Visualizing Query Occurrence in Search Result Lists

Tomi Heimonen and Natalie Jhaveri
Tampere Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences, University of Tampere, Finland
{tomi.heimonen, natalie.jhaveri}@cs.uta.fi

Abstract

The result lists of popular web search engines represent retrieved documents with only a title, a brief textual summary and a URL. Wepresent a novel approach that incorporates visualization into the conventional search result interface. For each resulting document, occurrences of the entire query are concisely depicted in the form of a small, document-shaped icon. An 18-participant user study was carried out to compare our design with the traditional search result list in terms of accuracy and task performance. Though statistically significant differences in performance were not observed, the participants' subjective ratings and opinions of the visualization's utility were positive. Despite the fact that the visualization introduces a new and somewhat complicated variable to consider when evaluating search result lists, the participants' performance did not fall below their performance level with the traditional interface. Our findings indicate the need to study such interfaces in a longitudinal setting.

1. Introduction

The conventional ordered search result list has prevailed as the principal search result presentation approach of most major World Wide Web (web) search engines. After years of practice, web users have become highly accustomed to selecting a web document based on its position in the search result list, along with its title, text summary, and Uniform Resource Locator (URL).

However, as users are not informed about the characteristics by which the results are selected and sorted, they must trust the search engine to be both optimal and impartial in its selection of results. In some cases, it is easy for the user to see that the search engine has produced highly relevant results for the given query. For example, when the user submits the query "bbc news" and the BBC News website is listed first among the search results, the user can be assured that his or her goal was fulfilled. However, with a more open-ended task, such as finding information about the features of various new digital cameras, it is more difficult to judge the relevance of the results. In these cases users must rely on different strategies for deciding which pages to view.

Information visualization techniques are widely used to provide insight into data [13, pp. **1-13].** Applying information visualization to information retrieval interfaces is referred to as document visualization [4, pp. 409-410]. Document visualization involves a trade-off between the simplicity of the visualization and the number of document features it addresses. Most designs tend to lean towards simplicity by selecting to convey only a few attributes of the data.

A logical and common choice in document visualization is to depict the distribution of the query terms; the terms used to express the information need. TileBars by Hearst *[6]* is a distinguished example of this approach. It represents a retrieved document using a horizontal bar to indicate its total length. The frequency and location of the query terms is shown by colour density inside the bar. **A** similar approach to visualizing the location of query terms within the actual result document was suggested by Byrd [3]. In Byrd's visualization, the scrollbar of the window displaying the result document is augmented with visual depictions of query term locations.

Another document visualization system that employs query term distribution was presented by Veerasamy and Belkin [14]. They visualize the retrieved documents as columns and query terms as rows. At the intersection, a solid bar varies in height to indicate the frequency of the query term within the document. Ogden ***et al.*** [9] introduced document thumbnail visualization, an approach that represents the retrieved documents with thumbnail images of the documents, which include highlighted, colour-coded query terms.

In addition to visualizing query term location, recent research suggests the combined use of textual and visual descriptors. Such a solution was proposed by Woodruff ***et al.*** [15] with the representation of search results as thumbnails and text summaries. Other more graphically-oriented systems have also been presented, such as xFIND by Andrews ***et al.*** [1] and Sparkler by Havre ***et al.*** [5]. Both systems provide the user with interactive tools for

1550-6037/05 $20.00 © 2005 IEEE

exploring the search result set. **A** comparison of textual and graphical search result interfaces can be found in [12].

Search result visualizations based on the focus+context and the overview+detail visualization strategies have also been proposed. Paek *et al.* [10] use a focus+context approach to provide additional textual context to a text summary when the user interacts with the result list. As the user clicks the text summary, she is presented with additional details of the retrieved document. An overview+detail approach, called visual bracketing, was presented by Roberts and Suvanaphen [11]. It combines several coordinated views to give both an overview of the search result list as well as more detailed information about specific results. The different views show the search results as a list of URL addresses, the search results in greeked text, and full search results in the traditional format.

Based on the existing body of work on visualizing search results, it can be concluded that a feasible visualization should include both textual and graphical elements, as well as provide details-on-demand facilities. However, as users begin to rely more heavily on mobile devices and services, screen space, bandwidth and interaction -related limitations must also be accounted for when designing search result visualizations. We believe it is now critical to evaluate more lightweight visualization solutions that can be delivered over the Web to a variety of devices. Such interfaces should be compact, informative and provide enough information for relevance assessment without requiring extensive interaction with the visualization.

2. Query Occurrence Visualization

As our research targets web search engines, drastic interface changes face problems of adoption. Rather than replacing the conventional interfaces that are used by millions of people daily, we propose a visualization that can be smoothly integrated into the search result list without distracting the users. It aims to provide insight beyond the title, text summary and URL format by indicating the distribution and frequency of the entire query in the content of the retrieved document. Instead of merely reporting the frequency and position of individual query terms, the indicator attempts to answer the question of "are my query terms appearing in the same context in the result documents?"

The visualization is depicted as a small document-shaped icon that is presented on the left side of the result items, as shown in Figure 1. Hereafter, this search result descriptor will be referred to as the indicator. The indicator illustrates the contents of the document in four equally sized sections (rows), irrelevant of the actual result document size. For each section, it indicates how often the individual query terms appear in close proximity of one another in the text of the result document. This is the case when all of the query terms occur within a fixed, 20-word range of one another. Here the algorithm to compute the occurrences is quite straightforward; the query terms can occur in any order as long as they all appear at least once within the specified "window".

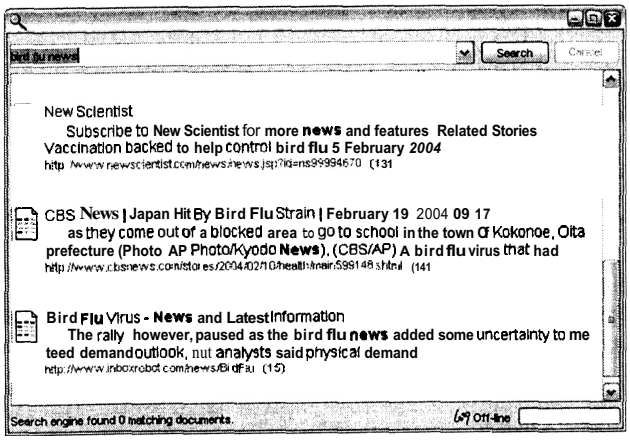

Figure 1. Search result list containing the visualization.

Figure 2 shows three possible indicators, from left to right: many occurrences of the query across the result document, no occurrences of the query in the result document, and query occurrence details unavailable for the document. Further, the first indicator reveals that the last section of the retrieved document only contains one occurrence of the query. In contrast, the first and the third sections contain three or more such occurrences, while the second section of the document contains no occurrences.

The overall effect of this visualization bears similarity to that of a population density map, which also depicts both frequency and distribution. Moreover, in the context of information visualization, the design falls into the category of suppression-based techniques [13, pp. 120-124].

Figure 2. The indicator with three different possible results.

Certainly, many variations of this kind of visualization are possible; however, these settings were selected for this initial evaluation of this approach with certain rationale. As the marks in the visualization are of the same colour and shape, the information can be processed efficiently by mechanisms of distributed attention, allowing the users to register the salient features automatically [8, p. 60]. Introducing more variables to the visualization would demand focussed attention from the user, whereby the user is forced to identify objects serially, one at a time [8,

p. 61]. Therefore, to avoid elevating the cognitive burden of the user, we chose not to represent each keyword of the query with separate colours or shape.

Clearly, this particular visualization does not offer a definitive solution for identifying relevant search results. The high relevance of a given document is not always indicated by a highly populated indicator. Nevertheless, we believe that the additional insight the visualization gives is of value in the relevance assessment process. This is especially useful in the case of exploratory and fact-finding search tasks [2]. In an exploratory task the user needs to obtain a general overview of the topic in order to narrow down the search. The indicator could be used either to dismiss results or to draw attention to results that contain a large number of query occurrences. In contrast, in a fact-finding task one occurrence is likely to be a sufficient indication of a prospective result document.

To assess the utility of the indicator we carried out a user study. The focus of the experiment was to find out how the proposed visualization affected user performance in information seeking tasks that involved relevance assessment, and to collect subjective observations from the participants regarding the visualization. We also wanted to examine how the visualization was used both with and without instructions on how to use it. Specifically, we address the following research questions:

1. How does the visualization-augmented interface perform when compared to the prevailing search result list interface?
2. How is the visualization subjectively received by the participants?

3. Method

3.1. Participants

Eighteen participants (10 male, 8 female; average age 27.4 years, *SD* 3.6) volunteered for the experiment. The participants were both researchers and students at the University of Tampere. All participants were experienced computer users (average 12.1 years, *SD* 5.2) and Web users' (average 7.1 years, *SD* 1.9). They also reported having extensive experience as search engine users. Of the 18 participants, 13 used search engines daily and 5 used them "many times a week". All participants used Google as their primary search engine.

3.2. Design

The experiment was organized as a within-subject design with one controlled variable, the search result list type. The participants were presented with three result list types:

- without visualization *(traditional)*
- with visualization – without instructions *(non-instructed visualization)*
- with visualization – with instructions *(instructed visualization)*

Since the non-instructed condition had to precede the instructed condition in order to retain internal validity of the test design, the order of presentation was only partially counterbalanced.

For each condition, the following dependent variables were measured:

- number of relevant, related and non-relevant results selected during each task
- task completion time
- time to first selection in each task

A total of 540 search tasks were completed, as each of the 18 participants completed 10 tasks with each of the three result list types. In each task, the result list contained 15 search result items. Each result item was categorized as relevant, related or non-relevant, based on the content of the actual result document. Relevant result items were those that provided the information required by the task at hand, related results were only partially relevant to the task, and non-relevant results were not relevant at all. To avoid introducing bias, the relevance judgement was carried out by a third party not affiliated with this study.

To retain statistical validity, certain steps were taken to eliminate variance that would be caused by differing search strategies between users (e.g. query formulation and iteration) and domain knowledge. Towards this end, we defined 10 task topics, such as international news, ecology, music, etc., and created three specific tasks for each topic. By ensuring that each condition included tasks of each of the 10 topics, we could eliminate the effect of domain expertise. For a similar reason, we provided predefined search queries (two to four query terms) for each task. This way, all participants were presented with a comparable quality level of results.

Finally, the search result lists and result documents were saved and loaded offline. In this way, we could eliminate the variance caused by varying loading times and the changing content of the search engine database.

3.3. Apparatus

The entire experiment, including the final subjective questionnaire, was conducted using a specialized search user interface testing application written in Java. The user interface of the application is shown in Figure 3.

A Pentium 4 desktop PC running Microsoft Windows XP Pro at 1280 × 1024 screen resolution and a 19-inch CRT monitor were utilized as the testing platform. The testing application logged the participants' interactions with the system for later analysis.

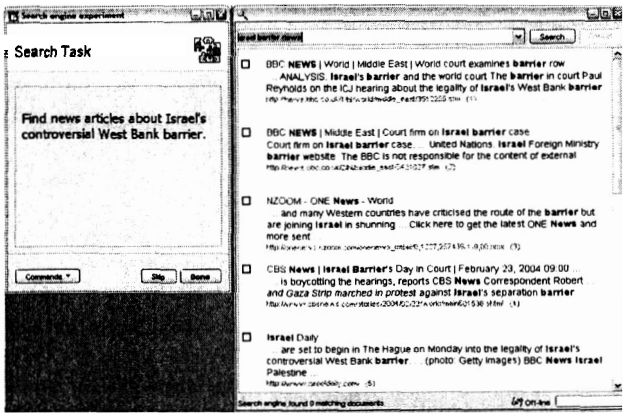

Figure 3. The testing application user interface.

The indicator was integrated into the testing application as an add-on component. The visualization component itself is comprised of three sub-components: a text extraction module, a data computation module and a visualization creation module.

The text extraction module retrieves and processes the result documents, removing all formatting and HTML tags to produce plain text files. For plain text extraction it employs parts of the HTMLParser toolkit [7].

The resulting text is then further processed by the data computation module, which searches the plain text file for instances of the query in the document text. First, the plain text file is split into for equally sized parts. Then, the sum of query occurrence instances is calculated for each individual part of the document.

Finally, the visualization creation module generates the visual representation of the query occurrence indicator and passes it to the testing application for on-screen rendering.

3.4. Procedure

In the introduction to the test, the participants were verbally instructed to "select as many relevant results as they can, as fast as possible". A relevant result was defined as one that "likely contains the information that you are looking for".

The participants were first given two practice tasks to familiarize themselves with the procedure and then the Participants continued to complete 3 sets of 10 tasks, one set for each result type interface. Each task was presented to the user textually as shown in the left-hand window in Figure 3. The task was started by clicking the "Start" button and finished either when the participant clicked the "Done" button or when 1 minute had passed, at which point the interface was disabled. During the task the participant selected relevant results by using the checkboxes to the left of the results.

Once one set of tasks was completed, the application indicated changes in the search result interface. Before beginning the 'instructed visualization' condition, the participant was given an explanation of the indicator and what its markings represent, using a clarifying illustration.

Upon completing the tasks, the participant was presented with the subjective questionnaire. Lastly, the experiment supervisor interviewed the participant about his or her usage approaches and impressions of the visualization.

4. Results

4.1. Performance Measures

To examine if there were any differences between the conditions, the performance measures (search result selections and performance times) were tested using repeated measures ANOVA. This analysis did not reveal any significant differences between the different conditions with respect to the measured variables.

A summary of the average number of search result selections by condition is shown in Table 1, while Table 2 lists the average performance times by condition.

	Relevant selections	Related selections	Non-relevant selections
Traditional	4.80 (SD 3.65)	0.96 (SD 1.27)	0.76 (SD 1.19)
Non-instructed	4,94 (SD 3.77)	1.03 (SD 1.36)	0.94 (SD 1.32)
Instructed	4,46 (SD 3.59)	0.94 (SD 1.28)	0.71 (SD 1.09)

	Task completion time (seconds)	Time to first selection (seconds)
Traditional	56.8 (***SD*** 6.5)	8.4 (SD 8.6)
Non-instructed	56.7 (SD 7.1)	8.7 (SD 8.4)
Instructed	56.4 (SD 6.9)	9.4 (SD 9.2)

The lack of difference in performance between the conditions can likely be attributed to the restrictive nature of the test design. As the participants only had 60 seconds in which to make the selections, it is possible they ran into a performance ceiling.

1.	I found the image-icon helpful in summarizing the contents of the listed web page.
2.	I found the image-icon helpful for spotting useless (i.e. garbage) results.
3.	I did not find the image-icon helpful for spotting useful results.
4.	I found the image-icon difficult to understand.
5.	I found the image-icon helpful in completing my tasks.
6.	I found the image-icon distracting when evaluating the results.

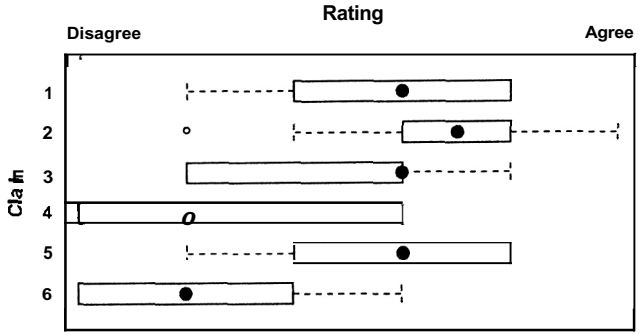

Figure 4. Subjective ratings by claim.

The subjective ratings indicate that the participants found the visualization somewhat useful for obtaining an overview of the contents of the result document (claim 1) and also somewhat helpful for completing the tasks (claim 5).

It also appears that the design of the visualization was successful in the respect that the participants did not find it too obtrusive (claim *6),* as also evidenced by the interviews. The participants also asserted that the meaning of the visualization was easy to interpret (claim 4).

However, the participants did not find the visualization particularly helpful in spotting useful results (claim **3).** On the other hand, this is offset by the fact that the participants found the visualization to be useful for spotting unsatisfactory results in the result list (claim **2).**

These observations seem to indicate that for the tasks used in this experiment, the participants did not find the visualization effective enough for discriminating between relevant and merely related results. This suggests that more relevance information should be incorporated to the visualization to make it a more comprehensive tool for assessing the results.

4.3. Observations

We made a number of interesting observations of the participants' actions during the experiment. In the non-instructed visualization condition, 11 of the 18 participants did not even notice the indicator. When asked at the end of the experiment whether they had used the indicator for that set of tasks, the participants replied that they had not noticed the visualization before it was specifically instructed. There is a certain amount of stress associated with experiments, especially when a time limit is enforced. This is perhaps one reason why over a half of the participants overlooked the visualization. Another possible reason is the fact that apart from the visualization, the user interface resembled that which the participants use on a frequent basis. Therefore it was easy for them to block out new aspects of the interface.

During the post-test interview we also discovered that the participants employed a variety of strategies while using the indicator. It was used to either immediately identify relevant results, immediately identify non-relevant results, or to provide additional insight into uncertain results.

However, few were able to truly develop trust in the indicator during such a short usage period, and without the possibility to view selected documents and judge also their relevance. At the end of the experiment, each participant was asked whether they chose to follow the indications of the visualization or the textual descriptors when there was a conflict in detail between the two. Only *5* out of 18 of the participants claimed to have trusted the indicator over the textual descriptors in such cases. Those participants who had trusted the indicator claimed to have used it to "cross-check" the textual descriptors or simply to filter out "poor" results. Comments from the post-test interviews, such as "I still need to learn to use it and gain trust", further revealed that trust in the indicator was not fully established by the end of the experiment session.

Despite this lack of trust, the participants did believe that more experience with using the visualization would

solve this issue. This is evident from the following comments: "With time, I might even formulate my queries to get better indication from [the indicator]" and "With time, e.g. after using it 20 times, I could learn to gain trust in it (or not) based on seeing the final pages." Undeniably, compared to using a more typical search engine result list, the inclusion of this indicator introduces some complexity to the relevance judgement process. However, the participants believed that with more practice, they could learn to benefit from the availability of such a resource.

These observations reveal the need for a subsequent longitudinal study. During such a study, the participants could define their own queries, view the resulting documents and accordingly incorporate the use of the indicator into their own search strategies. Such a study is likely to yield more interesting and informative results on the benefits and shortcomings of the query occurrence visualization approach.

5. Conclusions and Future Work

This paper presented a novel approach for visualizing search results and the user study that was carried out to assess its utility. First, we aimed to determine how the query occurrence visualization compares to the traditional search result list. The performance measures showed that there was no significant difference between the two interfaces in terms of accuracy or task completion time. This might be due to the overpowering familiarity of the traditional interface, as well as the time limit imposed by the test design. Second, we aimed to determine the participants' perceptions of the visualization. The subjective ratings indicated that the participants found the visualization to be unobtrusive, easy to understand and useful in spotting irrelevant results in the search result list.

The query occurrence visualization shows promise as a method for augmenting search result lists. It is encouraging to observe that even with minimal training the visualization can provide performance comparable to an interface that the users have had years of experience with. The users immediately understood the visualization and were also partially able to include it in their search strategies within a short period of time. With more time and free from any experimental constraints, we believe they could do this more effectively.

Next, we plan to conduct a longitudinal study in which the users will have more time to familiarize themselves with the indicator and more thoroughly include its use in their own relevance assessment approach. We will also develop the visualization further by adding functionality to show details-on-demand, e.g. actual excerpts from the result document that appear when the user points to the different sections of the indicator.

Acknowledgements

This work was supported by the National Technology Agency of Finland.

References

[1] Andrews, K., Gült, C., Moser, J., Sabol, V. and Lackner, W. Search Result Visualization with xFIND. In *Proceedings of UIDIS 2001*, IEEE Computer Society Press, 2001, pp. 50-58.

[2] Aula, A. and Kaki, M. Understanding Expert Search Strategies for Designing User-Friendly Search Interfaces. In *Isaias P. and Karmakar N. (Eds.) Proceedings of IADIS International Conference WWW/Internet 2003,* Volume II, 2003, pp. 759-762.

[3] Byrd, D. A Scrollbar-based Visualization for Document Navigation. In *Proceedings of ACM Digital Libraries '99,* ACM, 1999, pp. 122–129.

[4] Card, S., Mackinlay, J. and Shneiderman, B. ***Readings in Information Visualization: Using Vision to Think.*** Morgan Kaufmann Publishers, CA, USA, 1999.

[5] Havre, S., Hetzler, E., Perrine, K., Jurrus, E. and Miller, N. Interactive Visualization of Multiple Query Results. In *Proc. of the IEEE Symposium on Information Visualization 2001,* IEEE Computer Society, 2001, pp. 105-112.

[6] Hearst, M. A. TileBars: Visualization of term distribution information in full text information access. In *Proceedings of CHI '95,* ACM, 1995, pp. 59-66.

[7] `HTMLParser, http://htmlparser.sourceforge.net`

[8] Matlin, M. ***Cognition, Fifth Edition.*** Harcourt College Publishers, FL, USA, 2002.

[9] Ogden, W.C., Davis, M. W. and Rice, S. Document thumbnail visualization for rapid relevance judgements: when do they pay off? In *Proc. 7th Text Retrieval Conference (TREC-7),* 1998, pp. 528-534.

[10] Paek, T., Dumais, S. and Logan, R. WaveLens: A New View onto Internet Search Results. In *Proceedings of CHI '04,* ACM Press, 2004, pp. 727-734.

[11] Roberts, J.C. and Suvanaphen E. Visual Bracketing for Web search Result Visualization. In *Proceedings of the 7th international Conference on Information Visualization (IV'03),* IEEE Computer Society, 2003, pp. 264-269.

[12] Sebrechts, M., Vasilakis, J., Miller, M., Cugini, J. and Laskowski, S. Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99),* ACM, 1999, pp. 3-10.

[13] Spence, R. ***Information Visualization.*** Addison-Wesley, Harlow, England, 2001.

[14] Veerasamy, A. and Belkin, N. Evaluation of a tool for visualization of information retrieval results. In *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96),* ACM, 1996, pp. 85-92.

[15] Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J. and Pirolli, P. Using thumbnails *to* search the Web. In *Proceedings of CHI 2001,* ACM Press, 2001, pp. 198-205.