Chapter 4

Query Recommendation for Improving Search Engine Results

Hamada M. Zahera
Menoufiya University, Egypt

Gamal F. El-Hady
Menoufiya University, Egypt

W. F. Abd El-Wahed
Menoufiya University, Egypt

ABSTRACT

As web contents grow, the importance of search engines become more critical and at the same time user satisfaction decreases. Query recommendation is a new approach to improve search results in web. In this paper a method is proposed that, given a query submitted to a search engine, suggests a list of queries that are related to the user input query. The related queries are based on previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The proposed method is based on clustering processes in which groups of semantically similar queries are detected. The clustering process uses the content of historical preferences of users registered in the query log of the search engine. This facility provides queries that are related to the ones submitted by users in order to direct them toward their required information. This method not only discovers the related queries but also ranks them according to a similarity measure. The method has been evaluated using real data sets from the search engine query log.

INTRODUCTION

With the increase of size and popularity of the World Wide Web, many users find it’s difficult to get the desired information, although they use most efficient search engines (e.g. Google, Yahoo), (Tjondronegoro & Spink, 2008). Actually theses search engines allow users to specify queries simply as lists of keywords, following the approach of traditional information systems (Baeza-Yates & Ribeiro-Neto, 1999; Cambazoglu, 2010). But this list of keywords is not always a good descriptor of the needed information, therefore it was important to achieve user’s stratification of
Query Recommendation for Improving Search Engine Results

search engine results and make it easy to retrieve the required information (Hong, Siew, & Egerton, 2010; Höchstötter & Lewandowski, 2009). The problem of improving search engine results and obtaining the desired information from this huge amount of web contents has been processed by different ways such as clustering the search engine results in specific topics so the user can find the required results in selected category of search results (Chen, 2010; Caramia & Pezzoli, 2004). Although, the user doesn’t use the proper search words or search query while searching so this leads to a problem of getting un-required results and the user have to be familiar with specific terminology in a knowledge domain (Baeza-Yates, Hurtado, & Mendoza, 2004). This is not always the case of many users; they have only a little background about the information they are searching and unfortunately they didn’t get the required results. In order to overcome this problem, it’s not enough to use clustering search results method because the problem is not in obtaining the huge results but it’s in the keywords used in searching are not strongly related (Baeza-Yates, Hurtado, & Mendoza, 2004). Query recommendation suggests related queries for search engine users when they are not satisfied with the results of an initial input query, thus assisting users in improving search quality (Giacometti, Marcel, & Negre, 2009). Conventional approaches to query recommendation have been focused on expanding a query by terms extracted from various information sources such as a thesaurus like WordNet (Li, Otsuka, & Kitsuregawa, 2008; Chatzopoulou, Eirinaki, & Polyzotis, 2009).

The previous queries stored in query logs can be a source of additional evidence to help future users. A query recommendation system based on large-scale Web access logs and web page archive, and evaluate three query recommendation strategies based on different feature spaces (i.e., noun, URL, and Web community) has been presented (Li, Otsuka, & Kitsuregawa, 2008). The suggested Method aimed to help search engine users in finding their required results easily and quickly, this method suggests related queries beside the input query while the user searches so he can build a proper search query with the knowledge domain terminology which is important for search engine to get the related results. Also the additional time for improving the results must be unnoticeable by the user.

This paper discusses in details the proposed method for recommending list of queries are related to input query user which based on clustering process over web queries extracted from search engine query log file.

RELATED WORK

Yates (2004) has done a survey in to show the different improvements of search engine aspects (Wen, Nie, & Zhang, 2001) presented an algorithm for clustering search engine queries according to four notions according to: first, the context of the query; second, common clicked URLs between queries; third, Similar strings between the queries and fourth, the distance of the clicked documents in some pre-defined hierarchy. Beeferman and Berger (2000) suggested a technique for query clustering based on the third notion. Fonseca, Golgher, De Moura, and Ziviani (2003) presented a new method to discover the related queries which are based on association rules. The queries represent items in tradition association rules. The query log file is considered as a collection of transactions which represent a session in which the user submit all related queries in a specific time. The method showed good results, nevertheless arising of two problems. The first problem is the difficulty of determining which sessions of these queries that are belong to the same search process. The second problem the related queries which are submitted by different users can’t be discovered. This is because the support of a rule increases only of its queries are in the same session and they must by submitted by the same user.