Development of Framework for Efficient Personalized Web Search using Dynamic Profiling and collaborative filtering for Recommendation System

Dr. Mohd Ashraf

[1]*Associate Professor, Department of CSE, Maulana Azad National Urdu University, Hyderabad, India*

***Abstract*— In today's world, every information available on the World Wide Web is in digital form. Different Information retrieval (IR) technique is used for retrieving required information on this huge stack of information, but there is great difficulty in retrieving relevant information according to user preferences in return of simple short query with generic web search. So, we need to enhance the power of web search to retrieve relevant information. This research is an attempt to improve the search efficiency of web search. In this paper, we proposed a framework for personalized web search which uses a dynamic user profile to automatically update user profile and collaborative filtering for considering recommendation which helps to retrieve search result and relevant document to user according to its need and preferences by diagnosing its web search behavior according to previous search history. The Experimental result shows the proposed method of personalized web search is more efficient over generic search engine.**

Keywords- Collaborative filtering, Recommendation System, search Engine **Personalized web search, Information Retrieval, Apache Nutch, Dynamic user profile, Search history.**

I. INTRODUCTION

The information becomes an essential part of our daily life operations, but its huge quantity makes it difficult and time-consuming to retrieve relevant information according to the user's preferences. An important drawback of the existing system is that it's not flexible to user's need and interest [2]. Two major fundamental problems with web search are information mismatch and irrelevant document result. User's interaction with web search engine can be characterized as one size fit for all [3]. For example, One user may be a computer engineer who thinks 'mouse' as a computer peripheral device on the other hand another user may be a biological researcher who thinks 'mouse' as a mammal. The Semantic meaning of term 'mouse' differ according to individual user. This paper proposed a personalized web search framework to enhance the power of generic web search. This approach helps to solve with generic web search. The proposed framework work in a two-step strategy to improve information retrieval efficiencies. (1) In the first step, the search result will be drawn for each user with a small query submitted by the user, based on browsing search history of the user. (2) In the second step, the system uses the search result retrieved in the first step and follows with these steps set of categories to augment the query to conduct the web search. Specifically, the strategy provide are:

i) Evidence Analysis: gather the user's browsing, search history.
ii) Query Expansion: Expand query on the basis of user past search history.
iii) Construct a dynamic user profile based on the search history.
iv) Collaborative Filtering: To find apropos categories for the specific query in accordance with user's profile using k-nearest neighbor clustering. It's a prominent technique used for as the user profile learner.

This Paper contains seven sections. Section II describes the main contribution of the paper, Section III Literature Survey. Section IV describes Architecture of System; Section V describes user profile construction. Section VI describes proposed framework and algorithm, Section VII describes Experimental result.

II. MAIN CONTRIBUTION

The contribution of this paper is a novel approach for efficient personalized web search. Different already existing personalized web search has been suggested by different authors time to time, but the paper defines a an approach which tries to learn the behavior of user search to make search result relevant to the user [7]. In specific, the following has been achieved:

- An automatically construction and update user profiles has been defined.
- Clustering grouped the web pages into specific category it belongs.
- Dynamic user profile is built to adapt the changes in according to the interest of the user search.
- The trained user profiles to learn its behavior. It can also work as a recommendation system.

III. RELATED WORK

Fang Liu et al., [4] proposed an approach for personalized web search for intelligent information retrieval. In this paper, the authors proposed an approach to extract information from profiles from user's browsing histories. The user profile helps to enhance the information retrieval power of web search because it give information about the user search behavior and they provides method to decide which set of categories and based on the past browsing history of user. This approach uses two combining techniques of profile generation 1) semi-automatic 2) automatic.

Amruta Mantra et al., [1] proposed an approach to solving a problem with personalized web search to a certain extend by extending NUTCH open source search engine using users personalized information. The user information will be extracted from the social networking websites. In this paper, an author proposed a 'Profile Biasing Algorithm' to compute the final rank of each page using Profile Biasing *Index (PBI) and Page Rank (PR) as given by NUTCH. PBI calculation based on the number of keywords based on user databaseTitle and Author Details*

IV. ARCHITECTURE

A. Search Engine

Search is a technique used to find relevant information for the user on the internet or computer system. The search engine is computer software designed to search for desired information on WWW. As a demand for technology growing day to day we use search engine frequently. Search engine crawls the web pages according to its URL address and index the web page using Indexer which contains the keyword which matches with user query on specific web pages and shows the result. Search engine regularly updated index as internet growing at an exponential rate following Moore's law to give a result quickly and efficiently [1].

B. Apache Nutch

NUTCH is an open source search engine, and its efficiency can be extended as per the need of development.

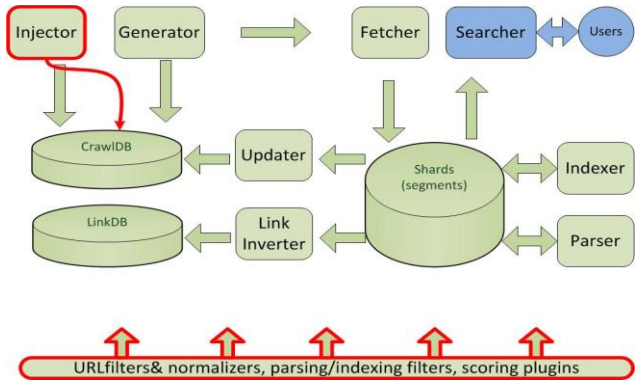

Fig.1. Architecture of Nutch

V. USER PROFILE CONSTRUCTION

The user profile is used to represent the user's interest and to infer their intention to user new queries. User profiles can be created in two modes: i) manually by user, ii) automatic profile generation using user search histories. Then the user query is matched with related category, where it belongs to stored local database. There are different clustering algorithm that can be to categorize the local database in different module so that user queries can be further matches with its profile and related category to show efficient search result.

VI. PROPOSED FRAMEWORK AND ALGORITHM

A: Algorithm

Input : Set of Result R=(R1, R2,….., RN) with original rank.

Output: Set of result with revised rank.

Step 1: Create a user profile.

Step 2: Gather all the user previous search history and categorizes the search domain into defined clusters using crawling into a particular domain of DMOZ categories.

Step 3: Input Query.

Step 4: Expansion of input query if required for efficient result based on query expansion algorithm [3].

Step 5: Dynamic user profiling is updated regularly on the basis of time duration.

$$W = C1^{*} W_{Long\text{-}Term} + C2 * W_{Short\text{-}Term}$$

$$W_{short\_term\ tk} = \frac{1}{N}\sum_{I=1}^{N}\left(\frac{tf(tk,ri)}{\sum_{j=1}^{T} tf(tj,ri)}\right)$$

$$W_{long\_term\ tk} = \frac{1}{N}\sum_{I=1}^{N}\left(\frac{tf(tk,ri)}{\sum_{j=1}^{T} tf(tj,ri)}\right) * e^{-(d2-d1)}$$

Wshort_term=(wshort_term t1, wshort_term t2,……, wshort_term tn) , vector of short term history.

Wlong_term=(wlong_term t1, wlong_term t2,……, wlong_term tn), vector of long term history.

No e^{-xt} : It is a decay factor that assumed that user interest is deviate if not browsed within a week.

N: Total number of page browsed by active user.

tf: Term frequency of term tk of ith result of page result ri.

In the decay factor if the d1 is the day when tk last occur and d2 following to that day d1 then life set to 14. If (d2-d1>7) then set life to 7.

Step 6: Compute similarity between user profile search result web pages rj and crawled web pages of DMOZ categories.

$$cosine(dj,u) = \frac{< rj * u >}{\|rj\| * \|u\|}$$

Step 7: Arrange the web pages based on decreasing order of cosine similarity.

Step 8: return set the result with a new rank

B: Architecture of Proposed Framework

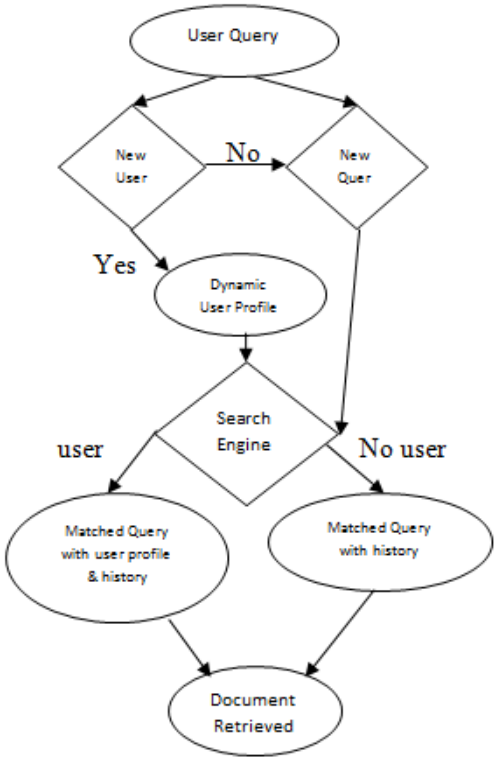

Fig 2. Flowchart of proposed system

VII. EXPERIMENTAL RESULT AND ANALYSIS

In our experiment we have used our own query set to measure the efficiency of personalized search engine over generic web search. Apache Nutch has been used to crawl the specified domain, of DOMZ categories. This will help in searching the query to search engine in specified category only. The keyword having same keyword weight is grouped into a specified category. As we can see in figure 4 different query given will be grouped in a separate category, whereas Flipkart and Yepme will be grouped in the same categories as its value approximately same. Other keywords are belonging to other categories.

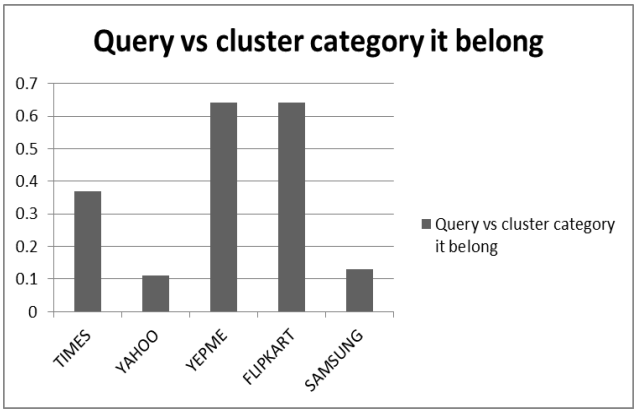

Fig.3. query vs cluster category it belong

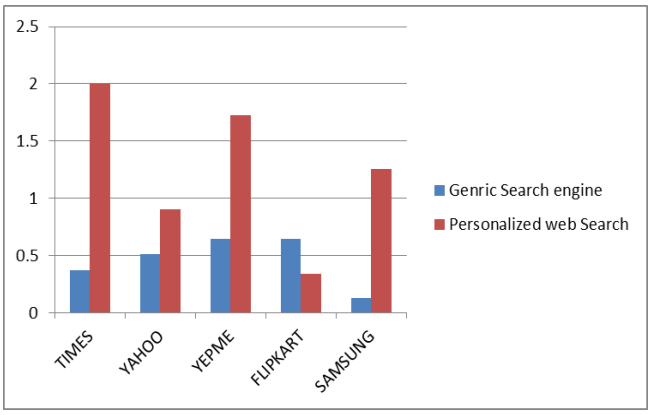

Fig.4. Performance of web search

In calculating the performance measure of personalized web search we found that it is leading to generic search engine result. So, results show that by using proposed approach the efficiency of search engines can be improved to a great extent and it gives result according to user need and relevance information.

VIII. CONCLUSION AND FUTURE WORK

The huge heap of information on the web has forced to develop an efficient information retrieval system for web search engine. Only with given small query to search engine and it searches the whole World Wide Web to provide relevant information to the user. In this paper, we proposed a framework that will help to enhance the power of generic search engines. The proposed framework has been implemented and the result is analyzed in certain parameters and we found personalized web search far better than generic search. Our work is significant in improving efficiency of search. In future, there is scope to use different ranking algorithm and clustering techniques. So, that efficiency can be further improved.

References

[1] Amruta Mantri, Priyanka Nawale, Trupti Pardeshi, Rajeshwary Shisode, Reena Pagare ."Profile Based Search Engine" *International Journal of Computer Trends and Technology (IJCTT)*,V4(2):164-168 Issue 2013 .ISSN 2231-2803

[2] S Ioannis A. Sarafis, "Data Mining Clustering of High Dimensional Databases with Evolutionary Algorithms", PHD. Thesis, Department of Computer Science, Heriot-Watt University, Edinburgh, Scotland, UK, September 2005.

[3] Namita Mittal, Richi Nayak, M.C. Govil, K.C. Jain, "A Hybrid Approach of Personalized Web Information Retrieval", *WI-IAT*, 2010, Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on 2010, pp. 308-313, doi:10.1109/WI-IAT.2010.270

[4] Fang Liu; Weiyi Meng; YU C: "Personalized web search for improving retrieval effectiveness," IEEE Transactions on Knowledge and Data Engineering, IEEE Service Center, Los Alamitos, CA, US, vol. 16, no. 1, 1 January 2004 (2004-01-01), pages 28-40, XP011106912 ISSN: 1041-4347.

[5] K.W.T. Leung, D.L. Lee, and W.-C.Lee, "Personalized Web Search with Location Preferences," Proc. IEEE Int'l Conf. Data Mining (ICDE), 2010.

[6] Mohammad Nauman, Shahbaz Khan, "Using PersonalizedWeb Search for Enhancing Common Sense and Folksonomy Based Intelligent Search Systems", *WI*, 2007, 2007 IEEE/WIC/ACM International Conference on Web Intelligence, 2007 IEEE/WIC/ACM International Conference on Web Intelligence 2007, pp. 423-426, doi:10.1109/WI.2007.44.

[7] Stephan Spiegel, "Personalized Information Recommendation based on Web User Profiles", Master Thesis, Department of Computer Science & Engineering, Shanghai Jio Tong University, China, July 2008.

[8] Ayesha Ameen, K.U. Rahman Khan , B.Padmaja Rani, "Semantic Web Personalization: A Survey", Information and Knowledge Management , Vol 2, No.6, 2012, ISSN 2224-5758 (Paper) ISSN 2224-896X (Online).

[9] P.Karuppusamy , Mrs. T. Lakshmipriya. "Personalized Query Based Search Techniques Using Association Rules". International Journal of Engineering Trends and Technology (IJETT). V4(10):4411-4417 Oct 2013. ISSN:2231-5381. www.ijettjournal.org. published by seventh sense research group.

[10] Ashutosh Kumar Singh, Ravi Kumar P, "A Comparative Study of Page Ranking *Algorithms for Information Retrieval"*, International Journal of Electrical and Computer Engineering 4:7 2009.