Session-based search with Querium

Gene Golovchinsky FX Palo Alto Lab Palo Alto, CA, USA gene@fxpal.com

Abdigani Diriye University College London London, UK a.diriye@ucl.ac.uk

ABSTRACT

We illustrate the use of Querium, a novel search system designed to support people's collaborative and multisession search tasks, in the context of the HCIR 2011 Search Challenge. This report demonstrates how a Querium's interface and search engine can be used to search for documents in an open-ended, exploratory task. We illustrate the use of relevance feedback, faceted search, query fusion, and the search history, as well as commenting and overview functions.

Author Keywords

Multi-session search, search interface.

ACM Classification Keywords

H5.3. Information interfaces and presentation: Group and organizational work.

INTRODUCTION

One important characteristic of many collaborative search systems is the notion that search activity spans multiple sessions. Examples of this include carrying out a background literature review, planning a holiday, and gathering information about a product to purchase [5]. Multi-session search tasks comprise a significant portion of most web search activity [5][9][6]; searchers undertaking multi-session search tasks tend to make heavy use of bookmarking features, re-searching to relocate previously seen content, printing, emailing and saving pages to tackle these search tasks [1]. A number of system features exist within experimental search systems and browsers to support such tasks, although studies (e.g., [6][2]) suggest that such features are rarely found in the wild, where people rely on bookmarks and email to manage their found information.

We based the design loosely on SACK [4], borrowing some of its querying ideas, but extend it with a number of communication options and a more streamlined interface. In the following section, we introduce the essential interface features of Querium, and then illustrate how the system was used to try to find an answer to the example task given on the HCIR 2011 web site.

THE QUERIUM SEARCH INTERFACE

Querium is designed to leverage the familiar style of interaction with web search engines while extending that interaction in important ways. The system includes a text entry field through which queries can be typed, a list of matching results is displayed for each query, and a left-hand sidebar displays additional controls and information.

Beyond these traditional interface components, Querium organizes search activities into tasks, and augments the search view by adding a resizable embedded document pane (but documents could also be opened in a separate window), rich query history displays, additional controls on the search results, a variety of filters, and a notepad/chat window (see Figure 1). In addition, a summary view is used to organize a team's activities with respect to some task.



Figure 1. Querium interface: (1) top area for new queries, (2) navigation sidebar, (3) main results, and (4) document view.

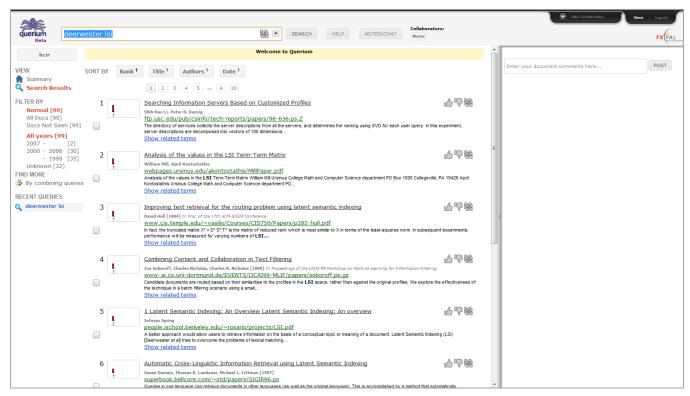
Querium uses the notion of a task as an organizing principle for managing search and collaborative activity. A task consists of one or more queries, and may involve one or more collaborators. A person may create as many tasks as necessary. Queries, comments, judgments of relevance, and other interactions with information are localized to a task; the same document retrieved in different tasks will have different assessments, and will be shared (or not shared) with different collaborators.

THE TASK

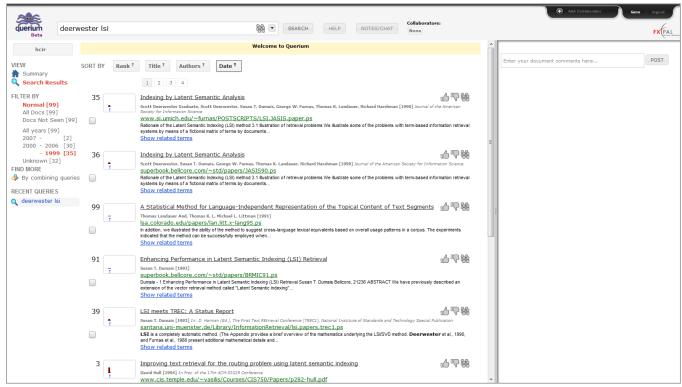
The HCIR 2011 search challenge sample task involved finding documents in support of the following information need using the CiteSeer collection [3] (CC BY-NC-SA 3.0):

Latent Semantic Indexing (LSI) is an indexing and retrieval method that uses a mathematical technique called Singular Value Decomposition (SVD) to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text. Deerwester et al. published seminal papers on LSI in 1988. Is there earlier work that anticipates some or part of this approach?

In the following section, we will describe the search process we used to explore the collection; novel interface features will be explained as they are used. We started with a simple query "deerwester Isi" Querium retrieved the top 99 matching documents, as shown below.

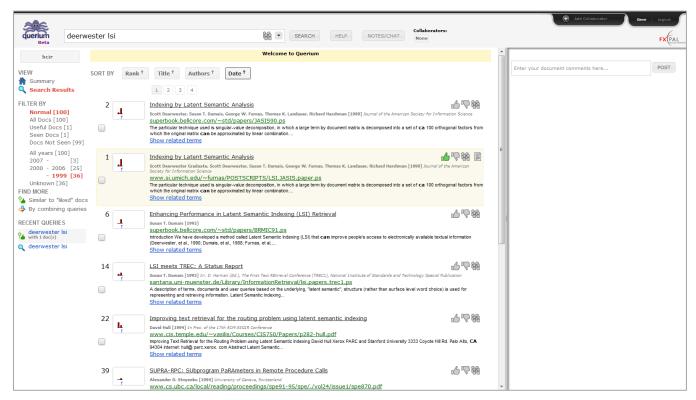


We now filtered the results to documents that were published prior to 1999, and sorted them by ascending date:

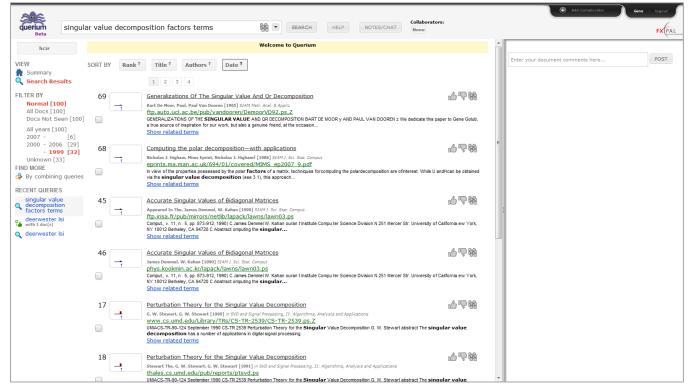


This produced a list of 35 documents, including some of the work by Deerwester's team. We marked that document as potentially useful by clicking on the "thumbs up" button, and then ran a relevance feedback search for documents

similar to the selected one. Relevance feedback was implemented using reverted indexing [7]. The figure below shows the results, filtered and sorted by date

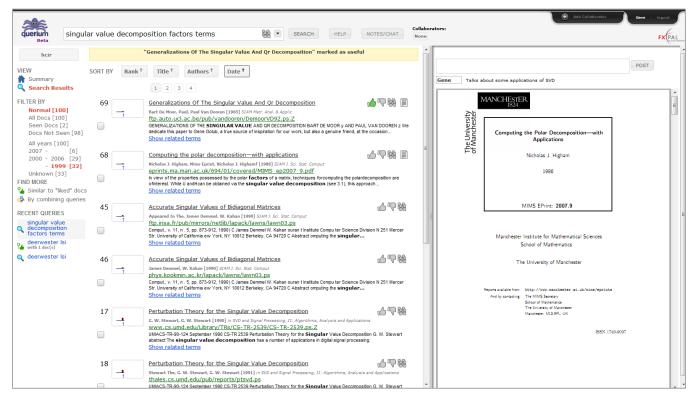


Browsing through these documents did not identify any obvious clues to answering the posed question, so we tried a new query "Singular value decomposition factors terms" The filtered, sorted results are shown below:



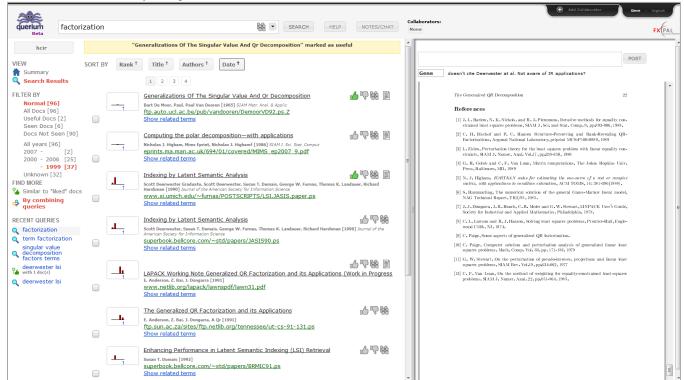
The top document seemed promising, so it was saved (marked with a "thumbs up") and a comment was added to the document. The second document, even though it

preceded the Deerwester et al paper, did not address any issues related to information retrieval.



Next we tried "term factorization" and "factorization", but these queries were too general. At this point, we performed a query fusion step. Query fusion combines the ranked lists of documents retrieved by the queries in the selected task;

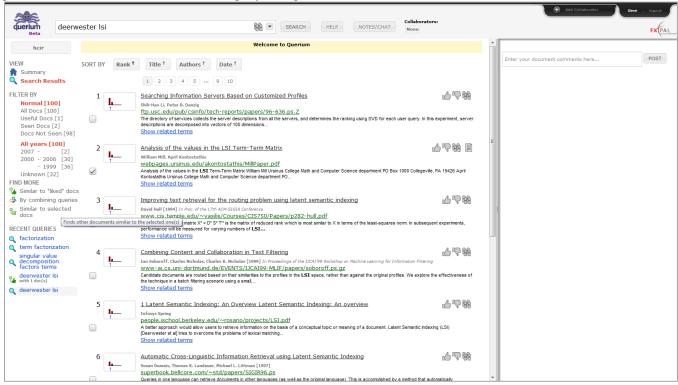
we used the same algorithm as described by Pickens et al. [8]. The combined results, filtered and sorted by date, are shown below:



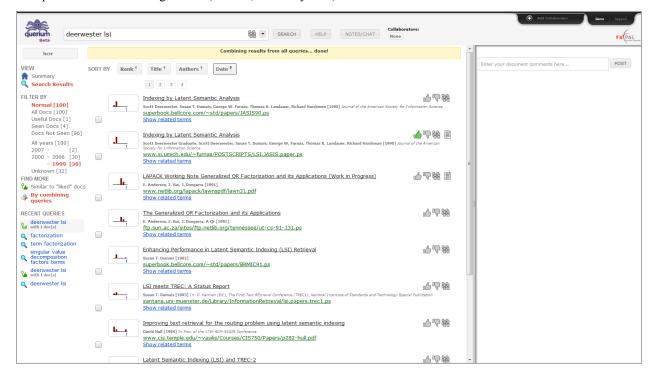
The first document seemed promising, but although it was published after Deerwester et al., it did not cite that work.

At this point, we went back to the first query, to try a different filter. One of the "features" of the dataset we were

using is rather noisy metadata, which makes it worthwhile to explore records with incomplete metadata. We then performed another relevance feedback query using the "Analysis of the values of the LSI Term-Term Matrix" document, as shown below:

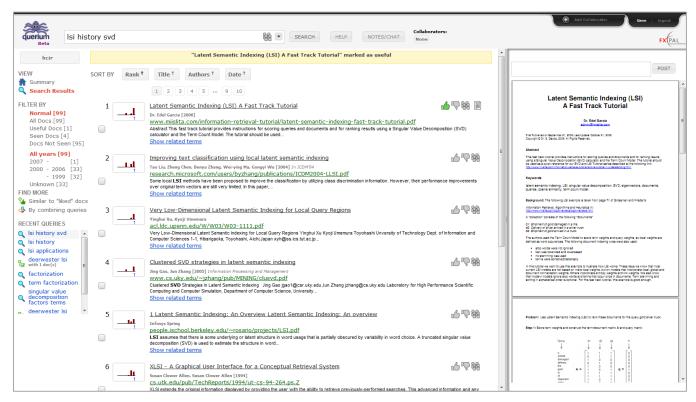


This produced the following results (filtered, sorted by date):



This returned no new interesting documents, so we tried a couple of additional queries, "lsi applications," "lsi history," and "lsi history svd". The first two queries were

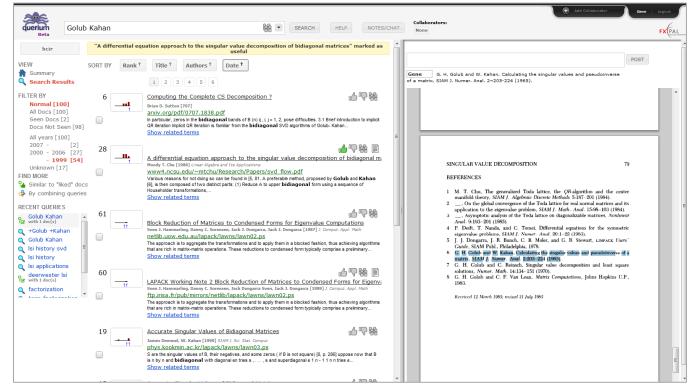
not effective, but the third was more promising. The results are shown below:



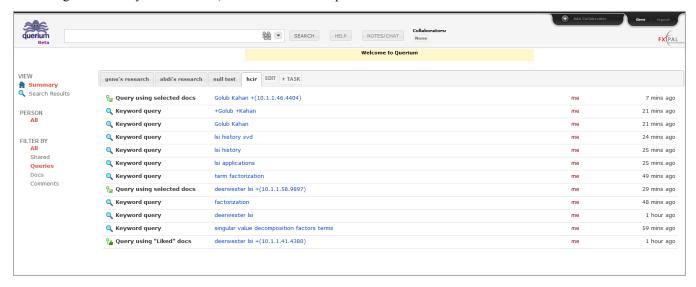
The first document in this list was a tutorial on LSI that included a link to LSI/SVD background. Following that link yielded the following quote:

In 1965 G. Golub and W. Kahan introduced Singular Value Decomposition (SVD) as a ... technique for calculating the singular values, pseudo-inverse and rank of a matrix.

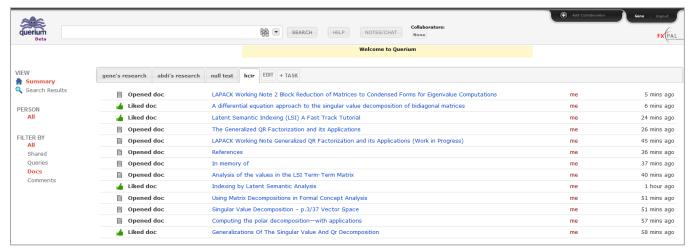
This suggested the queries "Golub Kahan" and "+Golub and +Kahan" as the next things to try. While these were not directly useful, relevance feedback from the first document produced the following filtered, sorted results. The second document seemed promising, and references the section cited a Golub and Kahan paper on SVD.



Reviewing the summary of the session, we ran a number of queries:



And we looked at a bunch of documents, commented on some, and marked some as useful (liked).



In the end, we concluded that the work of Golub and Kahan foreshadowed and may have informed the work of Deerwester, et al. We arrived at this result through a combination of keyword and relevance feedback queries, by filtering and sorting on facets of metadata, by looking at aggregate results for the entire session, and by exploiting different kinds of overviews. In performing this task, we did not exhaust the capabilities of Querium: query overviews, more direct use of histograms, and collaborative features were not used in this example. We expect to demonstrate them in the workshop.

REFERENCES

- [1] Aula, A., Jhaveri, N., and Käki, M. (2005) Information Search and Re-access Strategies of Experienced Web Users. In *Proc. WWW 2005*.
- [2] Capra, R., Marchionini, G., Velasco-Martin, J., and Muller, K. (2010) Tools-at-hand and learning in multi-session, collaborative search. In *Proc. CHI2010*.
- [3] CiteSeer. Available online at citeseerx.ist.psu.edu/

- [4] Golovchinsky, G., and Pickens, J. (2010) Interactive Information Seeking via Selective Application of Contextual Knowledge. In *Proc. IIiX* 2010. ACM Press.
- [5] MacKay B., Watters C. (2008) Exploring multi-session web tasks. In *Proc. CHI 2008*
- [6] Morris, M.R. (2008) A Survey of Collaborative Web Search Practices. In *Proc. CHI 2008*, ACM Press.
- [7] Pickens, J., Cooper, M., and Golovchinsky, G. (2010) Reverted indexing for feedback and expansion. In Proc. CIKM 2010, ACM Press
- [8] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. (2008) Algorithmic Mediation for Collaborative Exploratory Search. In Proc. SIGIR 2008. ACM, New York, NY, pp. 315-322.
- [9] Sellen, A., Murphy, R., and Shaw, K. How Knowledge Workers Use the Web. In *Proc. CHI 2002*, ACM Press, pp. 227-234.