# Personalised Web Search using User Profile Incorporation

**Sayali Phadke***                                    **Prof. Mr. C. O. Banchhor**
IT Department,SCOE                                    IT Department,SCOE
University of Pune                                    University of Pune
India                                                India

*Abstract— Search has become important aspect of day to day life. Search engines are the best way to get what user wants. But the necessity is to understand what user actually wants to search. Personalised search is best way to resolve ambiguity of the query terms given by the user. By considering three major parameters that will help to give the precise results to the user .User Profiling plays important role here, in the proposed idea we are considering user profile which contains details like interests, hobby's and educational background data. We will be also storing previous queries of the user as well.There are many other parameters taken into consideration while searching data on web for a given query. Hence we have to take into consideration user's profile and pattern for the search. As user enters the query the query will be re-formulated using re-formulated using the profile of the respected user. The context of user must be taken into consideration. This makes the search relevant to the user's requirements. Also the previously searched results will be taken into consideration for next probable search. This intelligence of search engine will help user to navigate towards the relevant or more precise search. We are providing easy way to search data on web by understanding user's context to provide more allied as well as personalized search. The result shows improvement in helping user's to find what they are looking for. More facilities like backup and restore are provided for enhancing searching experience. Relate download is the facility where user can find the related documents on the local machine. Also Backup and Restore facility is provided for facilitating fast search and a complete new search.*

*Keywords— Context, Use Profile, Personalized Search, Probable Search, Re-formulation*

## I. INTRODUCTION

Web search means searching intended data on web. When user submits the query to search engine he/she has some predefined goal or say expected output links in mind, but does search engine really provides the same? For most of the queries answer is 'yes' but for ambiguous queries it may fail. To avoid this we need to understand context of the search. Most of the time it is for information need. Search is an end to end process, it does not start randomly. It involves deciding goal to search till reaching the same. When user issues the query, the results can lead to much different kind of answers. For example, if the query is language then results can be foreign languages, computer languages and much more. Hence we have to specify explicitly the domain/context in which user is searching. Generally search engines consider some factor for matching the relevant documents. As term frequency, keyword position, inverse term frequency, term weighting etc. These all parameters are based upon the count of the given keyword. It may be the case where keyword appears many time but the document is not what user has intended for. So one has to consider exactly in what context user is searching. Here in this paper I would like to propose an idea of taking user profile into consideration in which domain of user can give idea about in what context user will search .User Profile will give some reflection of the domain/context of the user search. Here the concept is to reduce query refinements. As we know what will be the user's probable search , the refinements will be added by using some intelligence at back end. Also facilities like relate downloads, backup data , restore data are provided as to help user while searching and reduce the time and efforts needed for the search.

## II. RELATED WORK

Studies related to web searching have a long history. In one of the paper Abdelkrim Bouramoul[8] et. Al. have developed a tool for query re-formulation by using context based assistance and without assistance as well, it also gives three options for searching the data on web using three different api's.The search starts from user, specifying the query then search engine has some predefined factors to search it on the web. Jeff Huang [9] et. Al. have discussed reformulation of the query using query log analysis They have used reule based classifier using click through analysis. Eytar Adan [3] et al. has defined why we search that is every time when we search we have some pre-defined goal in mind, which can be used for visualizing and predicting the user behaviour. Generally when user searches data we can predict what can be the next search by seeing a browsing history of the user. Long term history can also be used to search data which can be most relevant to the user's interest. Personalised search has gain a lot of popularity in last few years.

An important aspect of personalised search is taking user's need into consideration while searching the data on the web. Sachin Kumar [7] et al. has shown the correct use of profile of the user into search, in this paper they have discussed techniques like re-ranking, indexing, clustering to provide most relevant links as a top results to the user. They have mentioned cluster formats that will ask user for needed format. Jin Young Kim et al. [5] have given an innovative way by which we can manage user profile based search by characterizing user interests, web contents and reading level of the user for the particular topic. Here in this paper they are considering RTL profile namely probability distribution of reading level and topic. RTL profile can be calculated by considering user clicking sequence, web contents, and users browsing history.

Only those links will be shown to the expertise user which contains some advanced data regarding the topic. As per this paper, query, website and user are always related. As user issues the query then visits the website. Hence there will be always a correlation between keywords submitted to the URL of the website. One more interesting approach called as disambiguation algorithm is given by Dmitri Kalashnikov et al. [1] which are specifically for people search. Here they are using extraction mechanism. To find relevancy of the specific page towards the user profile, extraction mechanism is used. It extracts name, keywords, hyperlinks, location, and organizations etc. to find how much the web content satisfies user's need. Daniel Rose and Danny Levinson [2] states that how navigational searches are less prevalent and how knowledge of user search can improve further user search. It also discusses framework for understanding the underlying goals of user search. Eugene et al. [4] have said that incorporating user behaviour data can significantly improve ordering of top results in real web searching. Implicit feedback model with independent results as well as incorporating feedback into the model has been discussed in this paper.

### III.     METHOD OF PERSONALIZATION

Personalization is very broad term as we think from the perspective of the web search. Personalised result is our main aim. As user searches data on web query is given to search engine, it will return the results on basis of some parameters. We are using search engine application interfaces. So once query is given on the term frequency weighting results are returned. As we are providing personalising we have context of the user from the profile of the user. Then the results can be re-ranked according to the context priority of the user. Or else the approach that we are adopting is why not to re-formulate the query before sending it to a search engine rather than re-ranking the results at the end. This can be achieved by putting little intelligence .As we are dealing with the user profile, which contains some basic information of the user and importantly domain of user and his /her area of interest. The background of the user can be considered as context in which user is searching. Hence we can reduce the work of mentioning area of search or say context by automatically considering it in user profile.

The profile of user will play an important role in search. We are also providing facility of filtration of the results. This can be provided for diverting user towards correct steps to reach the specific goal. As user has some predefined goal in the mind which will be seen from his profile, we can help user to follow the links which will lead to correct goal. Generally what happens when searching data is diversion from main intension of search, the reason can be some misleading words or difficulty in understanding words of the current document. Following facilities are provided to the user in this search tool.

### A.  *Previous query storage*

We are providing some enhanced facilities to the user. When user submits the query, we will be storing initial query given by the user. As well as we are maintaining count of the query submission. That is how many times user is submitting the same query. The profile will also keep the track of what user has already searched in previous search instance. So that when user enters the query somewhat related to the previous search instance we will be adding patch words from the previous search that may help user reaching to intended search. We are keeping track of following parameters to provide clearer search.

- Most frequently searched queries
  We are keeping track of frequently searched queries and the reformed queries for the same query, so that user can continue with the same search.
- Previously entered queries with refinements
  As user will enter the query and goes on refining the same to reach the intended search .So we are keeping track of the user refinements.

### B.  *Retrieving the results*

When user enters the query then we will re-form the query with the help of user profile .After this re-formed query will be sent to search engine like google to get results. This re-formed query answer will be those links which user has traversed in last search instance. Now if user changes the query that is refinement of the query automatically we will be getting that whether user wants to search on same path or search path becomes allied one. Many times it has been observed that results returned by any search engines do contain some links at bottom though they are very similar to what user wants. So in the process of re-forming we will be checking user's interest and new query will be sent to search engine. If user search is for the first time when user enters the query then we are checking the background of the user and then according to that we will be re-forming the query and then sending it to search engine. This re-forming of the query will be done by using the query re-forming algorithm discussed in next section.

Past searches means what user has searched in past few actions can reflect what he/she may search this time that is probable prediction. These links will be indexed according to their importance to the user current search.

### C. Query

As discussed in the example if search query is language results can be related to computer languages or natural languages. A user with computer science background will be only interested in languages C, C++ etc. Hence we it be computer only. If user is of computer science background he can check computer languages or else can may also go for natural languages. If user refines the query again to find natural languages then that part will be reflected in user profile. Also if user wishes to see more options for the refinements, he/she can add up interested domains.

### D. Backup and Restore

Another very important facility is backup and restore. Here as we are maintaining history and the areas of user search we are providing auto re-formulation of query. User may wish to start a new search and don't want previous instance of the queries to affect the current search.

So facility of back up is provided where all the previously entered data will be back up in excel and stored. So that user can start with fresh mind. If user wants previous data to continue search reverse facility that is restore is given. Restore facility allows user to set back the previous data and start from this data for search instance.

What happens generally when searching for the record is, when user enters the query it is dependent on what intension user has in mind .When entered query as 'rose' search may go for the flower 'rose' whereas user is from computer science background expecting 'rational rose' as a software. Now this can be very well captured from the user profile .As we can understand background of the user.

### E. Relate Download

When user searches data on web, he/she may download the documents. So the facility provided is finding out related downloads for the current query. When user enters the query all the downloaded documents will be shown to the user .Using this facility will reduce user's time to search for those documents on local machine.

## IV. ARCHITECTURE

The following diagram depicts the architecture of the intelligent search engine.
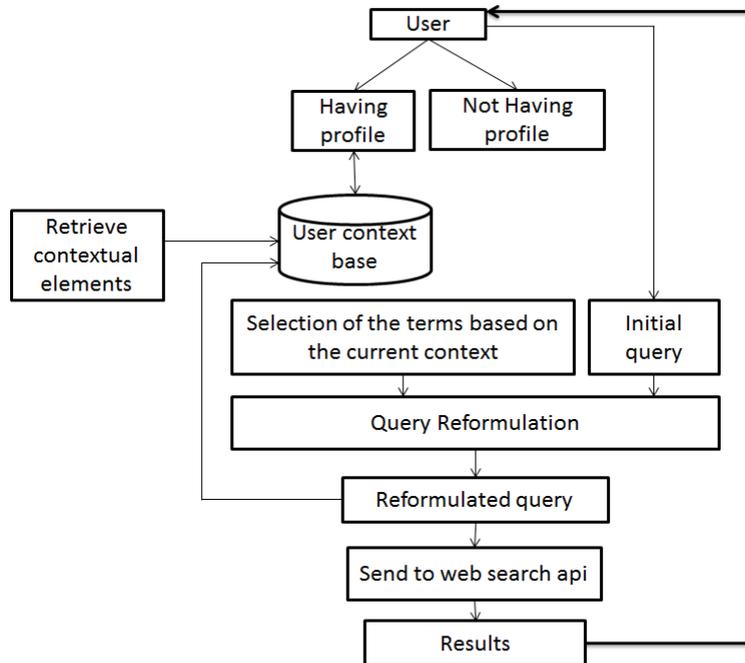


Figure 1: Flow of search engine

Here TriSearch is the search engine name which is actually doing all the tasks. When user submits the query, trisearch engine fetches the user profile from the database. User profile is entered by user at the time of registration If user does not entered the profile then general search that is not based on profile of the user. Then by using the profile and pre-requisites we will be reforming the query using query re-forming algorithm. Trisearch server actually brings all the links from the generalized search engine like google .Then at trisearch engine we are keeping track of query refinement. The flow goes like this, first user logins to the search engine. User creates the profile which contains basic information and domain of the user. Then user submits the query to search engine. Then by using query reforming algorithm, we add patch words then it is sent to search engine. Search engine brings all the relevant links from the web. Those links will be delivered to trisearch engine. Based on user profile query will be re-formed again. If user has computer science background then all those links will be at the top. It depends on what user interested in.

Also when user gets the final output we will be providing information needed to understand the document .The probable use of the same document in the future. This will be shown by considering application area of the user. By providing all these facilities we are providing more intended results to the user based on previously known information about the user.

## V.    ALGORITHM

As discussed in previous section, we will be re-forming the query before sending it to search engine. Hence we will be using new algorithm called as query re-forming algorithm.

### THE QUERY RE-FORMING ALGORITHM

1.  Take initial query 'q' from the user.
2.  Fetch the previous entered queries of the user from the profile
3.  If previously entered query matched then
4.  Compare the previously traversed instance and add some patch words to reach the target. (Patch words will be considered from recent instance of query search) And refine the query.
5.  Else
6.  Check the user profile and reform the query by adding some parameters from the user profile.
7.  Send the query to search engine
8.  If user refines the query repeat from 3
9.  End

Also we in the algorithm for re-forming the query we are making use of search operators to make it more relevant to user's need.

Step 1 takes input query from the user. We have user profile with us from which we will fetch previous search instance query and check whether it is somewhat related, if yes then add patch words to give the user intended result. If this is first search then check user profile and refine the query accordingly. If user changes the query refine or change data in accordingly. Here we are using concept of re-formulation of the query and avoid time for re-ranking results. If we provide re-formed query then results will be more relevant. Rather than re-ranking results after receiving from search engine initially re-form query by using user profile and algorithm given above.
.

## VI.    PERFORMANCE EVALUATION

The relevance of the query with the user interest is to be matched and it will be done by taking help of above algorithm as well as word matching technique .As user profile words matches with the words in web page that page will be considered more relevant to the user's interest for search. Generally when user enters the query it can be ambiguous that means it cannot exactly express user's intention for search. So we have to consider the relevance according to what user profile is.
For performance measure following criteria's are taken into consideration.

- Precision: Precision is also called positive predictive value is the fraction of retrieved instances that are relevant to the search. Precision takes all retrieved pages into account, but it can also be evaluated at a given cut off rank, considering only the topmost results returned by the system.
- Recall:  Recall is also known as sensitivity. Recall in information retrieval is the fraction of the pages that are relevant to the query that are successfully retrieved. Here we are considering two criteria CR1 and CR2 as follows:
  Criteria 1(CR1): Relevance of the first three results to the user.
  Criteria 2(CR2): Redundant results that are results from the same site.

Table 1.1 Performance Evaluation

| Criteria | Without Reformulation | With Reformulation |
|----------|----------------------|---------------------|
| CR1 | 5.55 | 6.29 |
| CR2 | 6.66 | 5.65 |

## VII.    CONCLUSION

Hence we are implementing a search engine tool approach that will provide some kind of advanced search by using reformulation algorithm. We are minimizing user's efforts to explicitly mention context of the search. User Profiling is used to give more personalized results to the user. As shown in the experiment section using re-formation using profile of the user increases relevancy measure of the results .As reformation of the query reduces user's efforts to add refinements to the existing query for the web search.

## REFERENCES

[1]    D. Kalashnikov, S. Mehrotra, Z.Chen, R. Nuray-Turan, N. Ashish, "Disambiguation Algorithm for People Search on the Web", ICDE, 1258-1260, Apr. 2007.
[2]    D. Rose , D. Lavinson, "Understanding User Goals in Web Search", Proceedings of world wide web, 13-19, 2004

Phadke et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(8), August - 2014, pp. 172-176

[3]    E. Adan, D. Weld, B. Bershad, S. Gribble, "Why We Search: Visualizing and Predicting User Behavior", WWW'07 Proceedings of 16th international World Wide Web conference, 161-170, May. 2007.

[4]    E. Agichtein, E. Brill, S.Dumais, "Improving Web Search Ranking by Incorporating User Behaviour Information", SIGIR'06 Proceedings of 29th annual international ACM SIGIR conference on Research and development in information retrieval, 19-26, New York ,NY,USA,ACM Press 2006 ,Aug 2006

[5]    J. Kim, K. Thompson-Collins, P. Bennett,S. Dumais, "Characterizing Web Content ,User Interests and Search Behavior by Reading Level and Topic", WSDM, 213-222 ,Feb 2012.

[6]    K. Sugiyama, K. Hatano, M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Efforts from the User.",WWW'04 Proceedings of 13th international conference on world wide web,  675-684, New York, NY, USA, ACM Press, 2004

[7]    S. Kumar, L. Ertoz, S. Singhal, B. Uygar, E. Han, V. Kumar , "Personalised Profile Based Search Interface With Ranked and Clustered Display", Technical Report

[8]    Abdelkrim Bouramoul, Mohamed-Khireddine Kholladi and Bich-Lien Doan," PRESY: A Context Based Query Reformulation Tool for Information Retrieval on the Web", Journal of Computer Science ,2010

[9]    Jeff Huang, Efthimis N. Efthimiadis," Analyzing and Evaluating Query Reformulation Strategies in Web Search log", CIKM'09, November 2–6, 2009