ABSTRACT

The web creates new challenges for information retrieval. Because of the phenomenal growth of web, searching for particular information consumes more time. Search engines are one of the most publicly visible realizations of information retrieval technology, because they critically help users in finding relevant information on the World Wide Web (WWW).

Traditional search engines obtain the results by keyword based techniques and finally retrieve thousands of result pages of which only a handful are relevant. In addition, WWW data and the web users are so dynamic and it is difficult to capture the user preferences and interests without interfering with the normal activity of the user. One common solution to the above mentioned problem is personalization, which customizes the Web environment for users and helps them search their information need easily by considering their interests and context. With the cost of running a large scale search engine already very high, it is likely that such a full-scale personalization is currently too expensive for the major web search engines.

Hence a layer of personalization above any existing search engine is a cost effective way of effective personalized web search. This research work proposes an integrated personalized search system architecture which utilizes a new search aiding index called User Conceptual Index (UCI) that provides a relation between search queries and pages, which matches the user’s information need. The UCI is specially designed to account for factors
that affect personalization like hit count, dwelling time, user actions and browsing behaviors to develop an effective personalized web search.

The textual features that represent the content of a page visited by the user are used for indexing purposes and such features are mapped with the search query to identify the context of search. This work models web search by utilizing graph based techniques in combination with implicit interest indicators like user actions. Hence this work also incorporates user’s actions in page re-ranking to provide a better user-oriented search.

While searching for information need, users tend to navigate across web pages through hyperlinks. By exploiting the relations between page-contents using link structures, best search paths can be recommended to the users. In addition, web users do not search for the same information everyday and such change in user interests highlights short-term interests which greatly affect the context of search. Classification of user’s interests into long and short-term interests is yet another significant contribution of our work which is further extended to evolve interest based user groups which aid in recommending pages to new users.

In spite of existing traditional evaluation measures, there are no specific benchmarks to evaluate the results of a personalized web search. Therefore, this thesis also proposes interest based evaluation measures for evaluating the personalized page recommendations.