

RankSpiral: Toward Enhancing Search Results Visualizations

Anselm Spoerri¹

School of Communication, Information & Library Studies, Rutgers University
4 Huntington Street, New Brunswick, NJ 08901, USA

ABSTRACT

This paper addresses the problem of how to enable users to visually explore and compare large sets of documents that have been retrieved by different search engines or queries. The RankSpiral enables users to rapidly scan large numbers of documents and their titles in a single screen. It uses a spiral mapping that maximizes information density and minimizes occlusions. It solves the labeling problem by exploiting the structure of the special spiral mapping used. Focus+Context interactions enable users to examine document clusters or groupings in more detail.

CR Categories: 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)*.

Keywords: Information visualization, meta searching, visual search tools, document visualization.

1. INTRODUCTION

Users searching the Internet face the problem of how to explore the many documents being retrieved or how to compare the documents found by different search engines. Search results tend to be presented as a ranked list, which has the advantage that users know where to start their search for relevant documents. However, only a small subset of documents is visible in a single screen. Most users only look at the first screen of search results [10]. Individual search engines only index 20% of the Internet [8] and therefore return different documents for the same query. Meta search engines address this limitation by combining the results by different engines. The automatic and effective fusion of different search engine results can be difficult [1]. Research has shown that documents found by multiple retrieval methods are more likely to be relevant [9]. Further, users may want to examine the top documents only found by a specific engine, since users may prefer some engines more than others. This paper addresses how to support users in the task of exploring large sets of search results.

2. RELATED WORK

Sparkler [6] 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 queries or search engines. Users need to examine individual document icons to be able to determine how many and which retrieval methods found them. NIRVE [4] contains a "Document Spiral" tool, which places the highest ranked document in the center. Subsequent document icons are placed and spaced along the spiral proportional to their relative document score. However, the placement of documents icons can lead to

spurious perceptual groupings and the display space is not optimally used. Torres et al. [12] use a similar spiral layout to display images based their similarity with the query. Carlis et al. [2] display the periodic data along an Archimedean spiral. Serial attributes are highlighted along the spiral axis and periodic ones along a ray emanating from the center. Several meta search engines visualize the retrieved documents. Kartoo [7] creates a 2-D map of the highest ranked documents and also displays the key terms. Grokker [5] uses nested circles or rectangles to visualize a hierarchical grouping of the search results. MetaSpider [3] uses a self-organizing 2-D map approach to organize the documents.

3. RankSpiral

The RankSpiral is part of the *MetaCrystal* toolset that enables users to control how to combine and filter the documents retrieved by different search methods [11]. *MetaCrystal*'s design is guided by the fact that documents found by multiple retrieval methods are more likely to be relevant [9]. Implemented in Flash using ActionScript, all its tools incorporate a "bull's eye" layout so that users can use the distance from a display's center as a visual cue of a document's potential relevance. All tools support Details-on-Demand to give users an immediate sense of a document's content; Focus+Context interactions enable users to examine document clusters in more detail without losing a sense of context.

Search engines commonly display their results as ranked lists, which can only show a limited number of document titles in a single screen. The *RankSpiral* overcomes this limitation by placing *all* the retrieved documents sequentially along an expanding spiral based on their decreasing total ranking score (see Figure 1). The *total ranking score* of a document is equal to the sum of the number of engines that retrieved it and the normalized average of its rankings by the different engines that found it. This causes documents retrieved by multiple methods to cluster toward the center and documents found by the same number of engines to be mapped into in the same concentric ring. The RankSpiral differs from a standard spiral mapping, because it places consecutive documents adjacent to each other so that they do not overlap. Specifically, they share a tangent that passes through the display center (see Figure 1). Thus, documents with the same total ranking score will not be mapped to the same location, but instead form a circular segment. A large jump in the total relevance score between consecutive documents causes a visual gap in the spiral. Shape, color and orientation coding are used to indicate which engines retrieved a document. Within in each concentric ring, size coding is used to visualize the relative ranking of the documents found by the same number engines. Documents with high (low) rankings by the different engines cluster are mapped toward (furthest away from) the center in their respective rings and the size of their icons is set the largest (smallest) value. This creates a subtle visual hierarchy and has these benefits: 1) more documents can be tightly placed along the spiral; 2) users can easily identify the top documents found by a specific number of engines.

¹ aspoerri@scils.rutgers.edu

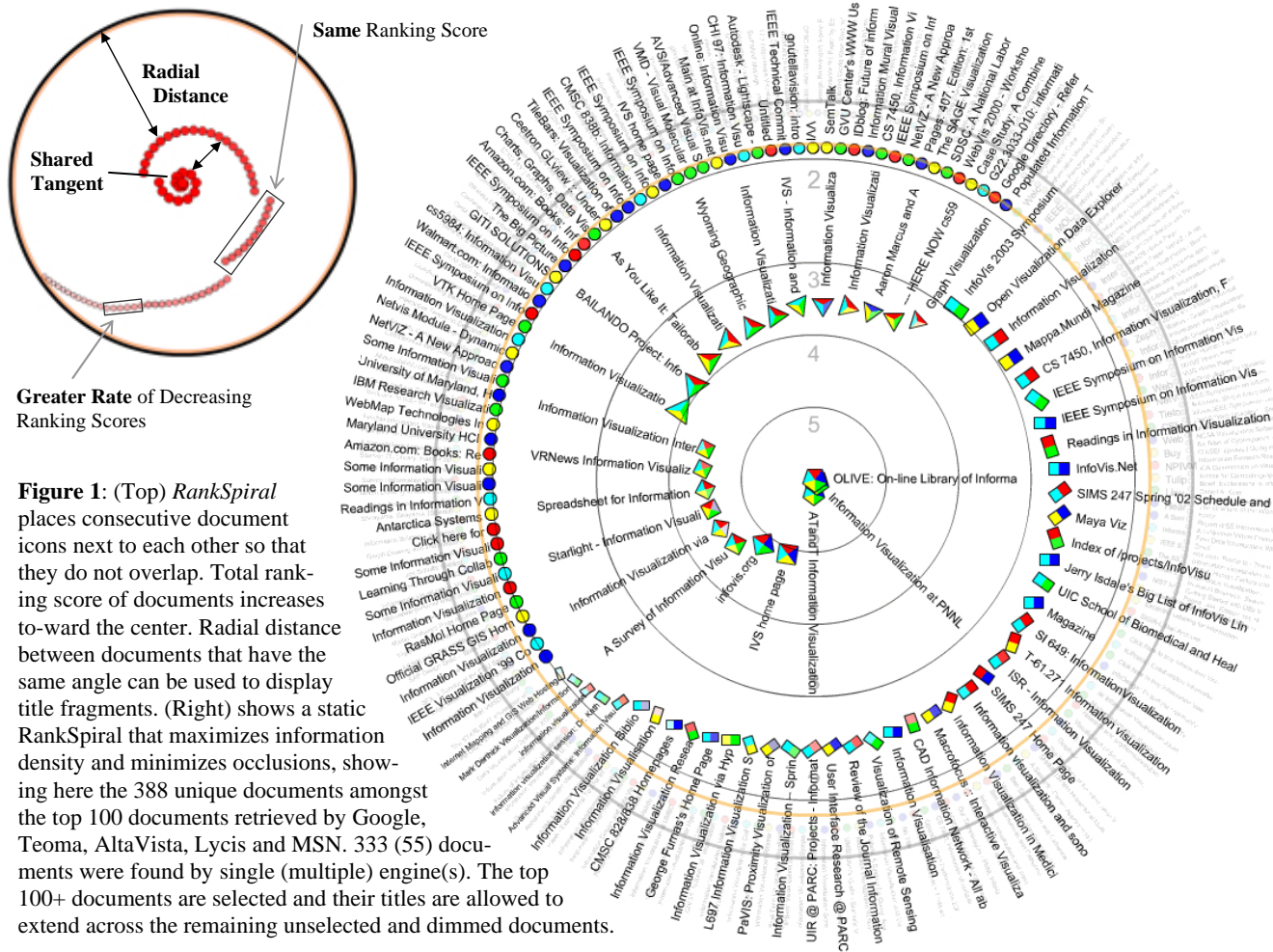


Figure 1: (Top) RankSpiral places consecutive document icons next to each other so that they do not overlap. Total ranking score of documents increases to-ward the center. Radial distance between documents that have the same angle can be used to display title fragments. (Right) shows a static RankSpiral that maximizes information density and minimizes occlusions, showing here the 388 unique documents amongst the top 100 documents retrieved by Google, Teoma, AltaVista, Lycis and MSN. 333 (55) documents were found by single (multiple) engine(s). The top 100+ documents are selected and their titles are allowed to extend across the remaining unselected and dimmed documents.

The RankSpiral solves the *labeling problem* by exploiting the structure of the spiral mapping used. For each document, the radial distance to the icon that has the same angle as the document in question can be computed (see Figure 1). This distance is used to display titles in such way as to not occlude any document icons. If users modify the focus area, the radial distances are recomputed and the visible portion of the titles is updated accordingly. If a document is selected, its title is allowed to extend across the unselected icons (see Figure 1). Thus, users can create a two-layered display, where the selected documents have longer title fragments and the unselected documents are dimmed and reduced in size. Users can rotate the spiral to read the titles displayed at different angles. They can dynamically highlight and magnify specific groupings of document titles. The RankSpiral is well suited to display skewed document distributions, such as meta search results, where most documents are found by a single method and few are found by several engines. Thus, the limited space of the spiral's center area is sufficient to display the few documents retrieved by many search engines.

In conclusion, the RankSpiral aims to enable users to rapidly explore many more documents than they normally would by offering them a high information density display that presents the search results in a compact and structured way. The next step is to conduct a formal evaluation and compare the RankSpiral with single and multi column displays that have comparable information densities and support focus+context interactions.

4. References

- [1] Callan, J. (2000). Distributed information retrieval. In Croft W. (Ed), Advances in Information Retrieval. Kluwer Academic Publishers.
- [2] Carlis, J.V., and Konstan, J.A. Interactive Visualization of Serial Periodic Data, Proceedings of ACM UIST'98.
- [3] Chen H., Fan H., Chau M. and Zeng D. MetaSpider: (2001). Meta-Searching and Categorization on the Web. JASIS, Volume 52 (13).
- [4] Cugini, J., Piatko, C. & Laskowski, S. (1996). Interactive 3D Visualization for Document Retrieval. Proc. of Workshop on New Paradigms in Information Visualization & Manipulation, CIKM '96.
- [5] Grokker – www.groxis.com
- [6] Havre, S., Hetzler, E., Perrine K., Jurrus E., and Miller N. (2001). Interactive Visualization of Multiple Query Results. Proc. IEEE Information Visualization Symp. 2001.
- [7] Kartoo – www.kartoo.com
- [8] Lawrence, S., & Giles, C.L. (1999). Accessibility of information on the Web. Nature, 400, 107–109.
- [9] Saracevic, T. and Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches and overlap. JASIS. 39, 3.
- [10] Silverstein, C., Henzinger, M., Marais, J. & Moricz, M. (1998). Analysis of a very large Alta Vista query log. Technical Report 1998-014, COMPAQ Systems Research Center, Palo Alto, Ca, USA.
- [11] Spoerri, A. (2004) MetaCrystal: A Visual Interface for Meta Searching. Proceedings of ACM CHI 2004.
- [12] Torres, R., Silva, C., Medeiros, C. & Rocha, H. (2003). Visual Structures for Image Browsing CIKM'03, November 3–8, 2003, New Orleans, Louisiana, USA.