Collaborative Query Previews in Digital Libraries

Lin Fu, Dion Hoe-Lian Goh and Schubert Shou-Boon Foo
Division of Information Studies, School of Communication and Information
Nanyang Technological University, Singapore 637718

Abstract

A novel method named collaborative query previews (CQPs) to augment information seeking in digital libraries is proposed. CQPs aim to help users make better sense of the collection of a digital library from two perspectives: query previews and recommendations of past queries. Query previews generate overviews about a collection of documents through a set of aggregate information on selected attributes. As a user issues queries, related queries submitted previously by other users are recommended. By obtaining overview information and utilizing other user’s experiences in retrieving information, a user is expected to formulate queries that better meet information needs. A prototype work environment incorporating CQPs has been developed and described in this paper.

Key Words
Information Seeking, Information Retrieval, Digital Library, Collaborative Query Previews.

Introduction

Information seeking is a broad term encompassing the ways individuals articulate their information needs, seek, evaluate, select and use information. In the course of a search, the individual may interact with people, manual information systems (such as libraries) or with computer-based information systems [13]. A major theme in the various information seeking models is that interaction and collaboration with other people is an important part in the process of information seeking and use (e.g. [2]).

In addition, various models reveal that information seeking typically requires the use of small set of “seed” information to locate large amounts of useful target information (e.g. [1]). In this
work, we term this “seed” information as **pre-query information** (PQI) – the information necessary to successfully carry out information retrieval. PQI is multifaceted and can be viewed as a synergistic combination of the information need, the information system used and the knowledge about the collection in the information system. The first aspect of PQI, the information need, is unpredictable as it varies between individuals and contexts. The type of information system used is the second aspect of PQI. This concerns the technical aspects of the system and once again, varies according to factors such as the implementation of the system, query languages used, ranking algorithms, and so on.

The third feature of PQI is the knowledge of the collection in the information system. This transcends the technical aspects of the system and instead refers to the skills and experiences acquired by individuals through the use of the system and through the interaction with other users. Thus for example, while all users may know the query syntax required of a particular digital library, only some users may know the “best” query formulation for an information need based on similar past experiences with using the system.

The information need (the first aspect) is beyond the control of an information system because it is highly dependent on the individual. Acquiring technical knowledge of the information system (the second aspect) on the other hand is relatively uncomplicated and usually involves user training. Of the three aspects of PQI, assisting users in gaining knowledge about the collection is perhaps the most challenging and will be the focus of this work.

In this paper, we propose a new approach named **collaborative query previews** (CQPs) to help users obtain knowledge about a collection in a digital library and in the process, improve the results of their search. To test this approach, a prototype digital library incorporating CQPs was designed and implemented. The digital library consists of digitized multimedia artifacts that reflect the rich history and heritage of Singapore and their corresponding metadata.

**Related Work**

Several techniques have been used to incorporate aspects of PQI into the information retrieval process. Two approaches most related to this work will be described here: query previews and collaborative filtering.

The query preview is a technique that provides an overview about the data distribution in a database [8]. Experiments have shown that query previews can help users reduce the chances of formulating queries with zero or large numbers of hits as well as the browsing effort by preventing the retrieval of undesired records [10]. However current implementations of query previews do not support communication and collaboration with other people during the process of gathering PQI since these previews only present information about the dataset. Recall that information seeking behavior models have indicated the importance of the human component since information seekers typically depend on others to gather PQI.
Collaborative filtering is a technique for recommending items to a user based on similarities between the past behavior of the user and that of likeminded people. Collaborative filtering is a beneficial tool for users in that it harnesses the community for knowledge sharing and selects high quality and relevant items from a large information stream [7]. Research in collaborative filtering however does not emphasize human information seeking behavior. For example, existing collaborative filtering systems do not take into account the importance of the knowledge about a document collection which is an essential component of PQI.

Collaborative Query Previews in Digital Library

To address the limitations of existing query preview and collaborative filtering systems, we propose an integrated approach to information retrieval through CQPs. CQPs combine query previews and collaborative filtering, and support collaboration and communication during the process of gathering PQI. Like traditional query previews, CQPs generate an overview about a collection of documents through a set of aggregate information on selected attributes. However unlike traditional query previews, CQPs overviews are dynamic and personalized, based on a user’s profile which describes his/her interests and which represents the information need. Such a profile may be generated by asking the user a set of questions or by monitoring the user’s query activity over time.

Consider a user in the digital library whose profile indicates interest areas related to software engineering, computer graphics and artificial intelligence. Given this profile (the information need), the CQPs facility would present the user with summary data on each area (the query previews). There might be 5000 items under the “software engineering” category, 1500 items under “computer graphics” and 3000 items under “artificial intelligence”. For each area, the user may drill-down to reveal several sub-areas. Here, under “artificial intelligence”, query previews may indicate 2000 items about “machine learning”, 500 about “fuzzy logic”, and 500 about “intelligent agent”. As the user is able to see the potential size of the query results before query submission, he is less likely to submit queries that return zero or large numbers of hits. For example, since the previews do not indicate any items on robotics, the user might decide against issuing queries that fall within that category.

Beyond providing overviews, CQPs introduce a collaborative aspect by providing recommendations of queries. That is, as a user selects attributes of interest on the query previews or issues queries, related queries submitted previously by other users will be displayed, giving the user a better understanding of the collection and type of information that can be retrieved.

If the user is not satisfied with the query results, he may either try to modify the query keywords, use the recommended queries, or browse the query previews to carry out further rounds of searching until he feels his information need is met. For example, he might select the recommended query “social activities” and use it for another round of information retrieval.
System Design and Implementation

The current prototype was developed at the School of Communication and Information at Nanyang Technological University, Singapore. It comprises a zoomable work environment (ZWE) [6] that provides access to a digital library of a subset of digitized multimedia artifacts that reflect the rich history and heritage of Singapore and their corresponding metadata. Besides basic keyword searching, ZWE augments the information seeking process through CQPs.

**Figure 1:** The Zoomable Work Environment

The ZWE interface is shown in Figure 1. The browsing tree provides access to the collection through a hierarchical organization of resources to an arbitrary depth. The query area supports a variety of search options such as Boolean and phrase searches and the query results are presented in an independent window. The work area employs a spatial metaphor and functions like a desktop in which users organize their resources. The size of the work area is arbitrarily large and panning and zooming is supported for navigation and visualization. More information about the features of the ZWE can be found in [6]. CQPs are displayed with query area in an independent window on the ZWE.

Figure 2 shows the architecture of the COP-enabled digital library system. The interface tier is the ZWE and includes the visualization of CQPs. ZWE is implemented using JAZZ [9], a Zoomable User Interface (ZUI) API, that allows developers to quickly and easily build zoomable information spaces.
The service tier provides the functionality of the digital library. It is an extensible, platform consisting of different application plug-ins that can process resources in the various resource repositories [5]. Existing application plug-ins include resource display, browsing, searching, annotating, query previews and query recommendation. For query recommendation module, we proposed a hybrid approach [3] [4] to cluster past queries and apply the algorithms to find similar past queries for a given query. Experiments show that our hybrid algorithm outperforms the existing query clustering approach [3] [4] [11].

The data storage tier consists of the multimedia repository, metadata repository, user profile repository and past query repository. The user profile repository stores information describing a user’s interests and are used to generate dynamic query previews. The past queries repository contains queries issued by users of the digital library and gathered through query logs. The multimedia repository uses the file system for storage while the Tamino XML server [12] is used to access and manipulate the metadata repositories since both metadata are represented in XML format.
Conclusions

By integrating query previews and collaborative filtering, CQPs support the information seeking process from two perspectives:

1. From direct previews of the data collection. CQPs employ query previews to provide an overview of the structure and content of a document collection. With this information, CQPs overcome the lack of indicators that would be helpful for users in formulating queries to meet their information needs.

2. From queries issued previously by others. Through collaborative filtering techniques, CQPs recommend related queries previously executed by other users to help the user make better sense of how the document collection met past information needs that coincide with the present information need. By harnessing other users’ experiences, a user can more accurately formulate queries and avoid or reduce those that result in zero or large numbers of hits.

With the initial prototype developed, the next phase of this work will focus on the evaluation of CQPs by users of the digital library. Continuing research is also being carried out to improve the aspects of query clustering by further investigating the use of hybrid approaches, including content-based, feedback-based and result-based approaches.

References


