

# Chapter 8

## Search Query Recommendations in Web Information Retrieval Using Query Logs

**R. Umagandhi**

*Kongunadu Arts and Science College, India*

**A. V. Senthil Kumar**

*Hindusthan College of Arts and Science, India*

### ABSTRACT

*Web is the largest and voluminous data source in the world. The inconceivable boom of information available in the web simultaneously throws the challenge of retrieving the precise and appropriate information at the time of need. The unpredictable amount of web information available becomes a menace of experiencing ambiguity in the web search. In this scenario, Search engine retrieves significant information from the web, based on the query term given by the user. The search queries given by the user are always short and ambiguous and the queries may not produce the appropriate results. The retrieved result may not be relevant all the time. At times irrelevant and redundant results are also retrieved because of the short and ambiguous query keywords. Query Recommendation is a technique to provide the alternate queries as a substitute of the input query to the user to frame the queries in future. A methodology was framed to identify the similar queries and they are clustered; this cluster contains the similar queries which are used to provide the recommendations.*

### INTRODUCTION

The exhaustive information available in the World Wide Web indeed, unfolds the challenge of exploring the apposite, precise and relevant data in every search result. The plentiful unstructured or semi-structured information on the web leads to a great challenge for the users, who hunt for prompt information. The scenario grows pathetic and distressing to provide a personalised service to the individual users from billions of web pages. At the end of the nineties the size of the web to be around 200 million static pages (Bharat, K. *et al.*, 1998). The number of indexable documents in the web exceeds 11.5 billion (Antonio, G. *et al.*, 2005). According to the survey done by Netcraft, Internet Services Company, England there

DOI: 10.4018/978-1-5225-0613-3.ch008

is 739,032,236 sites in September 2013 and 22.2M more than the month August 2013. Every year, millions of web sites are newly added in the information world. Hence a proper tool is needed to search the information on the web.

Search Engine retrieves significant and essential information from the web, based on the query term given by the user. The retrieved result may not be relevant all the time. At times irrelevant and redundant results are also retrieved by the search engine because of the query keywords which are short and ambiguous (Mark, S. 2008). The unpredictable amount of web information available becomes a menace of experiencing ambiguity in the web search. To prevent the web users from getting overwhelmed by the quantity of information available in the web, several strategies are proposed with the advent of data mining techniques.

Search engines are used to retrieve the information from web based on the query term given by the user in terms of web snippets. A web snippet denotes the title, abstract, and URL of a web page returned by the search engines. Apparently, in such instances of web-searching, Query Recommendations is the ultimate application in information retrieval. The Query Recommendation technique provides alternative queries to the user to frame a meaningful and relevant query in the future and rapidly satisfies their information needs. Search engine leaves the search information to the user for further references in the form of query logs. Query log is an important repository, which records the user's search activities. The mining of these logs can improve the performance of search engines. Query log file is a repository contains every query request and its navigation in the search engine and maintained either in the system desktop or in the proxy server. This Chapter deals the new approach for queries recommendations based on:

- The analysis on query log to observe the web users and their sessions, frequent access patterns.
- The proposed query recommendation technique is based on the combined similarity measure on various attributes. Both the positive and negative concepts preference help to explore the string of similarity between the concepts generated which in turn leads to cluster the users with similar intentions.
- The hybrid approach generates the time variant and invariant query clusters; this cluster contains the similar queries based on the attribute time which is used to provide the recommendations.
- The recommendation is based on the user's real search intention which is identified from the hybrid user profile and the recommended queries are prioritized and evaluated using the proposed technique.

The rest of the chapter is organized under background which deals with the review of literature, Basic terms which describes the terms and their definitions used in this chapter and Architecture which describes the methodology used in the proposed work. Next, Identification of similar users and queries identifies the similarity between the users' interms of their queries and Experimental results shows the evidence and the results for the proposed technique. Finally the chapter is concluded.

## **BACKGROUND**

Dupretet *al.* (Dupret, G., & Mendoza, M. 2006) addresses the non-trivial patterns prevalent in query log data as *Query Log Mining*. After the processing of log files, the data mined explicitly assumes three applications namely, *Query Recommendation*, *Document Recommendation* and *Query Classification*.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/search-query-recommendations-in-web-information-retrieval-using-query-logs/162895](http://www.igi-global.com/chapter/search-query-recommendations-in-web-information-retrieval-using-query-logs/162895)

## Related Content

---

### A Conceptual Tool for Usability Problem Identification in Website Development

Mikael B. Skovand Jan Stage (2011). *Web Engineered Applications for Evolving Organizations: Emerging Knowledge* (pp. 278-291).

[www.irma-international.org/chapter/conceptual-tool-usability-problem-identification/53065](http://www.irma-international.org/chapter/conceptual-tool-usability-problem-identification/53065)

### Power System Relay Protection Based on Faster R-CNN Algorithm

Yong Liuand Zhengbiao Jing (2023). *International Journal of Information Technology and Web Engineering* (pp. 1-15).

[www.irma-international.org/article/power-system-relay-protection-based-on-faster-r-cnn-algorithm/333475](http://www.irma-international.org/article/power-system-relay-protection-based-on-faster-r-cnn-algorithm/333475)

### A Deterministic Approach to XML Query Processing with Efficient Support for Pure and Negated Containments

Dunren Che (2006). *International Journal of Information Technology and Web Engineering* (pp. 49-67).

[www.irma-international.org/article/deterministic-approach-xml-query-processing/2618](http://www.irma-international.org/article/deterministic-approach-xml-query-processing/2618)

### A General View of Quality Models for Web Portals and a Particularization to E-Banking Domain

Mª Ángeles Moraga, Julio Córdoba, Coral Caleroand Cristina Cachero (2008). *Handbook of Research on Web Information Systems Quality* (pp. 113-129).

[www.irma-international.org/chapter/general-view-quality-models-web/21969](http://www.irma-international.org/chapter/general-view-quality-models-web/21969)

### Improved Algorithm for Error Correction

Wael Toghujaand Ghazi I. Alkhatib (2011). *International Journal of Information Technology and Web Engineering* (pp. 1-12).

[www.irma-international.org/article/improved-algorithm-error-correction/52802](http://www.irma-international.org/article/improved-algorithm-error-correction/52802)