International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

# Implementing the Supporting Privacy Protection in Customized net Search

**N. Rajendran[1], N. Prakash[2]**

[1]M.Tech, (Ph.d)., Student, Department of Computer Science & Engineering, Siddharth Institute of Engineering & Technology, Puttur, Chittoor Dt. (A.P)

[2]M.Tech, (Ph.d), Professor, Department of Computer Science & Engineering, Siddharth Institute of Engineering & Technology, Puttur, Chittoor Dt. (A.P)

**Abstract:** ***Personalized internet search (PWS) has in contestible its effectiveness in up the standard of varied search services on the net. However, evidences show that users' reluctance to disclose their personal data throughout search has become a significant barrier for the wide proliferation of PWS. we have a tendency to study privacy protection in PWS applications that model user preferences as hierarchic user profiles. we have a tendency to propose a PWS framework known as UPS which will adaptively generalize profiles by queries whereas respecting user specified privacy necessities. Our runtime generalization aims at hanging a balance between 2 prophetic metrics that judge the utility of personalization and also the privacy risk of exposing the neralized profile. we have a tendency to gift 2 greedy algorithms, specifically GreedyDP and GreedyIL, for runtime generalization. we have a tendency to additionally give an internet prediction mechanism for deciding whether or not personalizing a question is helpful. intensive experiments demonstrate the effectiveness of our framework. The experimental results ditionally reveal that GreedyIL considerably outperforms GreedyDP in terms of potency.***

**Keywords:** Privacy protection, personalized web search, utility, risk, profile

## 1. Introduction

The web program has long become the foremost important portal for standard individuals craving for helpful min formation on the net. However, users would possibly expertise failure once search engines come back orthogonal results that do not meet their real intentions. Such unconnectedness is essentially due to the large sort of users' contexts and backgrounds, oreover because the ambiguity of texts. Customized web search (PWS) could be a general class of search techniques aiming at providing higher search results, that square measure tailored for individual user wants. because the expense, user data has to be collected and analyzed to work out the user intention behind the issued question.

The solutions to PWS will usually be categorised into two types, particularly click-log-based strategies and profile-based ones. The click-log based mostly strategies ar straight forward-they merely impose bias to clicked pages within the user's question history. Though this strategy has been incontestable to perform systematically and significantly well it will solely work on perennial queries from constant user, which is a strong limitation confining its elevancy. In distinction, profile-based strategies improve the search expertise with complicated user-interest models generated from user profiling techniques. Profile-based strategies may be probably effective for nearly all kinds of queries, but are reported to be unstable beneath some circumstances.

### A. Motivation

To protect user privacy in profile-based PWS, researchers have to contemplate 2 contradicting effects throughout the search process. On the one hand, they arrange to improve the search quality with the personalization utility of the user profile. On the opposite hand, they have to cover the privacy contents existing within the user profile to position the privacy risk under management. some previous studies counsel that people ar e willing to compromise privacy if the personalization by activity user profile to the computer program yields higher search quality. In a perfect case, vital gain are often obtained by personalization at the expense of solely alittle (and less-sensitive) portion of the user profile, particularly a generalized profile. Thus, user privacy will be protected while not compromising the personalised search quality. In general, there's a exchange between the search quality and therefore the level of privacy protection achieved from generalization. Unfortunately, the previous works of privacy conserving PWS ar removed from best. the issues with the prevailing methods ar explained within the following observations

1) the prevailing profile-based PWS don't support runtime profiling. A user profile is usually generalized for only once offline, and accustomed change all queries from a same user indiscriminatingly. such "one profile fits all" strategy definitely has drawbacks given the range of queries. One proof according in is that profile-based personalization could not even facilitate to boost the search quality for some accidental queries, although exposing user profile to a server has place the user's privacy in danger.
2) The present strategies don't take under consideration the customization of privacy necessities. This most likely makes some user privacy to be overprotected whereas others insufficiently protected. as an example, in [10], all the sensitive topics square measure detected exploitation associate absolute metric referred to as disruption supported the information theory, assumptive that the interests with less user document support square measure a lot of sensitive. However, this assumption will be doubted with a simple counterexample: If a user encompasses a massive number of documents concerning

"sex," the disruption of this topic could cause a conclusion that "sex" is very general and not sensitive, despite the reality which is opposite. sadly, few previous work can effectively address individual privacy wants during the generalization

3) Several personalization techniques need repetitious user interactions once making customized search results. They usually refine the search results with some metrics that need multiple user interactions, such as rank evaluation average rank and so on. This paradigm is, however, unworkable for runtime profiling, because it won't solely create an excessive amount of risk of privacy breach, however additionally demand preventative process time for identification.

### B. Contributions

The on top of issues square measure self-addressed in our UPS (literally for User customizable privacy-preserving Search) framework. The framework assumes that the queries don't contain any sensitive info, and aims at protective theprivacy in individual user profiles whereas holding the irusefulness for PWS.

As illustrated in Fig. 1, UPS consists of a nontrusty search engine server and variety of shoppers. Every shopper (user) accessing the search service trusts nobody however himself/herself. The key element for privacy protection is associate online profiler enforced as a research proxy running on the client machine itself. The proxy maintains each the complete user profile, in an exceedingly hierarchy of nodes with linguistics,and the user-specified (customized). The framework works in 2 phases, specifically the offline and on-line part, for every user. throughout the offline part, a hierarchical user profile is built and customised with the user-specified privacy requirement

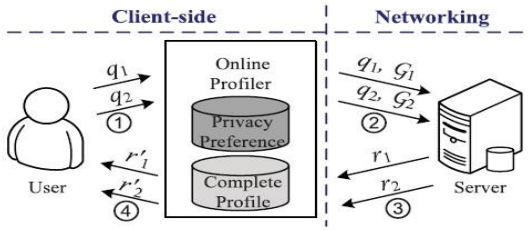


Fig. 1. System architecture of UPS.

1) once a user problems a question ch'i on the shopper, the proxy generates a user profile in runtime within the light of question terms. The output of this step may be a generalized user profile Gi satisfying the privacy requirements. The generalization method is radio-controlled by considering
2) Conflicting metrics, namely the personalization utility and also the privacy risk, both defined for user profiles.
3) Later on, the question and also the generalized user profile square measure sent along to the PWS server for personalized search.
4) The search results square measure customized with the profile and delivered back to the question proxy.
5) Finally, the proxy either presents the raw results to the user, or reranks them with the whole user profile

## 2. Related Works

In this section, we tend to summary the connected works. we tend to concentrate on the literature of profile-based personalization and privacy protection in PWS system

### A. Profile-Based Personalization:

Previous works on profile-based PWS chiefly concentrate on improving the search utility. the fundamental plan of those works is to tailor the search results by pertaining to, often implicitly, a user profile that reveals a personal information goal. within the remainder of this section, we review the previous solutions to PWS on 2 aspects, namely the illustration of profiles, and therefore the live of the effectiveness of personalization.

Many profile representations square measure out there within the literature to are made with existing weighted topic hierarchy/graph, like ODP1 Wikipedia , and so on. Another add builds the hierarchical profile mechanically via term-frequency analysis on the user information. In our projected UPS framework, we do not concentrate on the implementation of the user profiles. Actually, our framework will doubtless adopt any ranked representation supported a taxonomy of data. As for the performance measures of PWS within the literature, Normalized Discounted accumulative Gain (nDCG) could be a common live of the effectiveness of an data retrieval system. it's supported a humangraded relevance scale of item-positions within the result list, and is, therefore, known for its high price in specific facilitate totally different personalization ways.

### B. Privacy Protection in PWS System

Generally there square measure 2 categories of privacy protection problems for PWS. One category includes those treat privacy as the identification of a personal, as delineate in [20]. The other includes those contemplate the sensitivity of the info, particularly the user profiles, exposed to the PWS server.

| Symbol | Description |
|---|---|
| $\|T\|$ | The count of nodes of the tree $T$ |
| $t \in T/N \subset T$ | $t$ is a node ($N$ is a node set) in the tree $T$ |
| $subtr(t,T)$ | The subtree rooted on $t$ within the tree $T$ |
| $rsbtr(N,T)$ | The rooted subtree of $T$ by removing the set $N$ |
| $trie(N)$ | The topic-path prefix tree built with the set $N$ |
| $root(T)$ | The root of the tree $T$ |
| $par(t,T)$ | The parent of $t$ in the tree $T$ |
| $lca(N,T)$ | The least common ancestor of the set $N$ in $T$ |
| $C(t,T)$ | The children of $t$ within the tree $T$ |

The solutions at school 2 don't need third-party assistance or collaborations between social network entries. In these solutions, users solely trust themselves and can't tolerate the exposure of their complete profiles AN namelessness server. Krause and Horvitz use applied mathematics techniques to be told a probabilistic model, and so use this model to come up with the near-optimal partial profile. One main limitation during this work is that it builds the user profile as a finite

set of attributes, and therefore the probabilistic model is trained through predefined frequent queries. projected a privacy protection resolution for PWS based mostly on ranked profiles. employing a user-specified hreshold, a generalized profile is obtained in impact as a stock-still subtree of the entire profile. sadly, this work doesn't address the question utility, that is quality of PWS. For comparison, our approach takes each the privacy demand and therefore the question utility under consideration. A a lot of vital property that distinguishes our work from in Privacy-Preserving information Publishing (PPDP). an individual will specify the degree of privacy protection for her/his sensitive values by specifying "guarding nodes" within the taxonomy of the sensitive attribute. encourage by this, we tend to permit users to customise privacy desires in their ranked user profiles. Aside from the higher than works, a handful of recent studies have raised a stimulating question havefound that personalization could have completely different effects on different queries. Queries with smaller click-entropies, namely distinct queries, square measure expected to learn additional from personalization, whereas those with larger values (ambiguous ones) aren't. Moreover, the latter could even cause privacy disclosure. Therefore, the requirement for personalization becomes questionable for such queries. collect a collection of options of the question to classify queries by their clickentropy

## 3. Preliminaries and Problem Definition

In this section, we have a tendency to 1st introduce the structure of user profile in UPS. Then, we have a tendency to outline the bespoke privacy requirements on a user profile. Finally, we have a tendency to gift the attack model and formulate the matter of privacy preserving profile generalization. For easy presentation, Table one summarizes all the symbols employed in this paper.

### A. User Profile:

Consistent with several previous works in personalised internet services, every user profile in UPS adopts a hierarchical structure. Moreover, our profile is built supported the availability of a public accessible taxonomy, denoted as R, which satisfies the subsequent assumption. Assumption one. The repository R may be a immense topic hierarchy covering the whole topic domain of human information. That is, given any human recognizable topic t, a corresponding node

supR(t)=∑t'єc(t,R)supR(t')

Equation (1) is wont to calculate the repository support of all topics in R, looking forward to the subsequent assumption that the support values of all leaf topics in R square measure on the market. Assumption a pair of Given a taxonomy repository R, the repository support

In fact, Assumption a pair of is relaxed if the support values aren't on the market. In such case, it's still potential to "simulate" these repository supports with the topological structure of R. That is, supR(t') is calculated because the count of leaves in Based on the taxonomy repository, we have a tendency to outline a chance model for the subject domain of the human information. In the model, the repository R is viewed as a hierarchical partitioning of the universe (represented by the root topic) and each topic t a pair of R stands for a random event. The probability (s is Associate in Nursing relative of t) is defined because the proportion of repository support:

Thus, pr(t) is more outlined as
pr(t)=pr(t\root(R)),
where root(R) is that the root topic that has chance one.now, we have a tendency to gift the formal definition of user profile. Definition one (USER PROFILE/H)user profile H, as a hierarchical illustration of user interests, may be a nonmoving subtree of R. The notion nonmoving Definition a pair of (ROOTED SUBTREE). Given 2 trees S and T , S may be a nonmoving subtree of T if S is generated from T by removing a node set X T (together withsubtrees)from T A diagram of a sample user profile is illustrated in Fig. 2a, that is built supported the sample taxonomy repository in Fig. 2b. we are able to observe that the owner of this profile is principally inquisitive about engineering science and Music because the most important portion of this profile is created from fragments from taxonomies

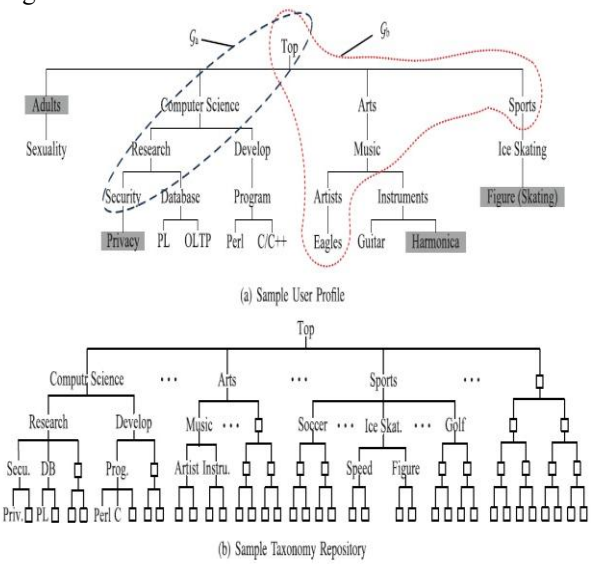


Fig. 2. Taxonomy-based user profile.

Although a user profile H inherits from R a set of topic nodes and their links, it doesn't duplicate the repository supports. Instead, every topic t a pair of H is labeled with a user support, denoted by supHðtÞ, that describes the user's preference on the various topic t. almost like its repository counterpart, the user support may be recursively aggregated from those nominal on the leaf

supH(t)=∑t'єc(t,H)supH(t')

### B. Customized Privacy necessities

Customized privacy necessities may be nominal with a number of sensitive-nodes (topics) within the user profile, whose disclosure (to the server) introduces privacy risk to the user. Given a sensitive-node s, its sensitivity, may be a positive price that quantifies the severity of the privacy escape caused by values expressly indicate the user's privacy issues, the foremost simple privacypreserving method is to get rid of subtrees nonmoving in the slightest

degree sensitive-nodes whose sensitivity values ar larger than a threshold. Such methodology is observed as forbidding. However, forbidding is way from enough against a a lot of sophisticated opposer. to obviously illustrate the limitation of forbidding, we tend to 1st introduce the attack

### C. Attack Model

Our work aims at providing protection against a typical model of privacy attack, particularly eavesdropping. As shown in to corrupt Alice's privacy, the listener Eve successfully intercepts the communication between Alice and the PWS-server via some measures, like man-in-themiddle attack, incursive the server, and so on. Consequently, whenever Alice problems Query q, the complete copy of Q together with a runtime profile

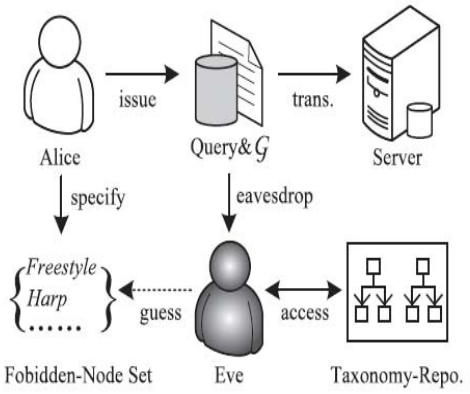


Fig. 3. Attack model of personalized web search.

Alice by convalescent the segments hidden from the initial H and computing a confidence for every recovered topic, relying on the information within the in public available taxonomy repository R.

Knowledge finite. The information of the adversary is proscribed to the taxonomy repository R. Both the profile H and privacy area unit outlined supported R. Session finite. None of antecedently captured infois available for tracing a similar victim in an exceedingly long duration. In alternative words, the eavesdropping are going to be started and over at intervals one question session programs for causing targeted (spam) advertisements to a large amount of PWS-users. These programs seldom act as a real individual that collects prolific info of a selected victim for a protracted time because the latter is way a lot of expensive. If we have a tendency to take into account the sensitivity of every sensitive topic because the cost of convalescent.

### D. Generalizing User Profile:

Now, we tend to exemplify the inadequacy of forbidding operation. In the sample profile in is such as as a sensitive node. Thus, solely releases its parent Ice Skating. sadly, AN mortal will recover the subtree of skating counting on the repository shown in Fig. 2b, where Figure may be a main branch of skating besides Speed. If the likelihood of touching each branches is equal, the adversary will have fifty p.c confidence on Figure. This may result in high privacy risk if is high. A safer solution would take away node skating in such case for privacy protection. In distinction, it'd be unnecessary to remove sensitive nodes with low sensitivity

The address the matter with forbidding, we tend to propose a technique, that detects and removes a collection of nodes X from H, specified the privacy risk introduced by exposing is often in check. Set X is often different from S. For clarity of description, we tend to assume that all the subtrees of H unmoving at the nodes in X don't overlap each other. This method is termed generalization, and the output G may be a generalized profile.

## 4. UPS Procedures

In this section, we tend to gift the procedures distributed for each user throughout 2 totally ifferent execution phases, namely the offline and on-line phases. Generally, the offline sectionconstructs the first user profile so performs privacy demand customization in line with user-specified topic sensitivity. the following on-line section finds the Optimal - Risk Generalization resolution within the search area determined by the custom-made user profile.

As mentioned within the previous section, the online generalization procedure is guided by the worldwide risk and utility metrics. The computation of those metrics depends on two intermediate information structures, specifically a value layer and a preference layer outlined on the user profile. the price layer defines for every node t a pair of H a price} value costðtÞ zero which indicates the overall sensitivity in danger caused by the disclosure of t. These value values are often computed offline from the user-specified sensitivity values of the sensitive nodes. The preference layer is computed on-line once a query alphabetic character is issued. It contains for every node t a pair of H a worth indicating the user's query-related preference on topic t. These preference values ar computed looking forward to a procedure known as question topic mapping. Specifically, every user should undertake the subsequent procedures in our solution:

1) offline profile construction,
2) offline privacy demand customization,
3) on-line query-topic mapping, and
4) on-line generalization.

Offline-1. Profile Construction. the primary step of the offline processing is to create the first user profile in a very topic hierarchy H that reveals user interests. we tend to assume that the user's preferences ar drawn in a very set of plain text documents, denoted by D.

Offline-2. Privacy demand Customization. This procedure first requests the user to specify a sensitive-node set S H, and also the individual sensitivity worth senðsÞ > zero for each topic s a pair of S

Online-1. Query-topic Mapping. Given {a alphabetic characteruery|a question |a question} q, the purposes of query-topic mapping ar 1) to work out a rooted subtree of H, that is named a seed profile, so all topics relevant to alphabetic character ar contained in it; and 2) to get the

preference values between alphabetic character and every one topics in H.

Online-2. Profile Generalization. This procedure generalizes the seed profile G0 in a very cost-based reiterative manner relying on the privacy and utility metrics. additionally, this procedure computes the discriminating power for on-line decision on whether or not personalization ought to be used.

## 5. Metrics:

### A. Metric of Utility

The purpose of the utility metric is to predict the search quality (in revealing the user's intention) of the question letter on a generalized profile G. the rationale for not measurement the search quality directly is as a result of search quality depends largely on the implementation of PWS programme, which is hard to predict. additionally, it's too pricey to solicit user feedback on search results. instead, we have a tendency to rework the utility prediction drawback to the estimation of the discriminating power of a given question letter on a profile G below the following assumption. Assumption three. once a PWS programme is given, the search quality is simply determined by the discriminating power of the exposed query-profile combine Although an equivalent assumption has been created in to model utility, the metric in this work can't be utilized in our drawback settings as our profile may be a graded structure instead of a flat one. Given a graded profile G and {a letteruery|a question |a question} q, we are able to intuitively expect additional discriminating power once . additional specific topics ar ascertained in the distribution of is additional focused on many topics in the topics in ar additional almost like one another. Therefore, an efficient utility metric ought to be consistent with observations and To propose

### B. Metric of Privacy:

The privacy risk once exposing G is outlined because the totalsensitivity contained in it, given in normalized kind. In the worst case, the initial profile is exposed, and also the risk of exposing all sensitive nodes reaches its most, namely 1. However, if a sensitive node is cropped and its ascendent nodes ar preserved throughout the generalization, we have a tendency to still have to evaluate the chance of exposing the ancestors. this will be done exploitation the value layer computed throughout Offline-2.

However, in some cases, the value of a nonleaf node may even be bigger than the whole risk aggregate from its children. as an example, within the profile Gb (Fig. 2a), the value of Music is larger than that of creative person since Music has sensitivity propagation from its sensitive descendent Harmonica.

$$Risk(t,\mathcal{G}) = max\left(cost(t), \sum_{t' \in C(t,\mathcal{G})} Risk(t',\mathcal{G})\right).$$

Then, the normalized risk may be obtained by dividing the unnormalized risk of the foundation node with the whole sensitivity in H, namely

$$risk(q,\mathcal{G}) = \frac{Risk(root,\mathcal{G})}{\sum_{s\in S} sen(s)}.$$

### C. On-line Decision:To Personalize or Not

The results rumored in [1] demonstrate that there exist a good amount of queries referred to as distinct queries, to that the profile-based personalization contributes very little or perhaps reduces the search quality, whereas exposing the profile to a server would needless to say risk the user's privacy. To address this drawback, we have a tendency to develop a web mechanism to choose whether to change a question. the fundamental plan is straightforward if a definite question is known throughout generalization, the entire runtime identification are aborted and also the query are sent to the server while not a user profile. We determine distinct queries exploitation the discriminating power (defined in Section five.1). Specifically, bear in mind that the personalization utility is outlined because the gain in refugee once exposing the generalized profile with the question. Thus, we consider the distinct queries as those with sensible refugee even when the shopper doesn't expose any profile. Given a questionq, is considered a definite question.

## 6. The Generalization Algorithms

We begin by introducing a brute-force optimum formula, which is well-tried to be NP-hard. Then, we have a tendency to propose 2 greedy algorithms, specifically the GreedyDP and GreedyIL.

### A. The Brute-Force formula:

The brute-force formula exhausts all attainable frozen subtrees of a given seed profile to seek out the optimum generalization. The privacy needs ar revered throughout the exhaustion. The subtree with the optimum utility is chosen as the result. though the seed profile G0 is considerably smaller thanH, the exponential machine quality of brute-force formula continues to be unacceptable. Formally, we have the subsequent theorem whose proof is given within the Theorem 1. The -RPG drawback (Problem 1) is NP-hard.

### B. The GreedyDP formula

Given the quality of our drawback, a additional sensible solution would be a near-optimal greedy formula. As preliminary, we have a tendency to introduce Associate in Nursing operator referred to as prune-leaf, which indicates the removal of a leaf topic t from a profile. Formally, we have a tendency to denote by the method of pruning leaf t from Gi to get . Obviously, the optimum profile The first greedy formula GreedyDP works in an exceedingly bottomup manner. ranging from in each ith iteration, GreedyDP chooses a leaf topic t two for pruning, trying to maximize the utility of the output of this iteration, namely . throughout the iterations, we have a tendency to additionally maintain a bestprofile- so-far, that indicates the having the very best discriminating power whereas satisfying the -risk constraint. The repetitive method terminates once the profile is generalized to a root-topic. The best-profile-so-far are the main drawback of GreedyDP is that it needs recomputation

of all candidate profiles (together with their discriminating power and privacy risk) generated from attempts of prune-leaf on all two. This causes significant memory needs and machine price

**C. The GreedyIL formula:**

The GreedyIL formula improves the potency of the generalization exploitation heuristics supported many findings. One vital finding is that any prune-leaf operation reduces the discriminating power of the profile. In other words, the refugee displays monotonicity by prune-leaf. Formally, we've got the subsequent theorem: Theorem 2. If may be a profile obtained by applying a prune-leaf operation on G, then Considering operation within the ith iteration, maximizing is like minimizing the incurred data loss, that is outlined as The higher than finding motivates United States of America to keep up a priority queue of candidate prune-leaf operators in down order of the data loss caused by the operator. Specifically, each candidate operator within the queue may be a tuple like wherever t is that the leaf to be cropped by op and indicates the IL incurred by pruning t from Gi. This queue, denoted by letter, allows quick etrieval of the bestso- far candidate operator. Theorem two additionally results in the subsequent heuristic, which reduces the whole machine price considerably Heuristic one. The repetitive method will terminate whenever –risk is happy. The second finding is that the computation of IL may be simplified to the analysis of the rationale is that, relating (the second term remains unchanged for any pruning operations until one leaf is left (in such case the sole alternative for pruning is that the single leaf itself). what is more, consider two attainable cases as being illustrated in t may be a node with no siblings, and t may be a node with siblings. The case is straightforward to handle. However, the analysis of IL in case2needs introducing a shadow sibling4 of t. Each time if we have a tendency to plan to prune t, we have a tendency to really merge t into shadow to get a replacement shadow leaf shadow0, beside the preference of t, i.e., PrðshadowFinally, we've got the subsequent heuristic, that considerably eases the computation of It may be seen that each oneterms in may be computed expeditiously. Heuristic two.

The third finding is that, just in case delineated higher than, prune-leaf solely operates on one topic t. Thus, it doesn't impact the IL of different candidate operators in letter. While in case , pruning t incurs recomputation of the preference values of its relation nodes. Therefore, we have Heuristic three. Once a leaf topic t is cropped, solely the candidate operators pruning t's relation topics ought to be updated in letter. In other words, we have a tendency to solely ought to recompute the IL values for operators making an attempt to prune t's relation topics. Algorithm one shows the pseudocode of the GreedyIL algorithm. In general, GreedyIL traces the data loss instead of the discriminating power. this protects lots of computational price. within the higher than findings, Heuristic one (line 5) avoids spare iterations. Heuristics two additional simplifies the computation of IL. Finally, Heuristics three reduces the requirement for IL-recomputation significantly. within the worst case, all topics within the seed profile have relation nodes, then GreedyIL has machinec However, this can be extraordinarily rare in observe. Therefore, GreedyIL is anticipated to significantly outdo GreedyDP.

## 7. Conclusions

This paper conferred a client-side privacy protection framework known as UPS for personalised internet search. UPS could probably be adopted by any PWS that captures user profiles in an exceedingly ierarchical taxonomy. The framework allowed users to specify bespoke privacy needs via the hierarchical profiles. additionally, UPS additionally performed online generalization on user profiles to safeguard the private privacy while not compromising the search quality. We proposed 2 greedy algorithms, particularly GreedyDP and GreedyIL, for the web generalization. Our experimental results disclosed that UPS might win quality search results whereas conserving user's bespoke privacy needs. The results additionally confirmed the effectiveness and efficiency of our resolution. For future work, we'll attempt to resist adversaries with broader information, like richer relationship among topics (e.g., snobbery, sequentiality, and so on), or capability to capture a series of queries (relaxing the second constraint of the human in Section three.3) from the victim. we'll additionally request additional refined methodology to build the user profile, and higher metrics to predict the performance (especially the utility) of UPS.

## References

[1] 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.
[2] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search viaAutomated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in InformationRetrieval (SIGIR), pp. 449-456, 2005.
[3] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence(WI), 2005.
[4] 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.
[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive WebSearch Based on User Profile Constructed without any Effortfrom Users," Proc. 13th Int'l Conf. World Wide Web (WWW),2004.
[6] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," Proc. 14th ACM Int'l Conf. Information andKnowledge Management (CIKM), 2005.
[7] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACMSIGIR Conf. Research and Development Information Retrieval (SIGIR),2005.