MetaCrystal:
Visualizing the Degree of Overlap between Different Search Engines

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

ABSTRACT
MetaCrystal enables users to visualize and control the degree of overlap between the results returned by different search engines. Several linked overview tools support rapid exploration, facilitate complex filtering operations and guide users toward relevant information. MetaCrystal addresses the problem of the effective fusion of different search results by helping users to visually combine and filter the top results returned by the different engines. Users can apply weights to the search engines to create their own ranking functions. They can control the degree of overlap by modifying the URL directory depth used to match documents or by changing the number of top documents being compared.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**] Information Search and Retrieval - *information filtering*. H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**] User Interfaces - *graphical user interfaces (GUI)*.

General Terms: Design

Keywords: Information visualization, meta searching.

1. INTRODUCTION
Users can choose from a multitude of Internet search engines, which tend to return different results for the same query [4]. Meta search engines address this limitation by combining the results by different engines. While some meta search engines visually organize the retrieved documents [1, 3, 4, 5], no meta search interface provides users with an overview of the *precise overlap* between the search engines. Meta searching can benefit from such a visualization, because: a) it is difficult to predict the quality of coverage for single search engines, which tend to index less than 20% of the Internet [6]; b) documents found by multiple retrieval methods are more likely to be relevant [2]; c) some engines are more effective than others, depending on the search domain; d) users prefer or trust some engines more than others.

2. METACRYSTAL
MetaCrystal consists of several overview tools to enable rapid exploration and complex filtering operations. The *Category View* groups and shows the number of documents retrieved by different search engine combinations. Modeled on the InfoCrystal layout [7], the number of documents retrieved by all engines is shown in the center (see Figure 1). Shape (size), color, proximity and orientation coding are used to visually organize the different

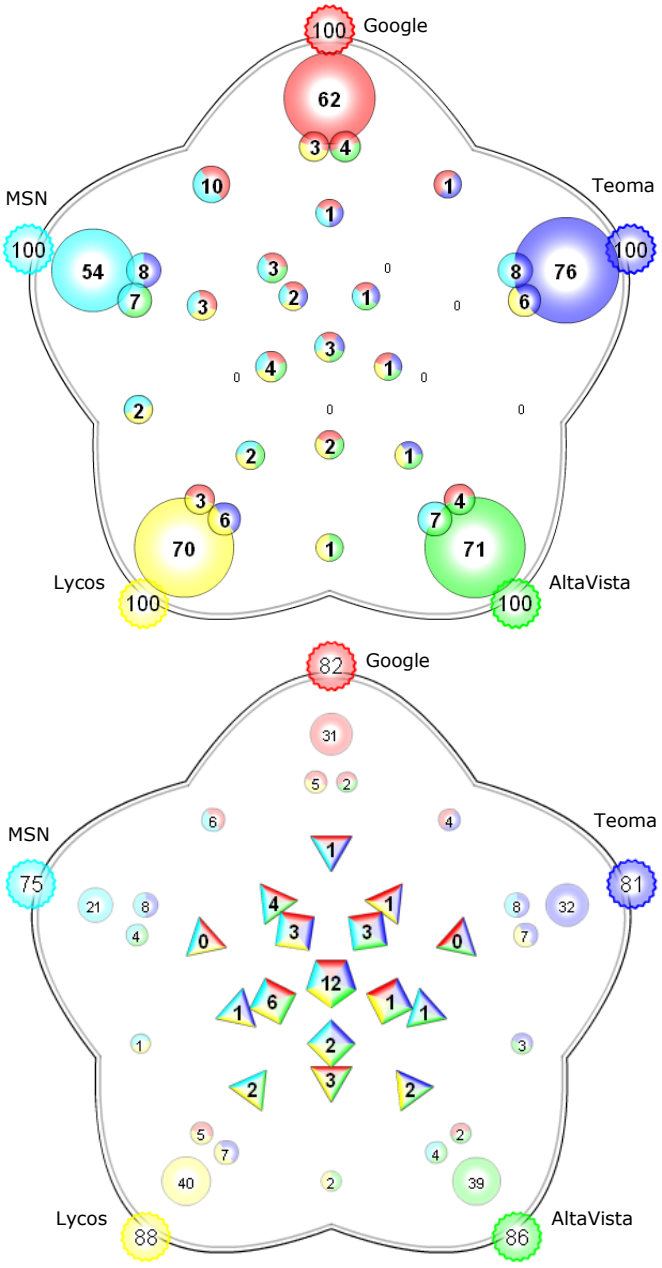

Figure 1: (Top) The *Category View* shows the overlap between the top 100 items found by Google, Teoma, AltaVista, Lycos & MSN, when searching for 'information visualization'. (Bottom) If a URL equal to the server address is used to match items, then the overlap increases. Icons for items retrieved by at least 3 engines are selected. Shape (size) coding is used for icons (not) selected.

WWW 2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

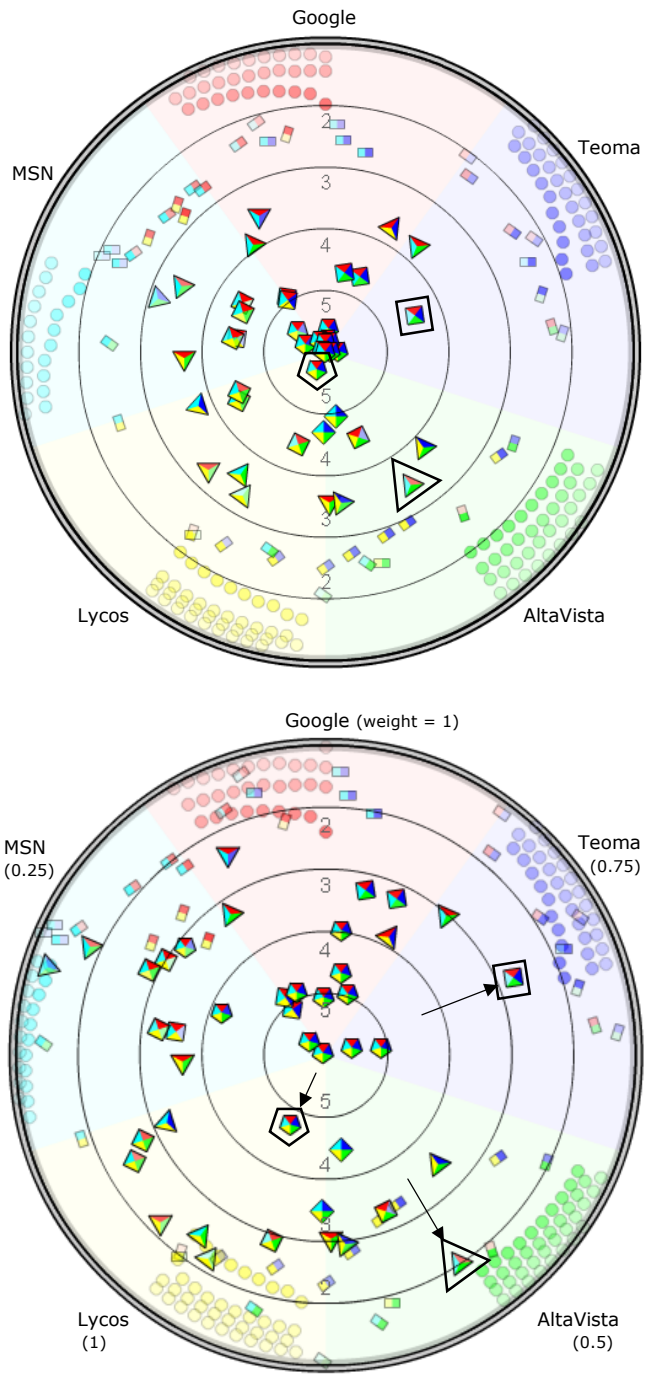

Figure 2: (Top) The *Cluster Bulls-Eye View* uses polar coordinates to place in close proximity documents with similar rankings by the different search engines. A document's *radius* is a function of the number of engines that retrieved it and the average of its different ranking scores. A document's *angle* reflects the relative ratio of its rankings by the different engines. (Bottom) Shows how all the documents cluster if the engines are assigned different weights of importance (1, 0.75, 0.5, 1, 0.25) and the radius is now equal to the *weighted average* of the rankings by the different engines. In both displays, the documents found by at least three engines are selected. Three documents are highlighted to show how their locations change when the weights are applied.

search engine combinations. The number of engines represented by a category icon decreases toward the periphery. Figure 1 (top) shows that 3 documents are found by all engines; 4 documents are found by Google, AltaVista, Lycos and MSN but not Teoma; and most of the documents are retrieved by only one engine. To improve the overlap, users can shorten the URL used to match documents. This reduces the number of unique documents in the result sets being compared, but it can increase the number of documents found by more than one engine (see Figure 1 (bottom)). The *Cluster Bulls-Eye View* shows *all* the retrieved documents; documents found by the same number of engines cluster in the same concentric ring and those with high rankings by the different engines cluster toward the center (see Figure 2). Shape, color and orientation coding show which engines retrieved a document. Users can apply different weights to the search engines to create their own ranking functions. Figure 2 (bottom) shows the resulting clustering of the documents if Google, Teoma, AltaVista, Lycos, and MSN are assigned the weights of 1, 0.75, 0.5, 1, and 0.25, respectively. Search engines commonly display their results as ranked lists, which can only show a limited number of documents. The *RankSpiral View* overcomes this limitation by placing *all* documents along a spiral based on their total ranking scores. *Details-on-demand* gives users an immediate sense of a document's content and how the rankings by the different engines contributed to its total ranking score. MetaCrystal is implemented in Flash using ActionScript. It enables users to perform complex *filtering operations* visually to create a short list of potentially relevant documents by: a) selecting only documents retrieved by a specific number of engines; b) specifying Boolean constraints (the Category View represents all possible Boolean queries in disjunctive normal form [6]); c) applying a threshold based on the total ranking score; d) selecting individual category icons or documents by clicking on them. Users can quickly scan the *filtered results* view, which shows a document's title, total ranking score and how the different engines contributed to its total ranking score. In summary, MetaCrystal addresses the problem of the effective fusion of different search results by enabling users to visually combine and filter the top documents retrieved by different search engines. Its complementary overview tools help users identify documents found by different engine combinations and at the same time scan the top documents retrieved by a single engine.

3. REFERENCES

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

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

[3] Grokker – www.groxis.com

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

[5] Kartoo – www.kartoo.com

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

[7] Spoerri, A. InfoCrystal: A Visual Tool for Information Retrieval. *Readings in Information Visualization: Using Vision to Think*. Card S., Mackinlay J. and B. Shneiderman (eds.) Morgan Kaufmann (1999), 140 - 147. [also Proc. VIS'93].