# ScholarWorks@UMass Amherst

**Title**

SEARCHING BASED ON QUERY DOCUMENTS

**Author**

**Youngho Kim**, *University of Massachusetts Amherst* Follow

**Document Type**

Open Access Dissertation

**Degree Name**

Doctor of Philosophy (PhD)

**Degree Program**

Computer Science

**Year Degree Awarded**

Fall 2014

## First Advisor

W. Bruce Croft

## Second Advisor

James Allan

## Third Advisor

Benjamin Marlin

## Subject Categories

Computer Sciences

## Abstract

Searches can start with query documents where search queries are formulated based on document-level descriptions. This type of searches is more common in domain-specific search environments. For example, in patent retrieval, one major search task is finding relevant information for new (query) patents, and search queries are generated from the query patents One unique characteristic of this search is that the search process can take longer and be more comprehensive, compared to general web search. As an example, to complete a single patent retrieval task, a typical user may generate 15 queries and examine more than 100 retrieved documents. In these search environments, searchers need to formulate multiple queries based on query documents that are typically complex and difficult to understand. In this work, we describe methods for automatically generating queries and diversifying search results based on query documents, which can be used for query vi suggestion and for improving the quality of retrieval results. In particular, we focus on resolving three main issues related to query document-based searches: (1) query generation, (2) query suggestion and formulation, and (3) search result diversification. Automatic query generation helps users by reducing the burden of formulating queries from query documents. Using generated queries as suggestions is investigated as a method of presenting alternative queries. Search result diversification is important in domain-specific search because of the nature of the query documents. Since query documents generally contain long complex descriptions, diverse query topics can be identified, and a range of relevant documents can be found that are related to these diverse topics. The proposed methods we study in this thesis explicitly address these three issues. To solve the query generation issue, we use binary decision trees to generate effective Boolean queries and labeling propagation to formulate more effective phrasal-concept queries. In order to diversify search results, we propose two different approaches: query-side and result-level diversification. To generate diverse queries, we identify important topics from query documents and generate queries based on the identified topics. For result-level diversification, we extract query topics from query documents, and apply state-of-the-art diversification algorithms based on the extracted topics. In addition, we devise query suggestion techniques for each query generation method. To demonstrate the effectiveness of our approach, we conduct experiments for various domain-specific search tasks, and devise appropriate evaluation measures for domain-specific search environments.

## Recommended Citation

Kim, Youngho, "SEARCHING BASED ON QUERY DOCUMENTS" (2014). *Doctoral Dissertations*. 218.
https://scholarworks.umass.edu/dissertations_2/218

Download

DOWNLOADS

Since November 12, 2014

Included in

Computer Sciences Commons

Share

COinS