Chapter 11

Human Computer Interaction for Effective Metasearching

Rashid Ali
Aligarh Muslim University, India

M. M. Sufyan Beg
Jamia Millia Islamia, India

ABSTRACT

Metasearching is the process of combining search results of different search systems into a single set of ranked results which, in turn, is expected to provide us the collective benefit of using each of the participating search systems. Since, user is the direct beneficiary of the search results; this motivates the researchers in the field of Human Computer Interaction (HCI) to measure user satisfaction. A user is satisfied if he receives good quality search results in response to his query. To measure user satisfaction, we need to obtain feed back from user. This feedback might also be used to improve the quality of metasearching. The authors discuss the design of a metasearch system that is based on human computer interaction. We compare our method with two other methods Borda’s method and modified Shimura technique. The authors use Spearman’s footrule distance as the measure of comparison. Experimentally, the method performs better than the Borda’s method. The authors argue that the method is significant as it models the user feedback based metasearching and has spam-fighting capabilities.

1. INTRODUCTION

Metasearching is the process of combining the search results obtained from a number of search engines in order to get better overall results. The need of metasearching arises from the fact that no search engine is comprehensive in its coverage of the Web. Also, no ranking algorithm can be thought of being universally acceptable. So, it is a good idea to combine results of different search engines into a better overall ranking that gives the combined advantage of different ranking techniques, which are employed by the participating search engines. The goal is to get a system that gives results better than the best of the participating search engines. In today’s world, where every
organization is putting its information on world wide web, the issue becomes more interesting as more and more metasearching will be needed to suit specific information search applications. We sometimes divide the metasearch engines into two categories namely external metasearch engines and internal metasearch engines. External metasearch engines take lists of results from popular public search systems and return an overall ranked list of results by combining them. In this case, the different ranked lists returned by the different search systems might contain different sets of documents. While in the case of internal metasearch engines, different rankings of the same set of documents, which are returned by the participating sub-engines, are combined into a single overall ranking. Sometimes, it is also called rank aggregation. On the other hand, in the case of external metasearch engines, the ranked lists from participating search engines are not permutations of the same set of documents. Therefore, the process of combining these ranked lists into a single ranked list may be called as rank aggregation of partial lists. If we take the participating search engines as voters and the ranked documents as candidates, then metasearching may be considered simply as developing a system that returns a group preference on these documents.

But, whatever the type of metasearching, the goal remains the same. That is to get a system, which gives results better than that of the best of the participating search engines. The results are better if they satisfy a user with his information need. For this, the user feedback should be taken. We can obtain user feedback by asking the user to fill in a feedback form after he has finished searching. But, it is too demanding for the user. Human Computer Interaction models the interaction between human and Web search engines to obtain user feedback implicitly. In this implicit feedback, we infer the feedback from the user by watching the actions of the user on search results presented before him in response to his query. On the basis of which, a weight is assigned to each result instead of the search engines. Then, the overall ordering of the results may be obtained by sorting the results in descending order of their weights. This overall ordering is in fact overall ordering by the user. But, the problem with the user feedback based method is that it is a time costly affair. Therefore, it is also not scalable. That means it can be performed with a small number of queries but not with a very large number of queries. For larger data sets, automatic method is the only answer. So, it motivated us to look for some automated mean to approximate user’s behavior.

In this Chapter, we discuss a method for effective metasearching based on human computer interaction. Our system learns ranking rules using rough set theory to estimate an aggregated ranking for the rankings obtained from the participating search systems. Our system learns the ranking rules on the basis of user’s ranking, which is available for a given set of rankings in the training set.

This chapter is organized as follows. In Section 2, we briefly look at the background and related work. In Section 3, we discuss the method of obtaining user feedback implicitly. Then, we briefly review the theoretical details of rough set theory and related terminologies. After that, we discuss the rough set based method for metasearching that models the user feedback based metasearching and present the metasearching algorithms. We show our results in Section 4. Finally, we conclude in Section 5.

2. BACKGROUND AND RELATED WORK

Here, we first discuss related work and then list some important definitions.

2.1. Related Work

In past, HCI researchers have shown interest in Web search. With the growth of World Wide Web, the importance of HCI aspects of search engines