IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 8, Aug 2018
www.ijasrm.com
ISSN 2455-6378

A Study to Web Search Personalization: Factors and its Optimization

Nidhi Saxena[1], Bineet Gupta[2]

[1] Information Technology, SRMGPC,
Lucknow, UP, India

[2] Computer Applications, SRMU,
Lucknow, UP, India

Abstract

Personalization is an attempt to construct a system that can provide search results based on user interest. It needs to be associated with searching in order to signify the importance and relevance of search results to a user. This paper is a study to different factors of personalization. It deals with the identification of problems associated with personalization approaches, challenges faced by them and necessity to build more efficient system which can overcome the problems to certain extent. A deep study of personalization is performed in order to find the necessity of an algorithm which can take positive approach of current methods of ranking along with some enhancement to provide user with more accurate and relevant results. The paper also provides a brief introduction of a new model which can act as an optimization technique and can enhanced personalization in order to satisfy user and to reduce complexity.

Keywords: *Personalization, implicit, explicit, Page ranking.*

1. Introduction

Personalization comprises of technology, which discriminate among individuals. Personalization considers user interest, browser history, page visited records for ranking the search results. Sometimes search depends on user feedback. In both cases profile is created which can be explicitly and implicitly.

- Explicit profile - In explicit profile creation user registration form need to be filled by user in order to specify priorities and concern. Users can modify the records by themselves.

- Implicit profile- Implicit personalization is done on the basis of user behaviour. Profiles are created on the basis of user search history. On the basis of earlier searched query and records in web log, user interest is acquired and search results are displayed.

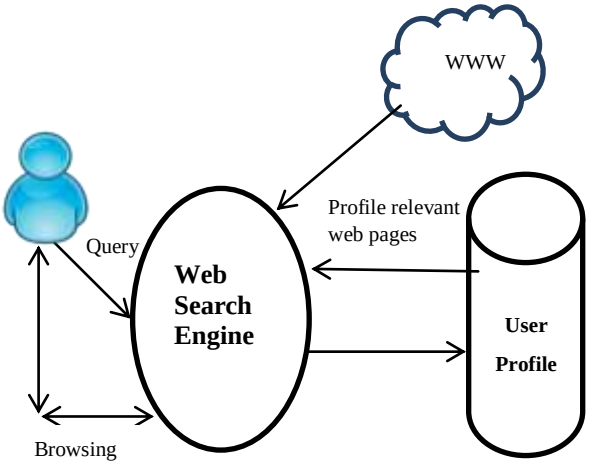

Fig. 1 Personalized Web Search

The given block diagram illustrate the working of Personalized search Engine. User search is recorded in the User profile, which further personalizes the search .User Profile either created implicitly or explicitly acts as a basis of most of searching and ranking methods.

Web personalization has improved the web search system to a great extent. Yet there exist certain challenges which need to be tackle. It is a demand of current scenario to handle each user more personally to provide accurate and relevant result with least effort. Therefore by considering all the factors associated with personalization a new model is introduced which can reduce many problems associated with personalization and can develop a secure, efficient and user friendly search engine.

2. Related Work

Previous work on personalization includes both explicit and implicit profiling techniques. Dynamic ranking (Brandt C, Joachims T, Yue Y & Bank J, 2011) introduces two algorithms which takes feedback from user and therefore involve user in order to find

the user interest. Query extension and taking social means for personalization method (Zhou D, Lawless S & Wade V, 2012) uses user profile generated by taking the tags and the web documents. A new form of probabilistic profile RLT profile (Kim JY, Thompson KC, Bennett P N and Dumais S T, 2012) was introduced which can be used to define key entries of web search. Work is carried on developing user profile based on semantic extraction from news article. It can be applied to Social web system and has impact on personalization (Abel F, Gao Q, Houben G J & Tao K, 2011).

There are many personalization techniques which collect information from user explicitly and create user profile for further search optimization (Srinvas C, 2012). Some algorithms considers document present at user device for personalization (Bhadoria R S, Sain D & Moriwal R, 2011). It is assumed that if a user keeps a document on his system it means that user is interested in those documents. There is a research based on search history which investigate three conditions (Bennett N P, White RW , Chu W, Dumais ST ,Bailey P , Borisyuk F & Cui X, 2012) –

Session: All earlier work in current search session.
Historic: All earlier work apart from current session.
Aggregate: All earlier work before the current query.

Various other algorithms introduced for personalization. Some keep track of user search history (Speretta M & Gauch S, 2005). Work was carried to Identify user interest automatically or implicitly (Qiu F & Cho J, 2006), (Shen X, Tan B & Zhai,C, 2005). Weighted association rule was introduced in this period for personalization (Forsati R., Meybodi M R, Neiat A G, 2009).

Later on various algorithms developed which keep record of each page visited by the user and time spend on that page (Liu H and Keselj V, 2007), (Khanchan R, Punithavalli M, 2011), (Matthijs N, Radlinski F,2011). These algorithms are effective for web page prediction. They provide an efficient web path traversal for various users based on their path traversal activities. More optimized version of earlier algorithm was when click event on the each page by the user is also stored and used for efficient ranking (Agarwal R, Arya K V & Shekhar S, 2011). Click event reflects the interest shown by user on each page. The algorithm shows that as the number of parameter increases, the effectiveness also increases. A new ranking algorithm Ratio Rank (Singh R & Sharma D K, 2013) was introduced in 2013 in which in links weights and out link weights are used with the consideration of number of visit count which is a better approach for personalization. New Enhanced page rank algorithm (Singh R & Sharma D K, 2013) considers mouse track along with previous factors Here the relevancy of the webpages resumed is high as the user behavior is thoroughly considered to rank the web pages. Along with frequency, time spent and mouse click event are considered for personalization.

Some algorithms were considering both implicit and explicit creation of user profile based on previous search query result as well as some personal information like the documents and e-mail user read (Teevan J , Susan T D & Horvitz E, 2017).

Several other algorithms were introduced recently for the purpose of personalization like algorithm using Embedding technique (Vu T, Nguyen D Q ,Johnson M, , Song D & Willis A, 2017), Personalized Search Utilizing Folksonomy (Zhou D , Zhao X W ,Lawless S & Liu J, 2017), Measuring and Predicting Search Engine Users' Satisfaction (Dan O & Davison B D, 2016) .All the algorithms designed earlier keep the concept of web log to provide required results.

3. Personalization Factors

As per the above studied work it is very clear that personalization reduces the length of query therefore reduces the complexity and search time.

Broadly we can divide personalization into three primary areas:

- Behavioral (search/surfing history)
- Geo-localized
- Social

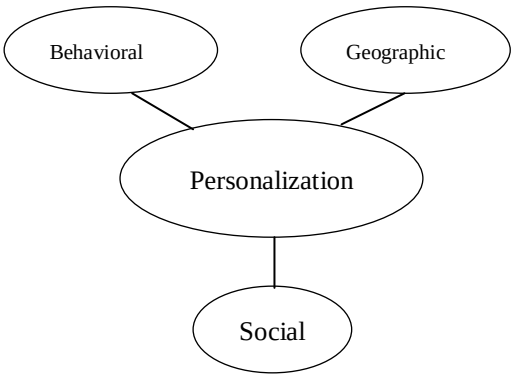

Fig. 2 Forms of Personalization

Personalization is usually provided based on following factors –

3.1 Country Specific

Under this personalization factor results are shown on the basis of Country to which IP address belongs so that results can be more specific and relevant. For

example someone in US searching for "football" will get results about American football. The Ranking of resultant web addresses will be according to place.

3.2 Locality Specific

Country specific personalization is further personalized on the basis of Locality or city for the ease of web search. For example if a person wants to buy some good he can easily get the results as per his choice with in his locality.

3.3 Personal History

This is the most important personalization factor and lot of work has been done on it as shown above. The basic elements for personal history are click event and time spends which provide clear picture of user interest in particular web site.

This personal history records are used in most of the shopping and social networking web sites.

3.4 Social Connections

This is one of the newer ranking factors to influence search results. Like what the others or friends thinks about a website. Based on their response the personal interest lies.

These connections act as an advisor for person. For example you are searching for a Restaurant. If your friend has already had a lunch there then your selection depends on his advice. All discussed factors can be broadly classified into-

- Traditional approach to Personalization
- New approach to personalization

Earlier human behavior was considered for keeping web log data. Search history and user web page behavior, surfing practices were the main key behind personalization. Sometimes web log contains the data being collected explicitly by user.

But new approach to personalization considers not only user behavior but other factors are also judged like communications with advertising, demographic and geographic data from different application like email and other.

4. Identification of problem-

Personalizing provides facility to retrieve the most relevant page for the query made by the user. All the methodologies used for personalization are considering user input for interest or create user search history based on user behavior while searching.

The explicit personalization method take user preference which needs user involvement but if we are using implicit personalization methodologies then for taking user preferences log history of previous searches are needed to be maintained on the system. The user is identified as a system and personalized search is according to the search made from the system. So it is still a challenge to provide search result to different users on single system with unique interest and to reduce the overall complexity of ranking to greater extent. Some major challenges in the area of Personalization are listed below-

- To reduce the number of inappropriate "hits" in Web searches. So it is a necessity to reduce the computational complexity of the overall ranking process which increases as search history increases on server or PC.
- To allow people, easily and precisely specify the information they want so as to reduce the inappropriate "hits" and to save time.
- To provides result based on lifestyle, demographic and other factors by associating a weblog record of individual with such data.
- To reduce the problems associated with improper searching like saving the link for later use or after getting required information went observing for more information. Leaving the browser in open state after searching is a general habit. Such type of carelessness leads to incorrect web history and results later.
- Top ranked pages are generally visited irrespective of their importance to user.
- If more than one person using same system for searching there can be difference in their interest. Each user should be able to define his/her area of interest in order to get personalized results.
- With the implementation of personalization there arises a problem of lack of privacy and security. It is difficult to have security and privacy along with personalization.

With the growing need of inclusion of new factors and to reduce the negative effect of personalization there is a necessity to develop an enhanced personalization based search system.

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 8, Aug 2018
www.ijasrm.com
ISSN 2455-6378

Therefore it is necessity to associate certain new factors with user profile while personalization in order to provide relevant result. It is necessary that a system should provide:

- Relevance Profile (having individual search results)
- Keyword Targeting/Phrase Strategies.
- Quality Content (with click event and time spend)
- Search Result Conversion
- Site Usability
- Analytics
- Security and Privacy

So, when looking user more personally, it is necessary to include these points as a bit of a check list. This checklist will allow understanding the requirement of user with more personal approach.

5. Optimized Web Search Model

A new approach to personalization is introduced which aims to identify each user and personalized their search uniquely. It is an extension to search approaches which are in use. It attempts to solve most of the problem with least effort and technology involvement.

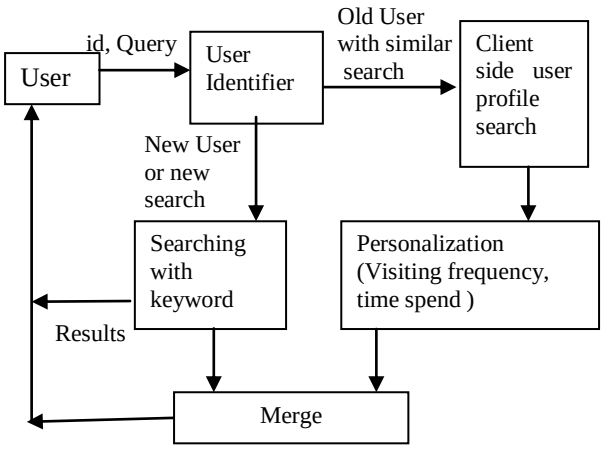

Fig. 3 Optimized Web Search Personalization Model

The given model combines the personalization based on user interest as discussed above with a new concept of keeping user identity and using it for searching and ranking.

The User is identified as New user and Old user and the keyword searched is also classified as New search and Old search. Separate Search log will be created for each user. If the user is new or the search is new then simple personalization approach will work which considers demographic detail before giving results but if the searched keyword is an old one by an Old user then result will be based on personalization factors like visiting frequency, time spent. The ranked web pages will be merged with other searched results in order to provide relevant and interested information to user.

6. Results and Evaluation

This approach is proposed keeping in mind the problem where single system is used by many users in home or office for accessing search engine. They may have different area of interest. Like if an old lady at home makes “RAM” as query, she will be expecting information about Lord RAM . On the other system same query “RAM” entered by a man from same machine. He requires data about Random Access Memory (RAM) but the result displayed will be the same if the users are not identified as different identity therefore identification before personalization provides user relevant results.

If we compare this approach with Google or any other search engine such unique identification is missing and therefore they provide personalized results based on machine not on person. The results will remain same for all users irrelevant of their interest. Such search engines require full query specifying the interest.

Certain merits of proposed model is listed below-

- Relevant results based on user interest
- It will allow people to easily and precisely specify the query.
- It will reduce the inappropriate “hits” and save time.
- Not much complicated internal algorithm required.
- It is easy to maintain privacy and security using this approach.

7. Conclusions

Personalization of web search is a requisite now-a-days to reveal user preferences in search results. In this paper, personalization with different approaches has been given. Various factors like Behavioral, Geographical and Social are examined. A detailed study of various factors which are prevalent now in search engines is provided in this paper. Besides examining the factors some of the drawbacks and missing area are depicted in challenges associated with personalization which lead to the necessity of search engine which can reduce the upcoming challenges to a great extent. As a solution to the problems associated with personalization a new technique is introduced which keep record of user

identity. If the identification factor if merged with other factors can provide more effecting results in an efficient manner. The proposed system will be useful in the area where more than one person is using same machine for searching but their area of interest differs.

References

[1] Abel F, Gao Q, Houben G J and Tao K, Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web .ESWC'11, Vol. Part II, Springer,2, Page No.- 375-389, (2011).

[2] Agarwal R, Arya K V and Shekhar S, An Efficient Weighted Algorithm for Web Information Retrieval System. In: IEEE International Conference on CICN, (2011).

[3] Bennett N P, White RW , Chu W, Dumais ST, Bailey P , Borisyuk F and Cui X , Modeling the Impact of Short- and Long-Term Behavior on Search Personalization. 35th Annual ACM SIGIR Conference, (2012).

[4] Bhadoria R S, Sain D and Moriwal R , Data Mining Algorithm for personalizing user's profile on Web, IJCTEE,1(2) , (2011).

[5] Brandt C, Joachims T, Yue Y and Bank J, Dynamic Ranked Retrieval, ACM WSM'11, Hong Kong, China, (2011).

[6] Dan O and Davison B D, Measuring and Predicting Search Engine Users' Satisfaction, ACM Computing Surveys, 49(1),(2016).

[7] Forsati R., Meybodi M R, Neiat A G, Web Page Personalization based on Weighted Association Rules, International Conference on Electronic Computer Technology, IEEE, (2009).

[8] Khanchan R and Punithavalli M, An Efficient WebPage Prediction based on Access Time Length and Frequency. IEEE 3rd International Conference on Electronics Computer Technology , (2011).

[9] Kim JY, Thompson KC, Bennett P N and Dumais S T, Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic. WSDM'11 ACM, Page No- 9–12, (2012).

[10] Liu H and Keselj V, Combined Mining of Web Server logs and web contents for classifying user navigation patterns and predicting user's future requests, Data and Knowledge engineering Elsevier, 61(2), Page No-304-330, (2007).

[11] Matthijs N and Radlinski F, Personalizing Web Search using Long Term Browsing History, ACM WSM'11 (2011).

[12] Qiu F and Cho J , Automatic Identification of User Interest For Personalized Search, 15th International conference on World Wide Web, ACM, (2006).

[13] Shen X, Tan B and Zhai,C, Implicit User Modeling for Personalized Search, 14th international conference on Information and knowledge management, ACM, (2005).

[14] Singh R and Sharma D K, Ratio Rank :Enhancing Impact of Inlinks and Outlinks. IEEE International Advance Computing Conference, (2013).

[15] Singh R and Sharma D K, Enhanced Ratio Rank Enhancing Impact of Inlinks and Outlinks.In: IEEE International Advance Computing Conference, (2013).

[16] Speretta M and Gauch S, Personalized Search Based on User Search Histories. International Conference on Web Intelligence IEEE /WIC/ ACM, (2005).

[17] Srinvas C, Explicit User Profiles for Semantic web search using XML. IJERA, 2(6):234-241, (2012).

[18] Teevan J , Susan T D and Horvitz E, Personalizing Search via Automated Analysis of Interests and Activities. 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 51(3), (2017).

[19] Vu T , Nguyen D Q ,Johnson M, , Song D, and Willis A, Search Personalization with Embeddings ACM 39th European Conference on Information Retrieval, ECIR , (2017).

[20] Zhou D, Lawless S and Wade V , Improving Search via Personalized Query Expansion using Social Media. Information Retrieval For Social media, Springer,15(3),Page No -218-242, (2012).

[21] Zhou D , Zhao X W ,Lawless S and Liu J, Query Expansion with Enriched User Profiles for Personalized Search Utilizing Folksonomy Data, IEEE transaction on Knowledge and Data Engineering, 29(7), (2017).