

# Improving Web Search for Information Gathering: Visualization in Effect

Anwar Alhenshiri  
Dalhousie University  
6050 University Avenue  
Halifax, NS, Canada  
+1 (902) 494-2093  
anwar@cs.dal.ca

Michael Shepherd  
Dalhousie University  
6050 University Avenue  
Halifax, NS, Canada  
+1 (902) 494-1199  
shepherd@cs.dal.ca

Carolyn Watters  
Dalhousie University  
6050 University Avenue  
Halifax, NS, Canada  
+1 (902) 494-6723  
watters@cs.dal.ca

## ABSTRACT

The nature of the Web implies heterogeneity, large volumes, and varied structures. Hence, finding results that best suit the needs of every individual in every type of Web task is a very challenging problem. This research presents an interactive Visual Search Engine (VSE) in which both query reformulation and results presentation are visualized. The paper presents the results of a user study in which the effectiveness of the VSE compare to Google is evaluated. The VSE was shown to be effective with respect to Web information gathering tasks.

## Categories and Subject Descriptors

H.3.3. [Information Search and Retrieval]: Search process, query formulation, clustering

## General Terms

Measurement, Performance, Design, Experimentation, Human Factors

## Keywords

Web information retrieval, relevancy, information gathering, search tasks, query reformulation, visual search, visual rendering.

## 1. INTRODUCTION

The process of results presentation in most conventional search engines presents few links per display with textual content attached to each link. The user may have to perform scrolling over multiple pages to find relevant results [11]. However, it has been shown that the majority of Web search users do not look beyond the first three results [5, 11]. At which point, users either modify the search query or switch to a different search tool [1].

Information visualization is suggested to improve users' performance by harnessing their innate abilities for perceiving, identifying, exploring, and understanding large volumes of data [3, 4]. Consequently, visualization has become of interest to researchers in Web information retrieval due to the large numbers of documents the Web contains in addition to the often overwhelming resultant

matches of Web search results. Visualization may permit the display of more results with connectivity features. Integrating visualization in Web search aims to combine computation and high bandwidth human perception [13, 14].

In previous works [5, 6], several visualization aspects were investigated in Web information search and retrieval. Some of the visualization research either achieved certain levels of success while being evaluated using search queries or usability case studies, or suffered from issues of delay and scalability concerns due to the user of sophisticated 3D visualizations. Evaluating visualization and clustering-based search interfaces may reveal different finding in the case of using the context of a complete task. The presented VSE aims to utilize the user's visual abilities to improve query reconstruction and search results exploration. The improvement is intended for the case of information gathering tasks in which uses locate, compare, and further locate Web documents for satisfying criteria described in the task [7].

## 2. DESIGN AND IMPLEMENTATION

The VSE uses Google as its underlying search service provider. Hence, the VSE is a visualized layer built on top of Google as a search interface with which the user interacts to submit and reformulate queries and also to explore search results. The VSE presents results as glyphs on an interactive interface. Each glyph contains the document title and a snapshot of the Web page as recommended in the work of Teevan et al. [12]. Edges between visualized glyphs represent content similarity between connected documents. The interactive interface of the VSE permits the user to see document statistics such as document size, Google's PageRank value, and documents' last updates. Along with the user original query, the VSE uses alternate queries provided by the semantic network WordNet [8, 9] for single term user queries and by randomly reordering query terms for multiple-term queries.

The VSE permits its users to reconstruct subsequent queries from a query reconstruction area on the display. The query

reconstruction area contains terms and phrases inferred from the top documents retrieved for the current query. The aim of this design is to assist users with perceiving more relevant results and related search queries. Figure 1 shows the interface of the VSE. A user study was conducted to evaluate the effectiveness of the VSE as reported in the following Section.

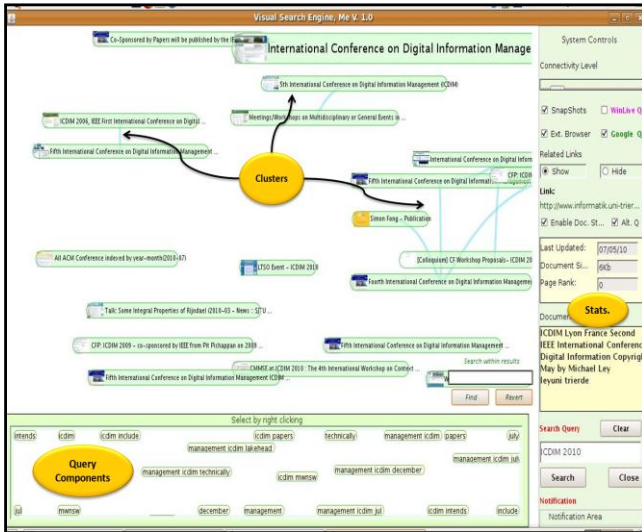


Figure 1. The VSE interface

### 3. EVALUATION

Fourteen participants took part in the study to evaluate the VSE by comparing it to Google. The type of Web tasks intended to be evaluated was information gathering since it represents approximately 61.5% of the overall Web tasks [10]. Information gathering tasks involve collecting information possibly of different types from different sources to achieve an overall goal identified in the task [7]. Hence, this type of task was considered for evaluation in this project. The other two types of search tasks (navigational and transactional according to Border [2]) imply seeking more specific results. Consequently, they were not considered in the study. An example of the information gathering tasks that were used in the study is the following.

*“Use the given search engine to gather webpages that include information about how to use the java programming language in transforming html documents into images. The pages you find should give someone a good idea about the task’s topic. You can submit up to five queries only, and you should not go beyond viewing one page of results for each query you submit. You can still view results of webpages in the Web browser”.*

In the evaluation study, users were asked to perform two information gathering tasks both on Google and the VSE. The study design was complete factorial within-subjects

and counterbalanced. After finishing each task on either search tool, the study asked participants to complete a post-task questionnaire in which they stated their confidence in the results in addition to other self-reported engagement measures. Study data included machine logged data in addition to data accumulated in the questionnaires. The study took under 30 minutes and was preceded by a short training session on the VSE. Participants were computer science students from Dalhousie University.

### 4. RESULTS AND DISCUSSION

Although the study was intended to evaluate only the effectiveness of the VSE compared to Google, the VSE was found to also be more efficient in performing searches for information gathering with an average time on task of 6.5 minutes compared to Google with 8.2 minutes. However, the difference was not significant in this case.

Regarding effectiveness, the VSE—compared to Google—required submitting fewer queries, opening fewer pages on the Web browser, and permitted its users to discover more relevant pages with closer numbers of types of information to the types required in the tasks. Google, on the other hand, required the participants to submit more queries, open more pages on the browser to locate fewer relevant results for the tasks. By comparing the study results, the t-test revealed significant differences between the VSE and Google with  $F = 45$ , and  $\alpha < 0.003$  with respect to above criteria. The quantitative results are shown in Table 1. For further illustration, a comparison of the VSE and Google with respect to the number of pages participants had to open on the Web browser to accomplish their tasks is shown in Figure 2. Figure 3 shows the difference between the two search tools regarding the number of queries used for achieving the tasks.

Table 1. Quantitative results

Where, ( $\mu$ ) is the Mean, and ( $\sigma$ ) is the Standard Deviation.

System	VSE	Google	
<i>Time (Mean)</i>			
	6.5	8.2	
<b>Submitted Queries</b>	$\mu$	2.5	3.5
	$\sigma$	1.5	1.6
<b>Pages opened on the browser</b>	$\mu$	2.1	9
	$\sigma$	1.5	11
<b>Relevant pages found</b>	$\mu$	6.5	4.5
	$\sigma$	2.7	2.1
<b>Covered topics</b>	$\mu$	3	2
	$\sigma$	1.4	1.3

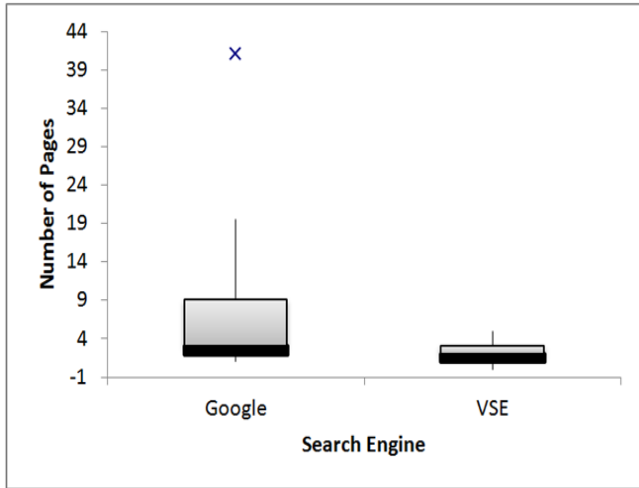


Figure 2. Pages browsed by participants

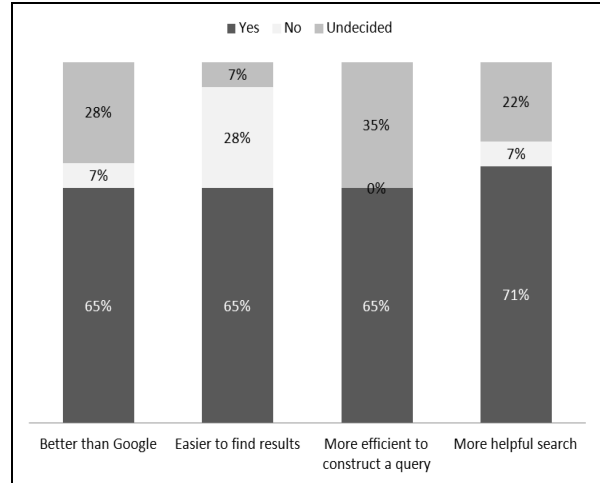


Figure 4. Post-study questionnaire ratings of the VSE compared to Google

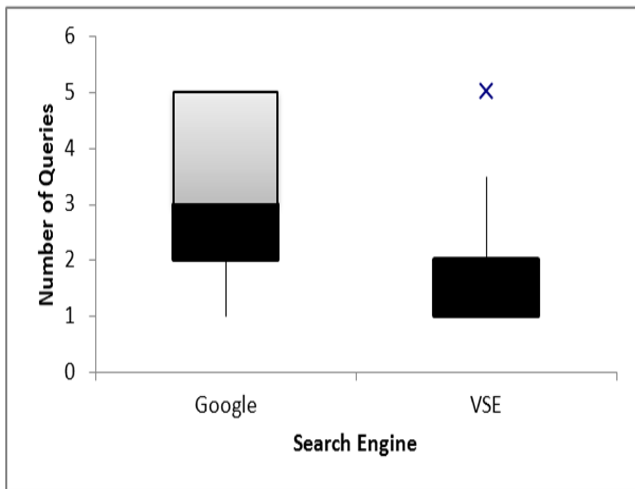


Figure 3. Queries submitted by participants

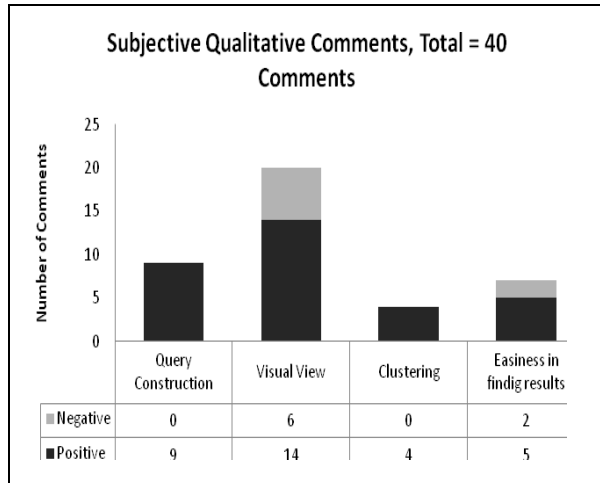


Figure 5. Subjective self-reported comments

Analyzing the results of the questionnaires showed that the VSE was considered better for information gathering than Google by the participants in the study. In addition, the VSE was regarded as more effective in reconstructing queries and more helpful in finding relevant documents. Figure 4 shows the results of the analysis of the post-study questionnaires. However, the one-tail z-test shows no significance difference between the two proportions of participants ( $z = 1.42, \alpha = 0.08$ ). In addition, the user qualitative comments are shown in Figure 5. Generally, 80% of the comments about the VSE were positive. According to the one-tail z-test, there was a significance difference between the two proportions of comments ( $z = 2.79, \alpha < 0.004$ ). Furthermore, the user confidence with the located results for the search tasks with both the VSE and Google is shown in Figure 6.

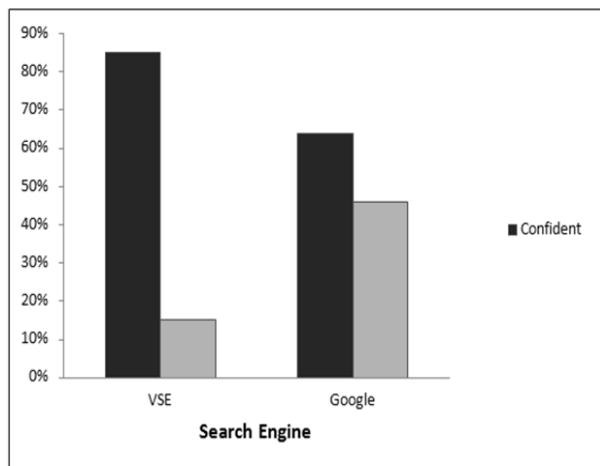


Figure 6. User confidence in the located results

Minor concerns were raised about the slow movement of the glyphs on the display. However, the study showed that the current Web search model suffers from ineffectiveness with regard to how search queries are reformulated and how search results are presented to the user. In the case of Web information gathering tasks, the limited document attributes shown by the search engine, the current approach for submitting and reformulating search queries, and the way search results are presented for comparing Web information and making decisions regarding the task requirements need further investigations.

## 5. FUTURE WORK

For information gathering tasks, users usually need to explore more results per session and to effectively perceive more features of the presented documents to be able to make effective decisions regarding the information gathered for the task requirements. In further research, different layouts will be evaluated for search results presentation. In addition, different clustering criteria will be investigated in information gathering with the use of visualization. The concepts of re-finding and Web information organization for information gathering will also be investigated.

## 6. CONCLUSION

The VSE demonstrated that exclusive textual presentations of Web search results would benefit from our visualizations. The VSE may help Web search users with finding relevant documents in the case of information gathering tasks. Future work will focus on this type of task by emphasizing its underlying subtasks for investigation.

## 7. REFERENCES

- [1] Alhenshiri, A. and Solis-Oba, R. 2007. Improving Results for Short Web Queries Using Preserved Query Knowledge. In proceedings of the 1st international Conference on Digital Communications and Computer Applications, Amman, Jordan, 634-649.
- [2] Border, A. 2002. A Taxonomy of Web Search. ACM SIGIR Forum. Vol. 36, Issue 2, 2-10.
- [3] Card, S., K., Mackinlay J. D., and Shneiderman B., Readings in Information Visualization. 1999. Using Vision to Think. Morgan Kaufman Publishers, San Francisco, CA, USA.
- [4] Friendly, M. 2008. Milestones in the History of Thematic Cartography, Statistical, Graphics, and Data Visualization. York University Archives, Canada. <http://www.math.yorku.ca/SCS/Gallery/milestone/>
- [5] Hoeber, O., and Yang, X. D. 2007. Visual Support for Exploration within Web Search Results Lists. In Conference Compendium of the IEEE Information Visualization Conference (October 28- November 01, Sacramento, California, USA), 157-167.
- [6] Jacso P.: SAVVY SEARCHING.2007. Clustering Search Results. Part I: Web-wide Search Engines, Online Information Review, 31(1), 85-91.
- [7] Kellar, M., Watters, K., and Shepherd, M. 2007. A Field Study Characterizing Web-based Information-Seeking Tasks. Journal of the American Society for Information Science and Technology, 58(7), 999-1018.
- [8] Miller G., A. 1990. WordNet: An On-line Lexical Database. International Journal of Lexicography, 3(4), 285-303.
- [9] Moldovan, D., and Mihalcea, R.2000. Using WordNet and Lexical Operators to Improve Internet Searches. IEEE Internet Computing, 4(1), 34-43.
- [10] Rose, D., E., and Levinson, D. 2004. Understanding User Goals in Web Search. In Proceedings of the 13<sup>th</sup> International Conference on World Wide Web (May 19-21, New York, NY, USA), 133-142.
- [11] Spink, A., Wolfram, D., Jansen, M., and Saracevic, T. 2001. Searching the Web: The Public and Their Queries. Journal of the American Society for Information Science and Technology, 52(3), 226-234.
- [12] Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., André, P. and Hu, C. 2009. Visual Snippets: Summarizing Web Pages for Search and Revisitation. In Proceedings of the 27<sup>th</sup> International Conference on Human Factors in Computing Systems (April 04-09, Boston, MA, USA).
- [13] Vincente, K., J., Rasmussen, J.1990. The Ecology of Human Machine Systems: Ii. Mediating "Direct Perception" in Complex Domain, Journal of Ecological Psychology, 2(3), 207-207.
- [14] Youssefi, A., Duke, D., Zaki, M. 2004. Visual Web Mining. In Proceedings of 13<sup>th</sup> International World Wide Web Conference on Alternate Track Papers and Posters (May 19-21, New York, NY, USA), 394-395.