A Literature Review on Personalized Web Search

K.Poornisri[1], S.Radha Priya[2]

[1]M.Phil. Research Scholar, PG & Research Department of Computer Science, Government Arts College (Autonomous), Coimbatore, India

[2]Assistant Professor, PG & Research Department of Computer Science, Government Arts College (Autonomous), Coimbatore, India

Abstract—Web mining is the application of data mining techniques to extract knowledge from Web. Web mining hasbeen explored to a broad degree and different techniques have been expected for a variety of applicationsthat includes Web Search, Classification and Personalization etc.With millions of pages accessible on web, it hasbecome difficult to access relevant information. One possible approach to solve this issue is web personalization. Web Personalization is viewedas an application of data mining and machine learning techniques to frame modelsof user attitude that can be applied to the task of concluding user needsand adapting future interactions with the ultimate intention of improved user satisfaction. In this paper, we present a survey on personalization on the web, strategies of personalization and their related works.

Keywords—personalization; web usage mining; user query; search engine; user profile.

I.   INTRODUCTION

The World Wide Web (Web) is a popular and collective medium to disseminate information today. People either browse oruse the search service when they want to find definite informationon the Web. Web mining techniques could be used to solve the data overload problems above directly or indirectly. The Web mining research is a converging research area fromseveral research association, such as database, Information Retrieval (IR), and AIresearch communities especially from machine learning andnatural language processing (NLP). Web mining is the use of data mining techniques to automatically determine and extracts information from Web documentsand services.

Web mining is the application of data miningtechniques to extract knowledge from web data, i.e., web content, web structure and web usage mining.

Web Content Mining: Web Content Mining is theprocess of extracting useful information from thecontents of Web documents. Content datacorresponds to the collection of facts a Web pagewas constructed to convey to the users. It may consist of text, images, audio, video, or structured recordssuch as lists and tables.

Web Structure Mining: The structure of a commonWeb graph consists of Web pages as nodes, andhyperlinks as edges connecting between tworelated pages. In addition, the content within aWeb page can also be coordinated in a treestructuredformat, based on the assorted HTML andXML tags within the page.

Web Usage Mining: Web Usage Mining is theapplication of data mining techniques to discoverattractive usage patterns from Web data, in orderto understand and better serve the demands of Webbasedapplications. Usage data captures thecharacter or origin of Web users along with theirsurfing behavior at a Web site [1].
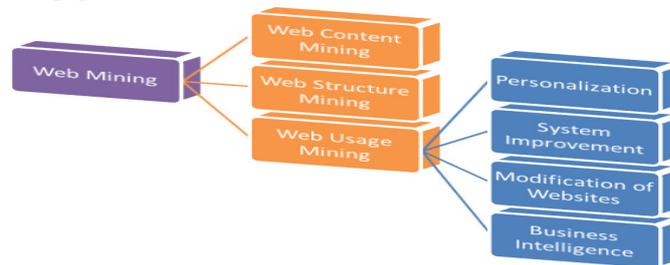
Figure 1: Classification of Web mining Techniques

With the growth of Internet, people arebecoming more and more reliant on the Websearch engines for their various information needs. Inspite of the extensive use, there are still severalchallenges for search engine. Particularly, when queriesare penetrated to a search engine, similar results are returned to various users [2]. To avoid this difficulty, personalized Web search has been established.In personalized search, how to efficiently achieve user'sreal-time information requirement is a key affair [3]. User'squery inputted to the search engine is the mostimportant source of checking information need.However, the query has the attributes of shortness,ambiguousness and incompleteness which control theclear definition of user's information requirements andthus influences the eligibility for personalized search.Personalization means persons would get the things or outcome according to their interests and assumptions without giving much more input. Personalization systems are a subclass of information filtering system that seek to predict the 'ratings' or 'preferences' that a user would give to an elements, they had not yet considered, using a model built from the characteristics of an item (content-based approaches or collaborative filtering approaches). Personalized Web search is to perform retrievalfor each user by taking his/her ownsituations/circumstances into account. As thecompetition in search market grows rapidly, somesearch engines have presented the personalized searchservice [4].

II. BACKGROUND

Personalized web search (PWS) is a one of thecategory of search techniques which provides better searchresults and the results which are tailored for individual userneeds. User specified information has to be collected andanalyzed to find out the user intention and goals behind theissued query. The personalization process can result in the dynamicgeneration of suggestions, the creation of pages accordingto the needs of the user, highlighting of existing hyperlinksthat are exactly required by the users. Most of the earlierresearch efforts in Web Personalization deal with WebUsage Mining [5].

For example, Google's Personalized Search grants users to specify the Web page categories ofinterest. Some Web search systems use relevancecomment to process user requirements or inspect usersto register their demographic information earlier so as topresent better service. Since these approaches requiresusers to employ in extra activities to specify theirchoice manually further than search, approachesthat are capable of implicitly recognizing usersinformation needs should be evolved. Since the need for personalized search is increasing, manyresearches have to be done to provide the relevantinformation by considering the users situations. The nextsection illustrates the various

personalized web searchapproaches proposed by many authors.
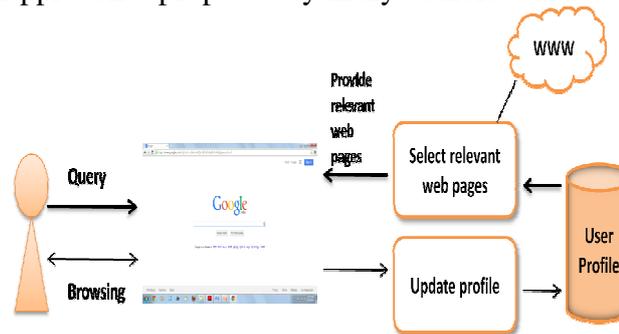
Figure 2: Process of Personalized Web Search

Mobasheret. al., presented most advanced system, "WebPersonalizer" [6]. It is a powerful framework for mining web log files to extract the useful data for the purpose of recommendations based on the surfingsimilarities of current user to earlier user. After collecting and cleaning of usage data, data mining techniques such as association rule mining, clustering, sequential pattern discovery, and classification are applied in order to determine interesting usage patterns.

Fang Liu et al., [7] recommended personalizedWeb search for improving retrieval effectiveness.Modern Web search engines are developed to serve allusers, independent of the particular requirements of anyindividual user. Personalization of Web search is toperform retrieval for every user according to theirinterests. In this paper, the authors presented a novelapproach to learn user profiles from users' searchhistories. The user profiles are then utilized to enlargeretrieval efficiency in Web search. A user profile and acommon profile are review from the search history ofthe user's and a category hierarchy, respectively. Thesetwo profiles are unified to map a user query into agroup of categories which corresponds to the usersearch objective and provide a context to disambiguatethe words in the user's query. Web search is performedaccording to both the user query and the association ofcategories. A number of profile learning and category mapping approaches and a fusion algorithm aregranted and evaluated. Experimental outcomes show that the approach to personalize Web search is veryeffective.

J. Lai et al., compared User profile results andpersonalized web search results [8]. The huge amount ofinformation is available on web. When user searchesanything, in some cases it provides same outcome fordifferent type of queries. So it is difficult for user to getrelevant and decisive results because it does not consideruser preferences as well as interest.Evaluation of users stated query search and [9] browsing activities are depends on searching queryinputted and clicking rate of each link in the response of query and theamount of time they used peculiar site. Solution to this is to compose user searching profile and approach fordocument profile construction. The review is takenplace that conclude how to use this model to combinedocument and user searching profile thatgives desiredpersonalized search outcome to user.

Kraft et al., states that the context, in its general mode, indicates to any additional information associated with the query in the search field, and also presentvariousmethods to implement the contextual search instead of modeling user profiles. If the context information is provided by an every user in any pattern, whether implicitly or explicitly, manually or automatically, search engines can use the context to custom-tailor search results. The process is titled as a personalized search. In this form, personalized web search could be either client or server-based [10].

Sugiyama et al., [11] suggested flexible web search based on user profile. The mainadvantage of this approach was user profile constructedwithout any effort or feedback from user. The main issue in previous approach needs continuous user communication. Thisapproach solved that problem. In this system, when a usersubmits a query to a search engine via web browser, the searchengine returns search results corresponding to the query.Based on the search results, the user may select a web page inan attempt to satisfy their information need. In addition, theuser may access more web pages by following the hyperlinkson their selected web pages and continue to browse. Theproposed system monitors the user's browsing history andupdates their profile whenever their browsing pagechanges. When the user submits a query the next time, thesearch results adapt based on their user profile.

Susmitha et al., targeted on exact domain selection, it proposed interface which takes only keywords [12]. For this reason, information is fragmented into various parameters related to specified domain only. Keywords entered by user are also related to definite domain which allows searching specifically what related data is expected by user. Devanget al., [13] focused on query grouping which grants search engine to understand user's session. Once query group is analyzed then search engine can easily identify the context behind queries and clicks in the corresponding query group. Query group is set of queries by same user that is related to common informational need. As user hits new query, query groups are dynamically updated and new groups may be builds over time. It explored long term search history which consists of past queries click through which can be used to enhance retrieve performance. It is also shown that such information can be used effectively for the effort of coordinating user search histories into query group.

Matthijs and Radlinski et al.,[14] gather web usage data that is URL of page, page session date and time, duration of page visit, length of the source HTML using Firefox. Peng et al. [15] built user profile by collecting search result used by users with reference of Google register. In this architecture tree is maintain and topics are linked in tree. Each topic that is search by user and stored in tree structure is manage in tree directory. Link Visited count is maintained and it displays degree of user interest. Charanjeet et al [16] present a general process of search outcome re-ranking that canbe used to re-order search results by using personalized ranking principle. Such criteria are typicallyexamined, learned, derived and then can be implemented from the user's search history log or simply from the modeling of user search attitude and interests.

III. PHASES OF WEB PERSONALIZATION

Web personalization is the process of customizing a Web site to the needs of anindividual user, considering the knowledge acquired from the analysis of the user navigational behavior in to the account in correlation with other information collected in the form of content, structure and user profile data [17]. The web personalization process can be categorized into different phases namely collection of web data, preprocessing of web data, analysis of web data, and finally decision making or recommendation. The Web personalization process can be classifiedinto four distinct phases [18, 19]:

➢ **Collection of Web data**

Implicit data built past clickstreams as recorded in Web server logs and/or via cookies or session tracking modules.Explicit data usually occurs from registration forms and rating questionnaires. Additional data such as demographic and application data can also be serviced. In some cases, Web content, structure, and application data can be added as additional origins of data, to shed more light on the next stages.

➢ **Preprocessing of Web data**

Data is frequently pre-processed to put it into a form that is compatible with the analysisapproach to be used in the later step. Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information according to the aim of analysis (example: automatically developed requests to embedded graphics will be recorded in web server logs, even though they add limited information about user interests), and accomplishing the missing links (due to caching) in incomplete clickthrough paths. Most importantly, unique sessions need to be identified from the various requests, based on a heuristic, such as requests sourcing from an exact IP address within a given time duration.

➢ **Analysis of Web data**

This step handles machine learning or Data Mining techniques to invent interesting usage patterns and statistical correlations between web pages and user groups. This step frequently concludes in automatic user profiling, and is typically applied offline, so that it does not add a concern on the web server.

➢ **Decision making/Final Recommendation Phase**

The last phase in personalization makes use of the results of the preceding analysis step to deliver recommendations to the user. The recommendation process typically associatesgenerating dynamic Web content on the fly, such as adding hyperlinks to the last web page requested by the user. This can be accomplished using a collection of Web technology options such as CGI programming.

IV. OVERVIEW OF PERSONALIZATION CATEGORIES

Memorization–Simplest and most widespread form of personalization, user information such as name and browsing history is stored (e.g.cookies), to be later used to recognize and greet the returning user. It implemented on the web server and can also threaten user privacy.

Customization–Takes as input a userpreferences from registration forms in order to customize the content and structure of a web page and it process tends to be static and manual or at best semiautomatic, implemented on the web server. E.g.Yahoo and Google.

Guidance or recommender systems– Tries to automatically recommend hyperlinks that are deemed to be relevant to the user interests, in order to facilitate access to the needed information on a large website. It is implemented on the Web server and it relies on data that reflects the user interest implicitly (browsing history as recorded in Web server logs) or explicitly (user profile as entered through a registration form or questionnaire).

Task performance support – A personal assistant executes actions on behalf of the user, in order to facilitate access to relevant information. This approach requires heavy involvement on the part of the user, including access, installation, and maintenance of the personal assistant software and it also has very limited scope in the sense that it cannot use information about other users with similar interests [1].

Web usage data mining personalization – The customer preference and the product association are automatically learned from click stream and in order to avoid the poor recommendations that will lead to disappoint customers, customers who are likely to buy recommended products are selected using decision tree induction.

Computational intelligent combinations– Provide the different information systems which have beendesigned to provide Web users with the information they search, without expecting them to ask for it explicitly.

Novel online recommender– It builds profiling models and offers suggestions without the user taking the lead.

Helping online customers decide through web personalization– The goal of a personalized website is to take advantage of the knowledge obtained from the analysis of the user's navigational behavior in

combination with other information collected, such as the user location, previous navigation patterns, and items purchased [20].

Automatic personalization based on web usage mining – In which the user preference is automatically learned from Web usage data, by using data mining techniques.

Caching – Efficiently delivering web content, i.e., caching and prefetching. Caching refers to the practice of saving content in memory in the hope that another user will request the same content in near future, while involves guessing at which content will be of interest to the user and loading it into memory [5].

V. CONCLUSION

Personalization of web search is a necessity now-a-days to reveal user preferences in search results. In this paper, a survey of personalization with different approaches has been given. The maximum numbers of web personalization methods are based user profile, search history, etc. Our aim will be increasing the search engine accuracy and reducing the time of user has to spend on it and ultimately users can get what they want in a crisp manner in shorter time and fewer clicks as well. In future, the concept of query keyword suggestion can be added and with the feature of query formulation and expansion, which helps the user at those times when users are not sure about the search queries terms, so that these feature will guide the user to get the desired information in a very specific context with comparably less effort and personalization can uses for more applications (e.g. Online shopping, etc.,) to retrieve information by user interests in shorter time.

REFERENCES

[1] A. J. Ratnakumar, "An Implementation of Web Personalization Using Web Mining Techniques", Journal of Theoretical and applied information technology, 2005.

[2]Nauman, M. Khan and Shahbaz, "Using Personalized Web Search for Enhancing Common Sense and Folksonomy Based Intelligent Search Systems", International Conference on Web Intelligence (IEEE/WIC/ACM), pp. 423 –426, 2007.

[3]   W.-C.Peng and Y.-C. Lin, "Ranking Web Search Results from Personalized
Perspective," The 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, The 3rd IEEE International Conference on E-Commerce Technology, pp.12,2006.

[4]   M.Eirinaki and M.Vazirgiannis, "Web Mining for Web Personalization", ACM Transactions on Internet Technology, 03(01), pp. 1-27, 2003.

[5]   B. Mobasher, R. Cooley and J. Srivastava, "Automatic personalization based on web usage mining", ACM Communication, 43(08), pp. 142-151, 2000.

[6] Albanese, A. Picariello, C. Sansone, L. Sansone, "A WebPersonalization System based on Web Usage Mining Techniques", in Proc. of WWW2004, 2004.

[7]   F. Liu, C. Yu and W.Meng, "Personalized Web search for improving retrieval effectiveness," IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 1, pp. 28 – 40, 2004.

[8]   J. Lai and B. Soh, "Personalized Web search results with Profilecomparisons," Third International Conference on InformationTechnology and Applications (ICITA 2005), Vol. 1, pp. 573 –576,2005.

[9]   M. Kutub, R. Prachetaa and M. Bedekar, "User Web SearchBehaviour," 3rd International Conference on Emerging Trends inEngineering and Technology (ICETET), pp. 549 –554, 2010.

[10]   R. Kraft, C. C. Chang, F. Maghoul and R. Kumar.Searching with context. In Proceedings of the 15th International Conference on World Wide Web (WWW'06), pp. 477–486, 2006.

[11]   K.Sugiyama, K. Hatano and M. Yoshikawa,"Adaptive Web Search Based on User Profile Constructed without any Effort from Users,"Proc.13th Int'l Conf. World Wide Web (WWW), 2004

[12] SusmithaDey, Siby Abraham, "User Interface For A Search Engine: A Customized and Multi-domain Approach", International Journal of Computer Information Systems and Industrial Management Applications.Volume 4, pp. 169-179, 2012

[13] DevangKaravadiya, Purnima Singh, "User Specific Search Using Grouping and Organization", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 1, Issue 4, 2012

[14] N. Matthijs and F. Radlinski. "Personalizing Web search using long term browsing history ". In Proceedings of the ACM WSDM Conference on Web Search and Data Mining, pp. 25–34.2011

[15] Peng, Xueping, ZhendongNiu, Sheng Huang, and YuminZhao."Personalized Web Search Using Clickthrough Data and Web Page Rating."Journal of Computers 7, no. 10, 2578-2584, 2012.

[16] CharanjeetDadiyala, Prof. PragatiPatil, Prof. GirishAgrawal, " Personalized Web Search", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, 2013.

[17]   JaideepSrivastava, Robert Cooley, MukundDeshpande and Pang-Ning Tan, "Web usage mining: discovery and applications of usage patterns from web data," SIGKDD Explorations, 1(12), 2000

[18]   Mobasher, B., Web Usage Mining and Personalization, in Practical Handbook of Internet Computing, M.P. Singh, Editor. 2004, CRC Press. pp. 15.1-37.

[19]   J.B. Schafer, J. Konstan and J. Reidel (1999).Recommender Systems in ECommerce, In Proc. ACM Conf. E-commerce, pp. 158-166.

[20]   D. Antoniou, M. Paschou, E. Sourla and A. Tsakalidis, "A Semantic Web Personalizing Technique The case of bursts in web visits," presented at IEEE Fourth International Conference on Semantic Computing, 2010.