Chapter II
Collaborative Querying
Using a Hybrid Content and Results-Based Approach

Chandrani Sinha Ray
Nanyang Technological University, Singapore

Dion Hoe-Lian Goh
Nanyang Technological University, Singapore

Schubert Foo
Nanyang Technological University, Singapore

Nyein Chan Soe Win
Nanyang Technological University, Singapore

Khasfariyati Razikin
Nanyang Technological University, Singapore

ABSTRACT

Collaborative querying is a technique that makes use of past users’ search experiences in order to help the current user formulate an appropriate query. In this technique, related queries are extracted from query logs and clustered. Queries from these clusters that are related to the user’s query are then recommended to the user. This work uses a combination of query terms as well as results documents returned by queries for clustering queries. For the latter, it extracts features such as titles, URLs, and snippets from the results documents. It also proposes an extended K-means clustering algorithm for clustering queries over a simple measure of overlap. Experimental results reveal that the best clusters are obtained by using a combination of these sources rather than using only query terms or only results URLs alone.
INTRODUCTION

Web search engines play a vital role in retrieving information from the Internet, but there are several challenges faced by users of information retrieval (IR) systems in general. Users are often overwhelmed by the number of documents returned by an IR system or fail to express their information needs in terms compatible to those used in the system. While new IR techniques can potentially alleviate these problems, specifying information needs properly in the form of a query is still a fundamental problem with users. Borgman (1996) attributed this to the lack of conceptual knowledge of the information retrieval process. Put differently, users need to know how to translate their information needs into searchable queries.

Research in information seeking behavior suggests an alternative approach in helping users meet their information needs. Studies have found that collaboration among searchers is an important step in the process of information seeking and use. For example, Taylor (1968) highlights the importance of the interaction between the inquirer and the librarian, while Ellis (1993) argues that communication with colleagues is a key component in the initial search for information. We use the term collaborative IR to refer to a family of techniques that facilitate collaboration among users while conducting searches. These techniques can be categorized based on the way people search for information into: collaborative filtering (Cohen & Fan, 2000), collaborative browsing (Lieberman, 1995), and collaborative querying (Setten & Hadidy, n.d.). In particular, collaborative querying helps searchers in query formulation by harnessing other users’ search experience or expert knowledge (Goh, Fu, & Foo, 2005).

Queries, being expressions of information needs, have the potential to provide a wealth of information that could be used to guide other searchers with similar information needs, helping them with query reformulation. If large quantities of queries are collected, they may be mined for emerging patterns and relationships that could be used to define communities of interest (e.g., Lycos 50), and even help users explore not only their current information needs but related ones as well. One can imagine a network of queries amassed from the collective expertise of numerous searchers, representing what people think as helping them meet their information needs. Query formulation and reformulation will then be a matter of exploring the query network, executing relevant queries, and retrieving the resulting documents.

There are many techniques for uncovering related queries to support collaborative querying, one of which is query clustering (Wen, Nie, & Zhang, 2001). Here, query logs of IR systems are mined, and similar queries are grouped and used as recommendations. Different query clustering approaches are distinguished based on the measure of the similarity between queries. Traditional term-based similarity measures borrowed from classical IR (Setten & Hadidy, n.d.) represent one example. However, the short lengths of queries (e.g., Silverstein, Henzinger, Marais, & Moricz, 1998) limit the usefulness of this approach. An alternative is based on the overlap of the top N results (such as URLs) returned from an IR system (Raghavan & Sever, 1995; Glance, 2001). This approach however might not be effective for queries with different semantic meanings, but leading to similar search results since terms are not considered during clustering. Hence, a hybrid approach was proposed by Fu, Goh, Foo, and Na (2003) which is a linear combination of the content-based and results-based approaches. Experiments showed that this approach produces better query clusters in terms of precision, recall, and coverage than using either the content-based or results-based method alone. Although these results show potential, there is still room for improvement by using additional information for measuring similarity. Examples include titles, snippets (document excerpts), or other metadata
Related Content

Virtuality and Reality: What Happens When the Two Collide?
www.igi-global.com/chapter/virtuality-reality-happens-when-two/29225?camid=4v1a

Recognised Creativity: The Influence of Process, Social Needs, and the Third Drive on Creative Individuals' Work through Social Media
www.igi-global.com/chapter/recognised-creativity/130420?camid=4v1a

Social Media as Positive Disruption in Education, E-Learning and B-Learning
www.igi-global.com/chapter/social-media-positive-disruption-education/60324?camid=4v1a

Knowledge Sharing for Cultural Heritage 2.0: Prosumers in a Digital Agora
Francesca Bertacchini and Assunta Tavernise (2014). International Journal of Virtual Communities and Social Networking (pp. 24-36).
www.igi-global.com/article/knowledge-sharing-for-cultural-heritage-20/121668?camid=4v1a