ISSN:0975-9646

A Survey on Personalized Search: An Web Information Retrieval System

Rutuja S .Lachake
M.E. Information Technology
PICT,Pune University

Prof. G. P. Potdar
Prof & Head of Computer Department
PICT,Pune University

Abstract— Nowadays, computers and internet has become inseparable parts of our life. Throughout the world, web has become the best source of abundant information. Search engines play a key role in finding out the information; they are enhanced with new advanced search technologies. Though search engines find much information with one key word, fail to provide the accurate, exact data that is required. Through 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 can be solved by personalized web environments. There is an increasing importance to the personalized web environment. This paper defines 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 Behaviour, Web Search, Personalization, Information Retrieval

I. INTRODUCTION

The main goal of the many search engine developers is improving the accuracy of the search results of the personalized websites with some other long term goals. To answer all the necessities of the browsers and to bring the best search engine applications; much importance is given to areas like better browsing, localization, question and answer methodology, visual results presentation and modified web search. Each technology relates to a different aspect of people's information behaviors; showing that different technologies should be developed to answer different demands.

Till now not much development is observed. Thus the web remained same. Also not much development and innovation is brought in this web personalization field because individual web search behavior has not changed much. Here in the case of web personalization, the main challenge is to read the mind of the users, imposing a very big challenge because the words used for any search are limited to two or three words. Similar difficulties are faced in localization and personalization of web information at the user acceptance side and software development side. So, by solving these difficulties, there arises the possibility of accurate information at the very first search. Following mentioned are some of the issues in Web searching:

- *Structuring Queries*- The difficulty faced by users are properly structuring queries, namely applying the rules of a particular system, especially Boolean operators (e.g., AND, OR, NOT) and term modifiers (e.g., '+', '!').
- *Spelling*- One reason for reducing the number of results retrieved is the users generally misspell terms in queries. It is often difficult to detect these spelling errors as these queries frequently retrieve results from large document collections. The user is unable to realize that the query contains a spelling mistake.
- *Query Refinement*- Many times the searchers do not refine their query, even if there may be other terms that relate directly to their needed information. Studies show that searchers rarely modify their queries, or do it incrementally, and then typically only for one or two times.
- *Managing Results*- There is trouble to the searchers in the management of the number of results. Mostly, the user queries are extremely broad, resulting in an unmanageable number of results. Few searchers view more than the first ten or twenty documents from the result list.

The introduced feature Personalization of a search engine will provide great utility and be a time saving tool for all of the browsers. It takes the search engine directly to the information and topics which had already been taken to by the user previously. Thus further helping the browser to collect the needed accurate information without being troubled by all the other unwanted data, which may be related to the topic but not relevant to the demands. But sometimes the situations occur where the browser doesn't want the search engine to have the knowledge about them resulting in rejection of this new application. Leaving these problems behind, Google and AskJeev are taking all the trouble to bring it intact.

A. *Making It Personal*

Till now, the browser is not given idea of any search engine resulting the browser doesn't know the requirements of any search engine. This shows that search engines need to be enhanced to understand the user in a better manner. One of the biggest hurdle is that computer software is not good at understanding people's information needs.

To give accurate information to the user, Google is trying to develop different techniques. The filtered results are

obtained if the user provides 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 question and answer technology, is trying to be the best even in personalization. Its new technology i.e. My Jeeves, 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 refined search result. The excessive data can't answer the demands of the user perfectly and exactly. Another problem is that, for all the queries same results are given, which may be similar but are asked by different users with varied needs. Personalization is one solution, which understands the needs of a particular user and gives filtered information; which comes along with the constraint of understanding the mind of a user. The analysis of the search behaviors shows that there are three techniques that are mostly used by users, first begin 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]; thus giving the filtered results.

Another technique which is browsing is used to find information. In this, the user goes ahead with hyper links by migrating through pages and selecting any one page. This is very helpful because one can get information from any part of the world and research scholars are much benefited. But due to large amount of information which is available, it becomes difficult to find the exact website. User has to spend a lot of time searching for the suitable one even with the most structured websites like Yahoo! Directory and the open Directory project 2. This implies that, the information is available, but accessing them is a difficult task.

When a query is given to a search engine, surfing is done with the help of Information Retrieval (IR). To answer a query, the search engine selects some websites, which can give the query answer. 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 said problem, one possible solution is Information Filtering (IF), which is modeled on the interests of the user. This technology believes that user's needs are static and at least do not change with the same rate as that of websites. The web is smartly updated, removed and added; so the IF can be the best way to reach information. Since it takes a user to be static, it doesn't trouble the user with abundant information but matches his interests with the information and continuous to update that related data. But one big challenge exists, which is 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.

Information Retrieval (IR) and Personalization are two different techniques 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 an innovative technique, more efficient than the previous ones. The paper focuses and elaborates on this new technology and its working system. To access any data in Internet, through surfing or a query, the key point is searching. Following any of the above said methods will confuse author with huge amount of information. Query directs a user to the relevant documents from many documents of different topics available [2]. Regardless of his requirements, these search engines will give the same results to any kind of user. This implies that these need to be improved for the sake of better search. Personalization is one such step that helps search engine to have unique aggregation of documents for any line of a topic which 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. Thus helping the user to 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 been evolved, but very few concentrate on user interaction. It is also observed by the users that it is a difficult task 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]. Few of its Cons are:

- The purpose, requirements and interest of a user will be saved in an external search system and this is not liked by many.
- Personalization of web search results is a computationally intensive tough task.
- A personalized search engine takes considerable time, which is done within a second by general browsers.
- To bring securitized results according to the interests of a user is not very easy causing an obstacle.

Present paper familiarizes and introduces a user with the procedures, methods and techniques of a personalized search engine. Personalization approaches aim to:

- Tailor search results to individuals based on knowledge of their interests

- Identify relevant documents and put them on top of the result list
- Filter irrelevant search results
- Maintains user's profiles representing the interest of users
- Make use of the context

II. RELATED WORK

A. *Personalization Based Web Usage Mining*

Web mining is nothing but a careful systematic search and evaluation of the documents available in World Wide Web. Web-mining is related to the information and its features are divided in the procedure given below:

- *Content Data*: These are the documents that are available to the browser. Content mining is deriving information from the material of the web pages.[4]
- *Structure Data*: Web structure mining is the way of selecting information from the structure of data.[5]
- *Usage data:* The data which is taken from the browser is connected to the web. As cited above [6] Web Usage Mining (WUM) is the exploration and evaluation of browser access to the web information system using the data available to customize the web for user was not any new idea but was suggested way back in year 1995. [7]

1) User-Interaction Tracking:

The data about the transactions 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, strict attention to minute details 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.

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) [8] has invented an "access pattern collection server" to overcome the above said problem which works only when user secrecy doesn't matter. Cookey et al. (1999) [9] has used referrer and agent fields of a server log to obtain the information about the stored references that are hit back. Spiliopoulor et al. (2003) [10] analyzed the output of many such methodologies. It is found that server and proxy logs are unable to provide the temporary 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 inspected 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 reliable 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. The documents are reorganized and grouped for a better understanding and analysis and should be divided according to the key words. In Shahabi et al. (1997) [11], 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 a few oversights. 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) [12] elaborated on this information collecting methods depending on the user-side data collecting idea.

1) Access Pattern Analysis:

Digging in all the usage data is not possible because they are enormous in amount. The basic method is that, the value or grade of a paper is estimated according to the number of hits that it has faced by the users. 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, 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 surroundings of web personalization applications. Yan et al.(1996) [13] and Lovene and logic (2000)[15] believe 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 which is explained is capable of combining the above said and many other qualities.

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 Mobasher et al [16] 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 employ a Markov method for this purpose.

Borges and Levene [17] explain a probabilistic regular grammar whose higher probability strings coincides to browsers selected access methods. Breese et al [18] 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 depend on the kind of application and wholeness of the usage data. Grouping to mine usage data methodology was initiated by Yan et al. [13]. With this method, browser terms are generally structured vectors. In the domestic 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 surroundings of WUM; Perkowitz and Etzioni [22] presented a new grouping algorithm, cluster miner, which is developed to answer particular web-personalization necessities; Fu et al. [19] employ BIRCH [13], an efficient hierarchical clustering algorithm; Joshi and Krishnapuram [20] prefer a fuzzy relational clustering algorithm for WUM because they believe usage data are fuzzy in nature; Strehl and Ghosh [21] propose relationship-based clustering for high dimensional data mining in the context of WUM. Paliouras et al [23], from the machine-learning society correlate achievement of cluster miner with two other grouping procedures which are vibrant in machine-learning research, for example, auto class & self organizing maps, and display that Auto-class is better than other procedures. Mobasher et al [16] 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 [24] 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 importance.

III. OVERVIEW OF PERSONALIZATION

It is difficult to personalize World Wide Web because web is a place for human to human communication whereas personalization requires software system to take part in interaction. Personalization system requires knowledge to be represented in a machine interpretable form which is not available in web. In semantic web we can develop languages for expressing information in machine process able form therefore semantic web is the most appropriate platform for understanding personalization [25].

A. *Semantic Web Personalization*

1) Objective of Semantic Web Personalization:

Two main objectives of semantic web personalization are to perform content-aware navigation and fruition of the resources. To identify the most appropriate resources, knowledge is used along with the descriptive keywords. The main advantage of using knowledge is that the exactness of the answers is increased. In semantic web the answers are always personalized or adapted so as to meet specific requirements characterizing the semantic web.

2) Advantages of Semantic Web Personalization:

The main advantage of semantic web is enriching web data, which is usually represented in HTML or other XML formats. It incorporates intelligent reasoning capabilities in web based systems. Semantic web based personalization has several advantages over web based personalization few of them are as follows:

- Uniformity in representing knowledge
- Domain models
- Resource Description Format (RDF) and formal reasoning

IV. CLASSIFICATION OF SEMANTIC WEB PERSONALIZATION

The main specializations of semantic personalized recommender systems are classified as follows:

- Vocabulary or ontology based system,
- Context-based recommenders
- Trust network-based recommenders,
- Rule based filtering,
- Content based filtering ,
- Collaborative filtering ,
- Hybrid Recommendations.

A. Ontology Based System or Vocabulary

Ontology based systems can be used to personalize the semantic web by using the concepts of domain ontology containing the information regarding the domain of interest in an ontology format. The following table shows the comparative study of ontology approach's.

TABLE I. COMPARISON OF ONTOLOGY APPROACH'S

Ontology Approach	***User Information***	***Method***	***Steps***	***Personalization***
Domain Ontology (Content) [26]	User Model	Classification Ontology	Restrictions are defined	Restrictions are combined with the user model
Graph Based Method (Search) [27]	User profile as Graph	Session Boundary recognition method	Conceptual correlation is calculated	Based on the user profile, the search results are re-ranked

TABLE II. COMPARISON OF CONTEXT APPROACH'S

Context Approach	***User Profile (Search)*** [29]	***User Activity (Search)***[30]	***User Search History (Search)*** [31]	***Concept Level (Content)*** [32]	***Context History (Content)***[33]
User Context	User ontologi-cal profile	User context-ual profile as classifier	User search history as weighted dominant keywords	Raw context abstracted into a set of concepts using fuzzy set theory	Users' profile, the current & past context, users actions for the past context and the services
Method	User context model	Docum-ent Profile	Learning algorithm based on correlation measure	Weight for each concept based with respect each user are determined	Users' preferences and association rules are used for calculating inference
Steps	Spreading activation	Similarity calculated using cosine measure	Relevant documents for each session	Items are represented as concepts, Aggregated context is calculated	Context history is used to stores and classify the user profiles services
Personalization	Re-ranking the search results based on user model and profile	Similarity between document profile and contextual profile	Statistical method induces beliefs on user context	Similarity between users current context and aggregated context is computed, correlation is performed	Reasoning users' preferences from context history ,infers the association rules

B. Context-Based Personalization

Context represents any information that can be used to characterize the situation of an entity where an entity can be a person, place, or object that is considered relevant to the interaction between a user and an application, including location, time, activities, and the preferences of each entity. In short it can be said that the user's aim for information acquiring attempt is represented by context [28]. The user context is automatically collected by context aware computing which extract information that is relevant to current context. Context awareness computes a broad range of contextual attributes such as the user's activities, current positions, and their surrounding environments to better understand what the user is trying to accomplish, and what services the user might be interested.

1)Advantages of Using Context Aware Systems in Semantic Web Personalization:

Semantic web personalization can use the benefits of existing ontologies which forms the backbone of semantic web. The several reasons for developing context-aware systems based on ontologies are:

- Context ontology sharing
- Ontology reuse
- Logic reasoning mechanisms

2)Personalized Web Search:

Personalization involves the process of collecting user-specific information during interaction with the user, which is further used to deliver appropriate results to the users based on their information needs. Personalized web search helps the user to find the information on web according to his/her preferences. The Table II shows the comparative study of the context approaches.

3)Trust Network Based Systems:

Semantic web is described to be a web of knowledge having properties such as heterogeneity, openness and ubiquity. In Semantic web environment everyone has the ability to contribute, trustworthiness of the people and their contributions are of great importance and value. Therefore trust plays a crucial role in bringing the semantic web to its full potential.

4)Rule based Recommendations:

In Rule based recommendation basically a set of rules are used to make personalized recommendations. Rule based system uses information stored in web logs to extract patterns of usages which are used to device the rule [34].

5) Semantic Content Filtering:

Semantic content based filtering is based on using semantic relations. Semantic content based filtering can make better content based recommendation by addressing the two most significant problems encountered during traditional content based filtering.

Cold start problem- This problem occurs when there are not enough user ratings. By using semantic content based filtering and by retrieving more semantically related concepts, this problem can be partially solved.

Over-specialization problem-We can provide more interesting or surprising recommendations for concepts by using combinations of content feature and semantic relations in semantic content based filtering, which can partially resolve the over-specialization problem which means that the user is restricted to getting recommendations those having a

strong resemblance to the ones he already knows or defines in the user profile.

6)Semantic collaborative filtering:

In this recommendation system we incorporate the semantic knowledge to enhance the performance of traditional collaborative filtering recommendations system. Traditional systems computes exact match to find similar users and items whereas semantic collaborative filtering system uses semantic match. The main advantages of semantic collaborative filtering are:

- To reduce item scattering problem in semantic collaborative filtering, users and items are mapped to a set of concepts in domain ontology.
- By recommending items that have high semantic similarity, it reduces the cold start problem.
- User preferences can be expressed as ratings.

7) Hybrid Recommendations:

The content based filtering techniques and collaborative filtering techniques and rule based filtering are combined to generate hybrid recommendations that take the advantages of all filtering techniques, thus producing high quality recommendations.

V. PERSONALIZED SEARCH OVERVIEW

A. *Personalization Techniques*

Search engines are mostly dependent on IR technology, for example Vector Space Model (VSM) [35], which 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 indication, which is formulated by using the information given in that web page. In the web world any web page will have a title name to be displayed which 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 [36]. If a user has no distributed knowledge about his need then he will not be able to use the appropriate correct word in his query resulting in slowing down of his search. Also, the reverse is equally possible [37]. But to the date, a research on search behavior about the usage of the accurate words shows that users usually restrict themselves to 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 example, one website organizer may a use a word which he likes most but a user for the same information may go for its synonym. Then tracing of that web page will not be possible for the search. Synonyms mean many words expressing same meaning and poli-semis mean one word with different meanings. Due to the above said language richness, the keyword method used by search engines suffers a lot [38]. There are maximum chances that a search gives unrelated results from billions of documents because title of these documents may have many synonyms and polysemis [39]. 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 rather than depending on key words. This understanding will enable a search engine to correlate well with the given query.

A Collaborating Method is the one in which a search engine considers even the profile of different users to present the search results [40]. The profile consists of the needs, context and interests of a user. If the above said match for any two different users then the results can be interchanged and thus helping to achieve accuracy. Dieberger et al [41] coined a new word for this approach i.e. Social Navigation. Social Navigation 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.

B. *User Modeling in Personalized Systems*

Modeling or profiling is a technique used in the personalization of a search engine. This technique stores the profile of a user and the kind of questions asked by him. To analyze users search behavior and to present relevant search results this collection is very useful. 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 user's needs. The more complex a personalized is, in a search engine the results will be much better. In this methodology interests of the user will be provided by him to get accurate results.

The methodology will have a format, with 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.

VI. DEPARTMENT OF INFORMATION

In this technique user is not troubled, since there is no necessity of profile. It is totally dependent on data but not on user. This technique uses the information in the web pages in order to give filtered results to a user.

A. Client Side Ontology Based Personalization

The methodology followed in the above said system is that according to the interests of the users, results will be grouped under different concepts. Here the psychology needs of the users play an important role as such we find in this technology different group of contents. The main job 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], this presents a content related relation between the search keywords 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.

B. Query Log Analysis

Many query logs has been analyzed [42] [43]. So many have been 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 [44]. The use of learning from impressions and clicks to improve the ranking is also an exciting direction [47].

VII. LOCATION BASED APPROACH

A. 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:

- 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.
- There are the chances that the user's location may be decided wrongly.

B. Personalization by Re-Ranking

As a more traditional approach, page re-ranking is put up to take the strategic advantage of the user's location. First taken are the top K documents which are rearranged 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.

VIII. CONCLUSIONS

Personalized search on the Web is a research field that has been gaining interest recently, since it is a possible solution to the information overload problem. The reason is quite simple: information plays a vital 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. 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. Ontologies 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)

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

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

[37] 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.

[38] 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.

[39] Eirinaki Magdalini and Vazirgiannis Michalis,"Web Mining for web personalization", *ACM Transactions on Internet Technology,Vol. 3, No. 1, (February 2003),*2-24

[40] Tvaroˇ Zek Michal,"Personalized navigation in the semantic web", Lecture Notes in Computer Science, Volume 4018, Adaptive Hypermedia and Adaptive Web-Based Systems,*Springer(2006), 467-472*.

[41] Daoud Mariam, Tamine-Lechani Lynda and Boughanem Mohand,"Using a graph-based ontological user profile for personalizing search", *,ACM. CIKM'08(October 26–30, 2008)*

[42] Hong Jong-Yi, Suh Eui-Ho and Kim Sung-Jin, "Context-aware systems a literature review and classification",*Elsevier Expert Systems with Applications 36 (2009) 8509–8522*.

[43] Sieg A., Mobasher B., and Burke R., "Learning ontology-based user profiles: a semantic approach to personalized web search", *IEEE Intelligent Informatics Bulletin, vol. 8, (2007).*

[44] Challam V., Gauch S., and Chandramouli "A.,Contextual search using ontology-based user profiles", RIA2007, (2007).

[45] Tamine Lynda, Boughanem Mohand and Zemirli Nesrine, "Inferring the user interests using the search history", *LWA 2006,(2006).*

[46] Shin Dongmin, Lee Jae-Won, Yeon Jongheum,and Lee Sang-Goo,"Context-aware recommendation by aggregating user context",*proceedings of 2009 IEEE Conference on Commerce and Enterprise Computing(2009).*

[47] Hong Jongyi,-Ho Suh A, Eui , Kim Junyoung ,and Kim Suyeon, "Context-aware system for proactive personalized service based on context history", *,Expert Systems with Applications Elsevier 36 (2009) 7448–7457.*

[48] Barla Michal, Tvaroˇzek Michal,and Bielikov M´Aria "A, Rule-based user characteristics acquisition from logs with semantics for personalized web-based systems" *Computing and Informatics, Vol. 28, (Sep-22 2009). 399–427*