ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Cold Start Problem in Personalized Web Search

Darshana Gupta[1], Vatika Tayal[2], Amit Thakkar[3], Kamlesh Makvana[4]

Master of Engineering Student, Dept. of Computer Engineering, NSIT, Jetalpur, Gujarat, India[1]

Assistant Professor, Dept. of Computer Engineering, NSIT, Jetalpur, Gujarat, India[2]

Associate Professor, Dept. of Information Technology, CHARUSAT, Changa, Gujarat, India[3]

Assistant Professor, Dept. of Information Technology, CHARUSAT, Changa, Gujarat, India[4]

ABSTRACT: In today's era, people are using different search engines to retrieve relevant information as per their interest. Web search engines retrieve results based on only query terms i.e. keywords. The user may also enter ambiguous query, a typical web search engine retrieves similar set of results without considering the need behind the query of users which is far optimal from individual user interest. A perfect search engine delivers the right information to right user at right time. Personalized web search delivers the right information to right user which is adaptable for individual user and satisfy users need by maintaining a user profile for an individual. In this paper, we have reviewed on many efficient personalized web search approaches which were proposed by many authors. At the end, we have introduced the cold start i.e. new user/new query problem in personalized web search. To the best of our knowledge, this is the first study that introduces the cold start problem in personalized web search.

KEYWORDS: Cold Start, Explicit Feedback, Implicit Feedback, Personalized Web Search, User profile

I. INTRODUCTION

Nowadays, the web has thus become a valuable and abundant information source. With the growth of Internet, data available on the www continues to increase exponentially. People are using different search engines to get relevant information as per their interest. The result of the search engine depends exclusively upon the query [10] i.e. keywords. The user may also enter ambiguous query, a typical web search engine retrieves similar set of results without considering the need behind the query of users which is far optimal form individual user interest. For example, if the user search for "*python*", search engine retrieves result related to both python *language* and python *snake*. At the same time if other user who is interested to know about python language he/she would also get both the result on search term "*python*". A perfect search engine delivers the right information to right user at right time. Personalized web search delivers the right information to right user which is adaptable for individual user and satisfy users need by maintaining a user profile for individual.

II. USER PROFILE

A. Definition

User profile stores the approximations of user's personal preferences of terms of user's browsing history, knowledge about the world, likes and dislikes etc. It is generated and updated by exploiting user related information. It may include demographic and geographical information including age, gender, country, language etc. User profile is an essential and necessary part of a Personalized Web Search.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

B. Generation of user profile

User profile can be generated into two ways:

1) Explicit Feedback: The users themselves do the profiling work by either specifying search preferences or by providing personal relevance feedback such as rating returned search results. This system increases the score of terms that occur in relevant documents and decreases the score of non-relevant documents.
2) Implicit Feedback: This can be achieved through automatically analyzing user's previous search history. This type of system obtains relevance feedback by monitoring search behaviour. To indentify interest of user on particular page implicit feedback will analyze bookmarked pages, user's search history, dwell time, query log etc.

C. Working of user profile in personalized web search

The user profile can be applied in several ways:

1) During the retrieval process
2) To process results of a query (post-processing)
3) To modify the query itself (pre-processing)

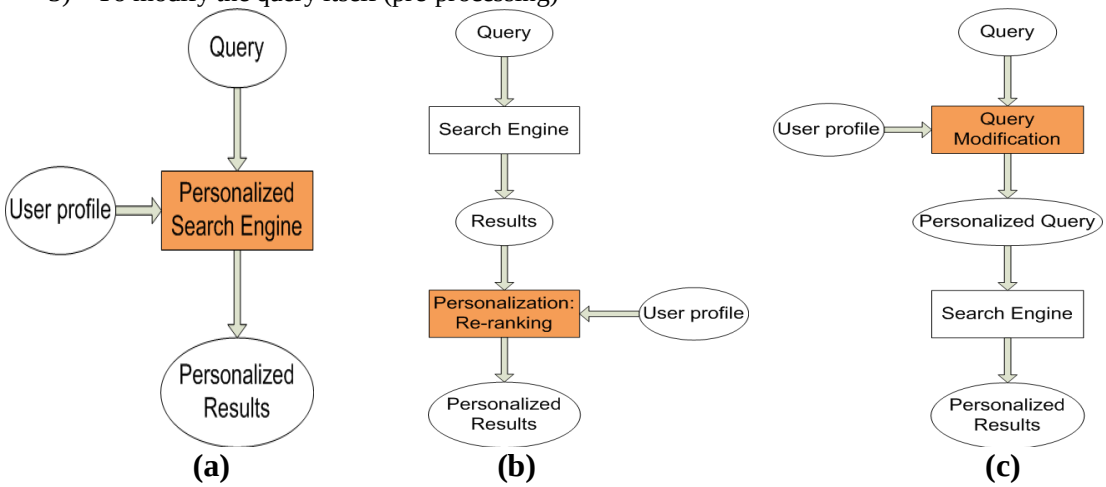

Figure 2.1: (a) Personalization process where the user profile occurs during the retrieval process, in a distinct re-ranking activity (b) or in a pre-processing of the user query (c) able to increase precision [9]

These several types of working of user profile can affect the personalization process and is as shown in **Figure 2.1.** The figure 2.1 states that the modern approach (figure 2.1(c)) is much better than that of traditional approach (figure 2.1(a) (b)) in context of query response and is drafted in figure level by level.

III. RELATED WORK

In this section, we demonstrate some of the works that have been done for personalize web search.

Makvana Kamlesh [1] proposed a novel approach that personalizes web search result through query reformulation and user profiling in form of Resource Description Format (RDF). Figure 3.1 shows the working of the framework. For re-ranking the search results, proposed approach uses Vector Space Model (VSM). It identifies number of user's previous query that is matched with current query and calculates interest value of each page associated with previous query. According to the interest value ranking of each page will be updated and will be displayed the result to user with higher preference. Interest value of each links can be calculated by [1]

$$W_c = \sum_{d=1}^{m} \frac{V_d}{r_c} \beta$$

Where, W_c = Interest value of each link
m = No. of Link in VSM
r_c = Actual rank of link
β = Equilibrium Factor

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

The framework also suggests some keywords that help user to incorporate him/her current interest.

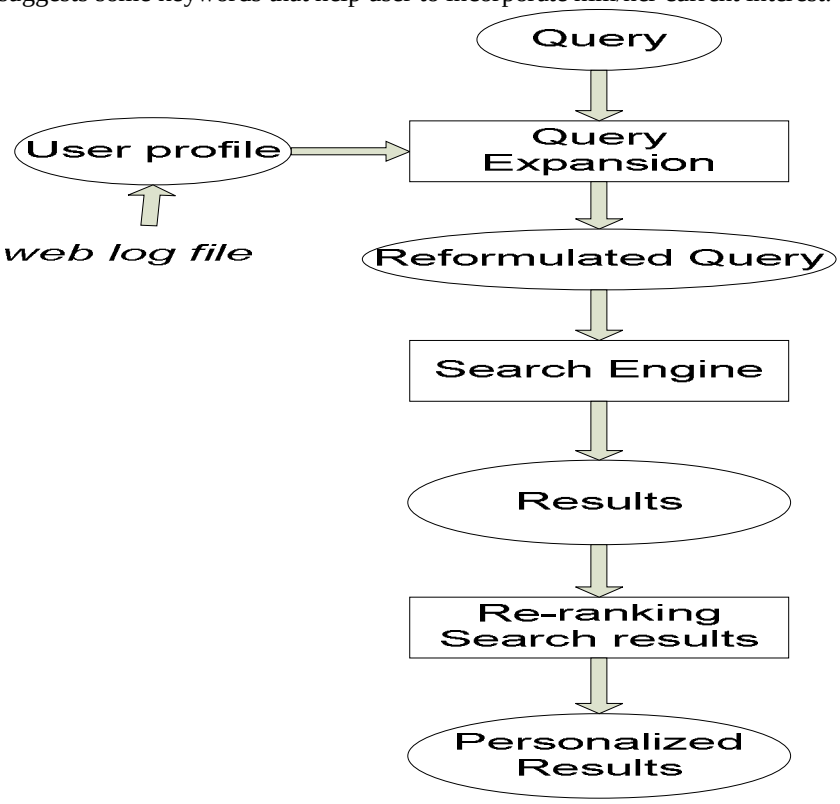

Figure 3.1: Framework of A novel approach to personalize web search through user profiling and query reformulation

Pannu [2] proposed web search personalization by using hybrid profiling i.e. a combination of implicit user profile and explicit user profile. Figure 3.2 shows the working of the framework. In explicit user profile users create their profiles manually or provide some kind of feedback to a search system, while in implicit user profile; the system creates profiles based on observed search history and browsing behavior. Keywords from results of search engine API are obtained and user profile is creating. Finally, documents are retrieved based on similarity between the documents and user profile.

To find the similarity between them the author used cosine similarity function and it is as followed by [2]

$$Sim(D,P) = \frac{D.P}{\| D \| \| P \|} = \frac{\sum_{i=1}^{m} d_i p_i}{\sqrt{\sum_{i=1}^{m} {d_i}^2}\sqrt{\sum_{i=1}^{m} p_i^2}}$$

Where, D = $(d_i , \dots , d_m)$ is the document vector

P = $(p_i, \dots , p_m)$ is a profile vector.

Further it can be simplified, If vectors D and P are normalised then ||D|| = ||P|| = 1 then formula can be:

$$Sim(D,P) = \sum_{i=1}^{m} d_i p_i$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

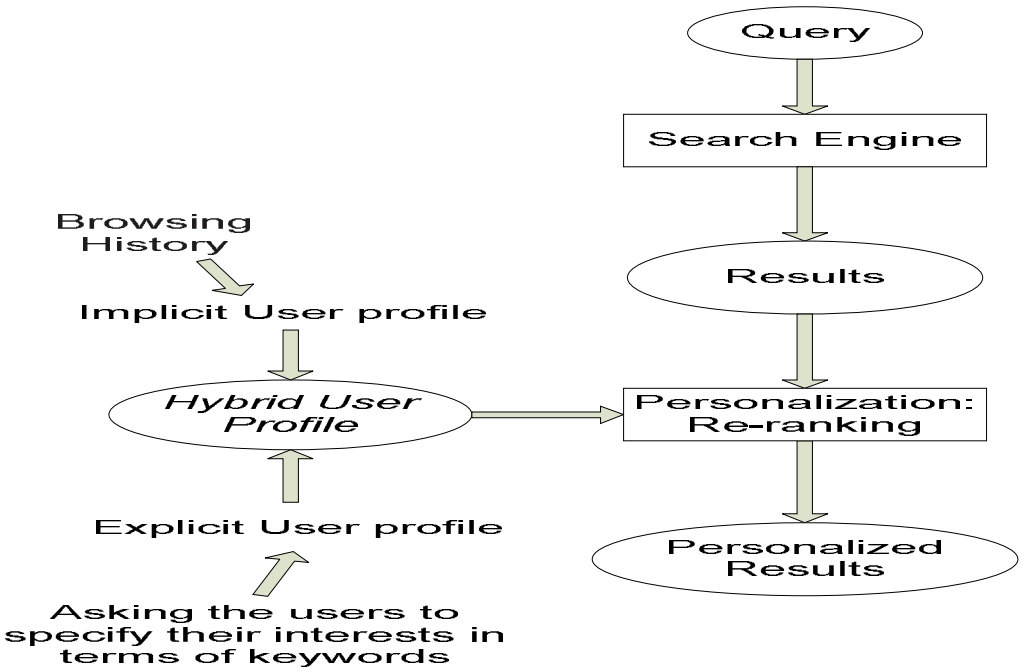

Figure 3.2: Framework of Hybrid profiling in information retrieval

Wu [3] developed a hybrid approach for personalized web search. Figure 3.3 shows the working of the framework.

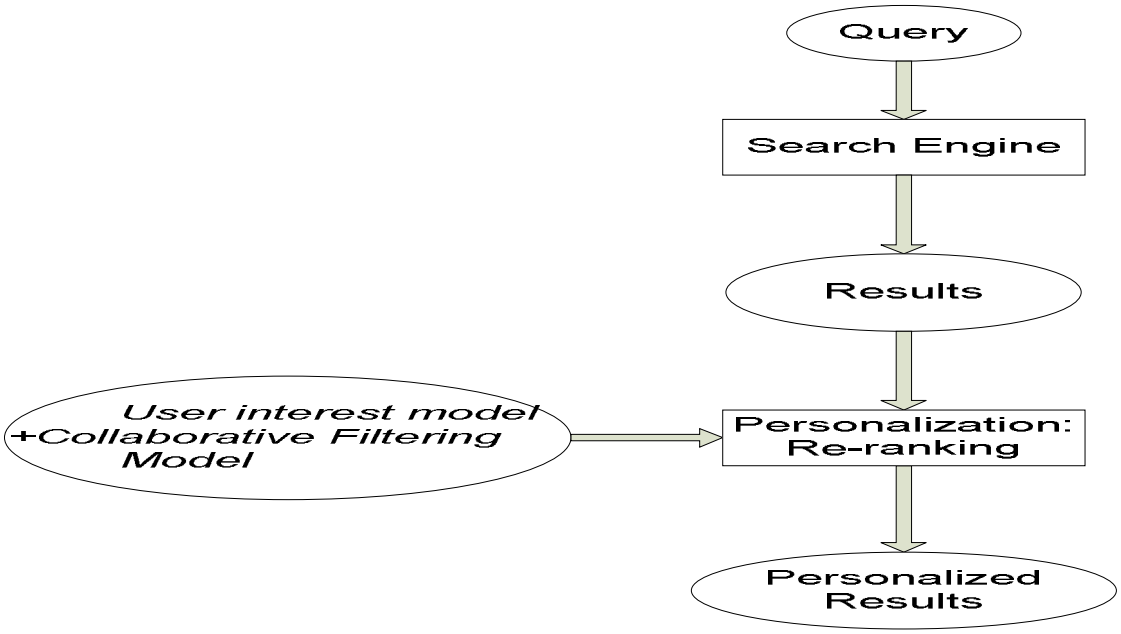

Figure 3.3: Framework of A hybrid approach to personalized web search

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

First, search results are retrieved based on the user's query. Then the interest value of each searching query based on the user's interest model and the recommendation value of each item based on collaborative filtering is computed. It can be computed by [3]

$$Score_{i,j} = \alpha \times I_{i,j} + (1 - \alpha) \times R_{i,j}$$

Where, $I_{i,j}$= User interest value for search results

$R_{i,j}$= Recommendation value of user U_i to r_j

To implement this system, the authors adopted 5 agents who were user interaction agent, controller agent, web information collector and processer agent, personalization reordering agent and feedback agent to retrieve search results.

Kenneth [4] recommended a Personalized Web search model with location preferences. Working of this framework is as shown in figure 3.4.

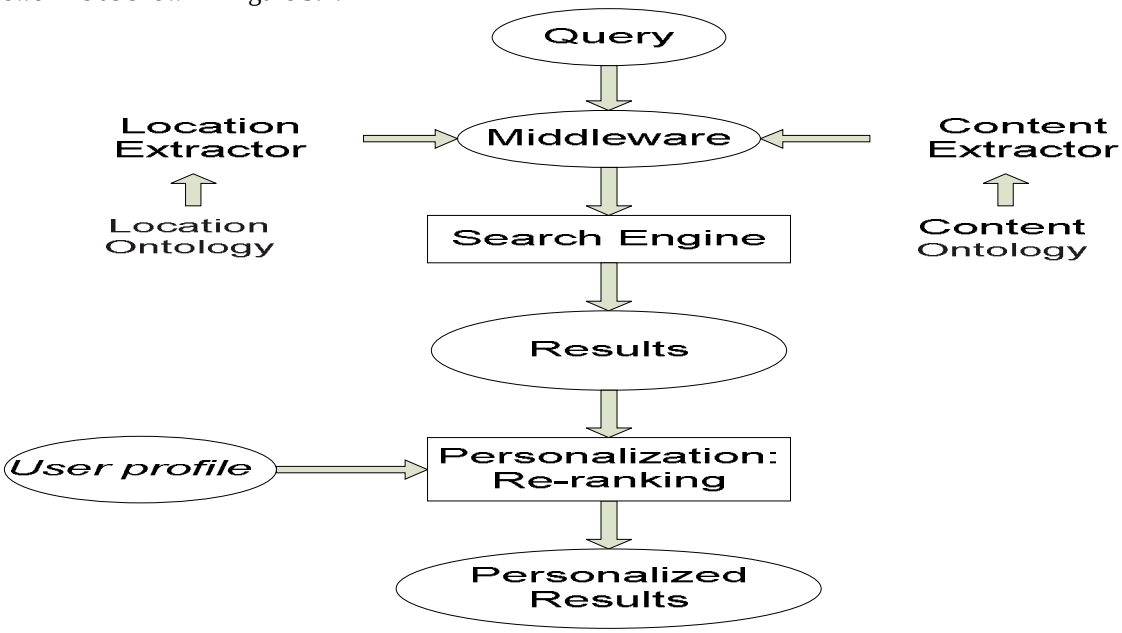

Figure 3.4: Framework of Personalized web search with location preferences

In this paper, authors have separated location and content concept and are organized into different ontology to make an ontology-based, multi-facet (OMF) profile which is captured by web history and location interest. With the help of this model we can make results by framing the concepts in respect to user preference. Location entropy is introduced for finding the degree of interest and information related to location and query by keeping the diverse interest of the users in mind. The content and location entropies of a query q (i.e. $H_C(q)$ and $H_L(q)$) as follows [4]:

$$H_C(q) = -\sum_{i=1}^{k} p(c_i) \log p(c_i)$$

$$H_L(q) = -\sum_{i=1}^{m} p(l_i) \log p(l_i)$$

Where, k = Number of content concepts

c_i = Number of search results containing the content concept

$p(c_i) = \frac{|c_i|}{|C|}$

m= number of location concepts

c_i = Number of search results containing the location concept

$p(l_i) = \frac{|l_i|}{|L|}$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

At last, SVM (Support Vector Machine) is used for re-ranking. The experiments show that the results produced by OMF profiles are more accurate than the ones which use baseline method.

Shafiq [5] proposed community aware personalized web search. Working of this framework is as shown in figure 3.5.

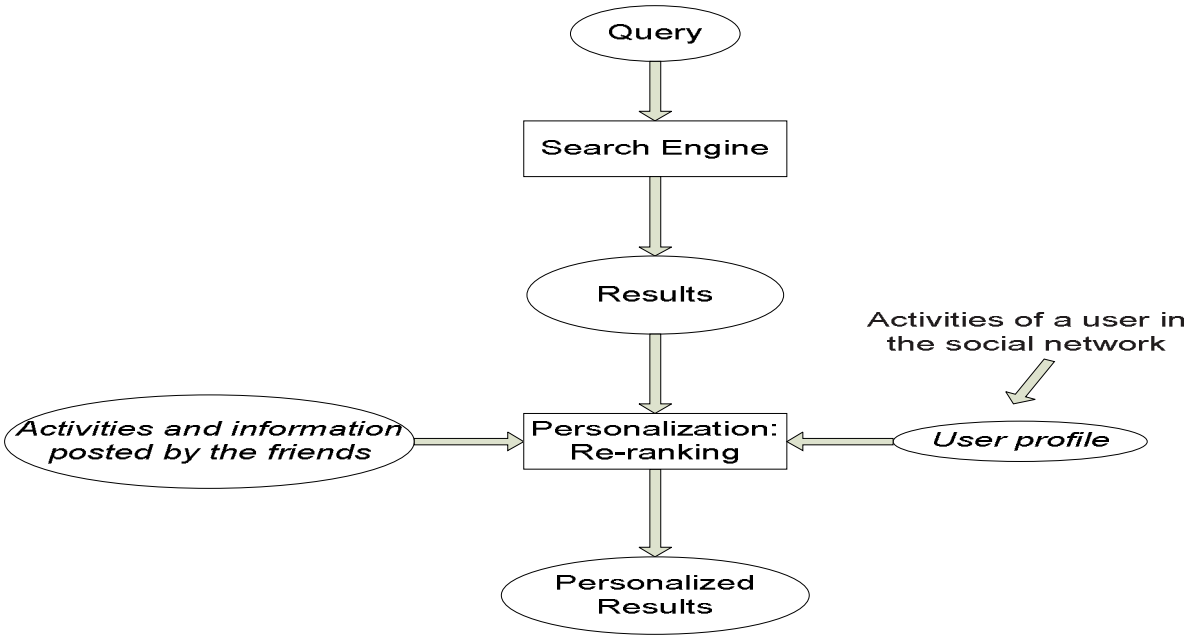

Figure 3.5: Framework of Community aware personalized web search

In order to find out personal interest and context, they follow a unique approach of (1) finding out activities of a user of his/her social-network, (2) finding out what information does the social networks (i.e., friends and community) provide to the user. Based on this information, they have developed a methodology that takes into account the information about social networks and prioritize search results from Web search engine.

Chen [6] proposed Rankbox, an adaptive ranking system for mining complex relationships on the semantic web. The objective of this paper is to provide an effective ranking method for complex relationship mining, which can 1) automatically personalize ranking results according to user preferences 2) be continuously improved to more precisely capture user preferences, and 3) hide as many technical details from end users as possible. Based on user's opinion, proposed system supports each user to give simple feedback about the current search results and employs a machine-learning based ranking algorithm to learn the user's preferences from his/her feedback. A personalized ranking function is then generated and used to sort results of the user's subsequent queries. The ranking function is defined as a scaler product of the feature vector $X(a)$ and a user-specific weight vector w_u:

$$f_u(a) = X(a) \cdot w_u$$

Where, w_u is a vector with the same length as $X(a)$, the subscript u denotes that the ranking function f_u together with the weight vector w_u specifically reflect user u's preferences.

The user can keep teaching the system his preferences by giving feedback through several iterations until he is satisfied with the search results. They evaluate the system on a large RDF knowledge base created from the Freebase linked-open-data.

Kumar [7] proposed a framework for constructing an Enhanced User Profile by using user's browsing history and enriching it using domain knowledge. Working of this framework is as shown in figure 3.6. The cosine formula used for the similarity of the URL u in User Profile to each web pages d_j in Domain Knowledge is as follows:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

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

Where, d_j = Document vector

u = Profile vector

The Enhanced User Profile specifically used for suggesting relevant pages to the user.

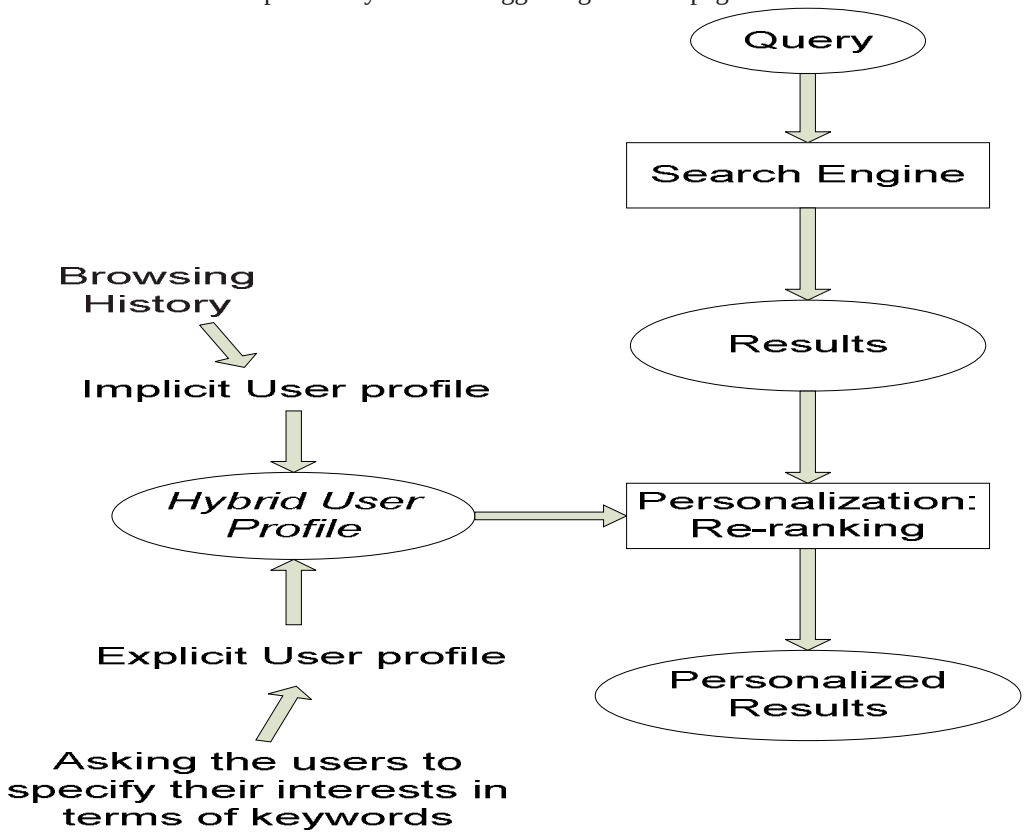

Figure 3.6: Framework of Personalized web search using browsing history and domain knowledge

Mittal [8] proposed a hybrid approach for Personalized Web Information Retrieval. Proposed system works in to two phases. The first phase includes the standard information retrieval using a search engine. Working of second phase is as shown in figure 3.7. The second phase uses the documents retrieved in first phase and steps forward using three modules. First, ontology is used for retrieval of user's context. Second, user profile is updated according to users' browsing behavior. For that they considered two aspects of user preferences: (a) Short term preferences, and (b) Long term preferences. In a User Profile, Weight of a query topic is constructed such that [8]

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

Where, *C1* and *C2*=constants satisfying C1+C2=1

The author use *C1*=0.75 and *C2*=0.25 in their experiments as Long term weight has higher precedence.

After retrieving the results from search engine, collaborative filtering is used for considering recommendation of similar users. Similarity between active user a and user u_i is computed using Pearson correlation coefficient and followed by [8]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

$$S_{a,ui} = \frac{\sum_{j=1}^{T}\left(W_{a,j} - \overline{W_a}\right) \times \left(W_{ui,j} - \overline{W_{ui}}\right)}{\sqrt{\sum_{j=1}^{T}\left(W_{a,j} - \overline{W_a}\right)^2 \times \sum_{j=1}^{T}\left(W_{ui,j} - \overline{W_{ui}}\right)^2}}$$

Where, $W_{a,j}$ = Weight of term j for user a computed based on term frequency in a browsed Web page
W_a = Mean term weight regarding user a
$W_{ui,j}$ = Weight of term j for user u_i
W_{ui} =Mean term weight regarding user u_i
T = Total number of terms
Then, re-ranking is applied by using dot product. Finally, user gets personalized search results.

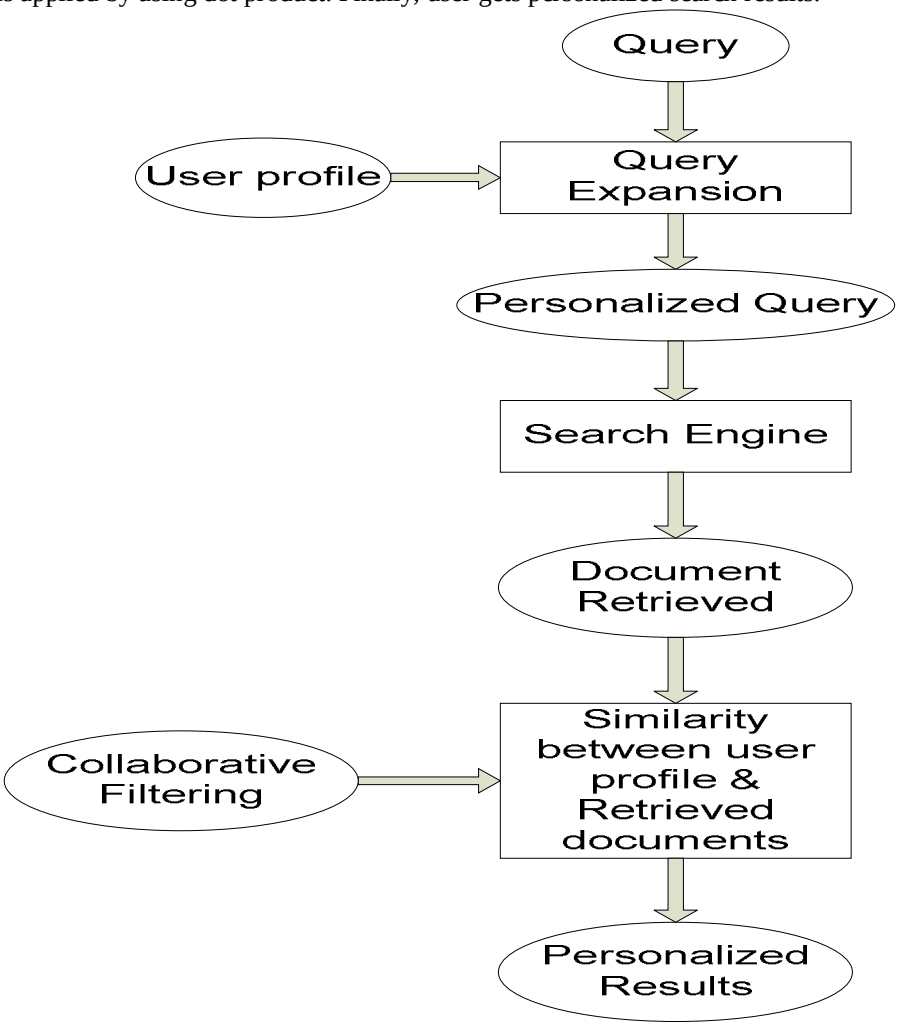

Figure 3.7: Framework of A hybrid approach of personalized web information retrieval

A. *Comparative Study*
This section describes comparative study of various existing framework is shown in Table 1. Table includes the existing framework, techniques used in that framework, advantages and limitations of that particular framework. These are taken from studied research papers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

Table 1. Comaprision of Various Existing Framework

Existing Frameworks	Technique	Advantages	Limitations
A novel approach to personalize web search through user profiling and query reformulation [1]	User profiling + Query reformulation	• Better identifies the current interest of user • Re-order the user's search result based on their preferences	• More computational cost when more user are accessing at a same time • Cold Start problem
Hybrid profiling in information retrieval [2]	Hybrid (Explicit + implicit) user profiles	• Improving the web search effectiveness in terms of precision and recall	• User Interest Diversification • Cold Start problem
A hybrid approach to personalized web search [3]	User interest model + collaborative filtering	• Good effect when the query is ambiguous	• Cold Start problem • Number of partaking user and the quantity of queries are very limited
Personalized web search with location preferences [4]	Content concepts + Location concepts	• Does not require users to explicitly define their location interest manually	• User Interest Diversification
Community aware personalized web search [5]	Finding out activities of a user of his/her social-network + Finding out what information does the social networks	• Improves the User Interest Diversification	• Increase complexity and expense • Cold Start problem
Rankbox: An adaptive ranking system for mining complex semantic relationships using user feedback [6]	Feedback to current result + Machine learning algorithm to identify user's interest from feedback	• Improve the ranking quality by sending more feedback to the system. • Able to identify changing behaviour of user	• Scalability
Personalized web search using browsing history and domain knowledge [7]	Enhanced User Profile (By using user's browsing history + domain knowledge)	• Better performance than those which are obtained through the simple user profile	• Cold Start problem
A hybrid approach of personalized web information retrieval [8]	Ontology + Dynamic user profiles + Collaborative filtering	• Improving the web search effectiveness in terms of precision and recall	• User Interest Diversification • Cold Start problem

IV. PROBLEMS AND DIRECTIONS

Many Personalized Web search frameworks have been discussed in the related works. The problems with the existing frameworks are explained in the following observations which guide the researchers to make efficient framework for personalized web search:

1. The existing frameworks give efficient personalized search results by maintaining the user profile which stores the approximations of user's personal preferences in terms of user's browsing history, knowledge about the world etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2016

Bottleneck of existing frameworks is that they do not provide efficient personalized search results to the new user/query.

2. Cold Start Problem: Cold start problem comes when a new user/query or both just enters the system. Three types of cold start problems could be identified: 1) personalization on existing queries for new users; 2) personalization on new queries for existing users; 3) personalization on new queries for new users. It is very difficult to provide personalization in case of new user because we have least information about user.
3. User Interest Diversification: User information needs may changes as time passes. User has different needs at different times based on their circumstances.

To the best of our knowledge, this is the first study that introduces the cold start problem in personalization of web search. The main aim of paper is to work on an area that will improve the cold start i.e. new user/new query problem and introduce new technique that improves it.

V. CONCLUSION

Personalized web search improve search quality and provides the personalized search results to individual user by maintaining a user profile for individual. From this survey, we understand that the main problem in the existing frameworks is that how to personalize the cold user/new user which just enters in to the system because we have no browsing history or past data. This review will help the researchers to develop an efficient technique for Personalized Web Search.

REFERENCES

1. Makvana Kamlesh, Pinal Shah and Parth Shah. "A novel approach to personalize web search through user profiling and query reformulation "Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on IEEE, 2014.
2. Pannu, Mohinder, Rachid Anane, and Ashish James. "Hybrid profiling in information retrieval." *Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on*. IEEE, 2013.
3. Wu, Xiyuan, et al. "A hybrid approach to personalized web search." *Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on*. IEEE, 2012.
4. Leung, Kenneth Wai-Ting, Dik Lun Lee, and Wang-Chien Lee. "Personalized web search with location preferences." *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*. IEEE, 2010.
5. Shafiq, Omair, Reda Alhajj, and Jon G. Rokne. "Community aware personalized web search." *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*. IEEE, 2010.
6. Chen, Na, and Viktor K. Prasanna. "Rankbox: An adaptive ranking system for mining complex semantic relationships using user feedback." *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*. IEEE, 2012.
7. Kumar, Ravindra, and Aditi Sharan. "Personalized web search using browsing history and domain knowledge." *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*. IEEE, 2014.
8. Mittal, Namita, et al. "A hybrid approach of personalized web information retrieval." *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. IEEE, 2010.
9. Micarelli, Alessandro, et al. "Personalized search on the world wide web." *The adaptive web*. Springer Berlin Heidelberg, 2007. 195-230.
10. Jay, Patel, et al. "Review on web search personalization through semantic data." *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*. IEEE, 2015.