www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 10 Oct. 2016, Page No. 18225-18231

Personalized Web Search with Custom Privacy Preservation

Swapna Desale[1], Prof. Jyothi Rao[2]

[1]K. J. Somaiya College of Engineering, Mumbai, India,
swapna.d@somaiya.edu

[2]Asst. Professor, K. J. Somaiya College of Engineering, Mumbai, India,
jyothirao@somaiya.edu

Abstract: ***One of the regular job done by users on the Internet is searching. Information is rapidly growing on the world wide web, hence users are facing difficulties to get effective and efficient search results for their searching terms. Personalization technique is used for providing more effective and relevant search results to user. We study Personalized Web Search engine which refers to the results based on the users' preferences. But users are not ready to share their personal information while searching process, which has become a main hurdle for large use of Personalized Web Search. Proposed system generalizes profile containing all interest areas related to query from complete user profile while satisfying the privacy requirements defined by users. Here system focuses to maximize the profile utility and minimize privacy risk. Click-log is used for user profile updating and providing page recommendations.***

Keywords: Personalized Web Search, Web Taxonomy, DMOZ, Lucence index, Stemming, Generalized Profile, Utility, Risk, Click-log.

1. Introduction

Information retrieval is a process of obtaining the information which is relevant to the required information from the source of information. World Wide Web is a biggest and frequently used source of information. Web pages are growing rapidly on the internet as days are passing. The web search engine is a popular way for searching the useful information on the internet. Web structures are sophisticated as well as large and users often fail to extract the specific data and receive irrelevant search results which may be ambiguous when they try to find out the information from these web structures. When user issues a query on search engine, many engines give the similar results to all users without considering the intension of users in which they submit their query. Search engine normally follow the rule of match-extract-return i.e. the matching of keywords to the desired documents. There are some considerations in search process: one, there are many words having multiple meanings in different situations so ambiguity of query should be firmly determined to get proper results on the internet. And another is, user should use the proper searching keywords which express the user's requirements in given situation. Multiple users may have different need of information while issuing exactly the same query. Example, user looking for computer hardware parts may give a query "mouse" to get the information about computer peripheral input device, while biologist who is finding the information about rodents can also issue the same query. Generally, search engine returns the result set which will be the mixture of all topics related to issued query term. So it will take some time for user to select the exact needed information. Small queries are uncertain, providing less knowledge about what user wish to get than that of the long queries. Sometimes, search engines provide search results which do not satisfy the goal of user. Such type of issues occurs since the large variety of users' interest, background and uncertainty of query. Hence, Personalization is a best solution to provide relevant and effective results to users.

Personalized Web Search returns effective search results which are customized for individual user needs. PWS refers to the results by user's preferences and interests. Some information about user is collected by explicitly or implicitly in personalized search and refine results to make more relevant. Nowadays, search with use of personal data of users is increasing to profile its user. Such information is collected by implicitly from browser history data, query history, bookmarks, click-through records, personal documents and many more ways. This collected personal information can reveal the user's private data, which can divert users from using these services and make upset to data-publishers to offer PWS-services. When user profile is combined with query term and sent to the server, users privacy can be corrupted by attackers. As attacker can obstruct the communication between user and server by invading server or man in middle attack. Attacker will try to capture user's sensitive data from profile. Hence users are not ready to disclose their personal information during search. Privacy concern has become a main hurdle for large use of PWS services. Hence, here we focus on providing preservation to user privacy requirements in personalized searching. In Web search engine, implicit information collection from user query history or click log data also improves the search results relevancy.

2. Literature Review

Personalization has been an active research field nowadays. Users are looking for better search engine which can provide the results to fulfill their information needs. Here, we focus on the literature of personalized search, related issues, PWS solution, user profile and privacy protection.

2.1 Google personalized Search

Google Search provides a feature of Personalized Search. Browser cookies are maintained for all searches on the google search. When searches are performed on google, results are

based on the relevance to web pages as per the search keyword as well as the websites visited by user or anybody else using same browser in past. This improves the users experience with search engine by providing relevant search results. But such personalized search can take place after many searches have been stored in browser history. This feature exploits the information to other users of same computer or browser, what other people have been searching for or making filter bubble [2]. Google's personalized search is processed by gathering search history in its database. For non-registered users search, google goes for the browser history on user's browser log to find related pages from the database. For registered users, after they logged into browser, their search history is used to find the content in which they are interested in and results are presented based on these records. User profile is created according to their web history with their interest, age, languages, location or gender and so on. But the drawback of this feature of personalized search is that it biases the results on what users have already viewed so no new information can be found. Results are filtered using browser history for all users who have not logged in and are using same computer. It can provide irrelevant search results as their results are personalized with others search history. [2]

Personalized web search can be implemented on server side or client side. [3] Server side implementation is followed by creating, storing and updating user profiles in search engine only. Benefits of this implementation are that user information can be used in initial search processing, personalization takes place without user efforts and search engine is able to use all the resources which are present at server side. Some general search engines like google search engine uses such architecture. But this architecture brings high cost of storage and computation and raises some privacy concerns as user profiles are stored on server. Client side implementation means user data collected and stored on client side. Rich user profile is constructed on client side as it stores information about users search behavior, web pages viewed, documents, bookmarks. So privacy concern is also decreases and computation and storage cost is distributed among the clients. Only the personalization algorithm is not able to use the server side knowledge.

2.2 Related Issues:

For designing an efficient and effective search engine, to get best relevant search results as per user requirements, need to consider the following issues:

2.2.1 User Profile:

User information used in personalization should be related to given query. Whole profile of user cannot be used for all sort of queries for getting personalized results. Profile-based personalization may not improve search quality even if exposing the profile of user to server can put privacy in danger. [11]

2.2.2 Profile privacy:

Privacy of user profile is an issue. Users do not want to share their personal or secret data to others hence privacy is provided to user's profile. But sometimes users profile is overprotected while others insufficiently. Profile can reveal a large amount of information which is a serious issue. This could make users nervous and afraid while using PWS. [5]

2.2.3 User involvement:

While refining the search results to make them more relevant, many system requires iterative user interaction. Which would be tedious job for people and could posed the privacy risk. While runtime profile generalization user interaction in generalization process can create a big privacy issue.

2.2.4 Dynamic search:

Users are not statics and their information need is also hard to infer. User data need change over time. Users may have different information need at different times according to current situation. Users web history is important for tracking user's behavior on internet. Using this click-log, users profile can be updated periodically.

2.3 PWS Solution

Personalized web search engine can be built by using profile-based or click-log-based methods. Click-log based techniques uses clicked pages in user's web history. Profile-based techniques make the use of user-complicated interest models which are made from user profiling methods. Click-log based methods perform good but these methods work on repeated queries only [4]. Profile based methods work for all sort of queries but sometimes are not stable in some situations [4]. There are some flaws in both strategies, but profile based methods are more efficient to provide improved quality results for searching. Profile-based methods make the use of users personal and behavior information which is collected from user's search history, bookmarks, documents, browsing history and so on [7].

Leung, Lee and Wang-Chien Lee et. al [9] has proposed a personalized web search technique that collects user's interests and preferences in terms of concepts by mining results and their click-log. Then by separating concepts into location concepts and their contents after that organizing these concepts into ontologies to make an OMF i.e. ontology-based multi-facet profile by collecting the user's location interests as well as their content. OMF improves the precision significantly as compared to baseline is proved by experimental results.

S. Preetha [12] has proposed a technique to enable large-scale personalization search evaluation. User's preferences are used in the clustering process to get the benefit of personalization. The information retrieval process has target to give results for users search to match their intention. The proposed framework working follows in two steps:

- In first section, users' feedback is collected implicitly from their click-log records and mapped into pseudo documents. Means of keywords are used for clustering these feedback sessions. One cannot be known about exact number of search objectives of user in earlier, so many parameters are examined and required value is selected from the bottom section.
- In second section, actual search results are reconstructed from the first part of search objectives of user.

2.4 User Profile

For profile-based personalization search, user preferences are collected to profile user by using multiple profiling techniques. User-related information and their interest may include:

- Information about user's gender, age, location, education, languages, nation, interest topics, and many more information.
- Web cookies containing clicked pages and past all queries.
- Personal documents, bookmarks, visited pages, emails etc.

User profile can be stored and represented into list. Vectors, bag of words [4]. Recent work focuses on hierarchical structure of profiles due to its good scalability, high access efficiency and stronger descriptive ability. Many hierarchical structured profiles are generated with existing hierarchy or graphs such as Open Directory Project [11], Wikipedia, WordNet etc. [5] construct the profile hierarchy by their own from user information with term frequency analysis.

C Liang et al.,[8] provide effective way to construct user profiles based on user interest and preferences. Construction of user profiles proposed three approaches: support vector machine methods, k-nearest neighbor method and Rocchio method. But K-nearest neighbor approach is better than others is proved from experimental results which are taken from constructed dataset.

Peng et al. [10] build user profile by collecting search result which are used with reference of Google directory by user. Proposed framework maintains a tree in which topics are linked. Every topic which is searched by user and stored in tree structure, is available in tree directory. Link Visited count is stored which shows the degree of interest.

2.5 Privacy Protection in PWS

Privacy for profile-based personalized web search categorized into two classes. In first category, identity of person is considered as a privacy concern [16]. Another category considers the privacy of the data. Literature done for category one provides solution like pseudo identity, nil identity, group identity and without any personal information. Answer to nil identity and no personal information is not practical because of high cost of cryptographic communication. Pseudo identity is proved to fragile. Existing work focuses on group identity type privacy protection. Direct relation in between search queries and user is broken. In category second, users do not permit their complete profile exposure to server because they trust themselves only. Xu et. al [5] provides a personalized web search with privacy protection on the basis of tree structured profile. Generalized profile is extracted from complete user profile by using a user defined threshold. But it does not focus on utilization of query which is critical for service quality of PWS. For analogy we consider the both query utility and privacy need into account. Xiao and Tao [6] has introduced the personalized privacy protection concept firstly by PPDP i.e. Privacy-Preserving Data Publishing.

3. Proposed System

Personalized Web Search provides better search results and these are filtered for user's need. PWS system try to extract the most relevant results to user's a query. This paper combine both profile-based and click-log based personalization solution for improving effectiveness of PWS system. System try to find out some information about user to get their interest areas and preferences by either implicitly or explicitly.

3.1 User Profile Construction

Complete user profile is constructed in a hierarchical structure using different topics that tells user interest areas. Proposed PWS adopted Open Directory Project, named ODP or DMOZ (Directory Mozilla) web directory as repository taxonomy, R. DMOZ repository is a large hierarchy of topics which covers almost whole topic domain of people understanding and the most important internet directory ever use. If anyone wants to achieve top rankings in the search engine, he or she must be included in the DMOZ Directory. The DMOZ web taxonomy is generally accessible to all and can be used by many people as a background theory. When user interest areas are collected explicitly from user, each interest topic is mapped with DMOZ data in proposed work. Different categories are extracted from DMOZ database and user-specified categories are added into their profile. User profile is a collection of user interest and preferences which is represented in hierarchical structure and is also a rooted subtree of Repository R. Rooted subtree can be said as user profile is generalized from R by removing some nodes from it. Each topic from user profile exist in R has a repository support expressed as $sup_R(t)$ which counts how many times the respective topic is occurred in available web taxonomy. Sample user profile made up of multiple user interest areas is illustrated in Fig.1. which is constructed from Dmoz taxonomy with their proper categories specified by user.

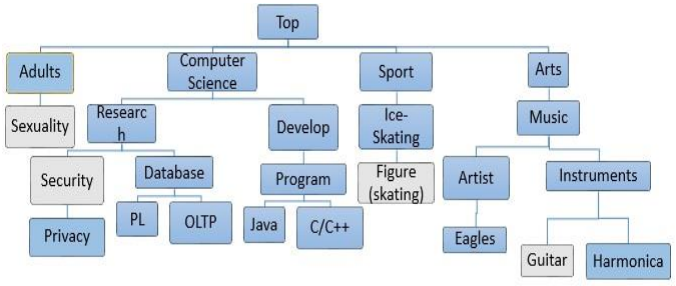

Figure 1: Sample User Profile

3.2 User Privacy Need

User defined privacy requirements are sensitive topics in user profile. Sensitive topics are those nodes from user profile whose disclosure to server put privacy at risk for user. Sensitive node set is defined for ensuring that topics which are specified as sensitive will not cause a privacy risk. In a sample profile shown in Figure 1, sensitive topics are sexuality, security, figure-skating and guitar shaded in grey color. Users privacy concern changes over every sensitive topic. Some topics can be over sensitive while others are less sensitive whose disclosure can be acceptable for getting more benefits from service. User may never wish to share their interest in sexuality and also don't want to disclose the interests like figure skating or guitar only to avoid advertisements. Hence, it asks user to define sensitivity values for respective sensitive topic available in sensitive node set. At last, the cost value for every topic that exist in user profile is computed to get the cost layer of complete user profile.

3.3 Stemming

Stemming is a process of linguistic normalization, which produces common form of word from various forms of a words. In short, the Stemmer transforms a word into its root form. Stemming is the conflation of the variant forms of a word into a single representation said as the stem. For example, the

presentation, presenting, and presented keywords could be stemmed to word "present". Stemming is used in Information Retrieval System, to combine a word to its many forms to avoid mismatches between the query which is asked by the user and the words available in the database. For example, if a user wants to search, "How to cook" and return a query on "cooking" he or she may not get all the relevant documents. If the query is stemmed, so that "cooking" keyword becomes "cook", after that retrieval will be successful. Stemming has been largely used in Information Retrieval Systems to improve its performance. Hence, Stemming is performed at various stages in proposed Personalized Web Search. While constructing user profile for preferences, all preferences are stemmed before getting their categories from Dmoz database. Stemming algorithm is performed on all searching keywords and stem words are used for generalizing runtime profile. These stem words and generalized profile are sent to server for information retrieval. At last, while updating user profile with new interest areas, stemmer applies on new interest areas for getting proper categories and better profile construction. [22]

3.4 Query-category Knowledge

The purpose of query-category knowledge is to capture the subtree rooted at user profile. This subtree can be called as seed profile which covers all topics from profile, related to the search term. When query is issued on client proxy, stemming of words takes place and all the topics related to query are extracted from repository. All extracted topics should be nonoverlapping and these topics with their ancestor nodes of repository produces query relevant trie. By overlapping complete user profile with this query relevant trie, the intersection will produce seed profile which will be a rooted subtree of user profile. Seed profile contains all the categories of entered query which are available in user profile.

3.5 Runtime-Profile Generalization

Set of nodes are detected and removed from user profile so that exposing the generalized profile will keep privacy risk under control. Seed profile is considered as input to this phase and some topics are pruned to avoid the risk of exposing private data. Generalization technique avoids unnecessary privacy disclosure as well as removes noisy topics from constructed profile. Runtime profile generalization considers two metrics, utility of profile and privacy risk. Generalization process takes place after query issuing to avoid irrelevant topic branches and sensitive topics disclosure. This procedure decides whether personalization should be done or not. Generalization is done in a cost based iterative manner to build a seed profile relying on utility and risk metrics. Personalization technique contribute less or reduces search quality for distinct queries while exposing profile may risk to privacy. Hence online mechanism allow user to decide for personalization. Runtime decision of personalization increases search quality and avoids unwanted user profile exposure.

3.6 Click-log Records

Web cookies are major implicit feedback from users which helps to improve efficiency of information retrieval system. Users feedback can provide important information for improving performance of IR systems. But due to user's reluctance to provide explicit feedback, provided approach focus on implicit feedback to collect information. Here users are not required to answer the questions, system observes their behavior on web. Proposed framework stores history log for all users consisting clicked pages, searching term, username, IP address and access timestamp. This click-log is used for updating users profile with new interest areas periodically which are not exist in original profile. Proposed system uses a ranking process for click-through data which provide recommendations to user and user is able to see their history logs with all details. Click-through data helps to personalized web search systems to improve their search quality and users search experience. Click-log is useful for dynamic search and making user profile strong to get newly information about users to improve the performance of system. Click-through records can be considered as triplets of query, ranking and clicked links or views pages by users. Users do not click randomly on link; they make some guesses or informed choice. Click-log data is collected at very low cost and no overhead to user.

4. System Architecture

Search system hosted on the client machine provides the feature of online profiler which is a key factor for privacy protection. Proposed personalized search system maintains the complete hierarchy of user profile as well as privacy requirements specified by user as a set of sensitive nodes. Dmoz web directory is used for getting human knowledge topics reference for profile construction. As shown in Figure 2 framework working is distributed in two stages named as profile construction stage and search stage for all users. In profile construction part, user profile hierarchy is built and customized with user's privacy.

4.1 First, user register on proxy server with personal details and local information of that user will be stored in local database to identify the behaviour of that user, which will help to system to provide relevant result to the user as per his/her interest areas.

4.2 After registration, users proper interest areas are collected with help of Dmoz ODP data and stored in hierarchical structure.

4.3 In this step, user defines sensitive topics on collected interest areas and sensitivity value for respective sensitive topics whose disclosure to server can cause a risk.

4.4 Sensitivity values are calculated and stored in local database for all rest of the topics in user profile for further use.

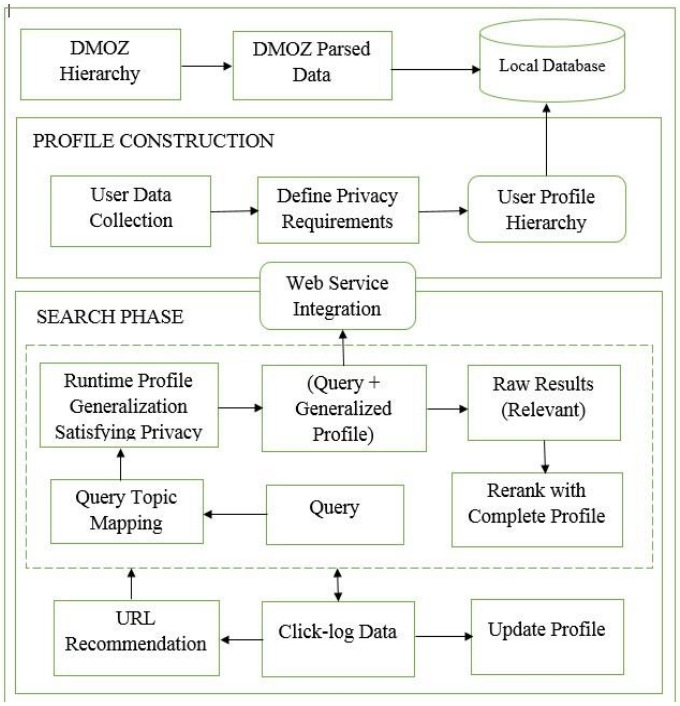

Figure 2: PWS System Architecture

4.5 User issues a search term q on the client machine and stemming of words algorithm is performed on the entered keywords for getting stemmed word by removing suffixes.

4.6 Stemmed words are taken as a new query and proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile G satisfying the privacy requirements. The generalization process is followed by two conflicting metrics, profile utility and privacy risk, which are defined for profiles.

4.7 The search term q and generated user profile both are given to the server for personalized search.

4.8 PWS server personalizes search results with provided profile and results are sent back to the search proxy on client machine.

4.9 At last, the system shows the general results, personalized results and recommendations to user regarding search query with the help of click-log.

4.10 Proxy server maintains all these Clickthrough data for updating user profile with new interest area.

5. Algorithmic Strategy

Dmoz contains different topics, their categories and related description. We took all Dmoz data in RDF format representation and parsed it to

Dmoz is represented in RDF format containing topics, categories and their related descriptions. We first parse Dmoz to build the multiple files according to topics with title and corresponding topic description. Stemming operation and removing of stop words are done on these created files called as pre-processing step.

Algorithm 1: Dmoz Parsing

1. Begin
2. do
   3.1 select record (topic, title and description)
   3.2 pre-processing (record)
3. while (! end of file)
4. end while
5. End

Once pre-processing is done, Dmoz topic index is created. After single topic indexing, following algorithm is used to construct Lucene index from Dmoz topics.

Algorithm 2: Topic indexing

1. Begin
2. Total number of topics in Dmoz = n
3. for j = 1 to n
4. do
   4.1 Read record (j)
   4.2 Analyse record (j)
   4.3 Index record (j)
   4.4 Write record (j)
5. j++
6. end for
7. End

After successful user profile construction, proxy ask user for set of sensitive nodes which will be subset of user profile. Sensitive node set disclosure to server can expose risk to user. It also asks user to specify respective sensitivity values for all topics contains in S. Cost layer is made by calculating the cost value for all topics of user profile by using following algorithm:

Algorithm 3: Init sensitivity labels(t)
Input: User Profile (H), Sensitive-Node set (S)
Output: Cost layer of H

1. Begin
2. For □ sensitive-node in profile H, cost(topic) = Sensitivity(topic)
3. For all the topics which are leaf topics in H and not sensitive, cost (topic) = 0
4. For every internal node which are not sensitive, cost(topic) is recursively calculated by following formula:
   4.1 if $\|C(\text{topic}, H)\| > 0$ then //C ( )→Children of
   4.2 for each node $i \epsilon C(topic, H)$ do
   4.3 go to step 4
   4.4
   $$cost(topic) = \frac{\sum_{topic' \epsilon C(topic,H)} cost(topic')}{\|C(topic, H)\|}$$
5. End

Stemming is used with intention of improving information retrieval process performance which focus to index documents according to their topics because their terms are grouped by stems or query expansion to get more precise results. Here, personalized web search engine adopts porter stemmer for stemming of entered query to get stemmed words. These stemmed words will be used for getting proper categories from Dmoz data for search query, profile construction and updating profile. Porter stemmer utilizes suffix stripping method.

Algorithm 4: Porter Stemmer
Input: Keyword
Output: Stem

1. Begin
2. User entered keyword q
   2.1 Cleanse the plurals and suffixes like -ed or -ing

2.2 Change terminal y as i into stem when it exists another vowel
2.3 Deal with -ational, -ization, etc.
2.4 Takes off the finals like -ness, -full etc.
2.5 Eliminate the endings like -ence, -ant etc.
2.6 Remove end terminal e
3. Return stem keyword
4. End

Runtime profile generalization is an important step of this framework. When query is issued on client, the runtime profile is generalized in effect of query term while satisfying user privacy requirements. This technique avoids privacy exposure of user as well as takes off the noisy topics from profile which are not relevant to the user's query.

Algorithm 5: Profile Generalization
Input: Query(q), Profile(H), Privacy threshold(δ), Topic domain (R)
Output: Runtime Profile (G*) while satisfying Privacy Risk (δ)

1. Begin //Decision of personalization (yes/no)
2. Query-topic mapping → T(q)
   //T(q) is a relevant set of q from R
3. Trie as R(q) which is relevant to query with all ancestors //Overlap H & R(q)
4. H ∩ R(q) → Seed Profile (G)
5. if (G_i ! =root(R)) do
   5.1 For each leaf topic(t)
   5.2 While (risk(t_i)>δ) do
   5.3 Pop a prune leaf operation on t_i in G
   5.3.1 Parent(t_i,H)→ leaf
6. Else
   6.1 Return root(R) as G*
7. End

6. Experimental Results

We evaluate the search quality of proposed personalized web search by comparing with google search results. Complete user profile doesn't improve the search quality of PWS for all queries. Generated profile related to the current query works well by providing better results. At the evaluation time top 20 results from proposed PWS and Google are used for comparing the relevancy. Table 1 shows the improvemental change in search results and user satisfaction in terms of relevancy. Multiple queries are issued on both search engines and results are evaluated for relevancy. First 20 pages are taken and checked for relevant pages for a user profile shown in Figure 1. Personalization on proposed system generate better results. Reason for this can be consistent topic distribution on Dmoz which is used in our implementation. An experimental outcome present that our PWS system provides better search results as compare to traditional search engine.

Table 1: Relevancy

	Relevant Documents (out of 20 pages)	
Search Engine / Query	*Proposed PWS*	*Traditional Engine*
Eagles	15/20	7/20
Violin	17/20	12/20
Sketch	12/20	10/20
Php	17/20	12/20
Soya recipe	9/20	9/20
Speed skating	14/20	9/20
Internet	18/20	15/20
Adult images	14/20	12/20
Road accidents	11/20	10/20
IEEE Conference	8/20	9/20
Oracle	13/20	11/20
Total relevancy (%)	67%	52%

Figure 3 and Figure 4 shows the general search results and personalized search results respectively. Personalized results are provided to user according to his profile to fulfill his information need and satisfying privacy requirements. For query "Eagles" the results are presented. Figure 3 provides the general search results which mixes all the pages without categorizing them. Hence user requires some time to sort the documents which they need. As shown in Figure 3, it presents bird information, American football team, music band and many more to together. As shown in Figure 4, the results are filtered in PWS engine for query "Eagle" to cover all categories of user interests from his profile. PWS provides the only results about eagle music band due to user's interest in music. While generalizing user profile in terms of query, privacy requirements are also considered. Runtime profile is generated to maximizing user profile use and minimizing the privacy risk.

Figure 3: General Results

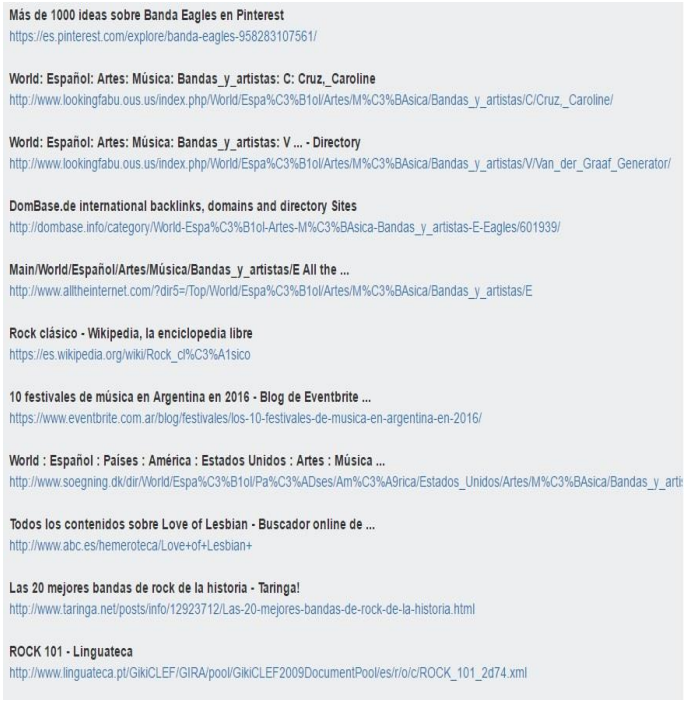

Figure 4: Personal Results

Figure 5 represents the 23 users experiences about PWS performance. User experiences to use proposed PWS are collected from different users as explicit feedback about relevancy and effectiveness in terms of good, satisfactory or bad performance.

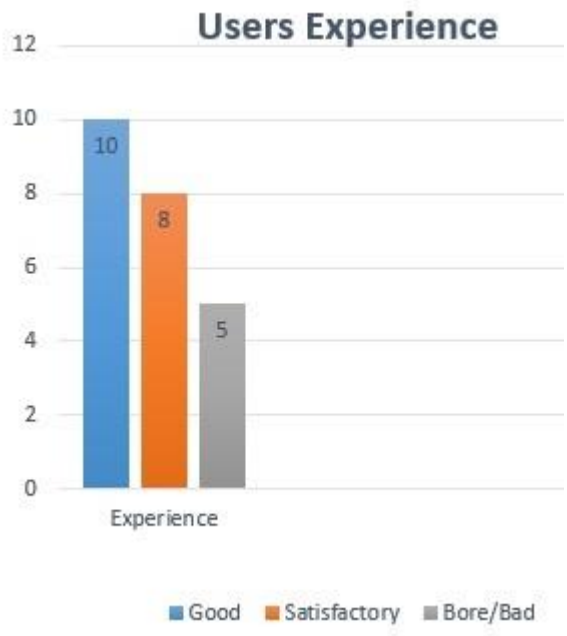

Figure 5: Users Relevancy Experience

7. Conclusion

This paper presents a personalized web search engine which provides effective search results according to individuals need. Proposed client side PWS engine provides privacy protection to user profile. User preferences are collected and stored in hierarchical format. User specifies which portion of profile should be protected and which content is exposed to server. System focuses on providing better search results by utilizing complete user profile while taking care of sensitive nodes of user profile. Users are provided with option whether to personalize or not for preventing unnecessary profile disclosure. Only the drawback can be said that it takes some time to generalize runtime profile. System maintains click-through data for every user. Browsing history size, storage and computation cost is less to generate user profile on each client.

8. References

[1] Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection in Personalized Web Search", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,2014

[2] https://en.wikipedia.org/wiki

[3] Ji- Rong Wen, Zhicheng Dou, Ruihus Song, "Personalized Web Search", Microsoft Research Asia, Beijing, China, 2009

[4] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.

[5] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600, 2007.

[6] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.

[7] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.

[8] C Liang, "User Profile for Personalized Web Search", International Conference on Fuzzy Systems And Knowledge Discovery, pp. 1847- 1850,2011

[9] K.W.T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering, pp. 70 I - 712, 2010.

[10] Peng, Xueping, ZhendongNiu, Sheng Huang, and YuminZhao."Personalized Web Search Using Clickthrough Data and Web Page Rating."Journal of Computers 7, no. 10 (2012): 2578-2584.

[11] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.

[12] S. Preetha, K.N. Vimal Shankar, "PERSONALIZED SEARCH ENGINES ON MINING USER PREFERENCES USING CLICKTHROUGH DATA", 2014.

[13] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.

[14] Jaap Kamps,"Visualizing WordNet Structure" July 2005

[15] Brett Spell, 'The documentation of the Java API for WordNet Searching (JAWS) specification', CSE, Southern Methodist University, July2005.

[16] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.

[17] K. Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, Aug. 2006.

[18] Y. Xu, K. Wang, G. Yang, and A.W.-C. Fu, "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.
[19] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.
[20] J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.
[21] Alessandro Micarelli, Fabio Gasparetti, "Personalized Search on theWorld Wide Web", 2007
[22] Deepika Sharma ,"Stemming Algorithms: A Comparative Study and their Analysis", IJAIS-2012