Search Engine Personalization Using Concept Based User Profiles

1. Assistant Professor (Department of Computer Science)
SRM UNIVERSITY, NCR CAMPUS, DELHI
2. sharma.cs06@gmail.com (student)
3. om.gupta0406@gmail.com (student)
SRM UNIVERSITY, NCR CAMPUS, DELHI

ABSTRACT
Most commercial search engines return roughly the same results for the same query, regardless of the user’s real interest. Since queries submitted to search engines tend to be short and ambiguous, they are not likely to be able to express the user’s precise needs. Personalized search is an important research area that aims to resolve the ambiguity of query terms. To increase the relevance of search results, personalized search engines create user profiles to capture the users’ personal preferences and as such identify the actual goal of the input query. And personalized ontology is constructed for specifying the user profiling knowledge. A good user profiling strategy is an essential and fundamental component in search engine personalization. The existing technology had several drawbacks like creation of single profile to all users and considers only the positive preferences. To overcome these problems, this project studied seven concept-based user profiling strategies that are capable of deriving both of the user’s positive and negative preferences. All of the users profiling strategies are query-oriented, meaning that a profile is created for each of the user’s queries. Moreover, we find that negative preferences improve the separation of similar and dissimilar queries.

Keywords: Personalization, Search Engine Personalization

1. Introduction
A search engine is a set of programs which are used to search for information within a specific realm and collate that information in a database. Personalization involves using technology to accommodate the differences between individuals. Web pages are personalized based on the characteristics of an individual. Personalization implies that the changes are based on implicit data, such as items purchased or pages viewed. A good user profiling strategy is an essential and fundamental component in search engine personalization. The term “preferences” is used in a variety of related, but not identical, ways in the scientific literature. Preferences could be conceived of as an individual’s attitude towards a set of objects. A user profile is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person’s identity. A user profile can also be considered as the computer representation of a user model. A profile can be used to store the description of the characteristics of person. This information can be exploited by systems taking into account the persons’ characteristics and preferences. Profiling is the process that refers to construction of a profile via the extraction from a set of data. User profiles behave like parameterizations of requirements statements, capturing regular variation in requirements for similar types of system. User profiling strategies can be broadly classified into two main approaches: document-based and concept-based approaches. Document-based user profiling methods aim at capturing users’ clicking and browsing behaviours. Different users may have different functional requirements, and so require different subsets of functionality to be evaluated, or they may have different non-functional constraints on functions. On the other hand, concept-based user profiling methods aim at capturing users’ conceptual needs. Users’ browsed documents and search histories are automatically mapped into a set of topical categories. User profiles are created based on the users’ preferences on the extracted topical categories.
2. Related Work

A major problem of current Web search is that search queries are usually short and ambiguous, and thus are insufficient for specifying the precise user needs. To alleviate this problem, some search engines suggest terms that are semantically related to the submitted queries so that users can choose from the suggestions the ones that reflect their information needs. In this paper, we introduce an effective approach that captures the user’s conceptual preferences in order to provide personalized query suggestions. We achieve this goal with two new strategies. First, we develop online techniques that extract concepts from the web-snippets of the search result returned from a query and use the concepts to identify related queries for that query. Second, we propose a new two phase personalized agglomerative clustering algorithm that is able to generate personalized query clusters [7]. The method proposed is based on a query clustering process in which groups of semantically similar queries are identified. The clustering process uses the content of historical preferences of users registered in the query log of the search engine. The method not only discovers the related queries, but also ranks them according to a relevance criterion. Finally, we show with experiments over the query log of a search engine the effectiveness of the method [4]. Query clustering is a process used to discover frequently asked questions or most popular topics on a search engine. This process is crucial for search engines based on question-answering. Because of the short lengths of queries, approaches based on keywords are not suitable for query clustering. This paper describes a new query clustering method that makes use of user logs which allow us to identify the documents the users have selected for a query. The similarity between two queries may be deduced from the common documents the users selected for them. Our experiments show that a combination of both keywords And user logs is better than using either method alone [10].

3. System Architecture and Method of Personalization

In the proposed system, it addresses both the problems of the user’s positive and negative preferences. All of the users profiling strategies are query-oriented, meaning that a profile is created for each of the user’s queries. It shows that user profiles which capture both the user’s positive and negative preferences perform the best among all of the profiling strategies. A new approach has been introduced in the proposed system is Personalized Ontology, which formally describes and specifies the user profile knowledge.

1. Extend the query-oriented, concept-based user profiling method proposed to consider both users’ positive and negative preferences in building users profiles.

2. Propose six user profiling methods that exploit a user’s positive and negative preferences to produce a profile for the user using a Ranking SVM (RSVM).

3. Proposed methods are based on users’ concept preferences. Users consider some concepts to be more relevant than others.

4. Proposed methods use an RSVM (Rank support vector Model) to learn from concept preferences weighted concept vectors representing concept-based user profiles. The weights of the vector elements, which could be positive or negative, represent the interestingness of the user on the concepts.
5. Evaluate the proposed user profiling strategies and compare it with a baseline proposed. Show that profiles which captures both the user’s positive and negative preferences perform best among all of the proposed methods also find that the query clusters obtained from methods are very close to the optimal clusters.

### 3.1 Proposed Algorithm

**Query Clustering Algorithm**

The following cosine similarity function is used to compute the similarity score of a pair of query nodes:

**Input:** A Query-Concept Bipartite Graph G  
**Output:** A Personalized Clustered Query-Concept Bipartite Graph Gp.

// Initial Clustering

1. Obtain the similarity scores in G for all possible pairs of query nodes using above Equation.
2. Merge the pair of most similar query nodes ($q_i, q_j$) that does not contain the same query from different users. Assume that a concept node $c$ is connected to both query nodes $q_i$ and $q_j$ with weight $w_i$ and $w_j$, a new link is created between $c$ and $(q_i; q_j)$ with weight $w = w_i + w_j$.
3. Obtain the similarity scores in G for all possible pairs of concept nodes using above Equation.
4. Merge the pair of concept nodes $(c_i, c_j)$ having highest similarity score. Assume that a query node $q$ is connected to both concept nodes $c_i$ and $c_j$ with weight $w_i$ and $w_j$, a new link is created between $q$ and $(c_i; c_j)$ with weight $w = w_i + w_j$.
5. Unless termination is reached, repeat Steps 1-4.

// Community Merging

6. Obtain the similarity scores in G for all possible pairs of query nodes using above Equation.
7. Merge the pair of most similar query nodes ($q_i, q_j$) that contains the same query from different users. Assume that a concept node $c$ is connected to both query nodes $q_i$ and $q_j$ with weight $w_i$ and $w_j$, a new link is created between $c$ and $(q_i; q_j)$ with weight $w = w_i + w_j$.
8. Unless termination is reached, repeat Steps 6-7.

### 4. Performance Evaluation

The performance evolution of this project is satisfying user profiles and minimizes the system resources in an efficient manner.

The above graph represents the accuracy results of personalized clustering using normal click based method and P Click Joachim’s method using the precision and recall value. The result graph compares the impact of performance to evaluate the effectiveness of the key components of personalization: attribute relations, user profile and the ontology creation. Click based method was compared against P Click Joachim’s methods, where some of the key components were utilized. Evaluating personalization that end has the most significant impact for all graphs; User Profile at the end of iteration had a significant impact only for complete graphs, confirming our findings from experiments discussed before. Finally in addition to selecting subsets from community merging capable of providing enhanced Search engine performance, this personalized search engine also has a faster runtime than many comparisons click based methods. Hence the “Search Engine Personalization using Concept Based User Profiles” achieved optimal solution.

### 5. Conclusion & Future Enhancement

#### 5.1 Conclusion

The design of search engine personalized can greatly improve a search engine’s performance by identifying the information needs for individual users. The system is proposed and evaluated through several user profiling strategies. The techniques make use of clickthrough data to extract from Web-snippets to build concept-based user profiles automatically. Preference mining rules is applied to infer not only users’ positive preferences but also their negative preferences, and utilized both kinds of preferences in deriving user's profiles. The user profiling strategies were evaluated and compared with the personalized query clustering method that is proposed previously. Apart from improving the quality of the resulting clusters,
negative preferences in the proposed user profiles also help to separate similar and dissimilar queries into distant clusters, which help to determine near optimal terminating points for the clustering algorithm.

5.2 Future Enhancement

We plan to take on the following two directions for future work. First, relationships between users can be mined from the concept-based user profiles to perform collaborative filtering. This allows users with the same interests to share their profiles. Second, the existing user profiles can be used to predict the intent of unseen queries, such that when a user submits a new query, personalization can benefit the unseen query.

6. References