, addressing Objective 2.

Objective 3: Methods, Results and Discussion

Analysis of Existing Visualization Techniques

This study analyzed 12 existing visualization techniques found in literature to determine how these techniques may be applied to shorten the average time needed to obtain a quality result using graphics-based search engines. As displayed in *Table 5*, this analysis focused on the visual presentations of a direct display of descriptive information, categorization of retrieved results and relevancy. Descriptive information, categorization and relevancy allow users to located desired information more efficiently.

Table 5 Analysis of Content Presented on Existing Graphics-Based Techniques

Technique	Direct Display of Descriptive information	Categorization	Relevancy
Cat-a-Cone	✓	✓	
Perspective Wall	✓		
Scatter/Gather		✓	
WebBook & Web Forager	✓	✓	
Envision		✓	✓
TileBars			✓
VIBE			✓
SPIRE		✓	✓
Bead		✓	✓
LyberWorld		✓	✓
Focus+Context		✓	

Of the 12 existing visualization techniques, three techniques present a direct display of descriptive information, such as a Web page's title and short summary (*Table 5*). For example, the Perspective Wall's 3D format consists of a center panel for detailed views of the retrieved document and two side panels for the other retrieved documents [10]. These attributes allow users to easily compare the retrieved results.

Eight of the twelve visualization techniques incorporate a visual categorization of retrieved documents, which separate documents into clusters for easier comparison (*Table 5*). For example, the Cat-a-Cone interface presents a rotating hierarchy of categories, known as a Cone Tree. Users can click on the desired category and automatically bring the category forward in the hierarchy along with its associated documents [7]. This visual categorization allows users to quickly narrow down their set of desired results.

Exactly half of the 12 existing visualization techniques rank the relevancy of the retrieved document based on a search query (*Table 5*). For example, TileBars displays document relevancy by a color hue, where darker colors represent a more relevant paragraph. The overall shading of each document allows users to compare the relevancy of all documents [6]. These ranking methods provide users with more relevant results for their search.

Based on the analysis of *Table 5*, only three of the 12 visualization techniques focus on presenting a graphically organized display of descriptive information, which was the main user preference found from Objective 1. *This analysis therefore confirms that there is a need for new*

techniques that will provide users with a more effective means to visually depict descriptive information in the search results, addressing Objective 3.

Objective 4: Methods, Results and Discussion

A. Preliminary Visualization Design

To create a better visualization for graphics-based search engines, *this research has developed novel visualization techniques*. The key concept of the design was to combine the user preferred descriptive content with graphical representations of categories found in Objectives 1, 2 and 3. Graphical categorization will give users a quick lead to the relevant group of information among the huge amount of search results, and descriptive content will help users to quickly find a specific document for their needs.

Figure 12 Preliminary Design Visualization

A set of preliminary visualization designs/mock-ups were created using Adobe Illustrator based on the new design concept, and later embedded in HTML pages using Microsoft FrontPage for simulations of searching results (*Figure 12*). All content was based on the top ten search results found on Google for each specific search query and their descriptive information was directly presented in the design. As shown in *Figure 12*, the retrieved results of the search for "Terra cotta warriors" are organized by the graphical category. These graphical categories and sub-categories, originally presented in a textual format on Google results, allow users to immediately focus on the set of results more closely related to the search task. Each categorical group branches into a related image, a visual technique used narrow down the search results.

Finally, each image is connected to the actual set of retrieved results, and each result directly presents a title, content summary, URL, and file size.

B. Preliminary User Study and Feedback

Table 6 Tasks for Preliminary Study

Task A	Find a store to buy a Terra cotta warrior
	Find the price for a 150 cm Terra cotta warrior
Task B	Find where Terra cotta warriors are made in China

To test the above model, a second user study was conducted to obtain user feedback on the initial visualization designs. A total of 12 subjects participated in this study. After entering a search for "Terra cotta warriors," subjects were asked to complete two tasks, Task A and Task B (*Table 6*). One group used Google and the other using the mock-up of the novel designs. The effectiveness of each set of visualization was measured by the average time needed for users to complete the task. Users were then asked to report which visualization they preferred, Google or the mock-up, and any additional comments.

Table 7 Average Task Completion Times in Seconds

Search Engine	Task A	Task B
Google	130	68
Novel Visualization	104	63

According to the study results shown in *Table 7*, users took less time, on average, to complete the assigned tasks using the mock-up. Based on the feedback, seven users selected the mock-up as their preferred visualization, while five users preferred the Google display. This user study of the initial designs indicated the promising aspects of this new visualization technique.

C. Design Refinement

Based on the results from the above second user study, it was evident that the initial visualization design needed to be improved in order to be more effective.

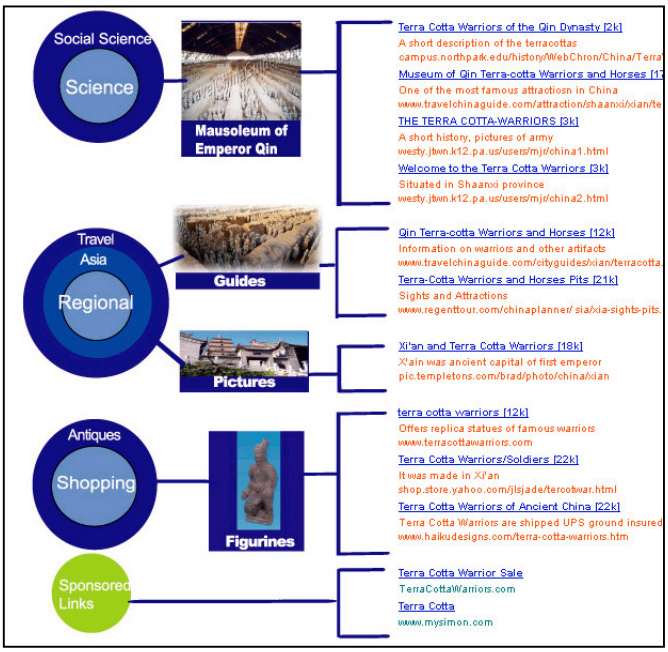

Figure 13a Categorized by Directory Categories

Therefore, a new design was created to incorporate the clustering of related documents in graphical brackets, presenting more clearly defined categories. A more apparent direct display of key information was presented by emphasizing category names and images (*Figure 13a*). Two novel sets of visualizations were created for the two different types of search tasks: (1) locating a general search topic and (2) locating a specific Web page (*Figures 13a, b, c, d and 14a, b, c*). Visualizations designed to locate a general search topic are sorted by directory categories, domain, country, and file format. The visualizations designed to seek a specific Web page are grouped by domain (*Figure 14a*), image, and match value of the search query. *These methods of categorization, which were originally presented on Google in a textual format, are emphasized visually to provide users with an immediate review of the retrieved results.*

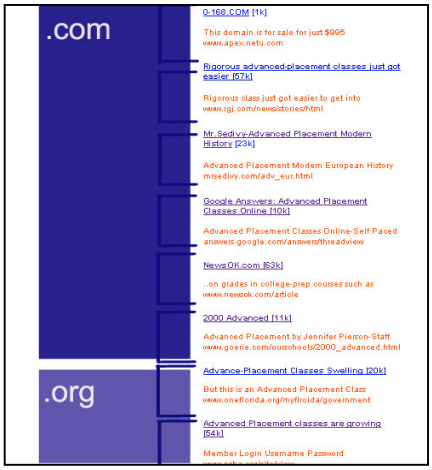

Figure 13b Categorized by Domain

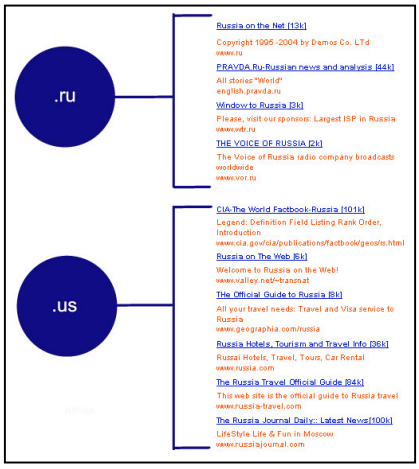

Figure 13c Categorized by Countries

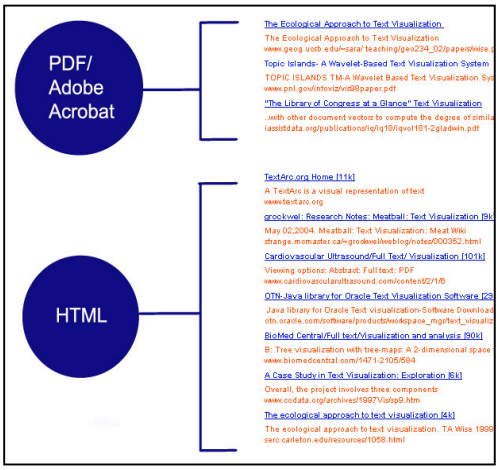

Figure 13d Categorized by File Format

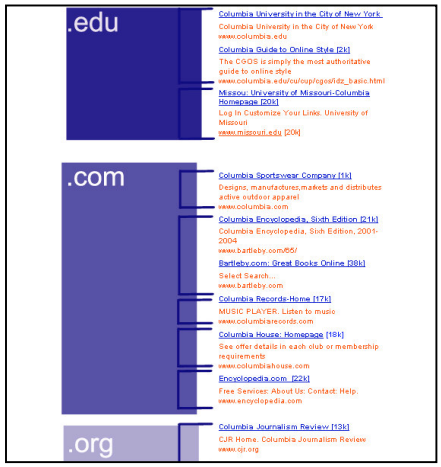

Figure 14a New Visualization Categorized by Domain

Figure 14b Categorized by Images

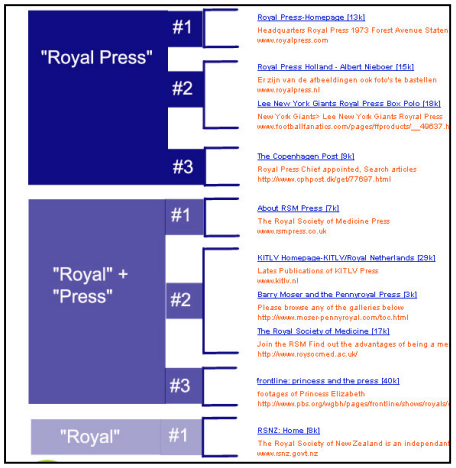

Figure 14c Categorized by Matching Text

D. Final User Study: Comparison Between Novel Visualizations and Google

To determine whether these novel visualizations were more effective than Google's current layout, a third user study was conducted. The identical 24 subjects from the first study participated in this study. This user study contained 16 different visualizations, eight novel visualizations and eight Google visualizations, Users were asked to complete eight different search tasks, four tasks using Google and four using the novel visualization (*Table 8*). Effectiveness of each visualization was assessed by total time needed to complete a task.

Table 8 Task Assignments for Final Study

#	Task
1	Find a website where you can purchase a terra cotta warrior
2	Find a website where you can take an online Advanced Placement class
3	Find a journal paper about text visualization
4	Find a United States guide for Russian Travel
5	Find the homepage for the University of Missouri-Columbia
6	Find the homepage for Royal Press Publishers
7	Find the Google page for information about online Advanced Placement classes
8	Find the NIMD homepage from ARDA

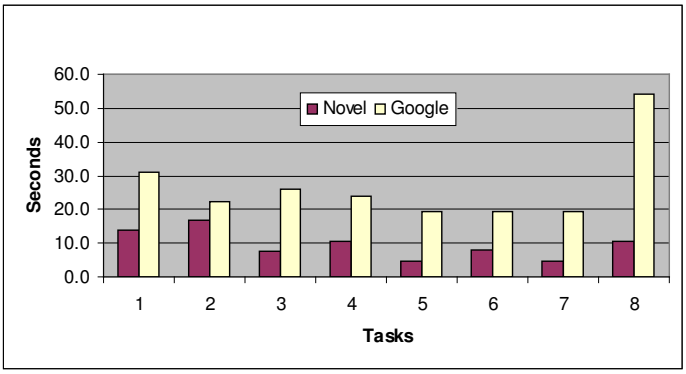

Figure 15 Average Completion Time for 8 Tasks

As shown in *Figure 15,* the average of all completion times using the novel visualization is 9.5 seconds while the average of all Google times is 26.8 seconds. *Therefore, it was evident that the average time to complete each task using the novel visualization is less than that using Google*.

To statistically confirm these findings, a *matched pairs T-test* was conducted to compare the average time to complete tasks using Google to the average time to complete tasks using the novel visualization. According to the results

H_O: μ Novel Visualization – μ Google = 0
H_A: μ Novel Visualization – μ Google <0
T-Test Results: T = -4.366 p=1.644 E-3
Because of the significantly small p-value ($p<0.05$), H_O can be rejected.
Therefore, the average completion time using the novel visualization is less than the average completion time using Google.

Figure 16 Statistical Analysis: Matched Pairs T-Test

shown in *Figure 16*, the p-value = 1.644 E-3 is less than $\alpha = .05$ and the null hypothesis is rejected. *It is statistically confirmed that the average task completion time using the novel visualization is less than the average task completion time using Google*.

These results show that the novel visualizations created in this research were more effective than the currently most preferred text based search results organization, Google. Furthermore, due to the visually organized, direct display of descriptive content presented in these novel visualizations, these visualizations outperformed Kartoo, which failed to be more effective than Google as shown in the study performed for Objective 1. *Therefore, these novel visualization designs implemented the key visualization properties needed in an effective graphics-based search engine, fulfilling Objective 4. Furthermore, these novel visualizations, created for both general and specific searches, were more effective than the current popular text-based search result organization.*

Conclusion and Significance

The novel findings and innovative concepts in this research indicate a great potential for graphics-based search engines to become the new mainstream search engine. This study has also created a new foundation for future research in information visualization for search engines. More effective visualization techniques have the potential to significantly improve the search experience for all users. This will enable all valuable information on the World Wide Web to be more accessible to all people, most importantly senior citizens, young children, and people with reading disabilities and language barriers.

References

[1] P. Andronico, M. Buzzi, B. Leporini, "*Posters: Can I find what I'm looking for?",* Proceedings of the 13th International Conference on World Wide Web, New York, New York, May 2004, pp. 430-431.

[2] S.K. Card, G.G. Robertson, and W. York, "*The WebBook and the Web Forager: An Information Workspace for the World-Wide Web*", Proceedings of ACM CHI '96 Conference on Human Factors in Computing Systems, Vancouver, British Columbia, April 1996, pp. 111-117.

[3] Google, 2003. How to Interpret Your Search Results. http://www.Google.com/help/interpret.html.

[4] Grokker, 2004. Help Desk Request. http://www.groxis.com/cgi-bin/perldesk/kb.cgi

[5] D. Fetterly, M. Manasse, M. Najork, and J. Wiener, "*A Large-Scale Study of the Evolution of WebPages",* Proceedings of the 12th International Conference on World Wide Web, Hungary, May 2003,pp. 669-678.

[6] M.A. Hearst, "*TileBars: Visualization of Term Distribution Information in Full Text Information Access*", Proceedings of ACM CHI '95 Conference on Human Factors in Computing Systems, May 1995, pp. 56-66.

[7] M.A. Hearst and C. Karadi, "*Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy*", Proceedings of 20th Annual International ACM/SIGIR Conference, Philadelphia, PA, July 1997, pp. 246-255.

[8] M. Hemmje, C. Kunkel and A. Willett, "*LyberWorld-A Visualization User Interface Supporting Full text Retrieval*", Proceedings of 17th Annual International ACM/SIGIR Conference, July 1994, pp. 249-259.

[9] Kartoo, 2004. How to User KartOO. http://www.kartoo.net/a/en/aide01.html

[10] J.D. Mackinlay, G.G. Robertson and S.K. Card, "*The Perspective Wall: Detail and Context Smoothly Integrated*", Proceedings of ACM CHI '91 Conference on Human Factors in Computing Systems, New Orleans 1991, pp. 173-179.

[11] L.T. Nowell, R.K. France, D. Hix, L.S. Heath, and E.A. Fox, "*Visualizing Search Results: Some Alternatives To Query-Document Similarity*", Proceedings of 19th Annual International ACM/SIGIR Conference, Zurich, August 1996, pp. 67-75.

[12] A. Ntoulas, J. Cho and C. Olsten, "*What's New on the Web? The Evolution of the Web from a Search Engine Perspective",* Proceedings of the 13th International Conference on World Wide Web, New York, New York, May 2004, pp. 1-12.

[13] Search Engine Watch, "Nielsen NetRatings Search Engine Ratings", July 2004, http://searchenginewatch.com/

[14] M.M. Sebrechts, J.Vasilakis, M.S. Miller, J.V. Cugini, and S.J. Laskowski, "*Visualization of Search Results*: A Comparative Evaluation of Text, 2D, and 3D Interfaces", Proceedings of 22nd Annual International ACM/SIGIR Conference, Philadelphia, PA, August 1999, pp. 3-10.

[15] D. Woods. The alarm problems and directed attention in dynamic fault management, Ergonomics, 1995