Appears in ***WWW'04 Workshop: Measuring Web Effectiveness: The User Perspective****, New York, NY, May 18, 2004*

Enabling Users to Visually Evaluate the Effectiveness of Different Search Queries or Engines

Anselm Spoerri
Department of Library and Information Science
School of Communication, Information and Library Studies
Rutgers University
4 Huntington Street, New Brunswick, NJ 08901, USA
aspoerri@scils.rutgers.edu

Abstract

This paper explores how information visualization can help users evaluate the effectiveness of different query formulations or the same query submitted to multiple search engines. MetaCrystal is used to visualize the degree of overlap and similarity between the results returned by different queries or engines. It enables users to identify documents found by multiple queries or engines. Such documents tend to be more relevant and users can use their number and distribution patterns as a visual measure of the effectiveness of their search.

Keywords: information visualization, search effectiveness measures, meta search.

1. Introduction

When searching the Internet, users are often confronted with an overwhelming number of potentially relevant documents to sift through. It is difficult for users to determine the overall effectiveness of their search. Researchers tend to use measures, such as *precision* and *recall*, to capture the effectiveness of an information retrieval system [12]. Searching the Web is a highly interactive process and additional measures are needed to capture the richness of user interactions and their experiences of ease of use. Spink and Wilson [14] have proposed that some of these new measures need to reflect how users progress through the different stages of their information seeking process and how their interactions change their understanding of their information problem. Spink [13] employed such a user-based approach to evaluate the meta search interface Inquirus. Her study showed that users experienced some change in their information problem and information seeking stage. The results also showed that search precision did not necessarily correlate with the user-based evaluation measures or change in the search process. Another study [17] evaluated the performance of four search engines by using these user-centered criteria: relevance, efficiency, utility, user satisfaction, and connectivity. This paper explores how MetaCrystal, a set of visual tools, can help users evaluate the effectiveness of their own searches. Documents found by multiple retrieval methods are more likely to be relevant [4, 11]. MetaCrystal enables users to identify documents found by multiple queries or engines.

Users tend to create short queries when searching the Internet and they rarely formulate advanced queries [15]. Eastman and Jansen [3] have shown that the use of most query operators in short Internet queries had no significant impact on the effectiveness of the search results. MetaCrystal can visualize this high degree of similarity between different formulations of simple Internet searches. It enables users to see that these related queries are not very effective in finding more relevant documents and that the relevance of the found documents can not be corroborated by these related queries.

Users employ meta search engines because individual search engines only index 20% of the Internet [10] and therefore return different documents for the same query. Meta search engines address this limitation by combining the results by different engines. MetaCrystal visualizes the precise overlap between the top documents retrieved by different search engines. It makes it easy for user to identify how many and which documents have been found by more than one search engine.

This paper is organized as follows: section 2 briefly reviews related visualization work. Section 3 describes the MetaCrystal toolset. Section 4 shows how the degree of similarity between different ranked lists can be visualized. Section 5 discusses how MetaCrystal enables users to get a visual sense of the effectiveness of their searches.

2. Related Visualization Work

Many visual tools have been developed to help users overcome the specific problems they encounter in the search process [8]. Several meta search engines have been developed that use visualization techniques. Vivísimo [18], Grokker [6], MetaSpider [2] and Sparkler [7] address the information overload problem. Vivisimo organizes the retrieved documents using the familiar hierarchical folders metaphor. Grokker uses nested circular or rectangular shapes to visualize a hierarchical grouping of the search results. MetaSpider uses a self-organizing 2-D map approach to classify and display the retrieved documents. Sparkler combines a bull's eye layout with star plots, where a document is plotted on each star spoke based on its rankings by the different engines. Kartoo [9] addresses the query reformulation problem. It creates a 2-D map of the highest ranked documents and displays the key terms that can be added or subtracted to modify the current query.

3. MetaCrystal Toolset

MetaCrystal consists of several linked tools that enable users to compare and combine the search results returned by different query formulations or different search engines processing the same query. The *Category View* displays the degree of overlap between the top result sets returned by different queries or search engines (see Figures 1 and 3). The *Cluster Bulls-Eye View* enables users to see how *all* the found documents are related to the different queries or engines, causing documents with high total ranking scores to cluster toward the center (see Figures 1). This view can also be used to visualize the degree of similarity between different ranked lists (see Figures 2 and 3). The *RankSpiral View* places *all* the documents sequentially along an expanding spiral based on their total ranking scores. This view overcomes the limitation of standard ranked lists which can show only a small number of documents at any given moment.

Implemented in Flash using ActionScript, MetaCrystal supports rapid exploration, enables advanced filtering operations and guides users toward relevant information. Its direct manipulation interface enables users to iteratively create crystals of increasing complexity that show the precise overlap between up to five search engines or queries. Users can apply different weights to the queries or engines to create their own ranking functions. Users can control the degree of overlap between the different engines or queries by modifying the URL directory depth used for matching documents or by changing the number of top documents compared.

The Category and Cluster Bulls-Eye Views will be described in more detail, because they enable users to identify documents found by multiple engines and to see how their rankings by the different engines are related. Users can use this type of visual feedback to get sense of the effectiveness of their search process.

3.1. Category View

In Figure 1, the Category View displays the precise overlap between the top documents retrieved by the search engines Google, Teoma and AltaVista, when searching for 'information visualization'. Modeled on the InfoCrystal layout [16], the interior consists of *category icons*, whose shapes, colors, positions and orientations encode different search engine combinations. At the periphery, colored and star-shaped *input icons* represent the different search engines, whose top 100 results are compared to compute the contents of the category icons. The icon in the center of the Category View displays the number of documents retrieved by all engines. The number of engines represented by a category icon decreases toward the periphery. Shape coding is used for a category icon if we want to emphasize the number of search engines it represents; size coding is employed to emphasize the number of documents retrieved by a search engine combination (see Figures 1 and 3).

3.2. Cluster Bulls-Eye View

In Figure 1, the Cluster Bulls-Eye View shows how *all* the retrieved documents are related to the different engines, because a document's position reflects the relative difference between its rankings by the different search engines. Documents with similar rankings by the different engines will be placed in close proximity. Shape, color and orientation coding indicate which search engines retrieved a document. The Cluster Bulls-Eye Vie*w* uses polar coordinates to display the documents: the *radius* value is related to a document's total ranking score so that the score increases toward the center; the *angle* reflects the relative ratio of a document's rankings by the different engines. The *total ranking score* of a document is calculated by adding the number of engines that retrieved it and the average of its different rankings by the different engines. This causes documents retrieved by the same number of engines to cluster and be contained in the same concentric ring (see Figure 1). Documents with high rankings by the different engines cluster closest toward the center in their respective concentric rings. Documents with low rankings cluster furthest away from the center in their respective rings. In addition, a document's position is influenced by the input icons. Although not shown in this view, the input icons act as "points of reference" that pull a document toward them based on the document's rankings by the different engines. The Cluster Bulls-Eye Vie*w* also makes it easy for users to scan the top documents that are only retrieved by a specific engine.

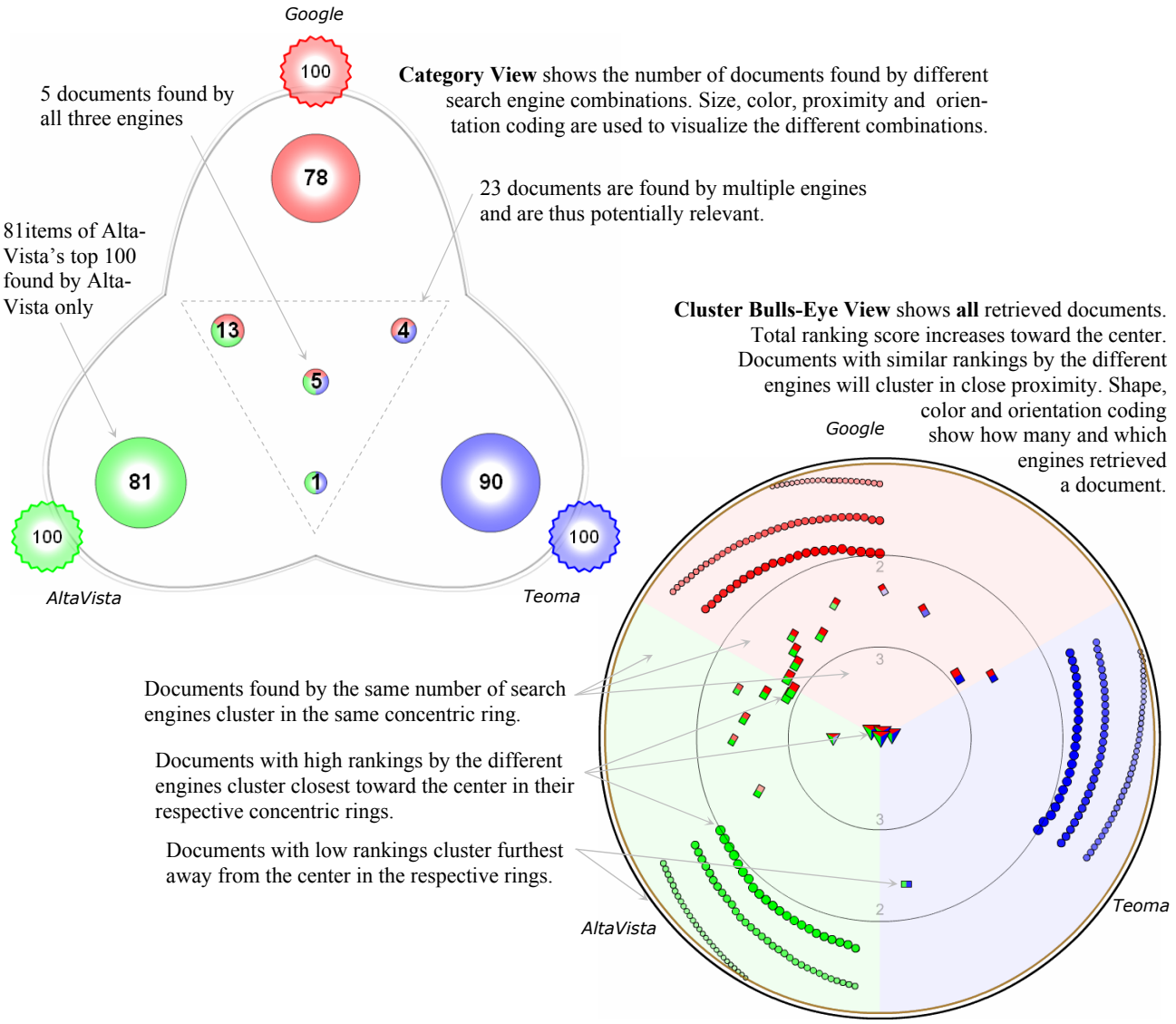

Figure 1: The *Category* and *Cluster Bulls-Eye* views provide users with complementary ways to explore the precise overlap between the top 100 documents found by Google, Teoma and AltaVista, when searching for 'information visualization'. These overviews make it easy for users to identify how many and which documents have been retrieved by more than one search engine.

4. Visualizing Degree of Similarity Between Multiple Ranked Lists

The Cluster Bulls-Eye View can be used to visualize the degree of similarity between different ranked lists. If a document is contained in all the search results being compared, then it will be placed inside the inner most circle of the Cluster Bulls-Eye View. The greater a document's rankings in the lists being compared, the closer toward the center its position will be. If a document has the same ranking in all of the results being compared, then its angle will be 90 degrees (see Figure 2 [i]). If the rankings of the documents are not correlated in the different result sets, then they will cluster as shown Figure 2 [ii]. If the rankings are identical for two of three results lists being compared, then documents will cluster along the line separating the lightly colored areas associated with each input query, as shown in Figure 2 [iii] a). If documents are only contained in two of the three result lists being compared and their rankings are identical, then they will cluster as shown in Figure 2 [iii] b). Figure 2 demonstrates that certain types of similarity relationships between ranked lists give rise to unique "visual signature" patterns in the Cluster Bulls-Eye View. Thus, it can be used to determine the degree of similarity between the ranked lists returned by different queries or search engines.

Visualizing Degree of Similarity between Three Ranked Lists

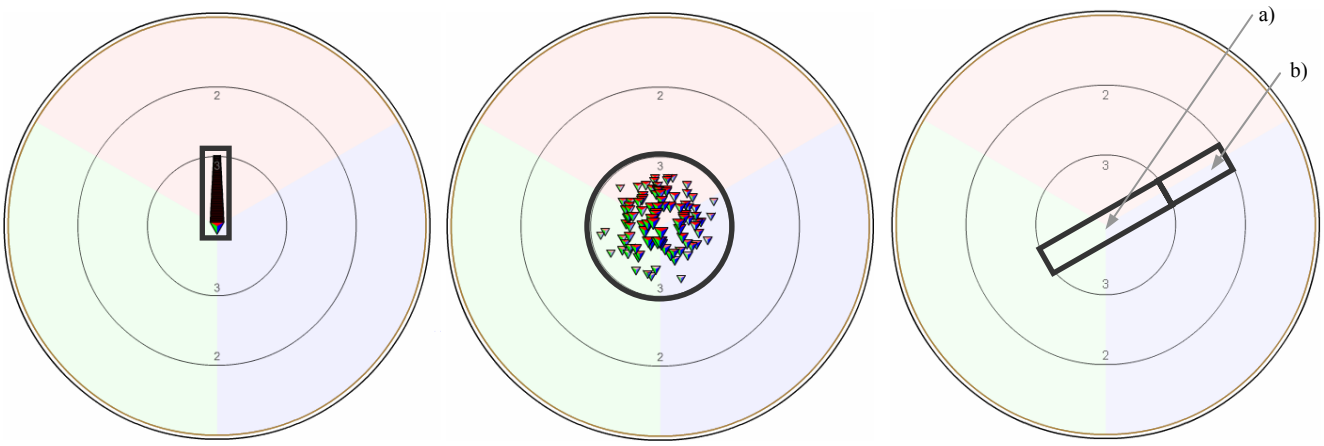

Figure 2: shows the "visual signatures" and how documents will cluster in the *Cluster Bulls-Eye View* if the three ranked lists being compared: [i] are identical; [ii] contain the same documents, but their ranking orders are randomized. [iii] a) shows the area where the documents found by all three queries will cluster if the rankings for the first two queries are identical; [iii] b) highlights the area where the documents found by only two queries will cluster if the rankings for these two queries are identical.

Comparing Results by Different Query Formulations

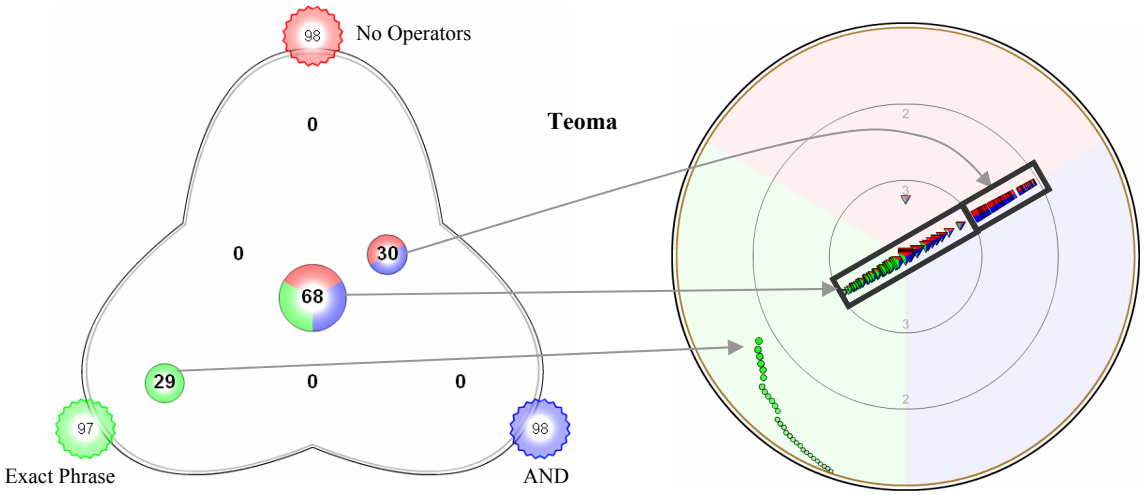

Figure 3: shows the degree of overlap between the top documents retrieved if we search for 'information visualization' and compare queries that employ "no operators", a "Boolean AND" or an "exact phrase" constraint, using the Teoma search engine. The *Category Views* on the left uses size coding for the category icons and shows that there is a great deal of overlap between the queries. In particular, the same documents are retrieved by the queries that use "no operator" or the "Boolean AND", because the category icons related to only one of these queries are empty. The *Cluster Bulls-Eye Views* on the right shows that the ranking order for the documents found by more than one query is identical or very similar. Specifically, the documents found by only two queries have identical rankings. The documents found by all three queries have identical rankings for the "no operator" and "Boolean AND" queries. Thus, these different query formulations do not provide "independent sources of evidence" to infer the potential relevance of the documents.

5. Visualizing the Effectiveness of Different Query Formulations or Meta Searches

The Category and Cluster Bulls-Eye views can be used to visualize and support the finding that the use of most query operators in short Internet queries leads to very similar search results [3], both in terms of the retrieved documents and their respective rankings. In Figure 3, searching for 'information visualization', we visually compare the search results returned if a query that uses "no operators", a "Boolean AND" or an "exact phrase" constraint, respectively, is submitted to the Teoma search engine. The Category View shows that there is a great deal of overlap between the different formulations. Research has shown that documents found by multiple queries are more likely to be relevant [4, 11]. However, the Cluster Bulls-Eye View shows that the queries which use "no operators" or the "Boolean AND", retrieve the same documents and their rankings are identical, because they cluster in Figure 3 in the same areas that are highlighted in Figure 2 [iii] a) and b), respectively. The high number of documents retrieved by more than one query formulation can not be interpreted in isolation. The relationship between the rankings by the different queries needs to be considered. This helps users determine if the different queries actually represent sufficiently independent "sources of evidence" to infer potential relevance. Because the different query formulations are submitted to the same Teoma database, the high number of documents found by multiple queries in the Category View and their highly structured distribution pattern in Cluster Bulls-Eye View indicate that the different query formulations are not very effective in finding more relevant documents.

Meta search engines combine the results by different engines, because individual engines tend to index 20% of the Internet [10] and thus return different documents for the same query. In meta search context, the fact that a document has been retrieved by multiple engines is significant, because each search engine uses a unique retrieval method and indexes different parts of the Internet. Hence, each engine can be understood as an independent "source of evidence" that can be used to corroborate the potential relevance of a document. MetaCrystal makes it easy for user to identify how many and which documents have been found by more than one search engine. The Category View groups all the documents found by the same combination of engines and displays the number of documents retrieved by different search engine combinations (see Figure 1). The Cluster Bulls-Eye View visualizes the relationship between a document's rankings by the different engines that retrieved it. Figure 1 shows that there is no systematic pattern of relatedness between the rankings for the documents found by multiple search engines. This indicates that the meta search has been effective in retrieving documents likely to be relevant.

The number of documents retrieved by more than one Internet search engine tends to be small [5, 10] (see Figure 1). MetaCrystal enables users to control the degree of overlap between the different engines by modifying the URL directory depth used for matching documents or by changing the number of top documents compared. Further, some of the top documents found by a single search engine are likely to be relevant, but users don't want to have to sift through a long list of documents to find them [8]. The Cluster Bulls-Eye View helps users identify documents found by different engine combinations and at the same time scan the top documents retrieved by a specific engine, especially since users may prefer some engines more than others.

Conclusions

This paper addressed how information visualization can help users evaluate the effectiveness of their search. MetaCrystal and its Category and Cluster Bulls-Eye views were used to visualize the degree of overlap and similarity between the results returned by different queries or engines. It was shown that certain types of similarity relationships between ranked lists give rise to unique visual patterns in the Cluster Bulls-Eye View.

Documents found by multiple retrieval methods tend to be more relevant [4, 11]. MetaCrystal enables users to identify how many and which documents have been retrieved by more than one method. This paper has shown that the relationship between a document's rankings by the different retrieval methods needs to be also considered to infer a document's potential relevance. The Cluster Bulls-Eye View enables users to visually examine and spot specific the distribution pattern of the rankings for all the retrieved documents. If there is a systematic pattern of relatedness between the rankings for the documents found by more than one method, then it is likely that the different retrieval methods are not effective, especially if they search the same database.

MetaCrystal was used to visualize and support the finding that the use of query operators in short Internet queries leads to very similar search results [3] and thus the different query formulations are ineffective in retrieving more relevant documents. The Category View showed that there is a great deal of overlap between the documents retrieved by different formulations of the same query, but little overlap when the same query is submitted to different Internet search engines. The Cluster Bulls-Eye View was used to show that the rankings for the documents retrieved by different query formulations are highly related, whereas the rankings in the meta search context are not related in a structured way.

References

[1] Belkin, N. & Croft, B. (1992). Information Filtering and Information Retrieval: Two Sides of the Same Coin. Comm. of the ACM, Dec., 1992.

[2] Chen H., Fan H., Chau M. and Zeng D. MetaSpider: (2001). Meta-Searching and Categorization on the Web. JASIS, Volume 52 (13), 1134 - 1147.

[3] Eastman, C. and Jansen, B. J. (2003) Coverage, relevance, and ranking: the impact of query operators on Web search engine results. ACM Transactions on Information Systems. 21(4), 383 - 411.

[4] Foltz, P. and Dumais, St. (1992) Personalized information delivery: An analysis of information-filtering methods. Comm. of the ACM, 35 (12):51-60.

[5] Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: The retrieval effectiveness of search engines. Information Processing and Management, 35 (2), 141–180.

[6] Grokker – www.groxis.com

[7] Havre, S., Hetzler, E., Perrine K., Jurrus E., and Miller N. (2001). Interactive Visualization of Multiple Query Results. Proc. IEEE Information Visualization Symp. 2001.

[8] Hearst M. (1999). User interfaces and visualization. Modern Information Retrieval. R. Baeza-Yates and B. Ribeiro-Neto (eds.). Addison-Wesley, 257-323.

[9] Kartoo – www.kartoo.com

[10] Lawrence, S., & Giles, C.L. (1999). Accessibility of information on the Web. Nature, 400, 107–109.

[11] Saracevic, T. and Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches and overlap. JASIS. 39, 3, 197-216.

[12] Saracevic, T. (1995). Evaluation of evaluation in information retrieval. ACM SIGIR ‘ 95.

[13] Spink, A. (2002). A user centered approach to the evaluation of Web search engines: An exploratory study. Information Processing and Management, 38(3), 401-426.

[14] Spink, A., Wilson, T. D. (1999). Toward a theoretical framework for information retrieval (IR) evaluation in an information seeking context. Proceedings of MIRA 99:

[15] Spink, A., Wolfram, D., Jansen, B. J., and Saracevic, T. (2001). Searching of the Web: the public and their queries. JASIS, 52 (3) (2001), 226 - 234 .

[16] Spoerri, A. (1999). InfoCrystal: A Visual Tool for Information Retrieval. In Card S., Mackinlay J. and B. Shneiderman (Eds.), Readings in Information Visualization: Using Vision to Think (pp. 140 – 147). San Francisco: Morgan Kaufmann.

[17] Su. L. T., Chen, H. L., & Dong, X. Y. (1998). Evaluation of Web-based search engines from an end-user's perspective: A pilot study. Proceedings of ASIS&T ‘98.

[18] Vivisimo – www.vivisimo.com