

## **ABSTRACT**

Web Search Engines (WSE) have become the main mean of initiating the interaction with the Internet due to the explosion of the number of documents available on the web. However, the problem of information resource location is resolved to a certain extent. The main limitation of these systems is that they do not consider the implicit intent behind the query and rely only on the query keywords. The user queries are limited to a few keywords and also not precise to retrieve the correct information they need. Query expansion assist users to get relevant results by formulating an enhanced query, by appending additional keywords to the initial search request. Leveraging the time dimension, ranking can improve the retrieval effectiveness of search engines.

Information is distributed across multiple pages connected via hyperlinks. The information relevant to a specific query is normally spread across various pages. Hence composed pages are generated to summarize the information from various relevant pages. This would enhance the search experience of the user as it is more personalized and customized for his requirement and interest.

A review of literature reveals the importance of concept based approaches in providing personalised search. The behaviour of users is implicitly derived and the preferences of the user are stored in concept based user profiles. Including negative preferences in the user concept preference profile helps to identify more related concepts for the query. To obtain negative preferences, three different strategies were identified namely Joachims-C, mJoachims-C and SpyNB-C. A Ranking Support Vector

Machine (RSVM) is used to learn user preferences from the concept preference pairs.

The main goal of this research work is to improve personalised search by providing query suggestions to the user by considering positive and negative preferences and re-ranking the search results based on temporal information.

The following are the major contributions of this research work:

- i. As search queries are ambiguous, effective methods for search engines to provide query suggestions on semantically related queries are studied. Personalised query suggestions are obtained for individual users based on their conceptual user profiles. User's clickthrough data is exploited to identify the positive preferences.
- ii. To capture the finer details of users' need, negative preferences are included in the user conceptual profile. User concept preference profile is built by considering both positive and negative preferences. Concept preference pairs are obtained from Joachims-C, mJoachims-C and SpyNB-C methods and the preferences are learnt using RSVM.
- iii. Evolutionary Algorithms (EA) like Genetic Algorithms (GA), Particle Swarm Optimisation (PSO) and Hybrid GA-PSO are investigated to form concept clusters. By using each of the user profiling strategies developed, concept clusters are obtained using these evolutionary algorithms thereby providing personalised query suggestions to the users.

To evaluate the methods, a yahoo middleware is developed for collecting user search details and clickthroughs. Users are

invited to search for information on the Yahoo search engine by making use of middleware. The performance of these techniques are evaluated using the metrics Precision, Recall, Mean Average Precision(MAP) and Precision@N. The performance of the evolutionary algorithms based query expansion techniques are evaluated using positive preferences alone and with inclusion of negative preferences also. The results are compared with those obtained from improved Beeferman and Berger's(BB) agglomerative clustering algorithm. It is found that clusters obtained from hybrid GA-PSO are found to be better compared to other algorithms in both cases of positive alone and positive and negative preferences of the user. In case of inclusion of negative preferences, user profile built with SpyNB-C method is found to predict the user preferences correctly

- iv. Re-ranking of search results is done based on temporal information. A time based ranking algorithm is presented which considered not only text relevance but also time relevance documents. By introducing update time and content time of documents into the ranking algorithm, the ranking results are reasonable and effective.
- v. Composed pages are generated by stitching together pieces from other documents. This is useful for providing query specific summarization. A search tool is developed to evaluate the performance of time based ranking. A survey was conducted to evaluate the quality of results. The users were asked to evaluate the search results with respect to specific queries. The ranking quality is measured using Discounted Cumulative Gains (DCG) metric. The proposed method

indicated a 20% increase in the DCG score when compared to the traditional search.

The experimental results indicate that the hybrid approach GA-PSO was found to yield the best results compared to improved BB's algorithm. The quality of concept clusters resulted from the hybrid approach provided personalized results for the unseen queries also. Improvements in DCG score shows that the generation of composed pages summarize the relevant information matching the user's search intents.