Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Survey on Web Search Results Personalization Techniques

Vijalakshmi Kakulapati[1]
Dept. CSE
JNTU, India

Dr. D. Vasumathi[2]
Dept. CSE
JNTU, India

Sudarson Jena[3]
Dept. CSE
Gitam, India

Abstract: ***Web has become the best source of information throughout the world. Search engines play a key role in finding out the information, which is abundant. As such search engines are enhanced with new technology. Though search engines find much information with one key word, fail to provide the accurate, exact data that is required. Through the research it has found that users will not have much interest if it is delayed and can't afford to spend time with queries. This time constrain can be observed in many situations. Hence the most significant point in the applications of the search engines is to find accurate information immediately. This aspect of accurate and immediate information for a search is well attended by personalized web environments. Since there is an increasing importance to the personalized web environment, which develops models of short-term and long-term users needs. This paper delineates many user personalization approaches and techniques, which are useful in web search domain and the systems that are at present used on the Internet.***

Keywords: Data Mining, User Behavior, Web Search, Personalization, Information Retrieval

1. Introduction

Improving the accuracy of the search results of these personalized websites is the immediate aim of the many search engine developers where as a better multimedia browsing, localization, question and answer methodology, visual results presentation and modified web search on devices other than PCs including cell phones are the other long term goals. All these areas are given much importance so as to bring the best search en gine applications, which will answer all the necessities of the browsers. The same point is reiterated by Anand Spink, a professor at the University of Pittsburgh's school of Sciences and a long time researcher of web search behavior, to quote, "Each technology relates to a different aspect of peoples information behaviors". These words show us that different technologies should be developed to answer variant demands. As per now not much development is observed because of which the Web remained same. Anand Spink enforces the same point in his book, 'Web Search': publishing searching of the web (Springer 2004), which analyses the development of search engines from 1997 to till date. His study suggest that not much development and innovation brought in this field as such there is much scope in the field of application development and result presentation experiments. This is possible because individual web search behavior has not changed much. Daniel Read's says that "Our challenge is to read a users mind, but this is a very big challenge because the words used for any search are limited to two or three words. Same kind of difficulties is faced in localization and personalization of web information at the user acceptance side and software development side. However, if these difficulties are resolved by R&D projects there is every possibility of accurate information at the very first search. Research has been started and one can see the same in the future plan for the development of Ask Jeeves and Google.

1.1. Making it Personal

Personalization of a search engine will be of great use and time saver to all the browsers. This new feature takes the search engine directly to the information and topics that it has already been taken to by the user. This makes browser reach the accurate information of his need without being troubled by all the other data, which may be related to the topic but not relevant to the demands. But sometimes the browser who doesn't want the search engine know so much about them does also not accept this new application. Problems apart, many including Google and AskJeev are taking all the trouble to bring it intact.

To date, browser is not given good knowledge of any search engine. Because of this browser doesn't know the requirements of a search engine. As such he can only give a hint of his demands in this own sphere without relating himself to the requirements of a search engine to do the job. This shows that search engines need to be enhanced to understand a user better. According to Spink "One of the biggest obstacles is that computer software is not very good at really understanding people's information needs. Using just implicit techniques is not effective, but explicit techniques need to be much more developed." To achieve the same Spink says a "Research suggests that you need to focus on a combination of implicit and explicit personalization".

Google is trying to bring two different techniques to give accurate information to the user. It gives filtered results if the user supplies it with the needed information voluntarily. The personalized web search which is still in beta testing (http://labs.google.com/personalized) helps the user to give a search profile so as to give filtered results which match the search profile supplied by user. It also updates him about new information in the web through email.

Ask Jeeves which is known for its excellence in question and answer technology, is striving to be the best even in personalization. Its new technology i.e. My Jeeves, which is still in beta testing helps a user to save his search results in his My Jeeves folder. This folder can be updated with any other personal information and this can be accessed from any PC. For a minimum of 1000 documents registration is not necessary and there is no limit if one is registered. The company is planning to connect this My Jeeves to its smart search techniques. For ex, if one is searching for 'NYC' at Ask Jeeves, he gets a structured smart search box, which contains links to maps, jobs, weather, local time, and the city's chamber of commerce. Through this they try to give a sophisticated search result. One can see their vision in Read's words "We want to combine the structured and unstructured data". "To me that kind of completes a circle in search". This is their first step whereas have got good plans to add the personalization feature to their search engines which at present works on amalgamation of structured and unstructured data. Through hyper textual space, user tries to find the information but is very much troubled by the teeming down of the information. This excessive data can't answer the demands of the user perfectly and exactly. Another problem is that, same results are given for all the queries, which may be similar but are asked by different users with varied needs. Personalization is one remedy, which understands the needs of a particular user and gives filtered information. But this has to be developed more for which understanding the mind of a browser is necessary. After a good analysis of the search behaviors it was found that there are three techniques that are mostly used by users, first being Surfing with the help of a query and a recommendation. Recommendation based search should be given some information about the user needs so that it can relate to the results of any other user with same interests. This is used for movies, music and products [1]. In this way it gives the filtered results.

Browsing is another technique, which is much used to find information. In this, user has to go through the pages, selects one and goes a head with hyper links. This is very helpful because one can get information from any part of the world and research scholars are much benefited. But due to copious information that is available, to strike the exact website is very difficult. Even with the most structured websites like Yahoo! Directory and the open Directory project 2, user has to spend a lot of time searching for the suitable one. Here information is available but accessing them is very tough.

The job of surfing is done with the help of Information Retrieval (IR) when a query is given to a search engine. When a query is asked a search engine selects some websites, which can answer the query. From millions of documents few are selected and given to the user. These results will be used as inputs i.e. the information found will help for further search. This is based on particular sources and the list of sources is not updated. Another problem with IR is that it depends on sequences of usually unrelated user queries. To resolve the above said problem one possible solution is Information Filtering (IF), which is modeled on the interests of the user. It has a complex representation of the user demands. This technology believes that user's needs are static and at least do not change in the same pace as websites. The web is vigorously updated, removed and added. Hence IF can be the best way to reach information. Since it takes a user to be static, it doesn't trouble the user with copious information but matches his interests with the information and continuous to update that related data. As such user will have exact information and one, which is updated. Hence it can be a best way to search information. But one big challenge to find whether the new data suits the user need or not. So to update the information much time is consumed. Due to this drawback IF prototypes are not preferred much by the user the following papers, will discuss the most important user modeling techniques developed for IF.

Information Retrieval and Personalization are two different techniques used that can be used to access data. In IR methodology, new information is updated with the help of re-retrieving and the Search Engines Document Index is changed accordingly. Personalization is a new technique, which is more efficient than the previous ones. This paper concentrates and expounds on this new technology and its working system. Searching is the key point to access any data in Internet. It can be through surfing or a query. Following any of the above said methods will confuse author with abundance of information. Query directs a user to the relevant documents from many documents of different topics available. [2]. But all these search engines will give the same results to any kind of user regardless of his requirements. As such these need to be enhanced for a better search. Personalization is one such step that helps search engine to have unique collections of documents for any line of a topic. This collection is brought with the help of the users. Here, a user is supposed to supply his requirements in a format and also the situation for which he needs information. This helps to match a user with any other user's results who has same interests. This helps a user reach the exact information required. The results will be according to the preferences, tastes, backgrounds and knowledge of a user.

Though many search engines have come very few concentrate on user interaction. It is also observed by the users that it is difficult to personalize them. As efficiency rises with personalization, difficulty also increases in using them. By personalization search can be made very efficient but it becomes very difficult to utilize it [3]. There are many reasons for this, one is that, the purpose, requirements and interest of a user will be saved in an external search system and this is not liked by many. Another problem is, that personalization of web search results is a computationally intensive is a very tough task. A personalized search engine takes considerable time, which is done within a second by a general one. To bring securitized results according to the interests of a user is not very easy. This is one big hurdle. To end, users may find difficulty with new technology that may get introduced for the personalization. Present paper familiarizes and introduces a user with the procedures, methods and techniques of a personalized search engine.

2. Overview On Personalized Search

Till now this paper focused on the necessity and importance of better search engine techniques. Now, personalization, which is a new development feature of search engines, will be delineated. First part of this discussion will deal with the new jargon and technical terms that are used. Later good information about the way a user details are maintained that are utilized to find a user requirements is provided and also the different sources that will be used to get the information.

Finally an overview of the various personalized search methods can be seen.

2.1 Personalization Techniques

Search engines are mostly dependent on IR technology, for example Vector Space Model (VSM) [4]. It shows that they are dependent on the data available. When a new web page is introduced it will be updated to a search engine with a particular denotation, which is formulated depending on the information given in that web page. Any web page in the web world will have a title name to be displayed. These names are to be from the content there in. These letters or key words are used by the search engines to give the results for an ad-hoc query [5]. If a user has no widespread knowledge about his need then he will not be able to use the appropriate correct word in his query. This slows down his search. Reverse is equally possible [6]. But till now, a research on search behavior about the usage of the accurate words shows that users usually restrict themselves 2 to 3 words.

Search Engines have got a problem with language as well. Any language is rich with synonyms and poli-semis. Let us say for ex. One website organizer may a use a word which he likes most but a user for the same information may go for a synonym of it. Then search engine will not trace that webpage. Synonyms mean many words expressing same meaning and poli-semis mean one word with different meanings. The key-word method used by search engines suffer much due to the above said language richness [7]. There is every possibility that a search gives unrelated results from billions of documents because title of these documents may have many synonyms and polysemis [8]. Due to synonyms even relevant information gets missed and with polysemis unrelated data comes into picture. This delays search and user will not be interested. Hence to provide better service, a search engine should go for the understanding of the information than depending on key words. This understanding will enable a search engine to correlate well with the given query. Figure 6.1.shows the principal of content-based personalization approaches.

If a search engine considers even the profile of different users to present the search results, it is known as Collaborating Method [9]. The profile of a user contains needs, context and interests. If the above said match for any two different users then the results can be interchanged and this will help to achieve accuracy as well. Dieberger et al [10] coined a new word for this approach i.e. Social Navigation. This refers to software that helps users to leave useful traces on web sites, such as reviews comments or votes utilized by others users while they are surfing or trying by query.

2.2. User Modeling in Personalized Systems

Modeling or profiling is a technique used in personalization of a search engine. This technique stores the profile of a user and the kind of questions asked by him. This collection helps to analyze users search behavior and can present relevant search results. The technique called "user-modeling component" is used while the information retrieval or filtering is done. The search engine with this technique can bring down filtered results according to the users needs. The more complex a personalized is a search engine the results will be better. In this methodology interests of the user will be given by him to get accurate results.

The methodology will have a format, which either a registration or a questionnaire, which will be simple and touches only basic interests of a user. Where as a complex one requires information in different angles like whether user has any previous idea about the subject, academic status and purpose of the information.

3. Information Dependent

In this technique user is not troubled, since no profile of him is required. It is totally dependent on data him but not on user. This technique uses the information in the web pages to give filtered results to a user. This is experimented by yan for Internet news articles, which he named as SIFT [11]. He established a profile for SIFT which contains the list of words that can and can't be used to search any article. If user wants any thing extra he must update this data. At least twenty articles are provided for any individual user. This information can be accessed either with World Wide Web browser or will be given to him through an electronic mail if he can't connect to web.

3.1. Client Side Ontology Based Personalization

According to the interests of the users, results will be grouped under different concepts, this is the methodology followed in the above said system. Here the psychology needs of the users play an important role as such we find in this technology different group of contents. The main task is the maintenance of user profile and updating groups accordingly. S. Sendhi Kumara & T.v. Geetha introduced a new personalized search index known as user conceptual index [UCI] [WI], this presents a content related relation between the search key words and the pages, which answers the users demands. This technique is dependent on the development of automatically identified user profile known as a Personalized Ontology and Page Ontology for the improvement of an existing personalized web search system based on the UCI.

3.2 Query Log Analysis

Many query logs has been analyzed in [12, 13]. So many been are published with relation to the distributed queries according to query length, query frequency, query type and topicality aspects. Many other works has focused on user behavior at the query session level, which shows the aspects of reformulation rates which can be relatively high.

Conceptualization personalization approach groups the queries by users that issue them. User profiles are created to achieve personalization from explicit or implicit participation behavioral of user feedback [14], and researchers have identified various differences such as short versus long term profiles, incorporating context, and different types of profiles such as content-based versus collaborative information [15, 16]. The use of learning from impressions and clicks to improve the ranking is also an exciting direction [17].

3.3 Personalization which Works on Hyperlink Technique

In this personalization is done at the page rank vectors. For this weighting links are used. These links are computed with the help of a URL, which in turn should match the profile of particular users. These profiles are presented in terms of Internet domain features taken from URLs. The interests of users are specified as binary vectors where every characteristic is connected to one or more DNS tree nodes. This technique is dependent on the features of hyperlinks, i.e. anchor terms or URL tokens. These hyper links are connected with profiles of users so as to filter information.

According to a profile vector, a weighted page rank is computed giving a weight to every URL in accordance with the match between the URL & the profile. This personalization also takes care that not only related but the best results are given. As such analysis of web pages is undertaken in two ways. One is being the procedure of ranking every web page with the help of a graph structure. This ranking of pages is done by a team of research people who concentrate on retrieval techniques for example a page is given rank according to the number of Hits [18] it has faced. Such ranked pages are given importance while giving results. Second kind of analysis is done with the help of hyperlinks, URL and anchor texts, which are techniques of hyperlinks, are used.

Page Rank technique is very useful in personalization. It helps to give potential results for a search. Page rank is one, which gives global ranking to the web documents. This ranking is given on the basis of hyperlinks [19]. If a link is preferred by many documents then that link will be given good rank. For example, if the link X directs a user to the link Y, then it is taken as X votes Y. In this way if many links direct user to Y, it gains importance. In addition page rank also considers the author of a particular web page. Pages, which have good authors, are given importance. Hence, it can be said that ranking of documents is depended on web graph structure. Web structure is very much used for page ranking by search engines like Google. These ranks are then used to grade the search results to develop precision. The value of a web page among billions of documents is estimated and given by the page-rank technique. But this page ranking is not without lapses, thee are discussed in [20, 21].

Considering the problems of page rank system, personalization of page-rank was under taken in a new way with the help of Hyperlink features such as Internet domains. In this new model, users provide their chosen profiles as binary feature vectors where a feature communicates to a DNS tree node or node set and pre compute the page rank score of each profile vector by giving a weight to each URL depending on the match between the hyperlink and the profile demands. Depending on URL Weights, a weighted page rank will be computed and given at query time to grade results.

Though above one is able to give the best results, has two problems at the implementation level. First, user is supposed to give his profile which he is not interested to and second, results are ranked according to the previous search track of the user. With the help of profile and past search track an unauthorized analysis of the mind of the user is done and accordingly his needs are estimated. After all this with previous track and present demand analysis a ranking methodology grades the documents and presents most appropriate results possible to a user.

3.4 Personalization Based Web Usage Mining

Web mining is nothing but exploration and evaluation of the documents available in www. An elaborated classification of web-mining domain is given in [22]. Right now a basic introduction to various domains that are related to Web-mining can be seen. Later WUM will be discussed in detail. Web-mining is related to the information and its features are divided in the below given procedure:

Content Data: Documents that is available to the browser are content data. Content mining is deriving information from the material of the web pages. [23].

Structure Data: Web structure mining is the way of selecting information from the structure of data [24].

Usage data: The data is taken from the browser is connected to the web. As cited above WUM is the exploration ands evaluation of browser access to the web information system [25] using the data available to customize the web for user was not any new idea but was suggested way back in 1995 [26]. A good analysis of the research in WUM is given in [27]. Tracking & analysis which are the present methods and work on two tiers of WUM are discussed in this part.

3.4.1 User-Interaction Tracking

The data about the dealings of a user with Internet is of great use for personalization. This connectivity data can be acquired in different ways: The web browser on the client side, web server logs, or representative server logs. As the importance of personalization rises, scrupulousness of tracking is of major importance and must be undertaken as the important feature in choosing a data source. There are many degrees of storage available in the web, especially to find out browsers access to much utilized page while browsing, user tend to refer back many a times data is directed with the help of web browser storage. In addition, proxy serves present a medium level of storage at the program level.

Nevertheless, cache hits are not totally saved at proxy server logs, which in return effect the analyzing of user preferences and search behavior. Lin et al. (1999) [27] has invented an "access pattern collection server" to overcome the above said problem but this works only when user secrecy doesn't matter. Cookey et al. (1999) [28] has used referrer and agent fields of a server log to obtain the information about such stored references that are hit back. Spiliopoulor et al. (2003) [29] analyzed the output of many such methodologies. It is found that server and proxy logs are unable to provide the temporal aspects of user communication. Time stamps stored in these logs for document demands will also have network-transmitting time. Because of the uncontrolled working of the network, the important information can't be scrutinized easily. Rather, if temporal characteristics are stored on the client side, hiding times of all user communications can be stored as promptly as needed. The data that is available with the user about the communication done with Internet is the most authentic and spatial. Since complete information is available with user, finding out the

URL or resource of a data becomes very simple. This is a very big challenge in case of proxy or server logs. Moreover previously collecting data about the web page usage is a single person job for a proxy, but now it is rendered to all the users.

This work is known as session identification and is efficiently done at the user side. Because of the stateless connection model of the HTTP protocol, documents asked for are logged automatically in the server or proxy logs. For a better understanding and analysis these documents should be divided and according to the key words, these documents are reorganized and grouped. In shahabi et al. (1997) [31], employed a remote agent that finds out browser communications on the user side. The information collected by every agent is saved as different semantic groups at the server so as to dismiss the user identification again. Nevertheless, collecting information at the client has few lapses. Java scripts or Java applets are employed to run the agents, which collect data from users. For this Java program must be incorporated in the browser of a client, which may not be liked by users. Shahabi et al.(2000) [32] expounded on this information collecting methods depending on the user-side data collecting idea.

3.4.2 Access Pattern Analysis

Exploring all the usage data is not possible because they are abundant. Few characteristics of the data must not be considered while structuring the access models. The value or grade of a paper is estimated according to the number of hits that it has faced by the users. This is a basic method. In addition, when a document is preferred that is selected first or after browsing few more documents among all the results. Aggregate tree and hidden Markov models [30], which are not independent, are utilized to find out this characteristic and to imagine future references. Along with spatial features temporal features like page view time are of much importance, especially in the milieu of web personalization applications. Yan et al.(1996) [33] and lovene and logic (2000) [35] opine that a paper can't be judged according to the time it is selected because sometimes some papers are not preferred due to its tough accessing process, Zipfin division and but this can be solved if the view time is combined with other characteristics , the present model that is explained in this article is capable of combining the above said and many other qualities. Features are considered for every group on which basis a session is formulated. Since many characteristics are available either complexity of the model or the accuracy can be sacrificed according to the given situation.

Scientists have presented many approaches and dates-mining methods to find out these features and to show browser access models. Hobasher et al have used the classical group regulation a priori algorithm to trace a frequent item sets depending on their patterns of occurrence at the browser sessions. They employ group regulations to trace connecter item sets to be proposed to the user depending on the items that are covered in the user sessions. Mobasher et al [36] display that grouping methodologies give better results when compared to group regulations when used in the personalization of a web. Other set of methods, which are not independent are used to imagine future reference depending on the previous selections of a browser. These methods understand and represent important similarities among page selections. Cadez et al [34] employ a Markov method for this purpose.

Borges and Levene [37] explain a probabilistic regular grammar whose higher probability strings coincides to browsers selected access methods. Breese et al [38] carry out an experimental evaluation of expected algorithms like Bayesian division and Bayesian networks in the framework of web personalization and show that the results of these algorithms rely on the kind of application and wholeness of the usage data. Grouping to mine usage data methodology was initiated by Yan et al.[33]. With this method, browser terms are generally structured vectors. In the native design of the vector structure, every part of the vector shows the importance of a feature, like hit-count, for correlating to the web page. A group algorithm is used to find the browser access methods. Active user terms are divided with the help of a definite application dependent on the similarity measure like Euclidean breadth.

Presently many Algorithms were tested to access the grouping achievement in the milieu of WUM; Perkowitz and Etzioni [42] presented a new grouping algorithm, cluster miner, which is made to answer particular web-personalization necessities; Fu et al.[39] employ BIRCH[33], an efficient hierarchical clustering algorithm; Joshi and Krishnapuram [40] prefer a fuzzy relational clustering algorithm for WUM because they believe usage data are fuzzy in nature; Strehl and Ghosh [41] propose relationship-based clustering for high dimensional data mining in the context of WUM. Paliouras et al[43], from the machine-learning society correlate achievement of cluster miner with two other grouping procedures which are vibrant in machine-learning research, for instance, auto class &self organizing maps, and display that Auto-class is better than other procedures. Mobasher et al[36] point out that a browser may exhibit features that one collected by various groups while he/she is to be divided as a single cluster. VerderMeer et al [44] examine anonymous WUM by taking dynamic profiles of browsers in association with static profiles. Dynamic clusters as a methodology to prepare the group model which can update the new developments in browsers behavior. A perfect similarity calculation, which can vary, is well estimated by the gap between partial user sessions and cluster representation is also a matter of important.

This methodology employs two step procedures to achieve web usage mining for search results personalization

- **Tracking:** This is the first step which traces the spatial and temporal qualities of browser interactions. To achieve it most of the times a remote ages is employs.
- **Analysis**: Assessment can be undertaken with the help of characteristics that are taken as matrices and employing them to present the information that is available periods. The qualities of these periods are the qualities of its parts at different levels. Hence at different levels of its parts information is gathered. The assessment of periods is done with the help of its different parts. As a last step similarity calculation is done which confirms the idea 'similarity'.

3.5 Adaptative Personalization

This method concentrates on personalization for the sake of necessities of a user. It is a predominantly for a user. At three levels this personalization is employed.

- Finding out the data that is already visited.
- Answering the present necessities accurately.
- Presenting the source for that temporary data need. Through test it is proved that this method has the ability to do those three tasks.

There are many ways of personalizing a web search and all are useful in some or other way. One among them is done with the help of browser's opinion on the accuracy of result, which is not directly expressed by user.
This model is not fixed on any feature but depends on the above said three areas hence it is flexible. Most of the personalization are not so flexible and depend on the search behavior and importance given to different papers by a user in the past search. According to these values, one is given to results so as to help the user with accuracy. Nevertheless, all these take sides with any one truck of personalization but not as flexible as the present one. They either gives importance to the ranking provided by general user entered that or consider the present need of the situation. But the present one amalgamates both features and uses them accordingly.
Some other way of personalization brings, importance and search results together. Then a similarity is brought between these two to present the search results [39, 42]. User interests inside the state of the out link are connected to the assessment techniques like page rank when it comes to HITS [38] they present a second chance to rank search results on the basis of user preferences. This method stands for the best quality where there are shifts from page rank techniques to pages of browser independent preferences. In [41] method is developed by understanding of topic preferences of a user from his search behavior.
This method of personalization believes and tries to answer variant search means a user might have. To achieve this three variant search models are taken.

3.5.1 Re-finding known Information

The usage of 'back' button by the users shows the demand for the results that were gone through once [42]. Through the given techniques one can go back and refer the search results but there are many situations where a user prefers to submit a query again for the same results.
If the first query in a period is the one that is used in the past search, present method understands that, user want to go back to previous search results. Three techniques one used to face such a need. First, it requests the local index to suggest the exact query so as to find the previous results. Here in this search agent takes the responsibility and modifies the query given to Meta search engine. Here in this technique browser is not without control and can stop the automatic reframing of a query at all levels. How the user improved the query in the previous search will be of great interest to develop the present one.
The final query of any search should have given the best results; hence it is considered much in this model. Second, the query is modified according to the content of the pages of clicked ones. Though there are few pages, which have better value than the selected ones, they are not considered. There can be two reasons for this rejection, one is after looking at the key words the user might have not liked it and felt not suitable. Second is he has a knowledge of that document. In either case they are not important to modify a query.
Third, hyper links that were reached by a user with the help of a link from results are retrained as additional click stream results.

3.5.2 Finding out about topics of user Interest

If a browser is a frequent user of internet, from his past search behavior and long term search the areas of his interests can be found .so, if such a person gives an inaccurate query, search engine can help him with good results keeping in view of his old track. This has best results when the same topic is dealt [43] when compared to a situation where new topic is taken.
In a situation where the query doesn't have any selection of documents and also when query is asked for the first time, results are given imagining the user interests. As such, this method presents personalized pseudo relevance feed back by expecting the first 10 pages that are taken from the user's local index are related. The words utilized to frame the query vector are chosen from the keywords or synopsis of the top 10 pages. More over the present search result will have the first 10 documents of his old search from the local index. This gives the user a better knowledge of his search record.

3.5.3 Serving an Ad-hoc Information Need

The focus of a user on different topics might be similar but there is every chance that a user changes his focus and may require some information for temporary needs. But the above said model can't turn like this and doesn't consider the context provided in that interaction except for the first query. Query is updated according to the selected pages in a search result.

1. When a result is selected, the browser profile of a query is enhanced with the key words of that particular selected paper. This helps to categorize the results given once again. For all the same kind of queries in that period, that query vector is enhanced to add key words from the selected ones.
2. When a page is selected the query vector of all queries that are nearer to the selected documents are enhanced to add key word from the selected one.
3. To end, if a browser has a processed query, the new query is established with words from the old similar queries within the present period. Again, for calculating query equalities queries are illustrated by the result sets centroid.

Presenting an illustration of the previous queries in the present search is very much important in a situation where the user comes back to a query and goes through its unlocked results, or when same query is submitted through tabbed surfers.

3.6 Limits of the Models Discussed Under Content Relation Based Approach

1. **Data restrictions**; IR techniques can be used only to a certain features of the data, like text and image, and extracted qualities can only find particular angles of the content.
2. **Over-specialization**: There is no chance for over specialization or to stretch away from the given profile, the results are assessed considering the content in it and are filtered. So it is restricted and limited to user profile.
3. **Communal/collaborative**: Till now profile and the previous results are used to personalize, the field 'context' is also of good use and is of a great research are these days.

Search with the help of more than one basis is encouraged by two ideas. First, these search engines with special features gather societies of the browsers whose needs are same and the context is not varied much. For instance, an infrastructure search engine is supposed to gather queries with the same theme, here jaguar queries are more related to infrastructure than vehicles.

Second, by gathering the selections of a group of users, it is very much easy to guess the query document reference as the probability that a page p_j will be selected for query q_t. The special quality of collaborative search is its strength to personalize search results for a group of users, rather depending on the old context analysis methods.

4. Collaborative Search

Collaborative search has no fixed rules about the way a search engine should be and is mostly can be utilized among the content types along with web pages, graphics and photos, audios and video. Finally its new method i.e. ranking metric is computationally good to scrutinize the number of search results and needs no other passing of result pages.

4.1 Collaborative Re-Ranking Of Search Results

This approach combines both the user and society profile to present the accurate search results. The profiles of these people play an important role and search is restricted to their interests. The combined profile of these users is proponed with the help & the content f interest's g the fast research. This profile helps for accuracy in results and in categorizing them. Moreover, the profile is enhanced if any documents are found to be relevant and important. This helps user to find good data. Modification of profile is in the form of µ ranking the words in that list. A profile which gives inputs for a search and Meta search engines can be developed with this collaborative method to rank the search results and reweighting the user and community profiles.

4.1.1: Features of This Architecture

Many features of this architecture magnify the collective interests of the user and community.

The features are:

- Much data collections are associated user and the community and categorize the pages accordingly
- More than one search engine or Meta search engine are made available.
- User information is saved.
- Community is made available.
- Maintaining the group.
- Pre processing for search context.
- Post processing for ranking the combined set of returned results.

Other than above wrappers are made available. That make the users to take information (or page details such as URL), user group and grading news from the page collections as well as wrappers that give the search pre-processor to present queries to(meta-) search engines and the search post-processer to get the results. This approach may use all or few features for personalization, which is explained in the following papers.

4.1.2: Preprocessing For Context Search

Among billions of documents, to give the accurate results many features are as concerned one among them is context. Here context is related to the user profile, his interests, his necessity and nature of the society that he belongs to. These details can be obtained from the user himself by giving a questionnaire and asking him to select the group that matches him most or otherwise the interest can be found if the query matches with any other similar query, which was employed by a group. Since the query is some interests can be the same.

4.1.3 Processes of Profiling Users

Through this step word-weight vector for the individual profile of a user is constructed. These words are selected after referring many documents to which a user might be suitable. This is a very tough job because many documents are available. As such an option is to ask the user provide with the pages that he has selected. But user may not be interested and even if he is interested may not submit all the documents. However, there is a chance that he presents the documents that are suitable to his demands. This challenge can be resolved when a set of pages from a group of people are considered i.e. that the documents from a group. The word-weight vector is estimated in a fixed procedure. For the user the vector consists the group of words { r t } with their weights u i w. If the word-weight vector is given a chance to use the documents, which are clicked and also ranked by the user, there, is every chance for a better profiling and ordering of these words. The user profile can also consider a user in a group. Here profile takes the profile of the user and only those

documents that are selected by the user and added to their group information. Here the problem is finding out the groups, which have similar set of collections. Anyhow if the page selection is a recommends system, the problems will not arise.

4.1.4: Process of Profiling Group

Like the above process the profiling of a group also consist of a word-weight vector. These words are selected from chosen and categorized documents of that particular group. The word-weight vector for the group is decided in a way equivalent to that deployed for users.

4.1.5 Processes of Managing Community

A community is supposed to match its aims, needs, and preference with the documents they selected. But this doesn't match always. This is one big challenge in profiling. Second, gathering information about the companies, which have similar interest and browser, enrolled in those groups. If a community recommends system, which has list of different, groups that are alike can rank them according to their similarities; it will be very useful to provide the group relatedness feedback system as a whole. Truly speaking the administrator is supposed to find the groups that are similar and the documents that suit a group. The matching of these areas can also be done automatically but this may not give the correct profile to a group. Data in community recommend system judges the profile of a group. This profile is dependent on the groups that match each other and the kind of documents they have selected. So this has to be done very carefully. It is of importance because this method can work even though group profile is not available. The strength of this is that it can work even with a single user where no group is available, but is though to deal with new users.

4.1.6 After Search Process (Post Processing For Context Search)

Post processer ranks the documents that have come as a search result. It can be done is two methods. One is matching either the content or the pointer of document with collection of documents already available with the group. If any of these matches, the search result is given importance and is ranked better. Another approach is with the help of word-weight vectors. Depending on the context (single browser, group or expert) the suitable profile is asked for. If the words in that profile match with the key words of a given document, according to the repetition of that word and the standard relevance feedback the search result is assessed.

4.1.7 Process of Updating User Profile Weight

With a query if a user is able to get the documents that are most appropriate, then they will be added to the collection of documents of that particular user or group. With such new pages even user profile can be rearranged and graded properly. In such situations, the query words or mostly used words in the response become the set of related words. To find the relatedness of a search for a query can be assessed with the standard Rochio technique, which depends on this Rel list. With this list, standard Rochio techniques is used other than relevance feedback system. The one big disparity between the present and old model is about the usage of non-relevant words. The present one doesn't consider these relevant words because there is no certain and standard source for this information. Because of rearrangements of words, the related terms from Rel will get their ranks improved in the user profile. A transition of user profiles will automatically enhance the group profile and ranks of the group experts.

4.2 Group Based Approach

4.2.1 Repetition and Regularity in Search Groups

It is usually thought that searching has to be done individually. But these are many techniques, which's take it a common task and indirectly the user in search. For example if a author is interested in understanding the techniques of work place, the portal designed will not only provide documents but also facilitates him with search boxes. These boxes will have documents that are already visited by the people of interest. So they can avail standard wed searches. Here one feature is ignored by these portals but is worth improving i.e. the number of hits a page faces can be used to rank the documents. Another feature is, there is a great chance of establishing of search groups. To illustrate the writer, sales man, or even a research school can come into contact with different kinds of people but with the same interest come in contact through these search boxes and can form a group. Such new groups are more useful due to the great chance that these will have same kind of search behavior. Figure 1 describes the output of an H-week research of the search models for a group of about to employees at a local software company. This research tested more than 20,000 individuals search queries and nearly 16,000 result selection.

It displays the percentage of query repetition at difference similarity thresholds and the least number of similar queries.

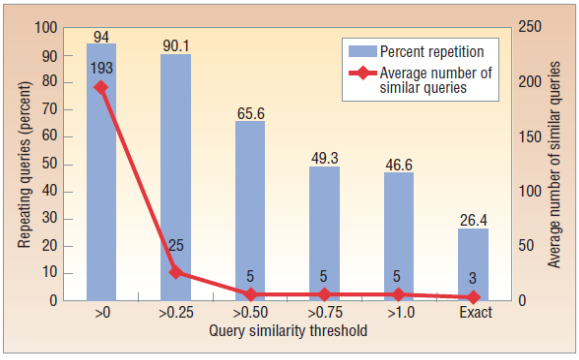

Figure-1 Search methods Statistics of a 17-week research of the group models for a group of about 70 employees at a local software company,

Figure-1 observes at the average equivalence between queries during the research. On average, just over 65 percent of given queries shared at least 50 percent (0.5 similarity threshold) of their query words with at least five other queries; more than 90 percent of queries shared at least 25 percent of their words with at least minimum of 25 other queries. Thus users within this temporary corporate search group found to search in the same way much more so than in generic search scenarios, which typically display less repetition rates of about 10 percent at the 0.5 similarity threshold.

This result, supported by similar studies of other search communities, shows that, in the context of communities of like-minded searchers, Web search is a repetitive and regular activity. As individuals search, their queries and result selections constitute a type of community search knowledge. This in turn suggests that it might be possible to harness such search knowledge by facilitating the sharing of search experiences among community members. As a simple example, when visitors to the wildlife portal search for "jaguar pictures," the collaborative search engine can recommend search results that other community members have previously selected for similar queries. These results will likely relate to the community's wildlife interests. So, without any expensive processing of result content, the search results can be personalized according to the community's learned preferences. This lets novice searchers benefit from the shared knowledge of more experienced searchers.

These percentages, which are reinforced by other research results display that when search done in the environment of like-minded users, it is repetitive and regular activity. This indirectly tells us that it might be likely to track such search erudition by availing the sharing of search practices among group members. To illustrate, when users to work place techniques portal look for images, the common search engine can suggest search results that other group members have previously choose for the same kind of queries. These results will be mostly related to the interests of the user. So, rather going for a long search processing of result matter, the search results can be personalized in accordance with the groups known importance. This helps novice users profit from the common practices of more experienced users.

4.2.2 Collaborative Web Search

The latent search knowledge created by search communities can be leveraged by recording the search activities of users and the queries been submitted and results selected at the community level. This data can then be utilized as the basis for a relevance model to direct the support of community-relevant results during regular Web search. A key objective here is to avoid replacing a conventional search engine, instead enhancing its default result lists by highlighting particular results that are especially relevant to the target community. For example, regarding the "jaguar" queries, it should be possible to promote some of the wildlife community's pages that relate to the wild cat ahead of those related to cats. Thus, the most relevant results from a community perspective can be promoted to the top of the default result list, while other results might simply be labeled as relevant to the community but left in place.

4.2.2.1 Capturing Community Search Knowledge

Capturing a community's search behavior means recording the queries submitted and the results selected for these queries, as well as their selection frequency. This can be conceptualized as populating the community search matrix, HC, called a hit-matrix, such that HC ij refers to the number of times that a result page, pj, has been selected for a query, qi. Thus, each row of a community's hit-matrix corresponds to the result selections that have been made over multiple search sessions by members of C for a specific query qi. In turn, the column of the hit matrix related to pj refers to the number of times that the community has selected pj for different queries.

***4.2.2.2* Making Relevant Promotions**

How it is possible that the present query, qT, be utilized to find out results from a groups hit-matrix as potential promotion candidates? To start with, any other group history with respect to qT must be identified to have any documents been clicked in the past for qT? Thinking that such pages are available the hit-matrix will have frequency selection data with respect to qT, and this data can be utilized to imagine the relevance of each such page. For example, Equation 1 computes the relevance of a result page pj with respect to the query qT as the relative proportion of selections that pj has got for this query:

$$Relevance^{C}(p_j, q_T) = \frac{H^{C}_{Tj}}{\sum_{\forall j} H^{C}_{Tj}} \qquad (1)$$

As it stands, this correct query-relevance method is restricted since it limits users taken for development to those pages previously selected for the specific target query (qT). Certainly, the results given in Figure 2 indicate that just about 25 percent of query requests in the test community accurately match previous request. A better easy method would allow for the consideration of pages that have been chosen for queries very similar to qT. For example, Equation 2 gives a direct way to estimate query similarity by computing the proportion of terms shared by qT and some other query qi.

$$Sim(q_T, q_i) = \frac{|q_T \cap q_i|}{|q_T \cup q_i|} \qquad (2)$$

This query-similarity metric can then be utilized as the ground for a modified relevance metric, as Equation 3 shows:

$$W\,Rel^{C}(p_j, q_T, q_1, \ldots q_n) = \frac{\sum_{i=1\ldots n} Relevance^{C}(p_j, q_i) \bullet Sim(q_T, q_i)}{\sum_{i=1\ldots n} Exists^{C}(p_j, q_i) \bullet Sim(q_T, q_i)} \qquad (3)$$

The relatedness of a page pj, with respect to some target query qT, is estimated by independently computing the exact query relevance of pj with respect to a set of queries (q1, ..., qn) taken to be sufficiently similar to qT; in practice only queries that share 50 percent of their terms with the target query need be considered. The overall relevance of pj with respect to qT is then the weighted sum of the individual exact query relevance values, with the relatedness of pj with respect to some qi discounted by the similarity of qi to qT. In this way, pages frequently selected for queries very similar to qT are preferred over pages less frequently selected for less similar queries.

4.3 Location Based Approach

4.3.1 Personalization By Query Rewriting

Elaborate the original query by user's to personalize the search results given user's location. For example, if the original query is "Chinese restaurants" and the user's location have been decided to be "newyork", a new search query may be framed as "Chinese restaurants Newyork" by relating original search query of the tenacious location. The new search query is then issued to the search engine to get the search result for the user. Because the location "newyork" is now added in the new search query, depending on which the search result is found, the search engine is more likely to get Chinese restaurants situated in newyork. This method, however, has lapses from two sources of errors:

1. The local goal might not be the only targeted of the query or even may not exist due to the limited precision of the general implicit local classifier.
2. The user's location may be decided wrongly.

4.3.2 Personalization By Re-Ranking

As a more traditional approach, page re-ranking is put up to leverage user's location. We first take top K documents and rearrange them in increasing ranks of those documents that match the user location with the original query. We also differentiate and weight user location matches for different sections of the top documents. It re-rank those document based on their present ranking score and text matching that tell if user location and its variation exists in certain document regions. The re-rank score sr(q; d; l) for query q, Web document d and user location *l* can be shown in the following way

$$s^r(q,d,l) = s(q,d) + \sum_i w_i I_i(d,l), \qquad (5)$$

Where, *s (q,d)* is the actual rank score before personalization, $I_i(d; l)$ is an pointer function that defines if user location *l* exist in document section *i* and w_i is the weight parameter for document section *i.* Supervised learning is utilized to assess the parameter w_i to maximize relevance after personalization.

4.4 Search Results Personalization Based On User Profiles

Research analyzed re-ranking for different abstraction levels by using user profiles, which can give different environments for personalization. Pretschner and Gauch [51] brought a system that allows for the automatic development of organized user profiles, which are developed based on a available category hierarchy. Operetta and Gauch [51] suggested establishing user profile based on weighted concept hierarchy, which is developed from the Open Directory Project (ODP). Siege et al. [56] also utilized ODP to know user profiles for personalized web search. ODP at present consists of more than 590,000 concepts or nodes, so they use only few top levels of categories in the ODP hierarchy. As search, the user profiles do not take in to account the low-level categories, which are more particular. Consequently, this may bring down the ranking quality for individuals with more particular leads, not shows as high-level categories in ODP. Also, using a prevailing hierarchy can be made the user profile consist and many irrelevant categories since all high-level categories are there in user profiles. To over come these levels of using an existing taxonomy or hierarchy, Kim and Chan [57] found an approach to construct a user interest hierarchy (UIH) by understanding from implicit user behavior. Later this has been developed by utilizing a scoring function as in [47, 48] for personalized ranking with support of the UIH and taken from bookmarks.

A page can be ranked depending on the user profile and the results given by a search engine. To develop the user profile, called UIH, web pages in user's bookmarks and the scoring function mostly based on DHC would be utilized. A UIH organizes a user's needs from general to particular. Near the root of a UIH, general needs are presented by bigger groups of terms while towards the leaves, more specific needs are presented by smaller groups of terms. The term refers to snippet that has one or more words. The root node contains all unique terms in the bookmarked webpage. The leaf nodes consist more related terms of interests to the user. The bonding of relationship between terms is understood with the help of their co-occurrence in the similar web pages.

4.5 Analysis of User Interests and Activities

Several attempts are made to to personalize the Web search. One such approach personalization is based on users describe general interests like Google Personal which asks users to build there profile by selecting the categories of interests. This profile can used to personalize search results by mapping Web pages in the similar categories. Most commercial use this approaches for information filtering systems, and it has been further explored by Gauch et al. [51] to personalize Web search results. Liu et al. [52] used same approach for mapping user queries to categories based on the user's search history.

User's intentions Information are collected at the time of query by means relevance feedback or query refinement techniques. Koenemann and Belk [53] examined several different interface techniques that diverse in their transparency for allowing users to specify how their queries should be expanded. McKeon et al. [58] describes the alternate methods for generating query refinements. Very short-term model of a user's interest are Relevance feedback and query refinement harness which require a query first be issued and then modified. Generally it was observed that people are

typically not unwilling to spend extra effort on specifying their intentions. Tee van et al. [54] study suggests that instead of fully specifying their search goals up front, some time people often browse to their targets via pages identified by less precise but more easily specified queries. People are motivated to spend additional effort on specifying their search intent, but they are not always successful.

An automated analysis of activities and interests for search results personalization has been proposed by Jaime Tee van et al [54]. They formulated some search algorithms that prior to users considers interactions with a long variety of content to personalize that user's current Web search. People are not relying on the unrealistic assumption that will precisely describes their interests while searching, this technique is attempting to leverage implicit information about the user's interests and using that to re-rank Web search results within a relevance feedback framework. The approach attempting to explore rich models of user interests which built from both search-related information which works on previously issued queries and previously visited Web pages, and other information created about the user such as documents and email the user has read. But the accuracy of this approach is strictly depending on sufficient representations of the user for personalization.

5. Conclusions

Personalized search on the Web is a research field that has been recently gaining interest, since it is a possible solution to the information overload problem. The reason is quite simple: information plays a crucial role for each user, and users are constantly challenged to take charge of the information they need to achieve both their personal and professional goals. The ability to filter and create a personalized collection of resources simplifies the whole searching process, increasing search engine accuracy and reducing the time the user has to spend to sift through the results for a given query. The novelty and liveliness of the personalization field suggests that, over the next few years, new and interesting algorithms and approaches will be proposed and probably transferred to the information systems with which users interact in everyday use, such as, search engines or desktop search tools. Anthologies and the Semantic Web are two important research fields that are beginning to receive attention in this context.

References

[1]. Montaner, M., Lopez, B., Rosa, J.L.D.L.: A taxonomy of recommender agents on the internet. Artificial Intelligence Review 19 (2003) 285-330

[2]. Savoy, J., Picard, J.: Retrieval effectiveness on the web. Information Processing & Management 37(4) (2001) 543-569

[3]. Khopkar, Y., Spink, A., Giles, C.L., Shah, P., Debnath, S.: Search engine personalization:An exploratory study. First Monday 8(7) (2003) http://www.firstmonday.org/issues/issue8_7/khopkar/index.html.

[4]. Salton, G., McGill, M.: An Introduction to modern information retrieval. Mc-Graw-Hill, New York, NY (1983)

[5]. Micarelli, A., Sciarrone, F., Marinilli, M.: Web document modeling. In Brusilovsky, P., Kobsa, A., Nejdl, W., eds.: The Adaptive Web: Methods and Strategies of Web Personalization. Volume 4321 of Lecture Notes in Computer Science. Springer-Verlag, Berlin Heidelberg New York (2007) this volume

[6]. Olston, C., Chi, E.H.: ScentTrails: Integrating browsing and searching on the web. ACM Transactions on Computer-Human Interaction 10(3) (2003) 177-197

[7]. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM 30(11) (1987) 964-971

[8]. Freyne, J., Smyth, B.: An experiment in social search. In Bra, P.D., Nejdl, W., eds.: Adaptive Hypermedia and AdaptiveWeb-Based Systems, Third International Conference, AH 2004, Eindhoven, The Netherlands, August 23-26, 2004, Proceedings. Volume 3137 of Lecture Notes in Computer Science., Springer (2004) 95-103

[9]. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. Proceedings for ACM conference on Computer supported cooperative work, New York, NY, USA, ACM Press (1994) 175-186

[10]. Dieberger, A., Dourish, P., H¨o¨ok, K., Resnick, P., Wexelblat, A.: Social navigation: techniques for building more usable systems. Interactions 7(6) (2000) 36-45

[11]. Kritikopoulos, A., Sideri, M.: The compass filter: Search engine result personalization using web communities. In Mobasher, B., Anand, S.S., eds.: Intelligent Techniques forWeb Personalization, IJCAI 2003Workshop, ITWP 2003, Acapulco, Mexico, August 11, 2003, Revised Selected Papers. Volume 3169 of Lecture Notes in Computer Science., Springer (2003) 229-240

[12]. A. Broder. A taxonomy of web search. SIGIR Forum, 36(2), 2002.

[13]. B. Jansen and A. Spink. How are we searching the web? a comparison of nine search engine query logs. Information Processing and Management, 42, 2006.

[14]. O. Madani and D. DeCoste. Contextual recommender problems. In Utility Based Data Mining Workshop at KDD, 2005.

[15]. M. Pazzani and D. Billsus. Learning and revising user profiles: The identification of interesting web sites. Machine Learning, 27, 1997.

[16]. M. Speretta and S. Gauch. Personalizing search based on user search histories. In IEEEWICACM Interntional Conferenceon Web Intelligence (WI), 2005.

[17]. Kleinberg, J.: Authoritative sources in a hyperlinked environment. Journal of the ACM 46 (1999) 604-632

[18]. Shannon Bradshaw and Kristian Hammond: Automatically Indexing Research Papers Using Text Surrounding Citations. In Working Notes of the Workshop on Intelligent Information Systems, Sixteenth National Conference on Artificial Intelligence, Orlando, FL, July 18-19

[19]. Page, L., Brin, S., Motwani, R., Winograd, T.: The Page Rank citation ranking: Bringing order to the Web. Technical report, Stanford University Database Group (1998)

[20]. Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., Raghavan, S.: Searching the web. ACM Trans. Inter. Tech. 1 (2001) 2-43

[21]. Langville, A.N., Meyer, C.D.: A survey of eigenvector methods of Web information retrieval. SIAM Review (Forthcoming)

[22]. Mobasher, B., R. Cooley, J. Srivastava. 1997. Web mining: Information and pattern discovery on the World Wide Web. Proc. of the 9th IEEE Internat. Conf. on Tools with Artificial Intelligence (ICTAI'97), Newport Beach, CA, 558-567.

[23]. Henzinger, M. 2000. Link analysis in web information retrieval. Bull. of the Technical Committee on Data Engrg., IEEE Computer Soc. 23 3-9.

[24]. Baumgarten, M., A. G. Büchner, S. S. Anand, M. D. Mulvenna, J. G. Hughes. 2000. Navigation pattern discovery from internet data. M. Spiliopoulou, B. Masand, eds. Advances in Web Usage Analysis and User Profiling, Lecturer Notes in Computer Science 1836 70-87.

[25]. Armstrong, R., D. Freitag, T. Joachims, T. Mitchell. 1995. Web-Watcher: A learning apprentice for the world wide web. AAAI Spring Sympos. on Inform. Gathering from Heterogeneous, Distributed Environments, Stanford, CA, 6-13.

[26]. Srivastava, J., R. Cooley, M. Deshpande, P. N. Tan. 2000. Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations 1 12-23.

[27]. Lin, I. Y., X. M. Huang, M. S. Chen. 1999. Capturing user access patterns in the web for data mining. Proc. of the 11th IEEE Internat. Conf. Tools with Artificial Intelligence, Chicago, IL, 22-29.

[28]. Cooley, R., B. Mobasher, J. Srivastava. 1999. Data preparation for mining world wide web browsing patterns. Knowledge and Inform. Systems 1 5-32.

[29]. Spiliopoulou, M., B. Mobasher, B. Berendt, M. Nakagawa. 2003. Evaluating the quality of data preparation heuristics in web usage analysis. INFORMS J. on Comput. 15(2) 171-190.

[30]. Cooley, R., B. Mobasher, J. Srivastava. 1997. Grouping web page references into transactions for mining world wide web browsing patterns. Proc. of KDEX'97, IEEE, Newport Beach, CA, 2-9.

[31]. Shahabi, C., A. M. Zarkesh, J. Adibi, V. Shah. 1997. Knowledge discovery from users web page navigation. Proc. of the IEEE RIDE97 Workshop, Birmingham, England, U.K., 20-31.

[32]. Shahabi, C., F. Banaei-Kashani, J. Faruque. 2001. A reliable, efficient, and scalable system for web usage data acquisition. WebKDD'01 Workshop, ACM-SIGKDD 2001, San Francisco, CA. http://dimlab.usc.edu/Research.html.

[33]. Yan, T. W., M. Jacobsen, H. Garcia-Molina, U. Dayal. 1996. From user access patterns to dynamic hypertext linking. Fifth Internat. World Wide Web Conf., Paris, France, 1007-1118.

[34]. Cadez, I., D. Heckerman, C. Meek, P. Smyth, S. White. 2000. Visualization of navigation patterns on web site using model based clustering. Technical Report MSR-TR-00-18, Microsoft Research, Microsoft Corporation, Redmond, WA.

[35]. Levene, L., G. Loizou. 2000. Zipf's law for web surfers. Knowledge and Inform. Systems 3 16-24.

[36]. Mobasher, B., H. Dai, T. Luo, M. Nakagawa, Y. Sun, J. Wiltshire. 2000a. Discovery of aggregate usage profiles for web personalization. Proc. of the Web Mining for E-Commerce Workshop WebKDD'2000, Boston, MA. http://maya.cs.depaul.edu/mobasher/papers/webkdd2000/webkdd2000.html.

[37]. Borges, J., M. Levene. 1999. Data mining of user navigation patterns. Proc. of Workshop on Web Usage Analysis and User Profiling (WEBKDD), ACM SIGKDD Internat. Conf. on Knowledge Discovery and Data Mining, San Diego, CA, 31-36.

[38]. Breese, J. S., D. Heckerman, C. Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. Proc. of ncertainty in Artificial Intelligence, Madison, WI, Morgan Kaufmann, San Francisco, CA.

[39]. Fu, Y., K. Sandhu, M. Shih. 1999. Clustering of web users based on access patterns. International Workshop on Web Usage Analysis and User Profiling (WEBKDD'99), San Diego, CA, 18-25.

[40]. Joshi, A., R. Krishnapuram. 1999. Robust fuzzy clustering methods to support web mining. Proc. of SIGMOD Workshop in Data Mining and Knowledge Discovery, Seattle, WA, 1-8.

[41]. Strehl, A., J. Ghosh. 2003. Relationship-based clustering and visualization for high dimensional data mining. INFORMS J. on Comput. 15(2) 208-230.

[42]. Perkowitz, M., O. Etzioni. 1998. Adaptive web sites: Automatically synthesizing web pages. Fifth National Conf. in Artificial Intelligence, Cambridge, MA, 727-732.

[43]. Paliouras, G., C. Papatheodorou, V. Karkaletsis, C.D. Spyropoulos. 2000. Clustering the users of large web sites into communities. Proc. of the Internat. Conf. on Machine Learning, Stanford, CA, 719-726.

[44]. VanderMeer, D., K. Dutta, A. Datta, K. Ramamritham, S. B. Navanthe. 2000. Enabling scalable online personalization on the web. Proc. of the 2nd ACM Conf. on Electronic Commerce, Minneapolis, MN, 185-196.

[45]. J. Kleinberg. Authoritative sources in a hyperlinked environment. In Proceedings of the ACM-SIAM symposium on Discrete Algorithms, 1997.

[46]. F. Liu, C. Yu, and W. Meng. Personalized web search for improving retrieval eectiveness. In IEEE Trans. on Knowledge and Data Eng., 2004.
[47]. F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In WWW, 2006.
[48]. B. Tan, X. Shen, and C. Zhai. Mining long-term search history to improve search accuracy. In KDD, 2006.
[49]. J. Teevan. The re:search engine - helping people return to information on the web. In UIST, 2005.
[50]. Y. Xu, B. Zhang, Z. Chen, and K. Wang. Privacy-enhancing personalized web search. In WWW, 2007.
[51]. Gauch, S., Chafee, J. and Pretschner, A. (2004). Ontologybased personalized search and browsing. Web Intelligence and Agent Systems, 1(3-4): 219-234.
[52]. Liu, F., Yu, C. and Meng, W. (2002). Personalized Web search by mapping user queries to categories. In Proceedings of CIKM '02, 558-565.
[53]. Koenmann, J. and Belkin, N. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In Proceedings of CHI '96, 205-212.
[54]. Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In Proceedings of CHI '04, 415-422.
[55]. Nielsen, J. Personalization is overrated. In Jakob Nielsen's Alertbox for October 4, 1998. http://www.useit.com/alertbox/981004.html.
[56]. A. Sieg, B. Mobasher, and R. Burke. A large-scale evaluation and analysis of personalized search strategies. In Proc. CIKM, pages 525-534, 2007.
[57]. H. Kim and P. Chan. Learning implicit user interest hierarchy for context in personalization. In Proc. Intl. Conf. on Intelligent User Interfaces, pages 101-108, 2003.
[58]. McKeown, K. R., Elhadad, N. and Hatzivassiloglou, V. (2003). Leveraging a common representation for personalized search and summarization in a medical digital library. In Proceedings of ICDL '03, 159-170.

