A Broad Study on Personalized Web Search

[1]A.Richlin Selina Jebakumari, [2]Dr.Nancy Jasmine Goldina, [3]Dr.Grasha Jacob

[1] Research Scholar, , M.S University, Tirunelveli,

[1]Assistant Professor Dept of Computer Applications, Sarah Tucker College, Tirunelveli,

[2]Assistant Professor Dept of Computer Applications, Sarah Tucker College, Tirunelveli,

[3]Associate Professor Dept of Computer Science, Rani Anna Govt College, Tirunelveli

Abstract - Internet is a global network connected world-wide with wealth in information is a boon to human. Desired information is retrieved through web search engines. Search engine is indispensable software that operates between the user and the information repository that retrieves relevant information according to the query given in an automated manner. Even though search engines do good job, may mislead users in many ways. Most of the time the users may find difficult to get their preferred information in precise. As, the utility of search engine is affected by inevitable factors there is a need to improve the utility of search engine. Web search personalization is the solution to provide users with their desired information. Web search personalization is a challenging process as the search engine has to efficiently acquire user's real time information. This paper focuses on various functions, issues, recent trends, approaches and evaluation techniques on personalized web search.

Key words - Personalized web search, Search Engine

I. INTRODUCTION

Internet a global communication network of computers connects millions of computers world wide as it is a repository of information. The content present on web in the form of hyper text which refer to the other text or hyper text. This collected works of documents around the world forms the web additionally referred to as World Wide Web (www). The pages in the hypertext documents are referred as web pages and its position is distinct by using a Uniform Resource Locator (URL). Web Search engines like Google is the single largest source of internet provider that allows people to fetch their chosen information.

II. WEB SEARCH ENGINE

A web search engine is a software system that is designed to search for information, on the arena internet. The search results are generally presented in a line of results, often referred to as Search Engine Results Pages (SERPs). In fact, the information is a mix of web pages, images and different forms of files. The two major functions of search engines are i) Crawling and building an index   ii) Providing the search users with a ranked list of the determined most relevant websites. Search engines like Google and Yahoo maintain dynamic information by running automated robots called as "Crawlers" or "Spiders," to reach the many billions of interconnected documents on the web[1]. Once the engines find these pages, they decipher the code from them and store selected pieces in massive databases, to be recalled later when needed for a search query. The key features helping to retrieve a good representation of documents/pages are as follows
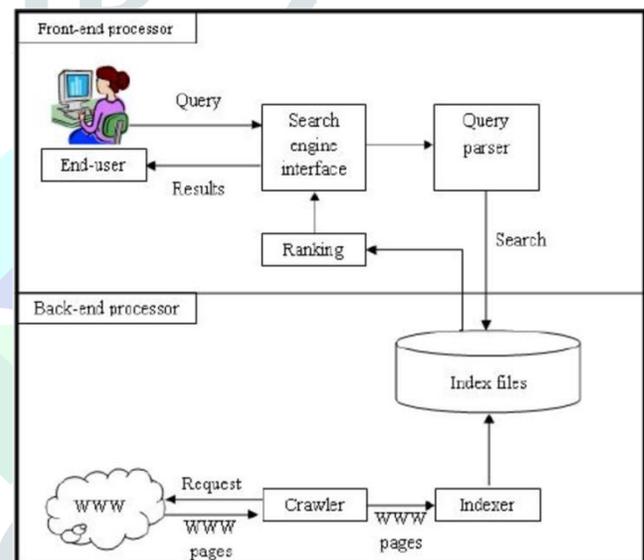
Search Engine Architecture

Term frequency - How often a query term appears in a document

Location of terms - Many search engines give preference to words found in the title or lead paragraph or in the metadata of a document [1].

Link analysis - Link analysis is based totally on how well each page is connected as defined by using Hubs and Authorities, where Hub documents link to large numbers of other pages (out-links), and Authority documents are the ones referred to by using many different pages(in-links).

Popularity - Popularity makes use of information at the frequency with which a web page is chosen by all users as a means of predicting relevance.

Length - In a choice among two documents both containing the same query terms, the document that contains a proportionately better incidence of the term relative to the length of the documentation is believed more likely to be applicable.

Proximity of query terms - When the terms in a query

arise close to each other within a document, it is more likely that the document is applicable to the query than if the terms occur at greater distance.

III. WEB SEARCH PERSONALIZATION:

Recently, major search engines have implemented Personalization, Personalization is an attempt to uncover most relevant documents using information about user's target, domain of interest, browsing history, query context and so on. That gives higher value to the user from the massive set of results. As the web content is growing exponentially, an increasing number of state-of-the-art methods are required to supply the relevant content to the individual user. The most common difficulties encountered when searching the web are i) Problems with the data itself ii) Problems faced by the users seeking to retrieve the records they need iii) Problems in understanding the context of search requests and iv) Problems with identifying the modifications in user's utility [4]. Consequently the significance of personalization is to customize the web for individual users through filtering out the inappropriate outcomes and pick out relevant outcomes.

IV. NEED OF PERSONALIZATION

The need for personalization arises due to the following facts: Firstly, distinct users have distinctive backgrounds and interests. For the identical query, they have one-of-a-kind information needs and goals [5]. Secondly, user information desire may change over time. Users may also have sort of necessities primarily based on the time and situations

V. PERSONALIZATION APPROACH

While applied to search, personalization would involve the following steps: 1.To acquire and constitute facts about the user with the intention to recognize the user's interests 2. Use this information to either filter the consequences back from the preliminary retrieval process, or directly include this information into the search procedure itself to select personalized results [5]. Web search personalization system use collected information about user from profiles, cookies and to conduct and revise the search to maximize the user satisfaction. The user profiles are created which specifies the user's interests, choices and information desires to better personalize the search results. Personalization of web search can be done at either server side or client side. Many issues arise on personalizing the web at server side like server should keep all the search history for each and every user. It also has to go looking the history of a specific user when a user submits any ambiguous query [5]. The overall performance of the server gets down when many clients submits the query at the same time. Consequently, most of the techniques employ client side approach as all the search histories and queries are maintained at the client system making the faster way to access the user profile.

VI. PERSONALIZED WEB SEARCH METHODS

i) Personalized Search Based On Content Analysis

On this approach, the content similarity between returned web pages and user profiles is calculated. Under content analysis, user profiles can be constructed using two ways: i) Topical categories ii) Keywords lists. In topical categories, a user profile is framed as a hierarchy of principles or topics [5]. In keywords lists, a list of keywords is used to expose the user alternatives.

ii) Personalized Search Based on Hyperlink Analysis

Generic search approaches rank documents depending at the link structure of the web. Accordingly, web page rank algorithms are being utilized in web search. Web Page Rank laid emphasis at the fact that crucial pages are connected by many crucial pages. The Page Rank of a page P is defined because the possibility that the surfer visited page P.

iii) Personalized Search Based on User Group

In this technique, the network of like-minded clients is formed. So, simplest the users are accountable to provide the information needed to shape the user profiles[5]. Search histories of users who have similar interests with the other user are used to refine the search results.

VII. PHASES OF WEB PERSONALIZATION PROCESS

The important steps of web personalization system include i) Web data preprocessing ii) User modeling in personalization iii) Recommending personalized page ranking strategies.

i) WEB DATA PREPROCESSING

Information preprocessing is the manner to convert the raw data into the proceedings ideas. A session information can be in someone of the subsequent forms (i) Web page (ii) Web structure (iii) Web utilization data: records preprocessing in personalization includes information cleaning, user identity, session identification, feature identification of visited pages and route identity[4]. Subsequently the input for the preprocessing step is a user session report that offers an specific account of who accessed the web page, what pages have been asked and in what order, and the way lengthy every web page become viewed. A user session is the set of the web page accesses that happen at some stage in a single visit to a website online.

ii) USER MODELING IN PERSONALIZATION

User modeling is an crucial a part of a personalized web seek. It's miles the technique of mounting personal options of the clients in terms of clients surfing history, expertise about the world, likes and dislikes and so forth.

iii) RANKING STRATEGIES

The research efforts on re-rating web search results are classified into the following class of strategies.
Explicit relevance judgments: The hassle-unfastened way to affirm whether a result retrieved for a query is relevant to the user is to explicitly ask that user.

Implicit relevance judgments: Implicit data can be generated by means of user's interaction with their service.

i) Content based implicit measures: This form of measure makes use of a textual illustration of user's interest to deduce the results which are relevant to their current need.

ii) Behavior-based implicit measures: This form of measure uses of humans' behavior together with their past interactions with search result lists, click-through data from the logs and so forth.

VIII. CLASSIFICATIONS OF APPROACHES TO PERSONALIZATION

Personalization system can be labeled based totally on the data they make use of, the learning paradigm used, the vicinity of the personalization and the method that the interaction takes with the user.

1. Individual Vs Collaborative

The term personalization impresses upon the individuality of users and the need for systems to conform their interfaces to the needs of the user. This requires data amassed on interactions of users with the system to be modeled in a user-centric fashion. Generally, data is collected by the business with which the user is interacting and consequently the business has access to data associated with all its customers [3]. A personalization system may additionally select to build an individual model of user likes and dislikes and use this profile to predict/tailor future interactions with that user.

2. Reactive Vs Proactive

Reactive approaches view personalization as a conversational system that requires explicit interactions with the user either in the form of queries or feed back this is incorporated into the recommendation process, refining the search for the item of interest to the user. Most reactive systems for personalization have their origins in case- based reasoning research. Reactive systems can be further classified based on the types of feedback they anticipate from the user [4]. Common feedback mechanisms utilized by these systems include value elicitation, critiquing / tweaking, rating and preference feedbacks. Value elicitation and tweaking/critiquing are feature based approaches to feedback. While in value elicitation the user must provide a rating for each feature of each recommendation object offered to the user Rating and preference are feedback tactics at the object level. In rating based totally feedback, the user need to rate all the recommendations offered to him, based totally on their "fit" with his necessities. In preference feedback the user is provided with a list of recommendations and is needed to choose one of the recommendations that excellent suits his requirement. The system then uses this feedback to present with user with other comparable gadgets. The iterations keep until the user reveals an object of interest or abandons the search.

3 .User Vs Item Information

Personalization system vary in the information they use to generate recommendations. Typically, the information used by these system encompass:

- **Item Related Information:** This includes content descriptions of the items being recommended and a product/ area ontology

- **User Related Information:** This consists of past preference ratings and behavior of the user and demographics systems that use item related information generally deal with unstructured data related to the items. As soon as this data has been processed, into relational form such as a bag-of-words representation usually used for textual data, a user profile is generated.

IX. PERSONALIZATION CATEGORIES

The details of personalization categories are

1.Memorization : Simplest and most widespread form of personalization, user information Which include name and browsing history is stored(e.g. using cookies) It is carried out at Web server and can also jeopardize user privacy.

2.Customization: Takes a user's input preferences from registration forms in order to customize the content and structure of a web page and procedure has a tendency to be static and manual or at best semiautomatic carried out on the Web server. E.g. My Yahoo and Google

3. Web usage data mining personalization: The customer preference and the product association are automatically learned from click stream. 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.[2]

4. Helping Online Customers Decide through Web Personalization: The intention of a personalized website is to take gain of the expertise obtained from the analysis of the user's navigational behavior in combination with other information amassed ,along with the user's vicinity, previous navigation patterns, and items purchased.

5. Caching: Effectively delivering web content, i.e., caching and perfecting. 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.

X . EVALUATION OF PERSONALIZATION SYSTEMS

Evaluation of personalization system remains a challenge due to the lack of understanding of what factors affect user satisfaction with a personalization system. It appears apparent that a system that accurately predicts user needs and fulfils these needs without the user needing to expend the same resources in accomplishing the task as he could have, in the absence of the system would be taken into consideration successful[3]. For this reason personalization systems have most commonly been evaluated is terms of the accuracy of the algorithms they hire. Recent user research have found that a number of issues can affect the perceived usefulness of personalization systems consisting of, trust in the system, transparency of the recommendation logic, ability for a user to refine the system generated profile and diversity of

recommendations. For a business deploying a personalization system, accuracy of the system which includes profits or qualitative metrics such as customer loyalty. Subsequently the evaluation of personalization systems desires to be accomplished along a number of different dimensions, some of which are better understood that others and have well established metrics available. The key dimensions along which personalization systems are evaluated consists of user satisfaction, accuracy, coverage, utility, explain ability, robustness, performance and scalability.

XI. CONCLUSION

Personalized web search is an active ongoing research field that focuses on the retrieval of the relevant web page results based on the user preferences. This paper covers the personalization mechanism in various dimensions. The motivation behind the personalization is to enhance quality of rankings. A lot more user studies need to be carried out to gain a better understanding of these issues to develop a Personalized Web Search framework. In this paper the essential key components of web search personalization is discussed and the most significant thing is to identify users information accurately.This observation will undeniably manual the researchers to apprehend the general web search personalization system to expand promising strategies for personalized web search approach.

REFERENCES

[1] Ms. Shital C. Patil1, Prof. R. R. Keole "The Role of Web Content Mining and Web Usage Mining in Improving Search Result Delivery" International Journal of Computer Science and Mobile Computing ISSN 2320–088X IJCSMC, Vol. 3, Issue. 3, March 2014, pg.7 – 14

[2] Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki "Web Personalization Using Web Mining: Concept and Research Issue" International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012

[3] K.Nirosha, V.Karthick, E.Jaya "An Evaluation of Personalization Systems using Web Mining Techniques" International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 2, Issue 2, pp: (265-270), Month: April-June 2014

[4] J.Jayanthi, DR.S.Rathi "Personalized Web Search Methods – A Complete Review" Journal of Theoretical and Applied Information Technology 30th April 2014. Vol. 62 No.3

[5] Himani Arya Jaytrilok Choudhary ,Deepak Singh Tomar "A Survey on Techniques for Personalization of Web Search" International Journal of Computer Applications (0975 – 8887) Volume 94 – No. 18, May 2014

[6] Jaytrilok Choudhary ,Deepak Singh Tomar "A Survey on Techniques for Personalization of Web Search Review" International Journal of Computer Applications (0975 – 8887) Volume 94 – No. 18, May 2014