

Evaluation of Profile Based Personalized Web Search Using KNN and ECC

Mr. S. Dhinakaran¹, Dr. J. Thirumaran²

¹Research Scholar, Department of Computer Science, Rathinam College of Arts & Science, Coimbatore, Tamil Nadu, India.

²Professor, Department of Computer Science, Rathinam College of Arts & Science, Coimbatore, Tamil Nadu, India.

Abstract

The profile-based PWS has situated more effectiveness in making upgrades to the exceptional of net search not too lengthy in the past, with developing utilization of individual and habits skills to profile its customers, which is usually gathered implicitly from question ancient prior, looking history, click-through information, bookmarks, person records, and many others. To shield individual privacy in profile-based PWS, advice a PWS framework often called us that will adaptively generalize profiles through queries at the same time respecting character distinct privacy requisites. It has to bear in mind two contradicting effects for the duration of the hunt approach. First, it makes an try to toughen the hunt excessive-great with the personalization utility of the user profile. Subsequent, it desires to hide the privacy contents current within the person profile to position the privacy hazard beneath manipulate.

To aid search high-quality, the quest system which is tailored from a cozy kNN system, after which exhibit how to tremendously toughen search satisfactory with the personalization utility of the person profile. In the comfy kNN scheme, Euclidean distance between an expertise records and a question vector is used to select okay-nearest database documents. To cover the privacy contents, Elliptic Curve Cryptography (ECC) algorithm is used for Encryption and Decryption process. Elliptic curve cryptography (ECC) is an approach to public-key cryptography centered on the algebraic constitution of elliptic curves over finite fields. Elliptic curves are moreover utilized in a quantity of integer factorization algorithms which have functions in cryptography.

Keywords: Personalized Search, Privacy, User Profile, and Cryptography, Nearest Neighbor.

INTRODUCTION

Privacy preserving data mining is a novel study direction in information mining and statistical databases, where knowledge mining algorithms are analyzed for the aspect-effects they incur in knowledge privacy. The primary consideration in privacy preserving data mining is twofold. First, touchy uncooked information like identifiers, names, addresses and so on must be modified or trimmed out from the original database, to ensure that the recipient of the data to not be competent to compromise another person's privacy. 2nd, sensitive talents which can be mined from a database via utilizing information mining algorithms will have to also be excluded, due to the fact such knowledge can equally well compromise knowledge privacy, as we will point out. The principal function in privacy keeping knowledge mining is to

advance algorithms for enhancing the common knowledge come what may, in order that the personal knowledge and personal skills remain private even after the mining procedure. The drawback that arises when private expertise can be derived from released information by using unauthorized customers is also mainly called the "database inference" obstacle.

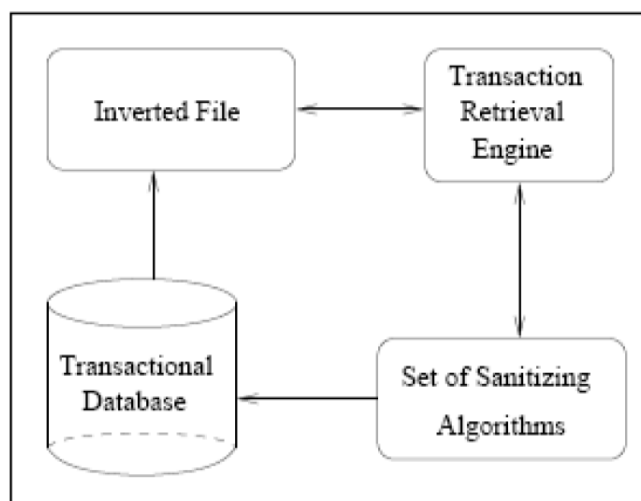


Figure 1: Framework of Privacy Preserving data Mining

Most of the time, privacy preservation happens in two fundamental dimensions: customers' individual knowledge and information concerning their collective activity. We discuss with the previous as person privacy renovation and the latter as collective privacy maintenance, which is involving company privacy in

- **Individual privacy preservation:** The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.
- **Collective privacy preservation:** Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which security control mechanisms provide

aggregate information about groups (population) and, at the same time, should prevent disclosure of confidential information about individuals. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to preserve strategic patterns that are paramount for strategic decisions, rather than minimizing the distortion of all statistics (e.g., bias and precision). In other words, the goal here is not only to protect personally identifiable information but also some patterns and trends that are not supposed to be discovered.

LITERATURE REVIEW

Search personalization is established on the truth that man or woman users tend to have specific preferences and that knowing the person's option can be utilized to fortify the relevance of the outcome the hunt engine returns. There have been many makes an attempt to customize internet search. These attempts more commonly fluctuate in

1. How to deduce the person choice, whether or not explicitly through requiring the user to denote information about herself or implicitly from the user's interactions?
2. What sort of expertise is used to infer the person's selection?
3. Where this knowledge is collected or stored, whether on the purchaser part or the server side, and
4. How this consumer choice is used to beef up the results' retrieval accuracy?

Any procedure supplying personalization services will must retailer some know-how concerning the consumer as a way to acquire its intention. Within the case of internet search this can be information that aids the approach in determining which internet sites are more likely to be of interest to the consumer. This information is referred to because the person profile. The simplest strategy to assemble one of these profile is to gather users' preferences explicitly, by using asking them to put up the essential information manually earlier than any personalization can be offered. This could be as common as checking a box to mark a field field as being of interest to them, or as distinctive as filling out lengthy types of private knowledge. Explicitly constructed profile knowledge is potentially of excessive first-rate and was once the option for among the earlier techniques supplying personalization like services.

Nauman et al., [1] used personalized web search for enhancing common sense and folksonomy based intelligent search systems. A huge division of the contemporary web is characterized by user generated content classified using collaborative tagging or folksonomy. It makes very tricky to search for appropriate content because of ambiguity in lexical illustration of concepts and variances in preferences of users. With additional services relying on tags for content classification, it is significant that search approaches progress to better suit the scenario.

Zhicheng, et.al, [6] proposed personalized search has been used for many years and many personalization strategies have been investigated, it is still unclear whether personalization is consistently effective on different queries for different users, and under different search contexts. The paper studies the problem and provides some preliminary conclusions. The paper present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies (including two click-based and three profile-based ones) using 12-day MSN query logs. By analyzing the results, it reveal that personalized search has significant improvement over common web search on some queries but it has little effect on other queries (e.g., queries with small click entropy).

P. Palleti et al., [3] developed personalized web search using probabilistic query expansion. The Web comprises of enormous amount of data and search engine recommends professional ways to assist navigate the Web to obtain the relevant information. Most common search engines, on the other hand, provide query results without taking user's purpose behind the query. Personalized Web search system intends to offer relevant outputs to users by considering user interests into account. In this approach, the authors developed a personalized Web search system applied at proxy which changes to user interests perfectly by generating user profile with the use of collaborative filtering. A user profile basically consists of probabilistic correlations among query terms and document terms which are utilized for providing personalized search results. Experimental outcomes prove that this proposed personalized Web search system is very effective and efficient.

Jordi Castella-Roca, et.al, 2010, [28] proposed the Internet is one of the most important sources of knowledge in the present time. It offers a huge volume of information which grows dramatically every day. Web search engines (e.g. Google, Yahoo...) are widely used to find specific data among that information. However, these useful tools also represent a privacy threat for the users: the web search engines profile them by storing and analyzing all the searches that they have previously submitted. To address this privacy threat, current solutions propose new mechanisms that introduce a high cost in terms of computation and communication.

O. Shafiq et al., [13] put forth community aware personalized Web search. Searching for the accurate information over the Web is not easy. In the era of high speed internet, elevated capacity networks, and interactive Web applications, it has turned out to be even easier for the users to publish data online. A huge amount of data is published over the internet; every data is in the form of web pages, news, blogs and other material, etc. In the same way, for search engines like Google and Yahoo, it becomes slightly hard to discover the accurate information, i.e., as per user's preferences; search results for same query would be different in priority for different users.

PROBLEM DEFINITION

Personalized search has been a fundamental study field and many systems have been developed and validated, still many

problems and challenges are but to be explored. This study concentrates on the analysis, evaluation and software of many customized net search techniques which might be being greatly used in these days. For that reason the motivation of this study work is directed closer to to recognize the web personalization processes, advantages, barriers and future trends.

Personalization is an effort to uncover most relevant documents making use of information about consumer's goal, domain of curiosity, looking history, question context and many others. That offers greater price to the user from the large set of results. As the web content material is developing exponentially, increasingly refined methods are required to give the imperative content to the character consumer.

The most customary difficulties encountered when shopping the net are [4]: i) issues with the data itself ii) problems confronted with the aid of the users trying to retrieve the information they need iii) issues in figuring out the context of search requests and iv) issues with picking the changes in user's expertise need. The predominant purpose in the back of all the issues is scale of the net that limits its utility. Scaling results in knowledge over load, internet customers spent more time on filtering out the important outcome, search engines like google won't in a position to furnish distinct results for people with exclusive intensions and context for the same question. For this reason the importance of Personalization is [1][2][3], to customise the web for person customers with the aid of filtering out the beside the point outcome and identify principal results.

Personalized internet search (PWS) has confirmed its effectiveness in improving the first-class of quite a lot of search offerings on the internet. However, evidences show that users' reluctance to disclose their private knowledge throughout search has become a foremost barrier for the large proliferation of PWS. We learn privacy safeguard in PWS functions that mannequin person preferences as hierarchical consumer profiles. We advise a PWS framework referred to as usathat may adaptively generalize profiles through queries whilst respecting person targeted privateness specifications. Our runtime generalization goals at putting a balance between two predictive metrics that evaluation the utility of personalization and the privacy chance of disclosing the generalized profile.

EXISTING METHODOLOGY

The prevailing profile-based personalized web Search does now not help runtime profiling. A person profile is more commonly generalized for most effective once offline, and used to customize all queries from a same user indiscriminately. Such "one profile matches all" technique without doubt has drawbacks given the form of queries. One evidence pronounced in is that profile-based personalization would possibly not even support to make stronger the quest first-class for some ad hoc queries, although exposing person profile to a server has put the user's privacy at chance.

The prevailing approaches do not consider the customization of privateness necessities. This mainly makes some person

privateness to be overprotected at the same time others insufficiently blanketed. For example, in, the entire sensitive subject matters are detected utilizing an absolute metric referred to as surprisal based on the information idea, assuming that the interests with much less person report aid are more sensitive. Nonetheless, this assumption may also be doubted with a easy counterexample: If a user has a tremendous number of documents about "sex," the surprisal of this topic may just lead to a conclusion that "intercourse" may be very basic and no longer touchy, regardless of the truth which is reverse. Regrettably, little prior work can conveniently address man or woman privateness needs throughout the generalization.

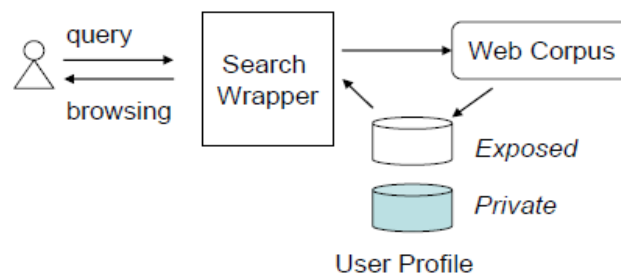


Figure 2: Existing System Structure

Figure 2, supplies an outline of the whole system. An algorithm is offered for the user to routinely build a hierarchical consumer profile that represents the person's implicit individual pursuits. Normal interests are put on a larger stage; precise interests are placed on a scale back degree. Simplest parts of the consumer profile will probably be exposed to the hunt engine according to a consumer's own privacy settings. A search engine wrapper is developed on the server side to include a partial person profile with the outcome lower back from a search engine. Rankings from each partial person profiles and search engine outcome are mixed. The custom-made results are delivered to the person with the aid of the wrapper.

The solution has three elements: First, a scalable algorithm automatically builds a hierarchical user profile from to be had supply information. Then, privateness parameters are supplied to the user to determine the content and amount of private understanding on the way to be published. 0.33, a search engine wrapper personalizes the quest results with the aid of the partial user profile.

EVALUATION OF PWS

Personalized search is a promising option to make stronger the accuracy of web search, and has been attracting a lot attention recently. Nonetheless, potent personalized search requires accumulating and aggregating user information, which probably raises severe concerns of privacy infringement for many customers. Certainly, these concerns have end up one of the crucial foremost limitations for deploying personalized search functions, and the way to do privacy-maintaining personalization is a excellent undertaking. In this paper, we systematically evaluate the hindrance of privateness

preservation in customized search. We distinguish and outline 4 levels of privacy security, and analyze quite a lot of software architectures for customized search.

When a person issues a question on the consumer, the proxy generates a user profile in runtime within the light of question terms. The output of this step is a generalized user profile enjoyable the privacy requirements. The generalization procedure is guided by means of considering two conflicting metrics, particularly the personalization utility and the privateness danger, each defined for consumer profiles. Subsequently, the question and the generalized consumer profile are sent together to the PWS server for personalized search. The hunt results are personalized with the profile and delivered again to the question proxy.

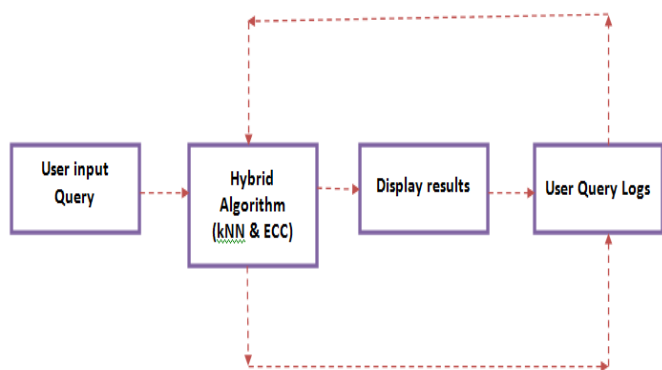


Figure 2: Framework of the Proposed System

Analyze the User Search

This module is for getting the query from the person. One GUI utility is needed to get the consumer important points for registration and getting the query. Each user wishes to register in that GUI software with their small print. And if the user gives the query within the search, every question is famous with the distinctive search time and it stored within the database. The present person is recognized by means of utilizing session attributes in jsp. And the enter is processed by way of discontinue phrases removing method to do away with the discontinue words within the person question, and provides the frequency of the words within the query.

Profile-Based Personalization

For this profile centered personalization, two foremost mechanisms have been developed: a profile generator that routinely creates user profiles representing the person preferences, and a content material-centered suggestion algorithm that estimates the user's interest in unknown content material by using matching her profile to metadata descriptions of the content. Both points are integrated right into a personalization method. Creating new information units for the looking profile founded personalization. Content are centered on the profile knowledge process.

Privacy Protection

The question and the generalized consumer profile are dispatched together to the PWS server for personalized search. The query, question logs, search results, clicked URLs, person preferences and consumer small print are will get encrypted and saved in Cloud Server. The encryption algorithm used is ECC (Elliptic Curve Cryptography), which offers privacy for all consumer small print. This presents privacy search for the customers.

kNN Algorithm

The algorithm can be summarized as:

- A positive integer k is specified, along with a new sample
- We select the k entries in our database which are closest to the new sample
- We find the most common classification of these entries
- This is the classification we give to the new sample.

ECC Algorithm

Encryption Algorithm:-

Suppose sender wants to send a message m to the receiver.

Step 1: - Let m has any point M on the elliptic curve.

Step 2: - The sender selects a random number k from [1,n-1].

Step 3: - The cipher texts generated will be the pair of points (B1,B2) where $B1 = k * G$ $B2 = M + (k * G)$.

Decryption Algorithm:-

To decrypt the cipher text, following steps are performed:-

Step 1: - The receiver computes the product of B1 and its private key.

Step 2: - Then the receiver subtracts this product from the second point B2 $M = B2 - (dB * B1)$ M is the original data sent by the sender

Web Data Results

This module retrieves the net information for the given consumer question and displays within the consumer interface reveal. Each question staff includes carefully associated and primary queries and clicks, it's predominant to have a compatible relevance measures in between the current question singleton crew sc and an existing question crew si. There are a quantity of viable approaches to determine the relevance between sc and si. A relevance measure that's powerful enough to determine an identical query agencies beyond the approaches that without difficulty rely on the textual content of queries or time interval between them. Our approach makes use of search logs with a purpose to verify

the relevance between query businesses more quite simply. In fact, the search historical past of a colossal quantity of users involves alerts related to query relevance, similar to which queries tend to be issued closely together (question reformulations), and which queries are likely to lead to clicks on equivalent URLs (question clicks). For query relevance identification, kNN algorithm is used, which calculates the Euclidean distance between the query and the web documents or URLs, the high-quality prime results are displayed in the influence page.

RESULTS % DISCUSSIONS

The following performance parameters are commonly used in privacy protection technique evaluation. The existing approach is compared with proposed scheme using these evaluation parameters. The performance of the TC process can be measured by one or more of the following methods:

Recall and Precision

They are two well known measures of effectiveness in text mining. While Recall is a measure of correctly predicted documents by the system among the positive documents, Precision is a measure of correctly predicted documents by the system among all the predicted documents. The system is evaluated in terms of precision, recall and Fmeasure.

Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents, while precision is defined as the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search.

$$precision = \frac{\text{number of correct results}}{\text{number of all returned results}}$$

$$recall = \frac{\text{number of correct results}}{\text{total number of actual results}}$$

F-Measure

F-measure combines precision and recall and is the harmonic mean of precision and recall.

$$F - \text{measure} = 2 \times \frac{precision \times recall}{precision + recall}$$

Several experiments were conducted with different query documents and the precision, recall and F-measure of the output was calculated. This higher improvement in precision value can compromise for the very small percentage of drop in the recall value. Moreover, the F-measure which combines precision and recall is much improved for similarity than existing system.

Eleven-Point average Precision

It is a measure based on recall and precision. Given a document, the system is allowed to achieve recall values of 0%, 10%, 20% . . . 100%, and the precision values at these points are computed. The resulting 11 point precision values are then averaged. The average precision values of all test documents are further averaged to obtain a global measure of the system performance.

Break Even Point

It is the point where recall equals precision. It is obtained by allowing the classifier to assign more categories. As a result, the recall increases and precision decreases until they become equal.

Training and Classification Efficiency

They were used by [5] to evaluate classifiers with closed performance. While the former calculates the average time taken to build a classifier for a certain category, the latter measures the average time taken to categorize a new document.

Results

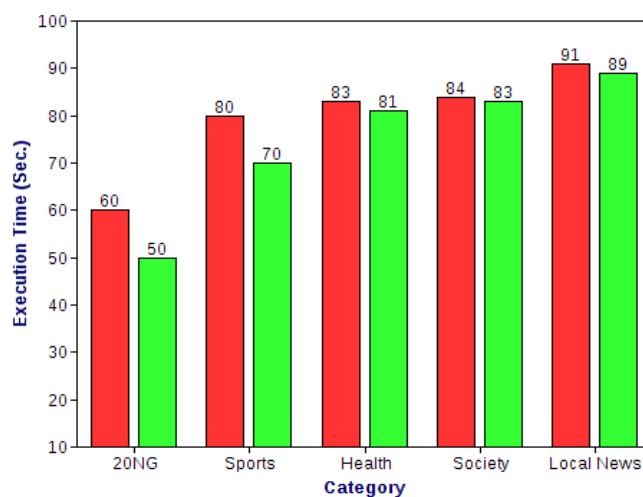


Figure 3: Evaluation of Recall using Hybrid Algorithm

In the above plotting, the red line represents the existing approach and the green line represents the Hybrid Algorithm for executing the user profile of the various users with various categories. The existing hierarchical link approach takes more time for extract the result from the dataset.

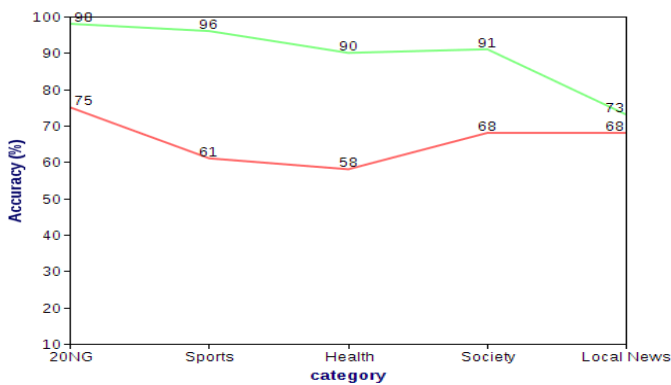


Figure 4: Evaluation of Precision using Hybrid Algorithm

The precision represents the accuracy of retrieval or categorizing the data. In the above result, the red line represents the existing approach and the green line represents the Hybrid Algorithm for executing the user profile. Existing approach accuracy level is poor compare with the Hybrid Algorithm of the proposed one.

CONCLUSIONS

The terrific progress of expertise on the web has compelled new challenges for the construction of effective search engines. This study work presents understanding on user customizable privateness keeping Search framework-U.S.For personalized internet Search. United states of America could probably be adopted via any PWS that captures user profiles in a hierarchical taxonomy. The framework allowed customers to specify custom-made privateness standards by way of the hierarchical profiles.

This system awarded a customer-aspect privacy safeguard framework for customized net search. This could probably be adopted with the aid of any personalized web search that captures person profiles in a hierarchical taxonomy. The framework allowed customers to specify customized privacy necessities through the hierarchical profiles. Additionally, this also carried out online generalization on person profiles to preserve the individual privateness without compromising the search first-rate. This method uses two algorithms, namely kNN (ok Nearest Neighbour) algorithm and ECC (Elliptic Curve Cryptography), for the generalization, search first-class improvement and privateness safety. The results published that this proposed would reap pleasant search results at the same time maintaining user's customized privateness requirements. The outcome also validated the effectiveness and efficiency of our solution

REFERENCES

[1] A. Zhang, X. Xie, K. Chang, C. Gunter, J. Han, and X. Wang. Privacy risk in anonymized heterogeneous information networks. In *17th International Conference on Extending Database Technology*, 2014.

[2] S. Wang, D. Agrawal, and A. E. Abbadi. Towards practical private processing of database queries over public data. *Distributed and Parallel Databases*, 32(1):65–89, March 2014.

[3] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisjuk, and X. Cui. Modeling the Impact of Short- and Long-term Behavior on Search Personalization. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 185–194, 2012.

[4] C. Eickhoff, K. Collins-Thompson, P. N. Bennett, and S. Dumais. Personalizing Atypical Web Search Sessions. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pages 285–294, 2013.

[5] A. Ghose, P. G. Ipeirotis, and B. Li. Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue. *Management Science*, 2014.

[6] Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection in Personalized Web Search", *Ieee Transactions On Knowledge And Data Engineering Vol:26 No:2 Year 2014*.

[7] M. Tibouchi. Elligator squared: Uniform points on elliptic curves of prime order as uniform random strings. *IACR Cryptology ePrint Archive*, 2014:43, 2014.

[8] Shou, Lidan, et al. "Supporting Privacy Protection in Personalized Web Search."(2012): 1-1.

[9] Y. Zhu, L. Xiong, C. Verdery, "Anonymizing User Profiles for Personalized Web Search", *Proc. 19th Int'l Conf. World Wide Web (WWW)*, pp. 1225-1226, 2010.

[10] A. Krause, E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services", *Artificial Intelligence Research*, vol. 39, pp. 633-662, 2010.

[11] L. Shou, H. Bai, K. Chen, G. Chen, "Supporting Privacy Protection in Personalized Web Search", *IEEE Transaction on Knowledge and Data Engineering*, vol. 26, no. 2, 2014.

[12] G. Chen, H. Bai, L. Shou, K. Chen, Y. Gao, "Ups: Efficient Privacy Protection in Personalized Web Search", *Proceeding. 34th International ACM SIGIR Conference Research and Development in Information*, pp. 615-624, 2011.

[13] Singh, A., Alhadidi, B. 2013. Knowledge Oriented Personalized Search Engine: A Step towards Wisdom Web. *International Journal of Computer Applications* 76(8), 1--9.

[14] Singh, A. and Sharma. A.: 2016. A Multi-Agent Framework for Context Aware Dynamic User Profiling for Web Personalization, In: *AISC Proceedings*, Springer.

[15] Veningston K and Shanmugalakshmi, R. 2012. Enhancing personalized web search re-ranking algorithm by incorporating user profile. In: *Proceeding of ICCNT'12 from 26th-28th July 20, Coimbatore, India*.