A Survey on Incorporating User Behaviour to Enhance the Web Search

S. Phadke [1], C. Banchhor [2]

1Research Scholar, Sinhgad College of Engineering, University of Pune, Maharashtra, India.
2 Professor, Sinhgad College of Engineering, Pune, India.

Abstract

Now-a-days millions of users search data on web daily. It has become an integrated part of everyone's life. Generally when we search data we have some intension to search. When we enter a query some links will be given by a search engine. But what about relevancy of those links to the user's actual need. Hence by incorporating user profile we can improve the searching process. By providing more satisfactory results to the user we will try to match results to the user's need. Many search engines are taking user profile into consideration, but some of them use user's interest hierarchy, many of them uses past history and so on. Re-ranking techniques are used to match links to the user's interest. Indexing and clustering is also used by many search engines to improve relevancy of the result.

1. Introduction

Search engine is one which when submitted a query searches data on web and returns the results to the user. Some of the popular as well as successful search engines uses simple graphical user interface where user enters the query which will be then undergoing process of removal of the words like and, or etc. and then will be submitted to the engine , it will crawl the web to fetch the matching links which will be given to the user. But some of the search engine takes user profile into consideration. User profile contains information about the user. It can be built by previous history of the user or user himself will enter all the details.

1. User Profile

User profile will be created by seeing what user has searched previously. That is query fired, links clicked and so on. It will be dependent on browsing history of the user. When user searches data he/she will follow some links and from those links we can extract the keywords which will be later considered as user's profile. This kind of mechanism can be called as feedback mechanism which is used to store the user's profile.

Secondly if such mechanism is not used then user only creates the profile where some personal details as well as details related to context of the user are taken. Context means user's area of interest which will be later used as a context to search data on the web.

2. Context to search

When the search considers the user's profile, one have to consider in what result user is interested. If user has some specific background, the interest will vary with the domain of the user. For instance, user searches for the languages then whether user is interested in computer languages or in natural languages can be understood from the user profile. If user belongs to computer science domain [2], he will be interested in C, C++ etc. Whereas if user is from arts background he will be interested in natural languages. Hence we need to re-rank the results that are returned by normal search engine. In first case, links that are having reference to C , C++ will be at the top.

2. Related Work

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. The profile can be generated by taking details from the user explicitly or by maintaining user's browsing history profile can be generated.

Eytar Adan [1] et al. has defined why we search that is intension for the search by 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. Sachin Kumar [2] 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. Personalised search can be

improved by using ranked and cluster display interface. They are providing filtration at each and every layer of search. As an example when searched for the language, depending on the user profile computer languages or natural languages will be shown. Even as an enhancement, if user goes for the computer language, then the filtered result option box will be displayed which contains some of the languages names that user has searched in the history. So user will be also navigated towards clearer goal. If user wants to see the whole cluster of the computer languages, it can also be viewed. The clusters will have the meaningful names which will depict what is inside the cluster. As an enhancement, if user is interested in combination of two or more clusters he/she can filter the results by clicking both the check boxes. This is how the profile can be incorporated by using re-ranking and clustering.

Jin Young Kim et al. [3] 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. Reading level can be predicted as expertise user and non-expertise user. So web content will vary as per level of the user. 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 co-relation between keywords submitted to the URL of the website.

One more interesting approach called as disambiguation algorithm is given by Dmitri Kalashnikov et al. [4] 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. Also as an enhancement they are using ontology based TF/IDF technique which will be applied to drive the concepts as example machine learning is grouped under machine_learning. Eugene Agichtein et al. [5] have discussed implicit feedback model in their paper which shows the importance of incorporating user behaviour information into web search. It says that implicit feedback model is particularly valuable for the queries with poor original ranking. Daniel Rose et al. [6] 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.

1. Techniques used in profile based searching
There are many different techniques used in profile based searching. Some of them are dependent on browsing history and some of them only consider the current profile of the user.

a. Based on user's browsing history
As discussed earlier searching can be personalised by considering what path user has followed while searching for the data. There are two types of methodologies first one is link personalisation and second is content personalisation. [7] Here in link personalisation, two different browsing histories are maintained that are persistent for long term preferences and ephemeral for short term preferences. In ephemeral method information related to user is gathered only from the current session whereas in persistent it is gathered from many other sources and it relies on different aspects of the user.

In content based personalisation, collaborative filtering is used based on user recommendation. It works as per user profile that means first links are filtered according to topic filter and then by user profile personal filter. Generally in this case, methods for user search prediction are used. Clustering, Rule based models, Bayesian networks are some of the techniques that are incorporated.

b. Based on user profile
Some of the searches are only based on current interest of the user. It is implemented by using clustered results. When user searches for a word which may belong to two different domains then the clusters of those domain containing the same word are shown to the user, from which user can select the needed domain. Filtration techniques are also employed here as user searches for a word, the more specific terms related to the same searches are shown to the user where user can give more context related search query.

c. Using search goal hierarchy
The concept of search goal hierarchy [6] says that what user knows while searching that means does user knows the URL, or only the keywords. If user knows what he/she wants but does not know the URL then the type of search will be navigational. Informational searches include closed search where user knows that what is to be searched. In undirected searches user wants to find answer for

some question. Undirected search means user has no knowledge about a topic but wants some information regarding the same. In resource based search user has a specific intension. The intension may be downloading, obtain information or interact with other resource or entertainment which may be only viewing the pages. This kind of search hierarchy will help hep to understand what user is searching, why user is searching that is intension behind the search and context of the search.

Hence by considering all the searching techniques which takes user profile into consideration, we can have more user oriented search rather than only showing many matching pages to the user. As search engine is now going to do maximum task of user, load on the user will be reduced. This is very effective way of searching that will reduce time as well as efforts needed for the searching on web. Domain consideration will reduce task of user as no need to mention the context of the search. Hence user can only submit the query without mentioning much about the context of the search.

As a summary following are the features or advantages of the search engine where user profile is incorporated,

- More focus on why user is searching rather than what user is searching. More consideration on context rather than topic.
- Less information is needed from the user at the time of search.
- More allied and efficient search can be achieved.
- In less time we can achieve more relevant data.
- User can be navigated through correct goal as search engine knows the context of the search.
- Users are just need to mention more of their context and less of the information about the topic.
- Search goal hierarchy will be very effective to understand what user will do next. That is the probable prediction of user behaviour.

3. Future Work

We can improve the normal traditional search by incorporating user profile. As an improvement what we can do is we can add filtration mechanism which will be very efficient. That is when user searches the data the URL's clicked and the path followed by the user will be stored as a reference to next search. Hence when user again searches the data we can use previous searched links as a guidance to provide relevant results.

As an addition we can provide the suggested sequence to the user by seeing his/her past browsing history and current profile. The main task will be navigating the user towards correct sequence of links which will make user to reach a goal in less steps yet with more relevancy to his/her interest. Also one innovative aspect will be by considering user profile automatically deciding difficulty of the query and only use implicit feedback for those difficult queries.

4. Conclusion

Hence we have seen many techniques by which we can improve the searching process. User profile plays an important role in optimizing search by incorporating user's interest, reading level and web content. As we are considering user's interest or domain the calculated results will be more relevant to what user needs. Hence we can enhance the techniques like re-ranking, indexing and clustering to improve the relevancy of results to match the expected results.

5. Acknowledgement

6. References

[1] 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, pp. 161-170, May. 2007.

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

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

[4] D. Kalashnikov, S. Mehrotra, Z.Chen, R. Nuray-Turan, N. Ashish, "Disambiguation Algorithm for People Search on the Web", *ICDE,* pp. 1258-1260, Apr. 2007.

[5] 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, pp. 19-26, New York ,NY,USA,ACM Press 2006 ,Aug 2006

[6] D. Rose , D. Lavinson, "Understanding User Goals in Web Search", Proceedings of world wide web, pp. 13-19, 2004

[7] 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, pp. 675-684, New York, NY, USA, ACM Press, 2004