Investigating the Effect of Visualization on User Performance of Information Systems

Xiaojun Yuan

College of Computing and Information, University at Albany,
State University of New York, Albany, NY 12222, US
xyuan@albany.edu

Abstract. This research investigated if and how users using an information visualization system perform differently from those using a traditional information retrieval system. A between-subjects experiment was conducted involving 32 subjects. One group of subjects engaged the visually oriented CiteSpace system; the remaining subjects searched the textual-based Web of Science system. The results indicated that subjects using the CiteSpace system spent significantly less time, felt significantly more satisfied, and performed significantly fewer mouse clicks than those using the Web of Science system. These results indicate that it would be helpful to consider different visualization methods to represent and organize information in the design of information retrieval systems.

Keywords: Searching, information visualization, user performance.

1 Introduction

Visualization is an important information retrieval (IR) technique [2] [18] that has been drawing greater attention in the field of IR, library and information science, and human-computer interaction [8]. Among other subjects, researchers have investigated (1) whether and how visualization features can improve the effectiveness and usability of information systems [18], (2) the impact of users' cognitive styles [19] or domain knowledge [20] on user performance of information visualization (InfoVis) systems, and (3) the usability of and user interaction behavior on different InfoVis systems, including VIBE [9], MetaCrystal [14], CiteSpace [1][15], and CiteWiz [7].

Different types of InfoVis systems have been devised to serve specific user groups. In this paper, we were particularly interested in how differently users perform when encountering InfoVis systems as opposed to generic text IR systems. Our hypothesis is that a system displaying graphical network visualization results leads to better user performance and user perception in completing search tasks than one displaying textual retrieval results display. To test this claim, we conducted an experiment exploring user information behaviors when interacting with either an InfoVis system or a textual IR system. For the InfoVis system, we chose a knowledge domain visualization system called "CiteSpace" (CS) [4], which features a graphic result output interface. For the contrasting system, we selected the Web of Science (WoS) , which displays a traditional ranked list of retrieval results.

A. Marcus (Ed.): DUXU/HCII 2013, Part IV, LNCS 8015, pp. 591–600, 2013.

In this paper, we report experimental results addressing the differences in user performance between the CS and the WoS systems. In the following sections, the related work in the field, the CS system, the WoS system, research methodology, data analysis, discussion, and conclusions are presented in sequence.

2 Related Work

InfoVis systems assign meanings to symbols, shapes, icons, nodes, and visual metaphors to help users intuitively understand complicated dataset.

Studies have researched how visual interface/system design or result presentation affects InfoVis system performance [16],[17]. Veerasamy and Belkin [16] compared visual interface/system design to other visualization/non-visualization interface(s)/system(s). Although they found a slight preference toward some visualization tools, the differences were not statistically significant. Woodruff, Faulring, Rosenholtz, Morrison, and Pirolli [17] tested the effect of thumbnails on task performance. Subjects used three different types of summaries (enhanced thumbnails, plain thumbnails, and text summaries) to search Web pages in order to find different types of information. The results were mixed and a strong effect of question category was found. For some questions, text outperformed plain thumbnails; for other questions, the result is opposite. Enhanced thumbnails (a combination of text summaries and plain thumbnails) were more consistent than either text summaries or plain thumbnails. Reiterer, Tullius, and Mann [13] found that most participants in their study did not like particular kinds of visualization systems such as bar chart and scatter plot formats. Dumais, Cutrell, and Chen [6] compared seven interfaces (including list and category interfaces) for integrating semantic category information with Web search results. User studies indicated that all the category interfaces were more effective than list interfaces, even when lists were augmented with category names for each result. The best category interface performance was achieved when both category names and individual page titles were presented. Likewise, Osdin, Ounis, and White [11] compared a visualized clustering approach in a Web context called HuddleSearch to a baseline search engine called PanOptic that used a conventional text list. In the study, 16 users completed eight tasks. The results on task completion time and user satisfaction were positive for the experimental, visualization system. The differences between the two systems were statistically significant on task type and task completion time.

3 Systems

3.1 The CiteSpace System

The CiteSpace system is a well-known, actively maintained, stable, and widely used knowledge domain visualization (KDViz) system. Also, it can be run on multiple computer platforms, thus making it convenient for researchers to evaluate. It was originally created to identify intellectual turning points [4] by constructing co-citation networks among highly cited articles and enables users to manipulate the resulting

graphical network. The visualization graph of the CS system is composed of nodes connected by lines. Nine types of nodes exist in the CS system (version 2.2.R1): authors, term, keyword, category, institution, cited reference, cited journal, cited author, and country. Correspondingly, nine visualization graphs represent the patterns of scientific literatures. Fig. 1 displays the resulting visualization graphs which correspond to the node "country." Other graphic result displays, based on different nodes, appear similarly to Fig. 1.

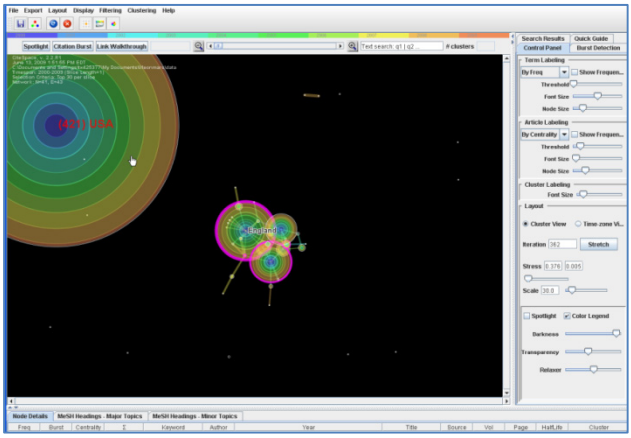

Fig. 1. Resulting Citespace graph showing node type for "country"

3.2 The Web of Science System

The Web of Science system (WoS) provides researchers with quick and powerful access to the world's leading citation databases (see Fig. 2). It displays the retrieved

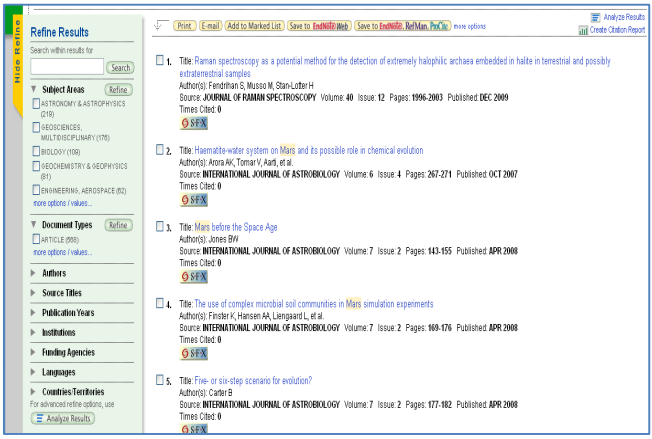

Fig. 2. The search results screen of the Web of Science system

results in a ranked list with information such as title, authors, source, publication year, number of citations. Through using the WoS system, users can find high-impact articles and conference proceedings, explore relevant information in related fields, recognize emerging trends of literature and research, and discover potential collaborators with significant citation records.

4 Tasks

The subjects were given scenario-based tasks with the topics related to "life on Mars." The tasks followed the model of simulated work task situations proposed by Borlund [3]. The tasks were categorized into two groups: aspectual and analytical search. Aspectual tasks required the user to identify as many different aspects as possible for a given topic and collect appropriate resources that cover all distinct aspects of that topic [5]. The following is an example aspectual task:

> *Scenario: As a graduate student, you want to write a paper about research on life on Mars. You are interested in how research has been done and what research has played an important role in this area during the past several years.*
> *Task: You want to identify all the countries which have many publications (>20) and also have collaborated with each other. Please put your answer on the answer sheet.*

Analytical search tasks demand more goal-oriented and systematic analytical strategies [10]. The following is an example analytical search task:

> *Scenario: As a graduate student, you want to write a paper about research on life on Mars. You are interested in how research has been done and what research has played an important role in this area during the past several years.*
> *Task: You need to collect some papers for the literature review. You know that some papers published by Edwards HGM would be very helpful. Please find the author who has the most collaboration with Edwards HGM, then put your answer on the answer sheet.*

5 Experimental Design

Data were collected in a between-subjects user experiment. Half of the subjects searched using the CS system; the other half engaged the WoS system. Each subject first completed the cognitive Extended CSA-WA test [12][1]. Then, they were asked to perform eight total tasks—four aspectual and four analytical search. Tasks were randomly assigned using a Latin-Square design, which ensured that no subject was given the tasks in the same order.

[1] This test and the partial related experimental results were reported in [19].

5.1 Subjects

Thirty-two graduate students from UAlbany, participated in the experiment[2]. They were enlisted by posting recruitment notices to departmental listservs and through announcements in classes at the university. In general, the subjects tended toward their 20s and 30s. They were fairly evenly split by gender and education level.

5.2 Dataset

The dataset was constructed by searching the topic "life on Mars," language "English," document type "Articles," and published between the years of "2000-2009" in the ISI Web of Science. In total, 857 records were retrieved from the Web of Science system. All documents were saved in a database specific to this experiment.

5.3 Measures

User performance was measured in terms of task completion time, result satisfaction, total number of mouse clicks, result correctness, and aspectual recall. These measures have been accepted and tested in prior studies (c.f. [18]). Time was measured starting with the user opening the visualization window and ending when the user finished typing the answers on the answer sheet. Result satisfaction was measured by asking each subject in a post-task questionnaire to rate his or her own satisfaction with the search results. Ratings were handled on a seven-point Likert Scale ranging from "Not at all" to "Extremely" (satisfied). The total number of mouse clicks metric encompasses a subject's actions while performing a task. It was measured by counting the total number of mouse clicks recorded in the logging software. Result correctness was measured as the external assessor's judgment of the subject's saved answer(s). Answers were judged on a two-point scale: Incorrect (0), and Correct (1). Aspectual recall was determined by pooling all aspects identified for each topic by all subjects. Aspectual recall, a measure developed in the TREC Interactive Track [5], is the ratio of aspects identified by the subject to the total number of aspects of the topic found by all subjects.

5.4 Conduct

The subjects read and signed a consent form, then filled out a questionnaire about their background, computer experience, and previous searching experience. Next, they were given a cognitive test (the E-CSA-WA test) to complete. Subjects were given brief instructions about the test and were asked to enter certain demographic data (name, age, etc.). Once the information was entered into the system, the test proper began. Each subject completed a tutorial on the system (CS or WoS). The subjects then completed two training tasks of each type, and four regular tasks of each type. Before each task, the subjects filled out a pre-task questionnaire. Then, they

[2] This study has been approved by the IRB at the University.

were given up to 10 minutes to complete that task. The interaction between the subjects and the system during the task was logged by the computer using TechSmith's Morae 2.1 logging software[3]). After completing each task, the subject filled out a post-task questionnaire. After the subjects finished all tasks, they were asked to complete an exit questionnaire. Finally, each subject was compensated $25 for their completion of the experiment. The experiment was conducted in a human-computer interaction lab at UAlbany. Each subject was tested individually.

6 Results

6.1 User Performance

ANOVA results (see Table 1) indicated that subjects using the CS system spent significantly less time than subjects using the WoS system, $F(1, 254) = 73.538$, $p = .00$. With respect to performance time, subjects using the WoS system spent more time than those using the CS system for a majority of the tasks. Qualitative results indicated that this performance variance could be attributed to "the graphical features of the (CS) system", which allowed visual grouping of literature while providing different choices for node type. Some subjects of the WoS system mentioned that it would be helpful if they could have inputted all the required information types in one search. Pearson Chi-square test showed that no significant relationship exists between system and result correctness, although subjects using the CS system found somewhat more correct answers than subjects using the WoS system, $\chi 2 = 2.723$, $df = 1$, $p = .099$. Wilcoxon signed-rank test results showed that subjects felt significantly more satisfied with their results using the CS system than when using the WoS system, $z = -3.731$, $p = .00$. ANOVA results showed that subjects using the CS system identified a bit more relevant aspects than those using the WoS system, but not significantly so $F(1, 126) = 0.658$, $p = .419$. ANOVA results indicated that subjects using the CS system had significantly fewer mouse clicks than those using the WoS system, $F(1, 254) = 17.448$, $p = .00$. Table 1 collects the performance results.

Table 1. Performance Measure (** significant at < .01 level)

Performance Measures	System: Mean (SD)	
	CS:	WoS:
Time (mins)	3.70 (2.08)**	7.05 (3.89)
Result correctness (0–1) (analytical search task)	0.70 (0.46)	0.56 (0.50)
Result satisfaction (1–7)	5.16 (1.88)**	4.30 (2.10)
Aspectual recall (aspectual task)	0.56 (0.37)	0.50 (0.41)
Number of mouse clicks	38.87 (24.71)**	58.64 (47.52)

[3] http://www.techsmith.com/morae.asp

Judgment of result correctness for topics in the analytical task type was accomplished by constructing the topics so that specific answers existed. Less variation occurred in the number of correct answers between different tasks for CS as compared to WoS. In particular, for task 3-"List two subject areas/categories that only authors from the USA are involved," none of the subjects found correct answers using WoS. On the other hand, they performed well for task 1-"Find the name of the university that has collaborated with Caltech in 2009 and published papers," and task 2-" Find the author who has the most collaboration with Edwards HGM." After further analyzing the qualitative data, we found out that subjects were somewhat overwhelmed filtering information by looking through the ranked list in the WoS system without clear indication about the relationship between authors, subject areas and country. To some extent, this difficulty explains why subjects using the WoS system did not find correct answers in task 3. Fig. 3 shows the distribution of correctness values for each task.

For most of the tasks, subjects using the CS system identified more aspects than those of the WoS system. The sole exception was task A—"Find all the institutions which collaborated on the topic in 2008." For the institution-related task, WoS system features are more supportive than those in the CS system. Fig. 4 summarizes the aspectual recall results.

6.2 Task Type Effect

By looking at the time, result satisfaction, and number of mouse clicks by task type, results indicated that the subjects spent less time on the analytical search tasks

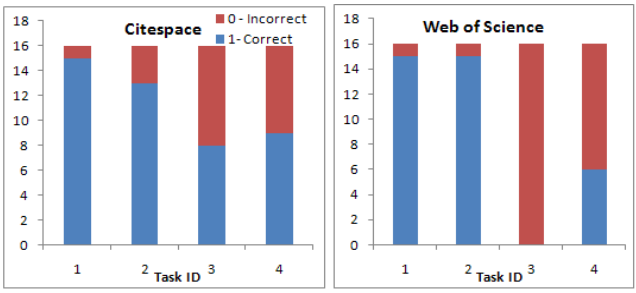

Fig. 3. Distribution of result correctness for each task

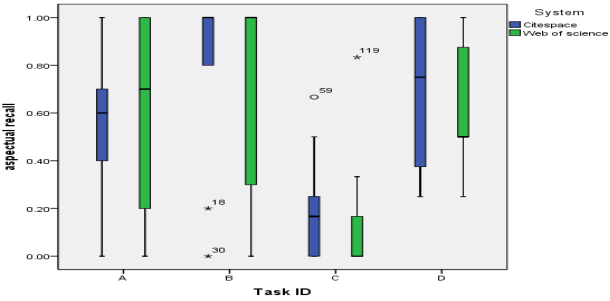

Fig. 4. Distribution of Aspectual Recall across systems

(M=5.03,SD=3.97) than the aspectual tasks (M=5.73, SD=3.02), but not significantly so, F = 2.538, p = .112. The result satisfaction was significantly higher for the analytical search tasks (M=5.07, SD=2.02) than the aspectual tasks (M=4.39, SD=2.00), with Z = –2.904, p = .004. The number of mouse clicks was roughly equal between the analytical search tasks (M=47.78, SD=46.30) and the aspectual tasks (M=49.73, SD= 30.33), F = 0.158, p = .691.

7 Discussion

The results indicate that CS system subjects performed better than did subjects using the WoS system.

Specifically, subjects using the CS system spent significantly less time in completing the tasks than did WoS system users. These results are similar to those reported by [11], which found that hierarchical clustering and summarization visualization techniques significantly helped users more quickly find the relevant documents than when using the baseline textual system.

CS system users employed significantly fewer mouse clicks than did those using the WoS system. This finding indicates that graphical visualization can reduce the interaction between the user and the system, in turn improving users' search performance. It further confirms the findings from [18] that a system relating different visualization techniques to tasks can improve user performance by reducing the interaction between the user and the system.

Subjects using the CS system felt significantly more satisfied with the results than did WoS system users. [11] reported similar, although not significant, results in user satisfaction in favor of a InfoVis system.

No significant difference was found in aspectual recall between these two systems. This finding is consistent with the experimental results comparing two textual IR systems [18]. Given the suggestion by Veerasamy and Belkin [16] that certain topics may be better suited to visualization, further exploration might clarify whether aspectual task is suitable for visualization. We infer that, because aspectual tasks require a certain amount of time for subjects to complete, the type of system (visualization vs. text-based) is less relevant to measures of performance time.

8 Conclusions

In this paper, a between-subjects experiment compared user performance and usability between an InfoVis system and a textual-based IR system. The results demonstrated substantial and significant advantages to the InfoVis system named CiteSpace in terms of user performance.

The results have implications for IR system design. For example, it may be helpful to organize the domain knowledge information or other general information in networking visualization graphs, which can show the pattern of such information and provide context. Also, it may be useful for users to identify patterns if a clear relationship between different types of information can be presented in networking graphs.

Limitations present in our experiment and issues concerning our conclusions indicate further investigation is warranted. We were constrained by a limited, and to some extent rather homogeneous, number of subjects, a limited number and type of search topics, and a limited database. In addition, the tested InfoVis system responded to only one type of visualization technique. It is challenging to evaluate InfoVis systems because the visualization display for each system is quite different (e.g., the system explored in [9] is different from that reviewed by [14]). These distinctions also make it hard to generalize the results from one study to another. Despite the limitations and unanswered questions associated with this study, we believe that our research takes an important step toward the evaluation of InfoVis systems. It contributes an insightful comparison of user behavior between an InfoVis system and a textual system on different types of tasks. In the future, we plan to perform more user studies to identify appropriate visualization techniques for different tasks and information-search strategies, and to relate different user groups to such techniques.

References

1. Allendoerfer, K., Aluker, S., Panjwani, G., Proctor, J., Sturtz, D., Vukovic, M., Chen, C.: Adapting the Cognitive Walkthrough Method to Assess the Usability of a Knowledge Domain Visualization. In: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, vol. 26 (2005), doi:10.1109/INFOVIS.2005.3
2. Belkin, N.J.: Intelligent information retrieval: Whose intelligence? In: Proceedings of the Fifth International Symposium for Information Science (ISI-1996), pp. 25–31 (1996)
3. Borlund, P.: Experimental components for evaluation of interactive information retrieval systems. Journal of Documentation 56(1), 71–90 (2000)
4. Chen, C.: Information Visualization: Beyond the Horizon, 2nd edn. Springer (2004)
5. Dumais, S., Belkin, N.J.: The TREC interactive tracks: Putting the user into search. In: Voorhees, E.M., Harman, D.K. (eds.) TREC: Experiment and Evaluation in Information Retrieval, pp. 123–152. MIT Press, Cambridge (2005)
6. Dumais, S., Cutrell, E., Chen, H.: Optimizing search by showing results in context. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 277–284 (2001), doi:10.1145/365024.365116
7. Elmqvist, N., Tisgas, P.: CiteWiz: a tool for the visualization of scientific citation networks. Information Visualization, 215–232 (2007)
8. Julien, C.-A., Leide, J.E., Bouthillier, F.: Controlled User Evaluations of Information Visualization Interfaces for Text Retrieval: Literature Review and Meta-Analysis. School of Information Studies 59(6), 1012–1024 (2008)
9. Koshman, S.: Comparing usability between a visualization and text-based system for information retrieval. Journal of Documentation 60(5), 565–580 (2004)
10. Marchionini, G.: Cambridge series on human. In: Computer Interaction: Information Seeking in Electronic Environments. Cambridge University Press, Cambridge (1995)
11. Osdin, R., Ounis, I., White, R.W.: Using hierarchical clustering and summarisation approaches for Web retrieval. In: Proceedings of the 11th Text REtrieval Conference (TREC). National Institute of Standards and Technology (NIST) (2002)
12. Peterson, E.R., Deary, I.J., Austin, E.J.: The reliability of Riding's cognitive styles analysis test. Personality and Differences 34, 881–891 (2003)

13. Reiterer, H., Tullius, G., Mann, T.M.: INSYDER: a content-based visual-information-seeking system for the Web. International Journal of Digital Libraries 5, 25–41 (2005)
14. Spoerri, A.: Visualizing Meta Search Results: Evaluating the MetaCrystal toolset. In: Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology (ASIST 2006), Austin, TX (2006)
15. Synnestvedt, M., Chen, C.: Design and evaluation of the tightly coupled perceptual-cognitive tasks in knowledge domain visualization. In: Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, Nevada (2005)
16. Veerasamy, A., Belkin, N.J.: Evaluation of a tool for visualization of information retrieval results. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1996, pp. 85–92 (1996), doi:10.1145/243199.243218
17. Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J., Pirolli, P.: Using thumbnails to search the Web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 198–205 (2001), doi:10.1145/365024.365098
18. Yuan, X.-J., Belkin, N.J.: Investigating information retrieval support techniques for different information-seeking strategies. Journal of the American Society for Information Science and Technology 61(8), 1543–1563 (2010)
19. Yuan, X.-J., Zhang, X.-J., Chen, C., Avery, J.: Seeking information with an information visualization system: a study of cognitive styles. Information Research 16(4), paper 499 (2011), http://InformationR.net/ir/16-4/paper499.html
20. Yuan, X.-J., Chen, C., Zhang, X., Avery, J.: Effects of Domain Knowledge on User Task Performance in a Knowledge Domain Visualization System. Accepted to HCIInternational 2013 (2012)