International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Privacy Preservation Protection for Personalized Web User by k-Anonymity with Profile Construction for Web Search Engines

Uma Maheswari.T[1], Dr.V. Kavitha[2]

[1]Department of CS, Sri Ramakrishna College of Arts and Science for Women, 395, Sarojini Naidu Road, New Sidhapudur, Coimbatore, Tamil Nadu 641 044, India

[2]Department of CS, Sri Ramakrishna College of Arts and Science for Women, 395, Sarojini Naidu Road, New Sidhapudur, Coimbatore, Tamil Nadu 641 044, India

Abstract: ***Personalized Web Search (PWS) of each user is varied from common web search, since personalized web search of the each user is majorly relying on the queries submitted by user and their user interests with their information needs. Though, substantiation shows that user's disinclination to reveal their private information throughout web search has become a most important barrier designed for the extensive proliferation of PWS. In order to overcome study to privacy protection in PWS relying on their user preferences in this work proposed a hierarchical tree structure for user profiles. Introduces a novel PWS framework named as User Customizable Online Privacy-preserving Search with K-anonymity (UCOPSK) with the intention of is able to adaptively simplify profiles with queries while regarding user specified privacy achievement in both online and offline searching. The proposed UCOPSK framework majorly consists of two phases as important like the offline and online phases, designed for each user. Throughout the offline phase, a hierarchical user profile is created and customized through the user-specified confidentiality requirements. K-anonymity is designed to each user to disclosure sensitive information of user, which is able to successfully prevent the information of each. K-anonymity additionally considers privacy for user through calculation K value. Experimentation evaluation results shows that the proposed UCOPSK achieves highest searching quality, less response time when compare to existing personalized search services, it shows that the proposed UCOPSK methods fully protect user privacy.***

Keywords: Web search engines, Privacy protection, personalized web search, web mining, utility, risk, profile, generalization and K-anonymity**.**

1. Introduction

Though search engines have been effectively organized to serve user's information needs, they are extreme beginning optimal. A most important deficit of existing personalized search engines is with the purpose of they go behind the representation of "one size fits each and every one" and are not adaptive to individual users. This causes inherent non-optimality as is seen clearly within the following 2 cases:

(1) Completely different users might use precisely the same question to go looking for various info, however existing search engines come back constant results for these users. (2) A user's info wants might modification over time. Constant user might use "Java" typically to mean the Java island in country and typically to mean the artificial language. Existing search engines are unable to differentiate such cases. So as to optimize search accuracy, should use additional user info and modify search results in line with every individual user [1]. To envision however personalized search might facilitate improve search accuracy; think about the question "Java" once more. In general, personalized search is taken into account joined of the foremost promising techniques to interrupt the limitation of current search engines and improve the standard of search results. Despite the attractiveness of personalized search, haven't nonetheless seen giant scale uses of personalized search services. This is often not as a result of such services aren't accessible, however doubtless as a result of users aren't comfy with the shortage of protection of user privacy [2-3]. Google, as an example, has deployed a customized search system one. However, to the simplest of our information, it's not been wide adopted by users nonetheless. PWS techniques, the profile-based PWS has incontestable additional effectiveness in raising the standard of net search recently, with increasing usage of non-public and behavior info to profile its users that is typically gathered implicitly from question history [4-5], browsing history [6], click-through information [7-8], user documents [9], then forth. Such implicitly collected personal information will simply reveal a gamut of user's non-public life. Privacy problems arising from the shortage of protection for such information, for example the AOL question logs scandal [11], not solely raise panic among individual users, and however additionally dampen the data-publisher's enthusiasm in providing personalized service. In fact, privacy issues became the most important barrier for wide proliferation of PWS services. To shield user privacy in profile-based PWS, researchers ought to think about 2 contradicting effects throughout the search method. On the opposite hand, they have to cover the privacy contents existing within the user profile to position the privacy risk in restraint. The major contributions of the work is summarized and defined as follows:

- Propose a PWS structure called User Customizable Online Privacy-preserving Search with K-anonymity (UCOPSK). Generalize profiles through K-anonymity designed for each query related to user-specified privacy requirements.
- The generalization procedure is pointed out and measured via two major metrics, specifically the personalization convenience and the privacy possibility, together definite designed for user profiles.

- Consider privacy for user given query by calculation of k anonymity for each query to determine the sensitive value for each query.

The remaining sections are organized as follows. In Section 2, the investigate privacy difficulty is studied based on the background study. In Section 3, procedure of UCOPSK framework is discussed. The experimental results of the existing and proposed methods are experimented in Section 4. Section 5 additional discusses a number of implementation issues of UCOPSK and concludes the paper of the privacy preservation in PWS.

2. Background Knowledge

In this section, study the existing methods user in the privacy protection for PWS and their major issues. This section majorly focuses on study the literature of profile-based personalization, personalized search methods and confidentiality protection in PWS system. In recent work [12], a user profile is created and constructed based on the hierarchy tree structure. User submitted queries and user selected topics is classified into the left and right nodes respectively in the hierarchies tree structure to construct user profile.

Shen et al [6] proposed a language schema to mine instantaneous investigate background and inherent feedback information. The language schema chooses suitable terms beginning connected previous queries and equivalent search results to enlarge the existing query. Teevan et al. [12] make use of rich schema designed to user interests, is constructed depending on together search-related information, and other information regarding the user.

Sugiyama et al. [13] proposed a modified collaborative filtering algorithm to create and build a user profiles to each personalized search. Sun et al. [14] developed a novel method named CubeSVD for each user personalized web search with examination of correlations between users, queries, and web pages in clickthrough data. Smyth et al. [15] proposed a novel collaborative web search be able to be well-organized in numerous search scenarios at what time usual neighbourhood of searchers be able to be identified.

Most recent works build profiles in hierarchal structures because of their stronger descriptive ability, higher quantifiability, and better access potency. The bulk of the hierarchal representations area unit made with existing weighted topic hierarchy/ graph, like ODP [11], Wikipedia [16], and so on. Another add [10] builds the hierarchical profile mechanically via term-frequency analysis on the user information. In our projected UPS framework, don't target the implementation of the user profiles. Actually, our framework will doubtless adopt any hierarchical illustration supported taxonomy of information.

Both [17] and [18] give on-line namelessness on user profiles by generating a gaggle profile of k users. Exploitation this approach, the linkage between the question and one user is broken. In [19], the Useless User Profile (UUP) protocol is projected to shuffle queries among a gaggle of users UN agency issue them. As a result any entity cannot profile a definite individual. These works assume the existence of a trustworthy third-party anonymizer, that isn't promptly on the market over the web at giant. Viejo and Castell_a-Roca [20] use inheritance social networks rather than the third party to supply a distorted user profile to the online programme. Within the theme, each user acts as a research agency of his or her neighbors. They'll commit to submit the question on behalf of UN agency issued it, or forward it to alternative neighbors.

3. Proposed User Customizable Online Privacy-Preserving Search With K-Anonymity (UCOPSK) Methodology

User Customizable Online Privacy-preserving Search with K-anonymity (UCOPSK) framework. In this paper to preserve privacy of the user, proposed UCOPSK framework makes an assumption the queries might not contain some perceptive information, and aspire on protective the confidentiality in single user profiles at the same time as retaining their effectiveness designed for Personalized Web Search (PWS). Figure. 1 shows and illustrates the procedure of the entire system architecture of UCOPSK. The proposed UCOPSK framework majorly consists of two phases as important like the offline and online phase, designed for each user. Throughout the offline phase, a hierarchical user profile is created and customized through the user-specified confidentiality requirements.

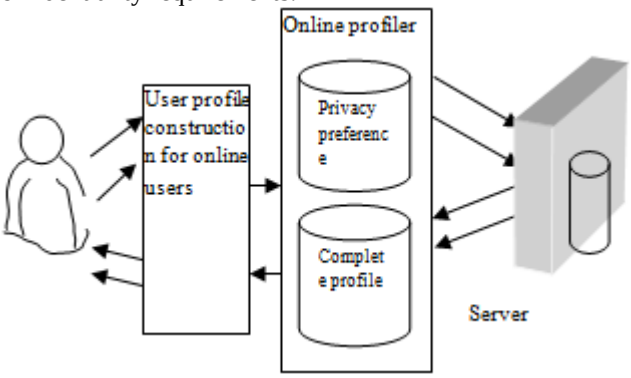

Figure 1: System Architecture of UCOPSK

The online phase queries are submitted to the user and 4 major steps are carryout during this process which is described as follows:

1) If user submits a query q to the client, the proxy server creates a user profile automatically in runtime and dynamic manner. The resultant of this step is generalized user profile G and simultaneously satisfying privacy requirements of single user. The generalization procedure is pointed out and measured via two major metrics, specifically the personalization convenience and the privacy possibility, together definite designed for user profiles.
2) Consequently, the user submitted query and the generalized user profile G are forwarded to the PWS server designed for personalized investigate.
3) The investigate results are personalized through the profile and send reverse to the query proxy.
4) In conclusion, the proxy moreover presents the raw results to the user through the entire user profile.

The major aim of the proposed UCOPSK work is to

determine the privacy protection alongside a typical model of privacy attack.

Knowledge bounded: The background information of the adversary is imperfect to the classification repository R. Together the user profile in the tree H and privacy are specified based on R. Privacy risk of each user is determined based on the total probabilistic to each sensitive nodes, which the adversary be able to possibly recover beginning runtime profile . For fairness amongst diverse users, be able to regularize the privacy risk through $\sum_{s\in S} sen(S)$.Purposely, every user has to assume the following procedures to solve the privacy protection problem in PWS:
Construction of profiles in online and offline
Privacy requirement customization
Online query-topic mapping, and
Online generalization.
Construction of profiles in online and offline

Construction of user profile in a topic hierarchy tree plays a most important role designed for personalization of web user information. Construction of user profile in offline manner will be very easy when compare to construction of user profile in online manner. Because the query searching of each user in online manner in changed dynamically every seconds. Therefore in this work proposed a new schema to construction of profiles in online and offline manner.

Each user profile in UCOPSK implements a hierarchical structure. User profile of each user is constructed relying on the accessibility of a public available taxonomy; it needs to assure the following statement.

The repository 'R' is a enormous topic hierarchy 'H' covering the complete topic area of human information. That is, known some human familiar topic t, a matching node be able to be found in Repository. Given a taxonomy repository 'R', the R' support is presented through itself designed for each leaf topic. The initial step is to construct the unique user profile in a topic hierarchy 'H' with the purpose of disclose user interests. The schemes presume with the intention of the users preferences are represented in text documents, take the following steps:

Detect the respective topic in Repository for every document $d \in D$. Thus, the preference document set D is transformed into a topic set T.

Construct the profile H as a topic-path prefix tree with T, i.e., H=trie(T)

Initialize the user support suph(t) for each topic $t \in T$ with its document support from D, then compute suph(t) of other nodes of H.

Topic Detection in R
The total number of user's click log is represented as $D=\{q_i, d_{i1},..\}$where q_i is denoted as the query in the log and d_{ij} is the document clicked by the user. The algorithm use can be summarized in the following steps:

Weight each and every one user through respect to similarity to the active user. This similarity among users is determined via the Pearson correlation coefficient among their term weight vectors in equation (1)

Choose n clusters with highest similarity which is greater than the user specified threshold value.
Calculate a prediction beginning a weighted grouping of the neighbor's ratings
In step 1, $sim_{a,u}$ is the similarity among users a and user u defined in Equation (1), and n is the total number of user

$$\mathrm{sim}_{a,u} = \frac{\sum_{t=1}^{T}(r_{a,t}-\overline{r_a})\times(r_{u,t}-\overline{r_u})}{\sqrt{\sum_{t=1}^{T}(r_{a,t}-\overline{r_a})^2\times(r_{u,t}-\overline{r_u})^2}} \quad (1)$$

Where $r_{a,i}$ be the rating value of the topic t through user a, and $\overline{r_a}$ be the mean value of $r_{a,i}$ and 'T' is the total number of topics which is searched by user.

In step 2, suitable users are selected based on their similarity value between users. The total number of selected users is set to n for any user. So it is named as static.

In step 3, Calculate a prediction beginning a weighted grouping of the query term weights by means of centroid vectors of clusters

$$\mathrm{P}_{a,t} = \overline{r_a} + \frac{\sum_{u=1}^{u}(r_{u,t}-\overline{r_u})\times Sim_{a,u}}{\sum_{u=1}^{n} Sim_{a,u}} \quad (2)$$

Where $\mathrm{P}_{a,t}$ is the denoted as the prediction value to each active user a designed for topics T. $Sim_{a,u}$ is the similarity among users a and user u defined in Equation (1), and n is the total number of user.

User profile construction for offline users. The initial step of the user profile construction is to construct a unique user profile in a topic hierarchy H with the purpose of disclose user interests. The proposed system makes an assumption that preferences of each user are denoted and characterize in a group of plain text documents. User profile construction steps for offline users are explained in the following steps:

Weight each and every one user through respect to similarity to the active user. This similarity among users is determined via the Pearson correlation coefficient among their term weight vectors in equation (3)
Choose n users with the purpose of largest similarity value to the active user.
Determine a prediction beginning a weighted grouping of the neighbor's term weights.
In step 1, $Sim_{a,u}$, is describes a the similarity value among users a and u, is determined via the computation of Pearson correlation coefficient is described as follows:

$$\mathrm{sim}_{a,u} = \frac{\sum_{t=1}^{T}(w_{a,t}-\overline{w_a})\times(w_{u,t}-\overline{w_u})}{\sqrt{\sum_{t=1}^{T}(w_{a,t}-\overline{w_a})^2\times(w_{u,t}-\overline{w_u})^2}} \quad (3)$$

Where $w_{a,t}$ is denoted as the weight value of the topic through query q depending on the user and determine the term frequency in a searched Web page described through Equation (4),

$$w_{qk}^{hp(cur)} = c^{hp(cur)} \cdot \frac{tf(q_k, hp^{(cur)})}{\sum_{s=1}^{m} tf(q_s, hp^{(cur)})} \quad (4)$$

and $\overline{w_a}$ is the mean value of the weight and 'T' is the total amount of topics

In step 2, suitable users are selected based on their similarity value between users. The total number of selected users is set to n for any user. So it is named as static.

In step 3, Calculate a prediction beginning a weighted grouping of the query term weights by means of centroid vectors of clusters:

$$P_{a,t} = \overline{w_a} + \frac{\sum_{u=1}^{u} (w_{u,t} - \overline{w_u}) \times Sim_{a,u}}{\sum_{u=1}^{n} Sim_{a,u}} \quad (5)$$

Where $P_{a,t}$ is denoted as the calculation of number of active user designed for query term weights, $Sim_{a,u}$ is the similarity among users a and user u defined in Equation (3), and n is the total number of user.

User profile construction for online users
Construction of user profile for online users consists of three major steps is described as follows.

Create and form a clusters depending on the number users, here clustering is performed based on the procedure of k-Nearest Neighbor algorithms. The similarity among user a and these clusters are determined via Pearson correlation coefficient.

Choose n clusters with highest similarity which is greater than the user specified threshold value.

Calculate a prediction beginning a weighted grouping of the query term weights by means of centroid vectors of clusters.
In step 1, $sim_{a,g}$ is represented as the similarity among users a and centroid vectors to each clusters g, via the use of Pearson correlation coefficient, defined below:

$$sim_{a,g} = \frac{\sum_{t=1}^{T} (w_{a,t} - \overline{w_a}) \times (w_{g,t} - \overline{w_g})}{\sqrt{\sum_{t=1}^{T} (w_{a,t} - \overline{w_a})^2 \times (w_{g,t} - \overline{w_g})^2}} \quad (6)$$

where $w_{a,t}$ is denoted as the weight value to each topic t of each user with the query term frequency by Equation (4), and $\overline{w_a}$ is the mean value of the weight and 'T' is the total amount of topics. The number of clusters are formed based on the similarity value between active user, and a weighted result . Calculate a prediction beginning a weighted grouping of the query term weights by means of centroid vectors of clusters:

$$P_{a,t} = \overline{w_a} + \frac{\sum_{g=1}^{u} (w_{a,t} - \overline{w_a}) \times Sim_{a,g}}{\sum_{g=1}^{n} Sim_{a,g}} \quad (7)$$

where $P_{a,t}$ is denoted as the calculation of number of active user a designed for query term weights , $Sim_{a,g}$ is the similarity among users a and centroid vectors to each cluster g in Equation (6), and n is the total number of centriod vectors.

To calculate the sensitive value for given query based on the user constructed profile. User can specify a parameter K for the user given query based on the user constructed profile that they can receive in K anonymous. It protects the privacy of individual's user with their specified user query. K-anonymity mainly focuses on the protection of privacy of individual user and their topics. For constructed user profiles satisfies the K Anonymity based on the following condition.

K-anonymity: A constructed user profile table UP_T assure K- anonymity in favour of each tuple $up \in UP_T$ there be present k-1 other tuples $up_{i1},..up_{i_{k-1}}$ such that $t[C] = t_i[c]$ designed for each and every one $c \in up$. Let $x_1,..x_K$ be a series of k self-determining topics and identically dispersed illustration of the query terms through equally distributed in the alphabet . Let be the several number of times the same query will be asked by user. K-anonymity can preserve and protect the privacy of individual user. For constructed user profile , generalization is forced at the topic level, is equivalent to the amount of different combinations of domains with the purpose of the topics in the constructed user profile table . Specified domain generalization is specified as for topics for user profile table is:

$$\prod_{i=1}^{n} (|DGH_{D_i}| + 1) \quad (8)$$

Offline-2: Privacy Requirement Customization:
Customized privacy requirements are able to be specified through an amount of sensitive-nodes in the user constructed profile. The sensitive nodes are a group of user précised sensitive topics. Sensitivity is a positive value that quantifies the rigorousness of the privacy leakage basis through disclosing the node. Taking into consideration the sensitivity of every sensitive topic as the cost of recovering it, the confidentiality risk is able to be specified as the whole sensitivity of the sensitive nodes. This schema determine the requests the each user to indicate a sensitive-node set $s \subset H$, and their sensitive value $sen(s) > 0$ used for each topic $t \in S$. Subsequently the cost value of each node $t \in H$ as follows:

$$cost(t) = \sum_{t \in C(t,H)} cost(t) \times Pr(t' | t) \quad (8)$$

For user query q the topic mining is achieved through the following two online procedures

Query-Topic Mapping
For user given, the major objective of query-topic mapping is to find the root of hierarchical tree , is named as seed profile, consequently with the intention of each and every one topics

appropriate to q are contained in it and to attain the preference values among q and each and every one topics in H. The procedure of Query-Topic Mapping is performed based on the following steps:

Discover the number of topics in R which is relevant to user specified query q. Then compute a relevance value through the query for each and every one topics in R. It is used to attain a position of nonoverlapping significant topics represented through T(q), specifically the appropriate set in R, include a query-relevant trie represented as R(q).It appears that T(q) are the leaf nodes of R(q).

Overlap R(q) through H to attain the seed profile G_b. which is moreover a deep-rooted subtree of H. The leaves of the seed profile G_0 structure a principally attractive node set among set T(q) and HIndicate it through $T_H(q)$ and observably have $T(q) \subset H$. Subsequently, the preference value of a topic is determined as following:

1) In the hierarchy structure H , topic 't' is considered as the leaf node and ,the preference value of the each topic depending on the query is represented as $\text{pre f}_H(t,q)$ with their user support $\sup_H(q)^3$ which can be obtained directly from the user profile.
2) If topic 't' is considered as the leaf node and $\text{pre f}_H(t,q)$
3) Orelse , t is not a leaf node. The user preference value of the particular is recursively summative beginning its child topics as,

$$\text{pre f}_H(t,q) = \sum_{t \in C(t,H)} pref_H(t',H) \quad (10)$$

At last, it is simple to attain the normalized preference designed for each $t \in C(t,H)$ as,

$$pr(t \mid q,H) = \frac{pref_H(t,q)}{\sum_{t' \in C(t,H)} pref_H(t',q)} \quad (11)$$

Purpose of the first step is to determine the compute relevance value $rel_R(t,q)$ through the query and be able to be used to form a conditional probability with the purpose of indicates how frequently topic t is enclosed with q:

$$pr(t \mid q) = pr(t \mid q,R) = \frac{rel_R(t,q)}{\sum_{t' \in T(q)} rel_R(t',q)} \quad (12)$$

Profile Generalization: Based on the conditional probability, profile of the each user G_0 is generalized in iterative manner depending on the two metrics namely privacy and utility metrics. In adding together, this process also calculates the perceptive power used for make a decision on whether personalization must be employed or not is illustrated in Figure 2.

Metric of Utility: The major objective of the utility loss is to calculate the searching quality of each user with query q on a generalized profile G. Since the investigation quality of the user is majorly relies on the performance of PWS investigate device, which is inflexible to predict. Additionally, it is moreover expensive to request user feedback on investigate results. On the other hand, we transform the efficacy forecast difficulty to the evaluation of the discerning power with user given query q on a profile G. Even though the equivalent statement has been done in [21], but it is not applied to utility loss measurement under hierarchical structure.

Metric of Privacy: The major objective of the privacy loss is to determine the sensitivity value of each user with query q on a generalized profile G. For generalized profile the risk value of exposing each and every one sensitive nodes reaches its maximum, specifically 1. Or else zero. This type risk value computation is done via the measurement of the cost during Offline-2. For known generalized profile the unnormalized risk of revealing it is recursively specified through,

$$Risk(t,G) = \begin{cases} \cos t(t) if t is leaf \\ \sum_{t' \in C(t,G)}^{\cos t(t)} Risk(t',G) otherwise \end{cases} \quad (13)$$

On the other hand, in a number of cases, the cost of a nonleaf node may even be higher than the total risk aggregated beginning its children. Consequently, (13) might undervalue the actual risk. Therefore it is modified for nonleaf node as
Then, the normalized risk be able to be attain through separating the unnormalized risk of the origin node through the entire sensitivity in H, specifically,

$$risk(q,G) = \frac{Risk(root,G)}{\sum_{s \in S} Sen(s)} \quad (14)$$

Particularly, every candidate operator in the queue is represented as op <t,IL(t,G_i)> to be reduced via the calculation of IL(t,G_i) indicates the IL sustain through pruning from G_i
Heuristic 1: The iterative process can terminate whenever δ risk satisfied.
The subsequent work of the second step is to determine the Information Loss (IL) to generated and generalized user profile by evaluating $\Delta PG(q,G) = PG(q,G_i) - PG(q,G_{i+1})$
Every time if effort to reduce t, essentially combine t into shadow to attain a original shadow leaf , simultaneously through the preference of t, i.e.,

$$\Pr(shd \mid q,G) = pr(shd \mid q,G) + pr(t \mid q,G) \quad (15)$$

In conclusion, have the following heuristic, which considerably straightforwardness the calculation of information loss IL(t).

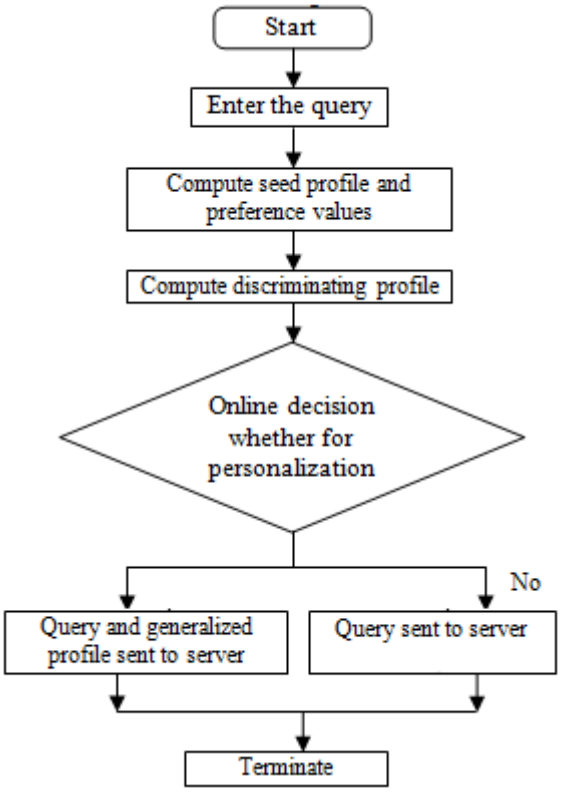

Figure 2: Query Topic Mapping and Profile Generalization

Heuristic 2:

$$IL(t) = \begin{cases} pr(t \mid q,G)(IC(t) - IC(part(t,G))), case1 \\ dp(t) + dp(shadow) - dp(shadow), case2 \end{cases} \quad \textbf{(16)}$$

$$dp(t) = pr(t \mid q,G) \log \frac{pr(t \mid q,G)}{pr(t)} \quad (17)$$

The final step of the work is to prune-leaf nodes topics based on solitary topic 't' that belongs to case . While in case, reducing the topics 't' acquire recomputation of the first choice values of its sibling nodes.

Heuristic 3: In the hierarchical tree structure once a leaf node topic't' is reduced, if and only if the candidate operators reduce t's sibling topics should toward be updated in 'Q'.

4. Results and Discussion

In this section, present the experimental results of UCOPS. In the primary experimentation, learning the complete results of the metrics in every iteration of the UCOPS and existing methods .Second, examine the results of the proposed and existing schemas under query-topic mapping. Third, examine the results of scalability between proposed and existing schemas in terms of response time. In the final stage of the experimentation analysis, learn the efficiency of clarity calculation and the hunt quality of UPS and UCOPSK. For experimentation work refer a topic repository make use of the ODP web Directory. To focal point on the pure English categories, remove taxonomies "Top/World" and "Top/Adult/ World." The log files of the each user are downloaded from online AOL query log. This log consists of more than 20 million queries and 30 million clicks of 650k users during the period of 3 months. The log file format of each user is described as follows: <uid; query; time [rank; url]>

Search Quality is defined as the relevant search results relying on the user query and the constructed user profile as per user's interests.

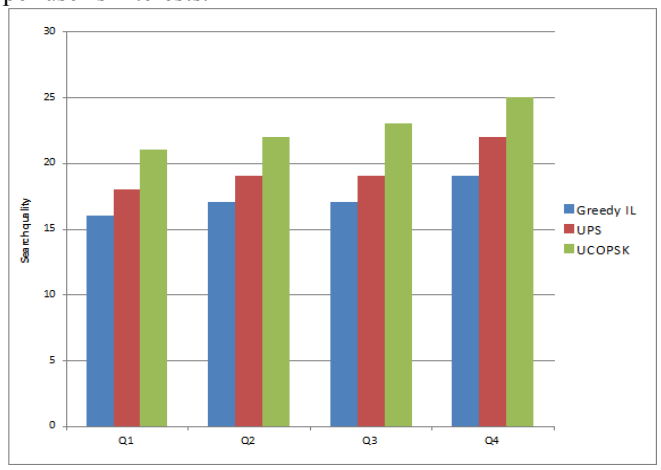

Figure 3: Performance Comparison based on Search Quality

Figure.3 gives the comparison of the existing system of GreedyIL relying on Quality of search, UPS framework and proposed UCOPSK. The number of query sets in the dataset is represented as Q1-Distinct Q2- Medium, Q3- Ambiguous Q4-Very ambiguous is denoted in X-axis and searching quality results are plotted in Y-axis. When compare to other methods UCOPS achieves 13% of improvement in search quality than the GreedyIL. The results are tabulated in Table 1.

Table 1: Evaluation Results for Search Quality

Query Set	Greedy DP Relevant URLs	UPS	UCOPSK
Q1	16.0	18	21
Q2	17.0	19	22
Q3	17.0	19	23
Q4	19.0	22	25

Figure. 4 shows the performance comparison results of the various schemas by varying the privacy threshold. Figure 4 gives the comparison of the existing system of GreedyIL, UPS and proposed UCOPSK based on the effectiveness of personalization. The Privacy threshold is plotted in X-axis and the average precision is plotted in Y-axis. Based on the privacy threshold value, the AVP varies through admiration to generalization. The UCOPSK achieves 15% of improvement in personalization than the GreedyIL

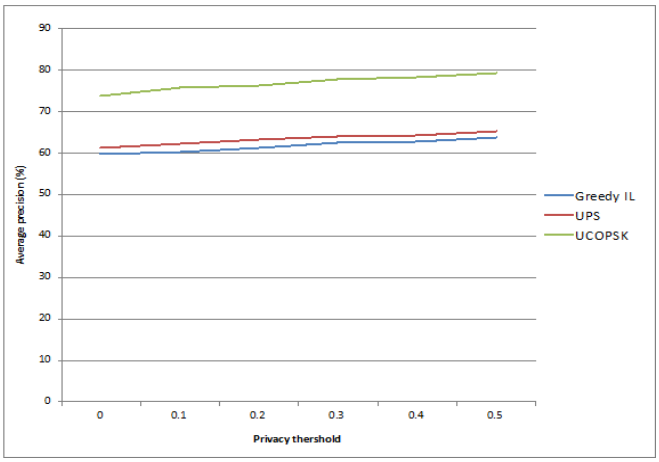

Figure 4: Effectiveness of personalization on varying

Response time : Response time is defined as the time required for generalization of profile following issuing the query relying on the privacy requirements of the user.

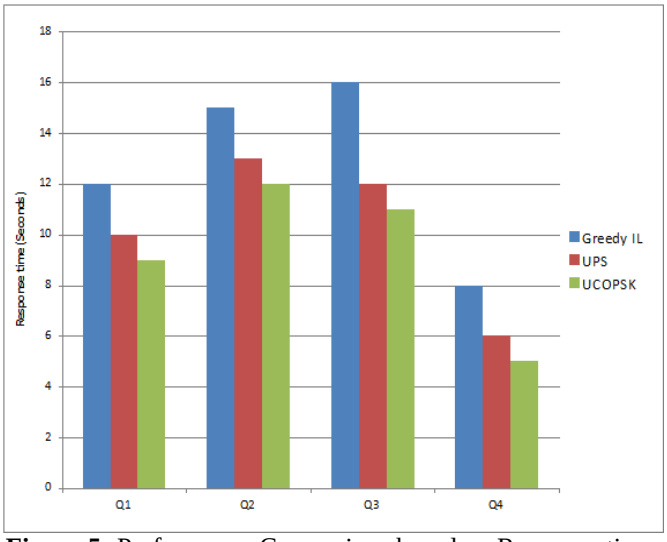

Figure 5: Performance Comparison based on Response time

Figure. 5 gives the comparison of the existing system of GreedyIL, UPS and proposed based on the response time taken by the query sets. The number of query sets in the dataset is represented as Q1-Distinct Q2- Medium, Q3-Ambiguous Q4-Very ambiguous is denoted in X-axis and average time results are plotted in Y-axis. The UCOPS achieves 12% of improvement in response time than the GreedyIL. The results are tabulated in Table 2.

Table 2: Evaluation Results for Response Time

Query Set	Response Time (sec)		
	Greedy IL	UPS	UCOPSK
Q1	12	10	9
Q2	15	13	12
Q3	16	12	11
Q4	8	6	5

Scalability: Scalability is defined as the system's capability to hold the rising profile size in a proficient manner or its capability to be distended to accommodate that growth.

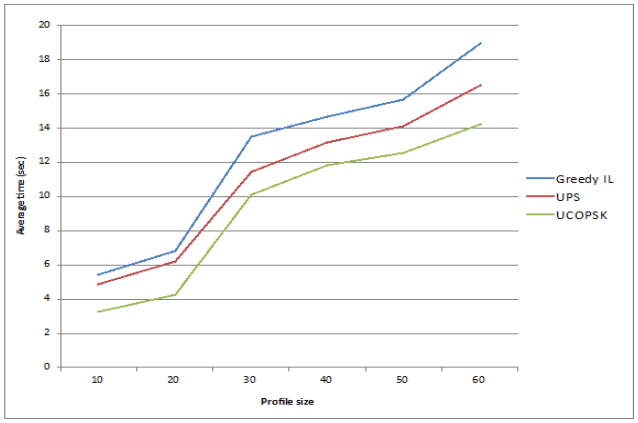

Figure 6: Performance Comparison based on Profile Size

Figure.6 gives the comparison of the existing system, GreedyIL, UPS and proposed UCOPSK based on the scalability of varying profile size. The Profile Size is plotted in X-axis and the average time is plotted in Y-axis. The UCOPSK achieves 11% of improvement in scalability than the GreedyIL; the results are tabulated in Table 3.

Table 3: Evaluation Results for Scalability of varying profile size

Profile Size (No of nodes)	Average Time (sec)		
	Greedy IL	UPS	UCOPSK
10	5.4	4.85	3.25
20	6.83	6.23	4.26
30	13.5	11.45	10.11
40	14.65	13.14	11.85
50	15.68	14.12	12.58
60	18.94	16.48	14.21

5. Conclusion and Future Work

In this work proposes a novel PWS framework named as User Customizable Online Privacy-preserving Search with K-anonymity (UCOPSK) with the intention of be able to adaptively simplify profiles with queries while regarding user specified privacy achievement in both online and offline searching. In this paper to preserve privacy of the user, proposed UCOPSK framework makes an assumption that queries might not contain some perceptive information, and aspire on protective the confidentiality in single user profiles at the same time as retaining their effectiveness designed for Personalized Web Search (PWS). The protect privacy for user the k anonymity is applied to each user and their topics, simultaneously assign sensitive value to each topic. A client-side confidentiality security UCOPSK is applied for personalized web search. The UCOPSK method allowed users in the direction of identify personalized privacy requirements by means of the hierarchical profiles. The privacy result of the proposed UCOPSK is compared to existing GreedyIL and UPS methods for the online generalization. The UCOPSK might attain high searching quality search results and preserve privacy requirements of user when compare to existing GreedyIL and UPS methods. The future work effort in the direction of defends against adversaries through broader background information, such as richer association in the middle of topics.

References

[1] J. Pitkow, H. Sch¨utze, T. Cass, et al, " Personalized search" , Communications of the ACM, 45(9):50–55, 2002.
[2] C.-M. Karat, C. Brodie, and J. Karat, " Usable privacy and security for personal information management", Communications of the ACM, 49(1), pp.56–57, 2006.
[3] S. Sackmann, J. Strker, and R. Accorsi , "Personalization in privacy-aware highly dynamic systems", Communications of the ACM, 49(9), pp. 32–38, 2006.
[4] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories", Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
[5] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy", Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.

[6] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search", Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.
[7] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.
[8] 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.
[9] 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.
[10] K. Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, Aug. 2006.
[11] Chirita P.A., Nejdl W., Paiu R., and Kohlschu¨tter C, "Using ODP metadata to personalize search", In Proc. 31st Annual Int. ACM SIGIR Conf. on Research and Development in InformationRetrieval, 2005, pp. 178–185
[12] Teevan J., Dumais S.T., and Horvitz E , "Personalizing search via automated analysis of interests and activities", In Proc. 31st Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 2005, pp. 449–456.
[13] Sugiyama K., Hatano K., and Yoshikawa M , "Adaptive web search based on user profile constructed without any effort from users", In Proc. 12th Int. World Wide Web Conference, 2004, pp. 675–684.
[14] Sun J.-T., Zeng H.-J., Liu H., Lu Y., and Chen Z , "CubeSVD: a novel approach to personalized web search", In Proc. 14th Int. World Wide Web Conference, 2005, pp. 382–390.
[15] Smyth B., Coyle M., Boydell O., Briggs P., Balfe E., Freyne J., and Bradley K , "A live-user evaluation of collaborative web search", In Proc. 19th Int. Joint Conf. on AI, 2005.
[16] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia", HP Labs, 2008.
[17] 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.
[18] 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.
[19] J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines", Computer Comm., 32(13/14), pp. 1541-1551, 2009.
[20] A. Viejo and J. Castell_a-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", Computer Networks, 54(9), pp. 1343-1357, 2010.
[21] A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services", J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.