A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces

ABSTRACT

There have been several studies that compare sequential search results versus clustered search results, and graphical presentations versus textual presentations. These studies have resulted in confirmed efficiency and preference of clustering over sequential lists. The studies between graphical and textual presentations have usually shown to be task dependant. This study shows a systematic evaluation of zoomable versus textual clustered search results. A controlled experiment with repeated measures design and within-subjects differences was performed with fifteen subjects, comparing Groxis, Inc.'s Grokker – their clustering product – a zoomable user interface, their textual clustering product and Vivisimo's textual clustering product. No significant differences were found for objective measures. However, there were significant differences for subjective measures. The textual clustering interfaces was preferred and elicited major satisfaction among the users. Results are summarized in both a quantitative and qualitative format.

Author Keywords

Information retrieval, textual clusters, graphical clusters, clustering, user test, evaluation, zoomable user interface, ZUIs.

ACM Classification Keywords

H.5.2 User Interfaces, H.5.4 User issues.

INTRODUCTION

Search engines are behind some of the most popular and widely accessed web sites because they allow the user to find the information they seek. However, with the explosive growth of information that is available through the World-Wide-Web, most queries result in many retrieved documents. [8] The most common search engines show the results as pages of scrolled lists, which are usually ordered by rank [16]. Users are forced to examine this long list of ranked documents in order to find relevant documents. [26] The list is in decreasing order in terms of the single measure of relevance used by the system. [10] The results on different topics or different aspects of the same topic are sometimes mixed together in the returned list. [2] This sequential list often comes with no indication of the system's evaluation of each document. There is no clear indication of the relationships that may exist among the retrieved documents. [10]

Another approach to present search results is by organizing them into clusters. Document clustering has been used in experimental information retrieval systems for many years. [21] Clustering is achieved by grouping documents into "topically coherent" groups. The topic coherent clusters can be used to identify promising subsets of documents, exemplars for relevance feedback or to eliminate groups of documents whose contents appear irrelevant. [8] The purpose of clustering is not to improve precision in search results, but rather an attempt to make search engine results easier to browse. [26] Clustering is useful for browsing large online collections of text as shown by Chen & Dumais. [5]

The standard practice of search engines is to provide brief textual summaries of the Web documents. [22] Users have to scan these summaries to find the information they are looking for. Efficiency has shown to be improved if similar documents are clustered together. A user will read the summary of the first document in a cluster and decide to inspect another if it appears to have irrelevant information. [21, 5]

Studies have been completed comparing search effectiveness between clustered retrieval and sequential retrieval. Clustered presentations have been shown to be more effective than sequential presentations, and are superior objective and subjective measures [21, 2].

Results given by the most popular search engines are displayed with a textual presentation. However, a graphical presentation of search results can also be achieved. With a graphical presentation, more information about more retrieved objects can potentially be conveyed in a small

space since each attribute of a graphical object can be used to convey document characteristics. [12] Graphically displaying document information has the potential to enable users to understand search results more quickly. [20]

There are many kinds of visualizations that might be used for search result representation. For example, document maps visualize the overall similarity structure of a corpus of texts, using a suitable metaphor which is reminiscent of geographical or astronomical cartography. Independent of the metaphor, the basis of the methods is that similar documents or document groups are located in neighbored areas in the respective map display. [2]

The addition of dynamic user control through direct manipulation principles allows the user to navigate large information spaces. [15, 18] The use of animated presentations and user-controlled selections enable users to explore large information spaces rapidly and reliably. [1]

Studies have been completed comparing search effectiveness between text and graphical interfaces. For example, Sebrechts et al. compared textual versus 2-D and 3-D interfaces [17]. Textual and 2-D visualizations were revealed to be more effective than 3-D visualizations. Another study by Becks et al. showed that the graphical overview offered significant benefits over a text-based access interface. [2] It must be noted, that for both, the results were task dependant. Certain tasks favored graphical interfaces and others favored textual interfaces.

Our interest is in visualizations that are based on Zoomable User Interfaces (ZUIs) [4]. These are systems that lay out information spatially on a 2D plane at different sizes, and use animated transitions to navigate through the information. ZUIs have been evaluated in a range of contexts and have been shown to have significant benefit for some domains and tasks [3, 7, 9]. The current study evaluates three different clustering interfaces. The first interface is called Grokker, a Zoomable User Interface (ZUI) produced by Groxis, Inc. [24]. The second interface is a textual version of the same clustered search results, also by Groxis, referred to as Grokker Text. The third interface is called Vivisimo and is another textual clustering interface [25]. This study attempts to provide a controlled comparison between these three interfaces on a number of typical information seeking tasks in an attempt to develop a better understanding of the potential that a ZUI based visualization may offer clustered search results. Since we are comparing specific instances of visual and textual approaches to clustered search results, we do not expect that this study will give a definitive answer as to whether ZUI-based or textual representations are better in general. But since we are analyzing apparently successful commercial solutions, we hope this study will offer some insight into the general approaches in addition to the specific systems.

PRODUCT DESCRIPTION

Grokker

Grokker is a graphical knowledge map. It provides a graphical representation of information such as search engine's results, databases, your computer's files, etc. The clustering mechanism is shown as a series of nested Venn Diagrams. Figure 1 illustrates a search for one of the tasks performed in the experiment: "Find in which city and on what date was the Super Bowl played for the 1998 season."

Users can navigate within a hierarchy by clicking on the circles, which represent the clusters and sub-clusters. When a smaller circle is clicked, the screen zooms in with an animated transition and sub-clusters within can be observed. When a bigger circle is clicked, the screen zooms (again animated). Filled circles indicate additional levels. Transparent circles indicate the lowest level. Once the user has reached the lowest level in the hierarchy, he/she can access the result by clicking on the figure and going to the browser window. This window can be viewed simultaneously with the graphical clustering interface.

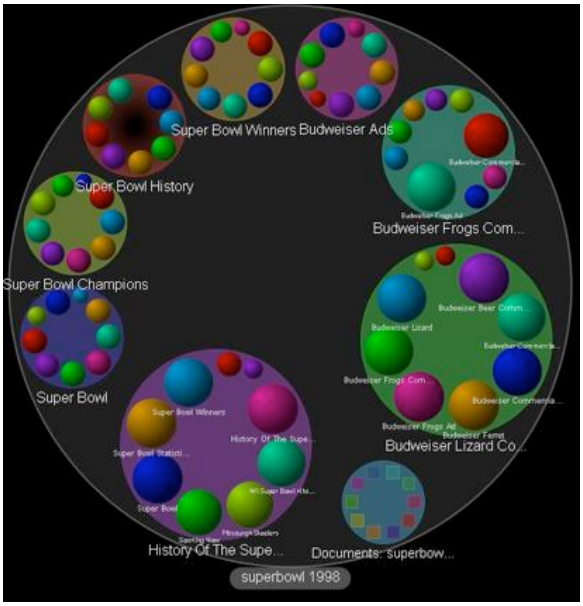

Figure 1: Grokker

Grokker Text

Grokker Text is another representation offered by Grokker. It provides the same results as the graphical clustering interface displayed in manner similar to Mac OS X's Finder. Figure 2 illustrates a search for one of the tasks performed in the experiment.

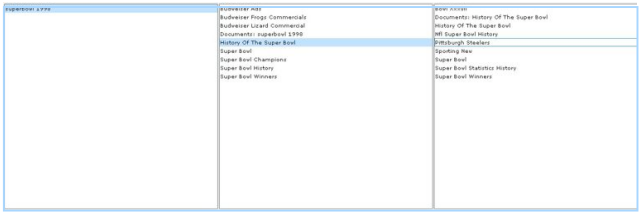

Figure 2: Grokker Text

Users can navigate within a hierarchy by clicking on the text and sub-clusters appear to the right. Once the user has reached the lowest level in the hierarchy, he/she can access the result by clicking on the text which loads that page in the browser window. This window cannot be viewed simultaneously with the textual clustering interface, so the user must manually switch to the browser window.

Vivisimo

Vivisimo is a textual clustering interface. It provides a clustering representation of information of Vivisimo's search engine results. The clustering mechanism is shown on the left frame. It provides the results displayed in manner similar to Microsoft Windows Explorer. Figure 3 illustrates a search for one of the tasks performed in the experiment.

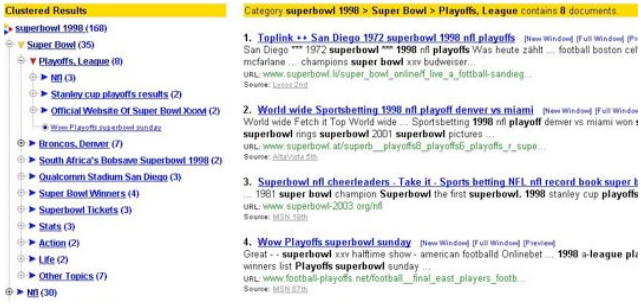

Figure 3: Vivisimo

Users can navigate within a hierarchy by clicking on the plus sign or the triangle icon. Results are shown at any level of the hierarchy. The user can select to view/load a result at any time at it can be viewed simultaneously with the textual clustering interface.

STUDY OBJECTIVE

The goal of this study was to determine usability differences between the three clustering interfaces: Grokker (ZUI), Grokker Text (textual) and Vivisimo (textual). In order to do so, a controlled experiment was performed with within-subject differences. Usability can only be determined once users attempt representative tasks using the interfaces. Thus, participants performed typical search tasks. Their performance was monitored and subjective preferences were recorded.

PARTICIPANTS

The pilot test was performed by two participants and the experiment by fifteen. They participated in a voluntarily basis with no payment for their involvement. Information of the surveyed was gathered, not to elaborate trends on this data, but as to determine the demographics of the sample. Two females, ages 24 and 32 participated in the pilot test. The experiment sample was composed of seven females and eight males, the age of the participants varied between 18 and 40 years. Seven participants' native language was not English. However, there were no differences in their performances when compared to the others. All participants were students (undergraduates and graduates) at the University of Maryland.

PILOT TEST

Prior to the testing, two pilot tests were performed in order to verify how the different metrics would be collected. The experimental design for the pilot tests had accuracy, completion time and satisfaction as dependent variables; and interface as the independent variable varying across the three interfaces. In order to obtain greater statistical power, a controlled experiment with within-subjects differences was designed.

Lessons Learned

Pilot Test #1

The same tasks were tested in all interfaces. It had been decided to test the same tasks to prevent task content effects. However, subjects lost interest and modified their search strategy over time.

There were three tasks per interface, which generated a long experiment. It was one and a half hours long, so the experiment was shortened.

The training and familiarization period given for Grokker created an advantage over the other interfaces, so training for all interfaces was equalized.

Pilot Test #2

Each interface was tested with two different tasks. This reduced the experiment to one hour.

For training, a brief explanation of how all the interfaces worked was given by the test administrator. The participant was guided through practice searches before performing the tasks for each interface.

Avoid between-subject variance

For both Grokker interfaces, the search took between 5-40 seconds to show all the available results. One subject waited for searches to complete and the other did not. For the actual study, we made all participants wait until search is complete.

When searching, one subject went to cnn.com through Grokker and then tried to use CNN's search function to complete the task. We thought this search method would not allow us to control the study appropriately. Thus, for the actual study, participants were not allowed to use websites' search engines.

Simplify the Experimental Design

Grokker has additional features that can be of assistance when performing a search. These features are filtering options, color-coding, shape coding and magnification. The first participant of the pilot test took advantage of these features, but the second did not. We decided that participants would not be allowed to employ these features because it would mean adding another independent variable to the experiment.

EXPERIMENT

Tasks

A list of six tasks was developed, each involved finding a specific piece of information. Each task had only one correct answer. (Example: Find the name of the city where Ecuador and Mexico played during last year's World Cup. Answer: Miyagi.)

A Graeco-Latin Square was used to establish task order for each participant and to confound task order effects.

Test Facility and Materials

The test apparatuses used included an IBM ThinkPad laptop (700 mHz, PIII, 768 Mb RAM) and a stopwatch. All participants were instructed to use the mouse; they were not allowed to use the laptop's stick pointer.

Test Administrator Tools

Pre-Test Questionnaire

A pretest questionnaire was submitted to obtain basic demographic information and to obtain the user's experience with computers and search engines.

Posttest Questionnaire

A posttest questionnaire was submitted to participants in order to gauge the approval of users with each interface. An excerpt from the Questionnaire of User Interface Satisfaction (QUIS) was used. Participants answered a 9-point Likert Scale.

EXPERIMENTAL DESIGN

The following list defines the dependent variables and the measures for which data were recorded for each set of conditions.

- *Completion Time*, measured from the start to the end of each task.
- Accuracy, tasks were rated as completed correctly, completed incorrectly or aborted and this condition was recorded.
- Satisfaction, subjective appreciation of the users was obtained by means of the post-test questionnaire.
- Preference, after completing the experiment, participants were asked which interface, among the three, they would use.

The independent variable was the interface, which varied between: Grokker (ZUI clustering), Grokker Text (textual clustering) and Vivisimo (textual clustering).

PROCEDURE (TEST PROTOCOL)

One person administered the experiment. All testing sessions were approximately sixty minutes long.

The participant was greeted and the general instructions of the test were provided. A pre-test questionnaire and consent form were then administrated. They were then given a brief explanation of the functionalities of the first interface. A search was performed with the test administrator's guidance. After the familiarization period, the task instructions were explained and the user performed the set of tasks. Once two tasks were performed, the user was presented with a Posttest Questionnaire to gauge the satisfaction for that specific interface. These actions were performed for each interface. Finally, the participant was asked which interface was preferred. To conclude the session, the user was thanked for his/her participation and dismissed.

Participant General Instructions

Each task was read aloud. The participant was asked if he/she understood the task to follow. If not, a further explanation was given. Once the participant was ready, he/she started to perform the task.

Task Rules

- 10 minutes time limit
- User can give up before 10 minutes
- Wait for search to be over before starting
- For Grokker and Grokker Text, only the Teoma database could be used
- User can navigate within a result's website
- User cannot use search function within result's website
- User can modify search keywords and perform new searches
- If you already know the answer, you still have to use the interface and find it
- Once the answer is found, inform the test administrator and show answer

Tasks

All tasks had only one correct answer. This answer could be obtained in different websites and by using different keywords.

Usability Metrics

Effectiveness metric

Effectiveness relates the goals of using the interface to the accuracy and completeness with which the goals were achieved. The measures of effectiveness used were:

- Accuracy – This was the percentage of tasks correctly completed per interface. A task could be completed correctly, completed incorrectly or aborted. All these occurrences were recorded. A correctly completed task is one where the user completes the task and obtains the required result. An incorrectly completed task is a task completed by the user, which did not obtain the required results. An aborted task is one that the user quits while performing the task.

Efficiency metrics

The efficiency was defined in this study as the average time taken to achieve the task.

- Completion time – The time taken to complete each task was recorded.

Acceptance

Acceptance describes the user's subjective level of satisfaction when using the interfaces. The measures of acceptance were:

- Satisfaction - was measured using posttest questionnaires based on the QUIS.
- Preference – users identified which interface they would select to use among the three.

Sources of Error

Subjects Recruited: The ideal subject composition would be that of a subset of the target market, with a proportional representation of the actual market. Additionally, subjects were asked to perform under conditions where all their moves were recorded and timed; this could affect what otherwise could be their real performance.

Sample Size: The sample size was not ample enough to maximize the statistical power of the study or minimize the potential for error.

Human Error: The measures recorded, especially that of time are subject to human error.

Databases: The experiment could not control the databases used in which the search query was performed. Grokker and Grokker Text had the same database (Teoma), but Vivisimo had a different database (Vivisimo). The results shown in these interfaces were usually different and a sure source of variability.

Task Content: Although the tasks specified for the experiment attempted to be of a similar nature. Their content could be an additional source of variability.

RESULTS

Data Analysis

The following section summarizes the performance data collected. Statistical analysis on the data set was performed in order to evaluate task duration, accuracy, satisfaction and preference. The General Linear Model Approach was used. The experimental design was a repeated measures design with three levels and within-subjects differences.

Completion Time

Completion time for Grokker was 304 sec; Grokker Text, 270 sec; and Vivisimo, 212 sec. Although there was a clear trend, there were no significant differences on completion time, $F(2, 28)=2.30$, $p=0.119$. Additionally, there were no significant differences for order effects, $F(2, 28)=1.91$, $p=0.319$; and no significant differences for task content effects, $F(5, 70)=1.054$, $p=0.393$. Figure 4 includes a summary of the data collected for the testing sessions.

Accuracy

All interfaces had approximately the same accuracy rates, where Grokker had 80.0% tasks completed correctly; Grokker Text, 76.7%; and Vivisimo, 86.7%. There were no significant differences for accuracy, $\chi^2(2)=1.015$, $p=0.602$. There were no significant differences for order effects, $\chi^2(5)=5.584$, $p=0.349$. However, there were significant differences for task content effects, $\chi^2(5)=17.768$, $=0.003$. One task had a 100% completion rate and another task had 46.7% completion rate. Figure 5 includes a summary of the data collected for the testing sessions.

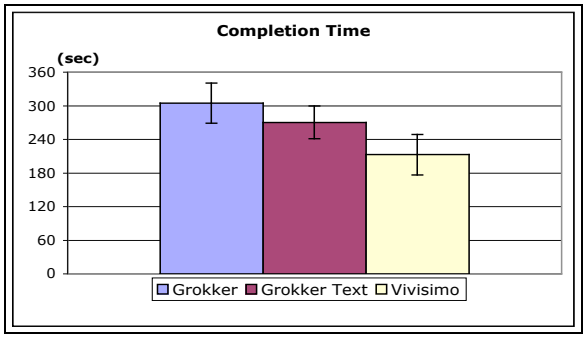

Figure 4: Completion Time

Satisfaction

Overall User Satisfaction: The results were obtained by calculating the mean of responses for all questions in the QUIS. Grokker had a rating of 5.64; Grokker Text, 5.47; and Vivisimo, 6.89. There were significant differences for overall satisfaction, $F(2,28)=5.45$; $p=0.01$ Vivisimo is significantly different from Grokker and Grokker Text, $F(1, 14)=10.04$; $p=0.00$ There were no significant differences for order effects, $F(2,28)=0.43$, $p=0.65$ Figure 6 includes a summary of the data collected for the testing sessions.

QUIS: The results are obtained by calculating the mean of responses per answer. Table 1 shows a summary of the satisfaction data.

Preference

Vivisimo was significantly preferred over the other interfaces, $\chi^2(2)=22.20$; $p=0.000$. Grokker had a preference of 13.3%; Grokker Text, 6.7%; and Vivisimo, 80.0%. There were no significant differences for order effects, $\chi^2(2)=0.600$, $p=0.741$. Figure 7 includes a summary of the data collected for the testing sessions.

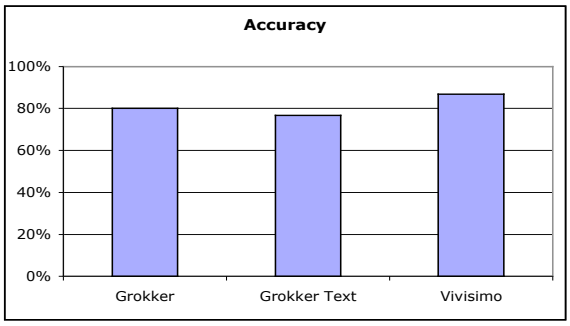

Figure 5: Accuracy

	Grokker	Grokker Text	Vivisimo	F (2, 28)	p
wonderful	5.53	5.53	6.47	1.65	
satisfying	5.27	4.60	6.13	2.43	
stimulating	7.80	5.00	5.60	13.53	p<0.001
easy	5.07	5.40	7.00	6.80	p<0.01
power	5.60	5.20	6.47	2.29	
flexible	5.13	5.13	6.73	3.77	p<0.05
read	4.67	6.27	7.73	9.03	p<0.001
logical	5.47	5.73	7.07	2.45	

navigation	5.67	5.67	7.07	2.36	
learn	6.60	5.67	7.27	2.83	
straightforward	5.40	6.00	7.73	8.57	p<0.001
easy complete tasks	5.53	5.20	7.13	7.83	p<0.01

Table 1: Posttest Questionnaire

Preference

Vivisimo was significantly preferred over the other interfaces, $\chi^2(2)=22.20$; p=0.000. Grokker had a preference of 13.3%; Grokker Text, 6.7%; and Vivisimo, 80.0%. There were no significant differences for order effects, $\chi^2(2)=0.600$, p=0.741. Figure 7 includes a summary of the data collected for the testing sessions.

Comments

Comments were not requested; however, several users commented after the experiment. These comments have been edited so that they are not replicated when users said something similar.

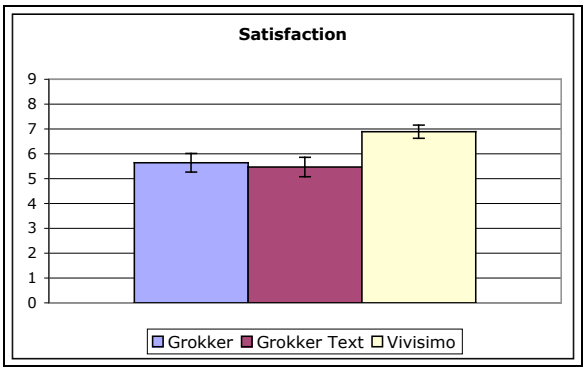

Figure 6: Satisfaction

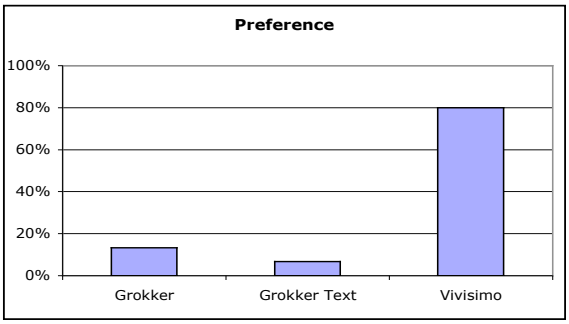

Figure 7: Preference

Regarding Grokker, users commented:

- "What is the clustering criterion?"
- "The system should have a better way of informing that there are NULL search results."
- "Search engine should eliminate files no longer accessible."
- "What do the colors mean?"
- "It was pretty but hard to use."
- "Very nice, cute."

Regarding Grokker Text, users commented:

- "The details do not explain when I'm on a cluster or on a leaf."
- "There were too many levels, I wanted to see possible results sooner."
- "I would use this for deep research, for fast research Vivisimo."
- "Sorry I'm changing keywords… I just remembered how bad this one was and wanted to be specific."
- "Much less exciting than the other (Grokker)."

Regarding Vivisimo, users commented:

- "The cluster window should be independent of the main window. I should be able to look at sub-clusters while a page is loading without being redirected elsewhere."
- "It's easier to switch from one branch to another."

OBSERVATIONS

Following is a list of salient occurrences observed by the experimenter.

Grokker

- NULL results were represented in one huge circle. This confused the users, who kept clicking on the circle trying to zoom in and get more results.
- Clusters were re-visited several times. There is no visual indication that a cluster has already been visited.
- Colors in Grokker could be mapped to indicate relevance or time. However when not mapped, users wanted to know if they had any meaning.
- A cluster labeled: "Documents" appeared in every single search and also as a sub-cluster. Not sure what it means, the results in it were of the same type as in other clusters. However, users avoided this cluster
- Cluster labels were cut-off some times preventing users to read them.
- The Navigation buttons in the browser window were usually overlooked, unless test administrator indicated where they were located.
- When there's one single cluster or when users reached the smallest cluster, they loaded all pages even though names were not meaningful for the task. This could suggest a navigation cost.

Grokker Text

- NULL results were displayed as normal cluster, users clicked several times before realizing there was nothing
- Users could not anticipate that they had reached a result unless they could no longer navigate down the hierarchy. Grokker shows results visually in two ways: no smaller clusters and browser window and clusters can be viewed simultaneously. In Grokker Text, users had to tab to the browser window to look at a result.
- Only three people paid attention to the detail portion, which occupies 50% of screen space. This portion gives information about the URL and sometimes when it was created. However it was inconsistent.

- See comments above for "documents" cluster, single and low level cluster

Vivisimo

- Users looked at results before going to clusters. This one of the advantages of Vivisimo. Results can be accessed immediately, in the other two interfaces, the user had to navigate to the lowest level before seeing a result.
- Short summaries under each result assisted users when deciding to load a page.
- The navigation menu for accessing sub-clusters is similar to Windows Explorer. Users click a plus sign to access lower levels. However, there was a triangle next to this sign that seemed performed the same function. The difference was that the plus sign allowed for the user to look at the hierarchy while a result page was loaded, while the triangle stopped loading the result page and loaded a cluster page. Users would always click one sign throughout the experiment and; those who clicked on the triangle, would become frustrated when the page stopped being loaded.

CONCLUSION

Effectiveness

There were no significant differences for effectiveness among the three interfaces. Users can perform a successful search using any of the interfaces.

Efficiency

There were no significant differences for efficiency among the three interfaces, although there was a strong trend. Users took an average of approximately 5 minutes to find information in Grokker, 4.5 minutes in Grokker Text and 3.5 minutes in Vivisimo.

Acceptance

Significant differences where found both for satisfaction and preference. Users where significantly more satisfied and preferred using Vivisimo when comparing it to the other two interfaces.

The results in this study did not show any task dependency for graphical/ZUI or textual interfaces. The ZUI caused an impression when first looked at and was considered "cool" by most of the participants. Although no significant differences in efficiency and effectiveness where found between the three interfaces, it is the level of satisfaction, preference and sense of accomplishment what can cause the success of a product. In this experiment, the textual clustering interface, Vivisimo, was the users' favored choice.

Design Analysis

Through several years of building and evaluating ZUIs, we have found in general that they are most likely to be successful when: 1) There is a natural spatial layout of the data; and 2) Good small representations of the data are available. The current domain, however, offers neither of these characteristics and so we are not surprised about the lack of positive results for ZUIs, despite this commercial interest – which we suspect has more to do with novelty than effectiveness.

Nevertheless, we do think this ZUI design could be improved. For example:

- Navigational cues could inform users which elements have already been visited.
- Hierarchical cues could tell users which level they are in and how deep the structure is.
- Clearer representation of the underlying web pages could help indicate to users what pages are there. The current circles are largely uninformative.
- Users should be able to see the navigation clusters as well as the actual website in the same window.

The authors recognize that there will be certain issues regarding the generalization of these results, due to the constraints of comparing specific visualization tools. However, the evaluation performed in this study can form a basis of recommendations and design considerations for future clustering search result visualizations.

ACKNOWLEDGMENTS

REFERENCES

1. Ahlberg, C. & Shneiderman, (1994) B. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. *In Human Factors in Computing Systems: Proc. of the CHI '94 Conference*. New York: ACM, 313-317.
2. Becks, A., Seeling, C. & Minkenberg, R. (2002) Benefits of Document Maps for Text Access Knowledge Management: A Comparative Study. *Proc. of the ACM Symposium on Applied Computing (SAC2002),* Madrid, Spain, 621-626.
3. Bederson, B. B., & Boltman, A. (1999). Does Animation Help Users Build Mental Maps of Spatial Information? In *Proc. of Information Visualization Symposium (InfoVis 99)* New York: IEEE, pp. 28-35.
4. Bederson, B. B., Meyer, J., & Good, L. (2000). Jazz: An Extensible Zoomable User Interface Graphics Toolkit in Java. *ACM Symp. on User Interface Software and Tech.*, (*UIST 2000),* CHI Letters, 2(2), pp. 171-180.
5. Chen, H. & Dumais, S. (2000) Bringing Order to the Web: Automatically Categorizing Search Results. *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'00),* ACM, 145-152.
6. Cutting, D.R., Krager, D.R., Pederson, J.O. & Tukey, J.W. (1992) Scatter/Gather: A Cluster-based Approach to Browsing Large Document Clusters. *Proc. of the ACM SIGIR '92 Conference on Research and*

Development of Information Retrieval, Copenhagen, Denmark, 318-329.

7. Gandhi, R., Kumar, G., Bederson, B. B., & Shneiderman, B. (2000). Domain Name Based Visualization of Web Histories in a Zoomable User Interface. In *Proc. of Eleventh International Workshop on Database and Expert Systems Applications - Second International Workshop on Web-Based Information Visualization (WebVis 2000)*, IEEE Computer Society, Los Alamitos, CA (2000), pp. 591-598.
8. Hearst, M. & Pederson, J.O. (1996) Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. *Proc. of SIGIR '96 Conference on Research and Development in Information Retrieval,* Zurich, August 1996, 76-84.
9. Hornbæk, K., Bederson, B. B., & Plaisant, C. (2002) Navigation Patterns and Usability of Zoomable User Interfaces With and Without an Overview, *ACM Transactions on Computer-Human Interaction*, 9 (4), 362-389.
10. Iwayama, M. & Tokunaga, T. (1995) Cluster-Based Text Categorization: A Comparison of Category Search Strategies. *Proc. of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieva*l, Seattle, US, ACM Press, New York, US, 273-281.
11. Korfhage, R.R. (1991) To See, or Not to See – Is That the Query? *Proceedings of the 14th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, Chicago, 134-141.
12. Mukherjea, S., Hirata, K. & Hara, Y. (2000) Using Clustering and Visualization for Refining the Results of a WWW Image Search Engine. *Proc. of the 1998 workshop on New paradigms in information visualization and manipulation*, ACM Press, New York, 29-35.
13. Nielson, J. & Mack, R.L. (1994) *Usability Inspection Methods,* John Wiley & Sons, Inc., New York, 1994
14. Nowell, L.T., France, R.K., Hix, D., Heath, L.S. & Fox, E.A. (1996) Visualizing Search Results: Some Alternatives to Query-Document Similarity. *Proc. of the nineteenth annual international ACM/SIGIR conference on research and development in information retrieval,* August 1996, 67-75.
15. Robertson, G.G., Card, S.K., & Mackinlay, J.D. (1993) Information Visualization using 3-d Interactive Animation, *Comm. Of ACM 36*, 4, 56-71.
16. Salton, G. (1971) Cluster search strategies and the optimization of retrieval effectiveness. In G. Salton, editor, *The SMART Retrieval System*. Prentice Hall, Englewood Cliffs, N.J., 223-242
17. Sebrechts, M.M, Vasilakis, J., Miller, M.S., Cugini, J.V. & Laskowski, S.J. (1999) Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. *22nd Annual Int'l ACM-SIGIR Conf. Research and Development in Information Retrieval*, ACM Press, New York, 1999. 3-10.
18. Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Second Edition*, Addison-Wesley Publ. Co., Reading, MA, 1992, 573 pages.
19. Scholtz, J. (1998) Kiosk-Based User Testing of Online Books. *Proc. of the Sixteenth Annual International Conference on Computer Documentation*, September 1998, 80-88.
20. Veerasamy, A. & Heikes, R. (1997) Effectiveness of a Graphical Display of Retrieval Results. *Proc. of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, Philadelphia, PA, 236-245.
21. Voorhees, E.M. (1985) The Cluster Hypothesis Revisited. *Pro. of ACM SIGIR*, June 1985, ACM Press, New York, US, 188-196.
22. Weippl, E. (2001) Visualizing Content Based Relations in Texts. *Australian Computer Science* Communications, v.23 n.5, January-February 2001, 34-41.
23. Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J. & Pirolli, P. (2001) Using Thumbnails to Search the Web. *Proc. of the Conference on Human Factors in Computing Systems*, CHI 2001, Seattle, WA, 198-205.
24. www.groxis.com
25. www.vivisimo.com
26. Zamir, O. & Etzioni, O. (1998) Web Document Clustering: A Feasibility Demonstration. *Proc. ACM SIGIR'98*, Melbourne, Australia, 46-54.