A Novel approach to personalize the web search

Chetan R. Rathod, A. J. Agrawal

Computer science & engineering, SRCOEM, Nagpur, Maharashtra, India
Email: Chetan.rathod777@gmail.com, avinashjagrwal@gmail.com

Abstract - Now a day, internet plays an important role in every individual's life. Information on the web is increasing at enormous speed. With growing amount of information user find it difficult to obtain the most relevant information that is hidden in unwanted information returned by the search engine. Every user has specific goal and different background while searching on the web. Current search engine produced best results to the given query, but unaware with individual's interests. Many approaches have been implemented to personalize the search such as feedback from the different user, collaborative filtering, and concept based profiling, etc. The proposed approach which uses user’s personal information like their background, interest, past experience etc. Proposed framework re-ranked the results obtained from existing search engine. Finally, these results are improved using user's personalized information.

I. INTRODUCTION

Over recent years, the World Wide Web has become a new communication medium thorough with Web information access. Users spend most of their time on internet which gives those lots of social, cultural, sports information to them. With the existence of various Search Engines like Google, Yahoo and many more, the users are tend to use them for retrieving their desired information from Web pages. But existing search engine cannot provide results according to individual user’s request. If user (programmer) search for “virus”, search engine retrieves result related to both computer virus and medical term virus. At the same time if other user who is interested to know about computer virus he/she would also get both the result on search term “virus”. Web users normally issue keyword queries to search engines to fetch relevant information on specific topics, as users may have different backgrounds and expectations for a given query; some search engines try to personalize their results to better match the overall interests and preferences of an individual user. In most of existing search engine, search result is retrieved by evaluating relative importance of links. Ranking algorithm calculate rank of links by their in edge and out edge links. Higher the in edges or out edges, higher the rank is assigned to link.

Many researches have been done to personalize web search result [6]-[13] that incorporate user profile by analyzing previous search history of specific user. But still problem of current personalized search approach is that the major approaches are uniformly applied to all users and queries [14] and there are not fit to all. It argued that queries should not be handled in the same general manner and expectation of user will be changed over a time on same query. It has been shown in [2] for some navigational queries like Google, Yahoo, YouTube, etc. there is no need to personalize web search results because users normally use uniform links from retrieved results. But for informational queries like “opera”, it is not quite easy to identify exact intention of user. Various personalization strategies which include [13]-[17] have been proposed that analyze individual user’s context and expand ambiguous query to best match with user’s context.

The rest of the paper structured as follows Second section describes the background and related work, third section describes our proposed work and final section describes conclusion.

II. RELATED WORK

In this section, some existing work done related to personalization of web search are presented like personalization using concept based profiling, collaborative filtering, etc.

In [1] T. Bibi proposed hybrid approach based on document based as well as concept based profiling. In this paper authors only re-rank the results for the given query obtained from existing search engine. The proposed approach not only re-ranks the results but also minimizes the results and also access the user profile to change their interest.

In [2] D. POO proposed approach based on the concept namely Static content profiling, static collaborative profiling, dynamic content profiling and dynamic collaborative profiling. Authors implemented this approach for domain specific search only. The proposed approach adopts some concepts from this paper.

System gets this data from search engines logs. As only click through data are considered knowledge based personalization gets restricted.

White et al [3] proposed method which is based on predefined taxonomies like ODP [4].

In INVAID [5] system gets explicit user feedback in the form of ratings and suggest pages of interest to the user based on the feedback. In the real world, it find difficult that all user give their ratings to the documents sincerely.

A user profile is normally structured as a concept/topic hierarchy [8], [9], [14], [15], [16], [18]. User issued queries and user-selected snippets/documents are categorized into concept hierarchies that are used to generate a user profile. When the user fires a query, each of the returned snippets/documents is also classified. The documents are re-ranked based upon how well the document categories match user interest profiles. The architecture of personalized web search algorithm based on content analysis that uses ontology to generate user profiles.

Query expansion aims at deriving a better formulation of the user query in order to enhance retrieval. It is founded on exploiting various social or collection specific characteristics in order to generate additional terms, which are added to the original input keywords before identifying the matching documents returned as output. Query expansion can be performed through three techniques mentioned in [19]. Automatic query expansion can be achieved by Relevance Feedback Techniques. The primary idea of Relevance Feedback (RF) is that useful information can be pulled out from the relevant documents returned for the initial query in the sense that the user was the one taking the relevant results, and then several methods were used to take out new terms, related to the query and the selected documents.

Another technique is Co-occurrence Based Techniques [20]; terms highly co-occurring with the issued keywords have been demonstrated to increase precision when appended to the question. Many statistical measures have been developed to best assess “term relationship” levels, either analysing whole documents or lexical affinity relationships.

Thesaurus [21] Based Techniques is broadly explored method to extend the user query with new conditions, whose substance is nearly connected to the input keywords. Such relationships are commonly distilled from large scale thesauri, as Word Net, in which several sets of synonyms, hyponyms, and so forth are predefined. But as for the co-occurrence methods, initial experiments with this approach were controversial, either reporting improvements, or even decreases in production quality.

III. PROPOSED APPROACH

By using literature review new approach proposed to personalize the web search. The proposed approach divided into two phase first is registration phase and in second phase minimization and optimization of the results taken place which is obtained from existing search engines.

A. Registration Phase

In the proposed system, user wants to register himself with the system to capture their information required for static content profiling. The user will have to fill the important information that will help the system to understand the user and provide appropriate results of user’s interest. The system uses model called W3 model [2] (fig. 1) to capture user’s vital information. It is generally used for static content profiling.

![W3 Model](Fig. 1. W3 Model)

W3 model contains three questions 1) “who is the user” 2) “what are the user’s objectives” 3) “where is the user”. Three questions used to determine information that will be include in the user’s profile. In each step, set of features $W_i = \{f_1, f_2, f_3, \ldots, f_n\}$ are listed. The final registration $R$ will be the union of all features taken from each step, $R = W_1 \cup W_2 \cup W_3$. Users have to fill their information that we have derived from this model. Each profile $P$ is stands for set of feature-value pairs. Profile begins with empty set but more and more interaction of user with system will add the new features to the profile. For the new user, profile is static. But after some interaction, system will automatically add feature to the profile using previous searches, user searching behavior, cookies of the browser. For this the system will modified by using concept called dynamic content profiling.

![Architecture of Proposed Approach](Fig. 2. Architecture of Proposed Approach)
First new users register himself with the system, then login with username and password provided in the registration phase. All information of the user will store in the database. User fires a query, before pass to the search engine; query will automatically expand by the system using user’s information. System gives suggestion to user to append relevant term with original query. System use thesaurus to give suggestion to the user. Reformulated query will pass to the search engine; search engine crawled and brings all the results to the system. We use Google API to crawl the results. Most of the documents are unwanted or similar in the result, so we find the similarity between the documents using Jaccard’s function. If the content of the one document is similar as the content of the other document then the system discard one document from the results, this process continue till we get all relevant document. Finally we match the user interested keywords with the ranked results. Document that contains most of the user interested keyword will give first rank in the final results.

IV. CONCLUSION

In this paper, first the systems reformulate the query and append relevant keyword to the original query that will help user to search with an appropriate keyword. The proposed approach will also provide related search that better identifies the current interest of the user. Finally system displayed most relevant link at the top of the retrieved results. By observing most of the work related to the personalization of web search, it has been conclude that this system will help user to save precious time and increase precision. During the research some observations made which causes future directions to personalized web search. The three important such observations are non-text information may also be used in this framework, personalization of web search may be done without implicit relevance feedback or current system may able to identify the semantic relationship between user’s pervious search and current search.

ACKNOLEDGEMENT

The authors are grateful to the Head, Computer Science and Engineering Department and the Principal, Shri Ramdeobaba College of Engineering and Management for providing adequate facilities to conduct a research. Authors are thankful to the faculty members of Computer Science and Engineering Department for their support and cooperation during this work.

REFERENCE


[4]. ODP. http://www.dmoz.org/


[21]. “An Introduction to Information Retrieval”, by Christopher D. Manning, Pg. no.189-193.

◆◆◆