www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 10 Issue June 2021, Page No. 25336-25346
ISSN: 2319-7242 DOI: 10.18535/ijecs/v10i6.4589

Designing Improvisation In Privacy Protection For Personalized Web Search Using Deep Learning Approach

Ms. Sushama K Bhandare[1], Dr. Avinash S. Kapse[2]

[1]Student Department of Computer Science & Engineering Anuradha Engineering College, Chikhli
Email: sushamabhandare99@gmail.com
Ph.D(CSE),M.E.(CSE),B.E.(CSE), Diploma (CT) Head of Department & Assistant Professor Department of Information Technology Anuradha Engineering College, Chikhli
Email: askapse@gmail.com

Abstract: Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user- specified privacy requirements. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Index Terms—Privacy protection, personalized web search, utility, risk, profile

1. Introduction

1.1 Introduction

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query.

The solutions to PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods are straightforward— they simply impose bias to clicked pages in the user's query history. Although this strategy has been demonstrated to perform consistently and considerably well [1], it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be poten- tially effective for almost all sorts of queries, but are reported to be unstable under some circumstances [1].

Although there are pros and cons for both types of PWS techniques, the profile-based PWS has demonstrated more effectiveness in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history [2], [3], [4], browsing history [5], [6], click-through data [7], [8], [1] bookmarks [9], user documents [2], [10], and so forth. Unfortunately, such implicitly collected personal data can easily reveal a gamut of user's private life. Privacy issues rising from the lack of protection for such data, for instance the AOL query logs scandal [11], not only raise panic among individual users, but also dampen the data-publisher's enthusiasm in offering personalized service. In fact, privacy concerns have become the major barrier for wide proliferation of PWS services.

1.2 Objectives

The above problems are addressed in our UPS (literally for User customizable Privacy-preserving Search) framework. The framework assumes that the queries do not contain any sensitive information, and aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS.

As illustrated in Fig. 1, UPS consists of a nontrusty search engine server and a number of clients. Each client (user) accessing the search service trusts no one but himself/ herself. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics, and the user-specified (customized) privacy requirements represented as a set of sensitive-nodes. The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as follows:

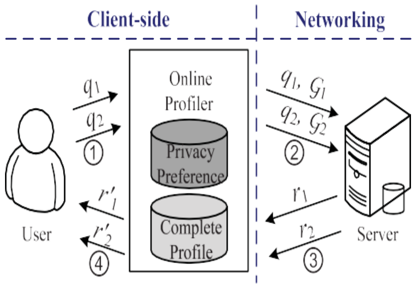


Fig. 1. System architecture of UPS.

When a user issues a query q_i on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile G_i satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics,

namely the personalization utility and the privacy risk, both defined for user profiles. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search. The search results are personalized with the profile and delivered back to the query proxy. Finally, the proxy either presents the raw results to the user, or reranks them with the complete user profile. UPS is distinguished from conventional PWS in that

- It provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements
- Allows for customization of privacy needs; an
- Does not require iterative user interaction. Our main contributions are summarized as following:

We propose a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements. Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as 6-Risk Profile Generalization, with its N P-hardness proved. We develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. While the former tries to max- imize the discriminating power (DP), the latter attempts to minimize the information loss (IL). By exploiting a number of heuristics, GreedyIL out performs GreedyDP significantly. We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile. Our extensive experiments demonstrate the efficiency and effectiveness of our UPS framework.

1.3 Privacy Protection In Personalized Search

In privacy protection, analytically observe the concern of privacy preservation in personalized search [10]. Here discriminate and describe four levels of privacy protection, and analyze numerous software architectures for personalized search. It shows that client-side personalization has advantages over the existing server-side personalized search services in preserving privacy in this situation; personalized web search cannot be done at the individual user level, but is possible at the group level. This may reduce the effectiveness of personalization because a group's information need explanation is used to model an individual user's information need. However, if the group is appropriately constructed so that people with similar interests are grouped together, it has much richer user information to offset the sparse explanation of individual user information requirements. Thus the search performance may essentially be improved because of the availability of more information from the group profile [11] and [12]. In this circumstance, personalized web search cannot be done at the distinct user level, but is possible at the group level. This may reduce the effectiveness of personalization because a group's information need description is used to model an individual user's information need. However, if the group is properly constructed so that people with comparable interests are grouped together, it may have much richer user information to offset the sparse explanation of distinct user information needs. Thus the search performance may really be better because of the accessibility of more information from the group profile

1.4 Implicit User Modeling For Personalized Search

In implicit user modeling for personalized search [2], explicated how to infer a user's interest from the user's search context and practice the conditional implied user model for personalized search. A decision speculative basis and develop methods for implicit user exhibiting in information retrieval. They developed an intelligent client-side web search agent (UCAIR) that can achieve eager implicit feedback, e.g., query development established on prior queries and instant result re-ranking established on search show that search agent can progress search accuracy over the popular Google search engine. In this paper, described how to make and update a user model based on the instant search context and implicit feedback information and use the model to improve the accuracy of ad hoc retrieval. In order to extremely benefit the user of a retrieval system through implicit user modeling, offered to perform "eager implicit feedback". Those is, as soon as experimental any new piece of evidence from the user, and update the system's certainty about the user's information need and respond with improved retrieval outcomes based on the updated user model. A decision-theoretic basis for enhancing interactive information retrieval based on eager user model updating, in which the system replies to each achievement of the user by choosing a system exploit to enhance an efficacy function. In a traditional retrieval model, the retrieval problem is often to match a query with documents and rank documents giving to their relevance values. As a result, the whole retrieval progression is a simple independent cycle of "query" and "result display". In the planned new recovery model, the user's search circumstance shows a significant role and the conditional implicit user typical is exploited directly to benefit the user. The novel retrieval model is thus essentially diverse from the traditional pattern, and is inherently more general.

2. Literature Survey

2.1Profile-Based Personalization

Previous works on profile-based PWS mainly focus on improving the search utility. The basic idea of these works is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. In the remainder of this section, we review the previous solutions to PWS on two aspects, namely the representation of profiles, and the measure of the effectiveness of personalization.

Many profile representations are available in the literature to facilitate different personalization strategies. Earlier techniques utilize term lists/vectors [5] or bag of words [2] to represent their profile. However, most recent works build profiles in hierarchical structures due to their stronger descriptive ability, better scalability, and higher access efficiency. The majority of the hierarchical represen- tations are constructed with existing weighted topic hierarchy/graph, such as ODP[1] [1], [14], [3], [15], Wikipe- dia[2] [16], [17], and so on. Another work in [10] builds the hierarchical profile automatically via term-frequency ana- lysis on the user data. In our proposed UPS framework, we do not focus on the implementation of the user profiles. Actually, our framework can potentially adopt any hierarchical representation based on taxonomy of knowledge. As for the performance measures of PWS in the literature, Normalized Discounted Cumulative Gain (nDCG) [18] is a common measure of the effectiveness of an information retrieval system. It is based on a human- graded relevance scale of item-positions in the result list, and is, therefore, known for its high cost in explicit feedback collection. To reduce the human involvement in perfor- mance measuring, researchers also propose other metrics of personalized web search that rely on clicking decisions, including Average Precision (AP) [19], [10], Rank Scoring [13], and Average Rank [3], [8]. We use the Average Precision metric, proposed by Dou et al. [1], to measure the effectiveness of the personalization in UPS. Meanwhile, our work is distinguished from previous studies as it also proposes two predictive metrics, namely personalization utility and privacy risk, on a profile instance without requesting for user feedback.

Generally there are two classes of privacy protection problems for PWS. One class includes those treat privacy as the identification of an individual, as described in [20]. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server. Typical works in the literature of protecting user identifications (class one) try to solve the privacy problem on different levels, including the pseudoidentity, the group identity, no identity, and no personal information. Solution to the first level is proved to fragile [11]. The third and fourth levels are impractical due to high cost in communication and cryptography. Therefore, the existing efforts focus on the second level. Both [21] and [22] provide online anonymity on user profiles by generating a group profile of k users. Using this approach, the linkage between the query and a single user is broken. In [23], the useless user profile (UUP) protocol is proposed to shuffle queries among a group of users who issue them. As a result any entity cannot profile a certain individual.

In privacy protection, analytically observe the concern of privacy preservation in personalized search [10]. Here discriminate and describe four levels of privacy protection, and analyze numerous software architectures for personalized search. It shows that client-side personalization has advantages over the existing server-side personalized search services in preserving privacy in this situation; personalized web search cannot be done at the individual user level, but is possible at the group level. This may reduce the effectiveness of personalization because a group's information need explanation is used to model an individual user's information need. However, if the group is appropriately constructed so that people with similar interests are grouped together, it has much richer user information to offset the sparse explanation of individual user information requirements. Thus the search performance may essentially be improved because of the availability of more information from the group profile [11] and [12]. In this circumstance, personalized web search cannot be done at the distinct user level, but is possible at the group level. This may reduce the effectiveness of personalization because a group's information need description is used to model an individual user's information need. However, if the group is properly constructed so that people with comparable interests are grouped together, it may have much richer user information to offset the sparse explanation of distinct user information needs.

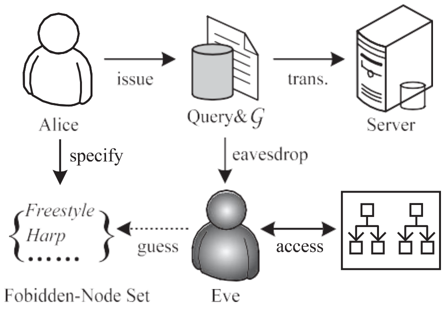


Fig. 2. Attack model of personalized web search.

Alice by recovering the segments hidden from the original H and computing a confidence for each recovered topic, relying on the background knowledge in the publicly available taxonomy repository Note that in our attack model, Eve is regarded as an adversary satisfying the following assumptions:

Knowledge bounded The background knowledge of the adversary is limited to the taxonomy repository R. Both the profile H and privacy are defined based on R.

Session bounded None of previously captured information is available for tracing the same victim in a long duration. In other words, the eavesdropping will be started and ended within a single query session. The above assumptions seem strong, but are reasonable in practice. This is due to the fact that the majority of privacy attacks on the web are undertaken by some automatic programs for sending targeted (spam) advertisements to a large amount of PWS-users. These programs rarely act as a real person that collects prolific information of a specific victim for a long time as the latter is much more costly. If we consider the sensitivity of each sensitive topic as the cost of recovering it, the privacy risk can be defined as the total (probabilistic) sensitivity of the sensitive nodes, which the adversary can probably recover from G.

3. Methodology

3.1 UPS procedure

Specifically, each user has to undertake the following procedures in our solution:

1. Offline profile construction,
2. Offline privacy requirement customization

3.1.1 Offline-1. Profile Construction

The first step of the offline processing is to build the original user profile in a topic hierarchy H that reveals user interests. Let's assume that the user's preferences are represented in a set of plain text documents, denoted by D. To construct the profile, take the following steps: 1. Detect the respective topic in R for every document d ϵ D. Thus, the preference document set D is transformed into a topic set T. 2. Construct the profile H as a topic-path trie with T, i.e., H = trie(T). 3. Initialize the user support supH(t) for each topic t ϵ T with its document support from D, and then compute supH(t) of other nodes of H with

$$sup_{\mathcal{H}}(t) = \sum_{t' \in C(t,\mathcal{H})} sup_{\mathcal{H}}(t').$$

…Equation (1)

There is one open question in the above process how to detect the respective topic for each document d ϵ D. C.Offline-2. Privacy Requirement Customization This procedure first requests the user to specify a sensitive-node set s, and the respective sensitivity value sen(s) > 0 for each topic s ϵ S. Next, the cost layer of the profile is generated by computing the cost value of each node t ϵ H as follows: 1. For each sensitive-node, cost(t) = sen(t); 2. For each nonsensitive leaf node, cost(t) = 0; 3. For each nonsensitive internal node, cost(t) is recursively given by in a bottom-up manner:

$$cost(t) = \sum_{t' \in \mathcal{C}(t,\mathcal{H})} cost(t') \times Pr(t' \mid t).$$

……Equation (2)

Till now, we have obtained the customized profile with its cost layer available.

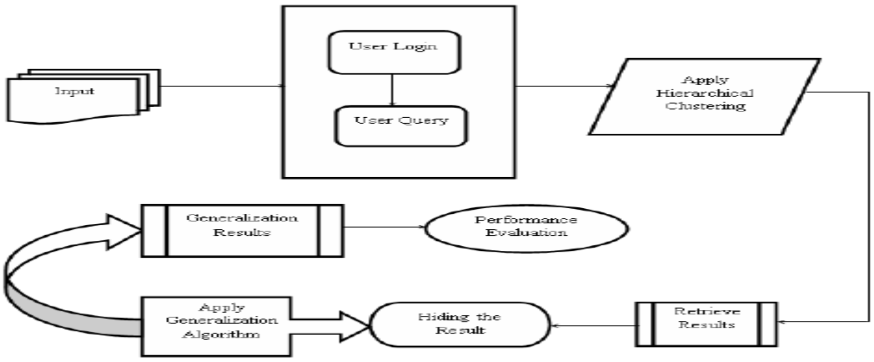


Fig. 3 System Architecture

3.2 Modules

- **Dataset preprocessing**

Most commonly a data set corresponds to the contents of a single statistical data matrix, or a single database table, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows. This module, choose input dataset. Chosen dataset has been loaded into the database. After loading the dataset into the database, we can view the dataset. By using the string matching algorithm we filter out unwanted values in the dataset and it has been preprocessed and store into the database.

- **User Login**

This is for user login page. In this module, users are entered by using the unique id and password. In this module, users are entered after registering. After registering each user has unique id. After login, user posts some queries which are based on our dataset which is loaded into the database.

- **Query Searching and Search Results Retrieval**

In this module, user submits query. Based on the query, relevant results has been displayed and also based on the submitted query some history results are displayed. Based on the query and already posted queries, we can calculate the similarity values between them. In that three types of similarity values has been estimated. From that, the result is retrieved which is based on the high relevant results by using the minimum range of similar values.

- **Estimate Relevant Results**

In this module, user posts query and sub query also. Based on the query and sub query, estimate the results based n string matching. Based on the relevant results and total number of data's in the dataset, we can estimate the support values.

- **Retrieve user profile in privacy manner**

In this module, adversaries to mine the history results means, only query time has been displayed. In this, other information such as query, query results, username are not displayed by using the background knowledge. First we generalize the table, and then suppress the values based on the generalized table. Generalized values are stored in the history results. When the adversaries' views the history result means, they can only view the generalized results. Finally, the performance can be evaluated by using the parameter such as time, cost and communicational and computational cost.

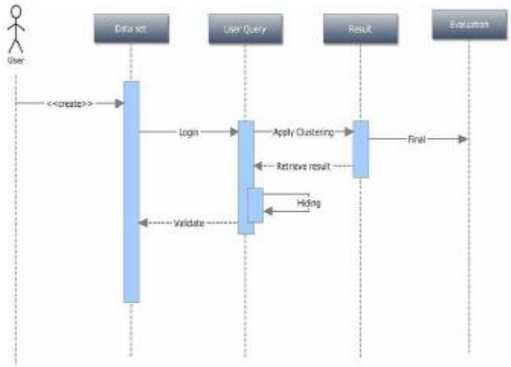


Fig. 4 Sequence Diagram

3.3 Existing System

- **Methodology of Existing System**

In the Existing Work, a client-side privacy protection framework called UPS for personalized web search was proposed. UPS could theoretically be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The context allowed users to stipulate customized privacy requirements via the hierarchical profiles. In addition, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality. In this they proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. In this for query mapping process it has various steps to compute the relevant items.

Most works on anonymization focus on relational data where every record has the same number of sensitive attributes. There are a few works taking the first step towards anonymizing set-valued or transactional data where sensitive items or values are not clearly defined. While they could be potentially applied to user profiles, one main limitation is that they either assume a predefined set of sensitive items that need to be protected, which are hard to done in the web context in practice, or only guarantee the anonymity of a user but do not prevent the linking attack between a user and a potentially sensitive item. Another approach to provide privacy in web searches is the use of a general purpose anonymous web browsing mechanism. Simple mechanisms to achieve a certain level of anonymity in web browsing include: (i) the use of proxies; or (ii) the use of dynamic IP addresses.

3.5 Explanation

Step1: Detecting & removal of unwanted symbols

Step2 Compute similarity calculation for user given word and word in database

Step3: In that similarity calculation, extract the features in the dataset.

Step4: Then estimate the ASCII difference for user given word and words in database

Step5: The estimate the similarity values.

Step6: Then retrieve the most relevant documents based on the similar values

Table no. 01 Comparison between previous and proposed system

Sr.No.	Parameters	Existing system	Proposed System
	Structure	Build in ODP not in hierarchical structures	Build in Hierarchical structure for stronger descriptive ability, better scalability, and higher access efficiency
	Graded	Based on a human graded relevance scale of item-positions which increases cost	Classification done in privacy protection problems so reduces cost
	Identity	Identity based designing not followed	Including the pseudo-identity, the group identity, no identity, and no personal information

4. IMPLEMENTATION

4.1 IMPLEMENTATION AND RESULT

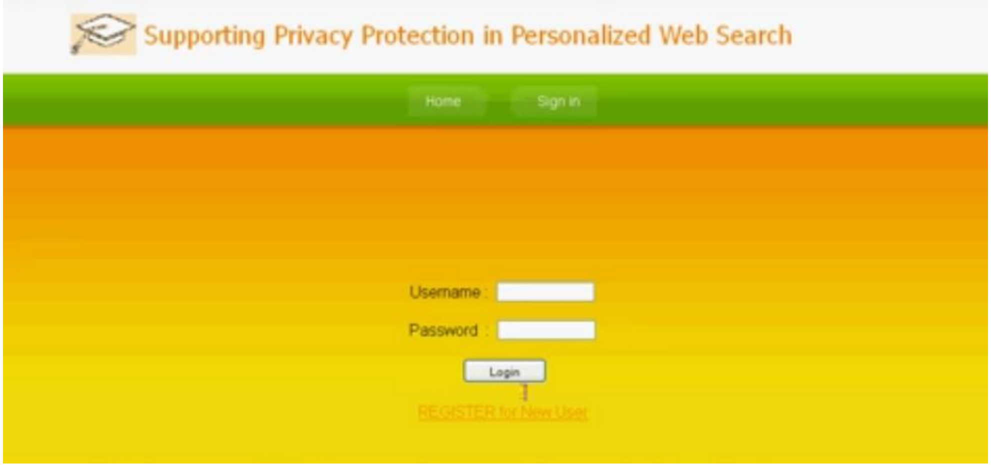


Fig. 5 Login Panel

Figure 5 depicts the user login page where individual login created to the registered user

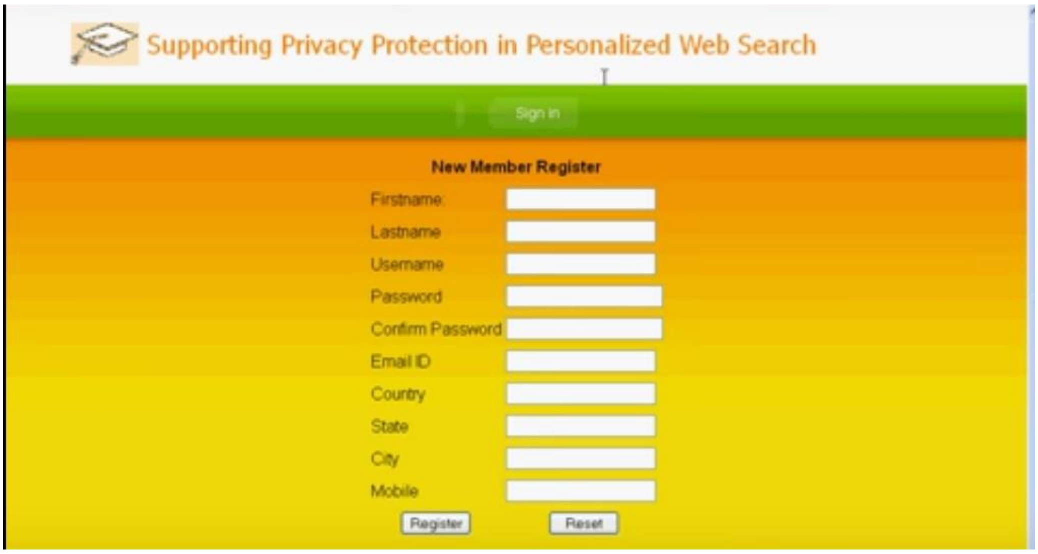
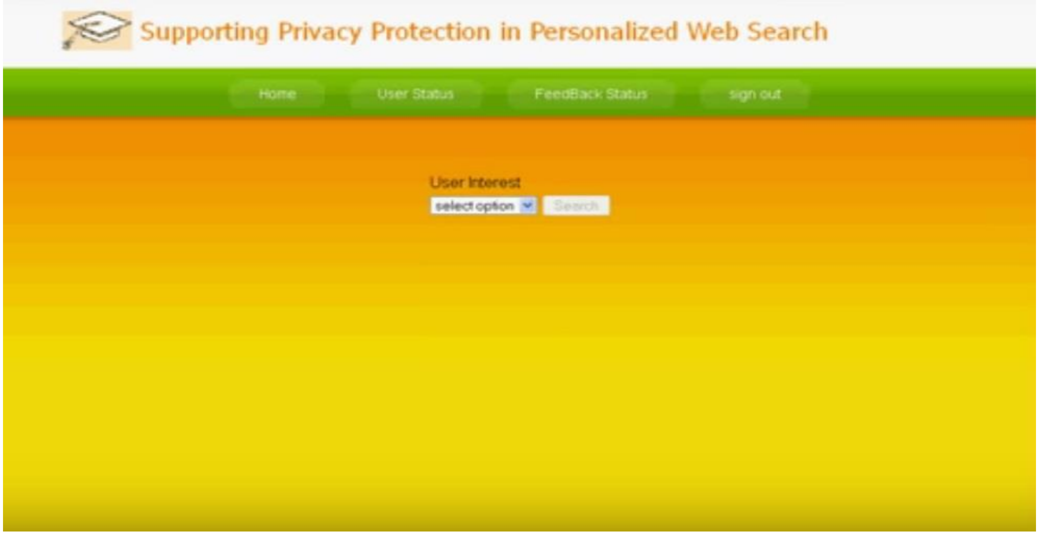


Fig. 6 New User Registration Page

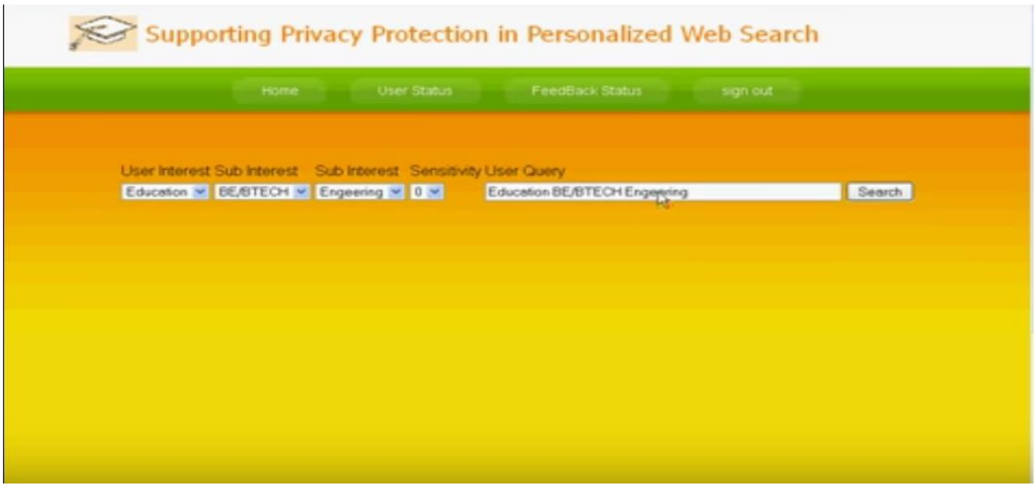
Figure 6 depicts the user registration page for authenticate login which provides security to some extent



User Interest: select option

Fig. 7 Interest wise search

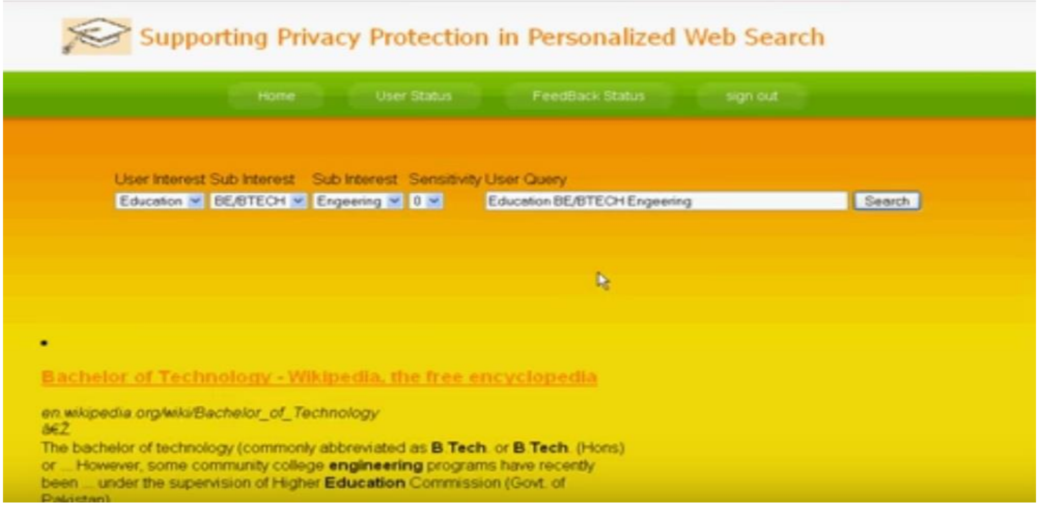
In this page the user search the details using interest of the user and then further filter is applicable according to its criteria



User Interest: Education
Sub Interest: BE/BTECH
Sub Interest: Engeering
Sensitivity: 0
User Query: Education BE/BTECH Engeering

Fig. 8 Filtering for search

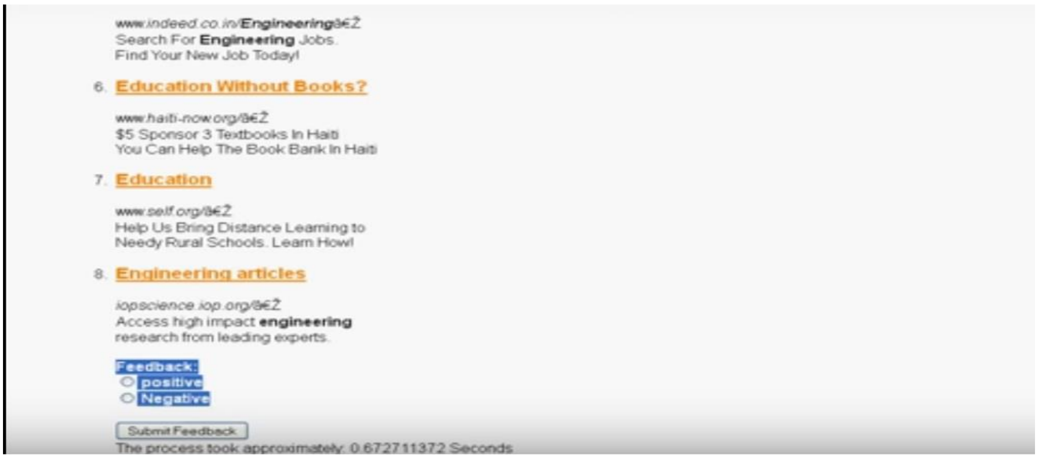
After the filter set for the interest value its then allow user to categorized the search to some extent and the result save in accordance with it



User Interest: Education
Sub Interest: BE/BTECH
Sub Interest: Engeering
Sensitivity: 0
User Query: Education BE/BTECH Engeering

- Bachelor of Technology - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Bachelor_of_Technology
â€Ž
The bachelor of technology (commonly abbreviated as **B.Tech.** or **B.Tech.** (Hons) or ... However, some community college **engineering** programs have recently been ... under the supervision of Higher **Education** Commission (Govt. of Pakistan)



www.indeed.co.in/Engineeringâ€Ž
Search For **Engineering** Jobs.
Find Your New Job Today!

6. Education Without Books?

www.haiti-now.org/â€Ž
$5 Sponsor 3 Textbooks In Haiti
You Can Help The Book Bank In Haiti

7. Education

www.self.org/â€Ž
Help Us Bring Distance Learning to
Needy Rural Schools. Learn How!

8. Engineering articles

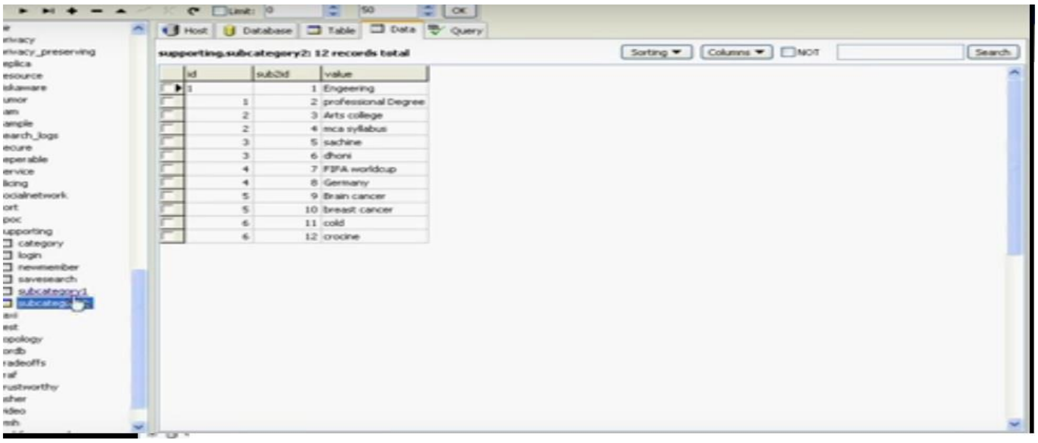
iopscience.iop.org/â€Ž
Access high impact **engineering**
research from leading experts.

Feedback:
○ positive
○ Negative

Submit Feedback
The process took approximately 0.672711372 Seconds

Fig. 9 Search Result with positive and negative feedback

When the complete categorization is done the search continues to provide the result and that result can be judged further as per negative or positive feedback so as to increases the search pattern depending on database

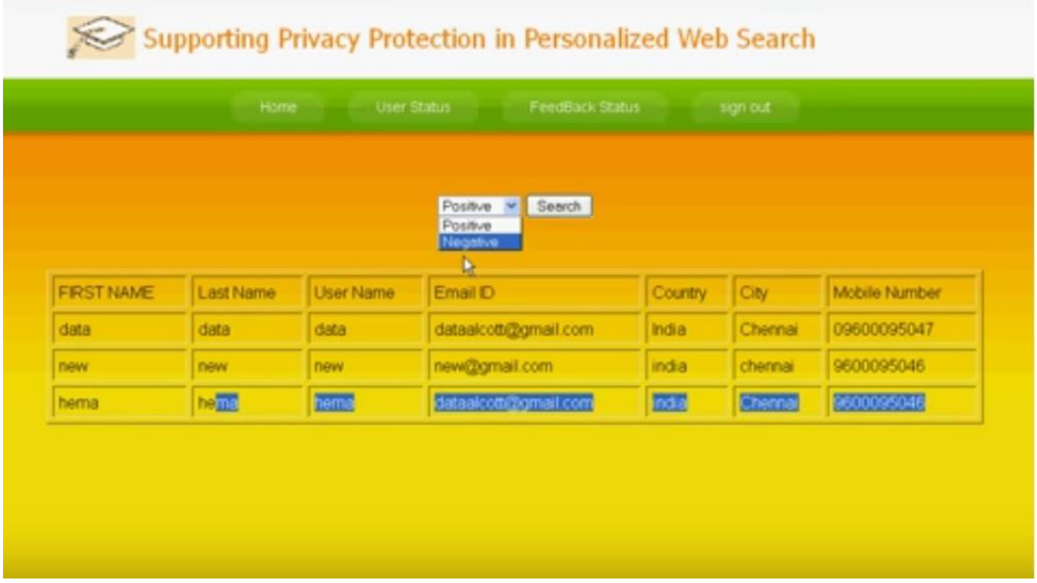


supporting.subcategory2: 12 records total

id	sub2id	value
1	1	Engeering
1	2	professional Degree
2	3	Arts college
2	4	mca syllabus
3	5	sachine
3	6	dhoni
4	7	FIFA worldcup
4	8	Germany
5	9	Brain cancer
5	10	breast cancer
6	11	cold
6	12	crocine

Fig. 10 Backend Database

The database pattern which helps for searching was stored in database for easy retrieval purpose



Supporting Privacy Protection in Personalized Web Search

FIRST NAME	Last Name	User Name	Email ID	Country	City	Mobile Number
data	data	data	dataalcott@gmail.com	India	Chennai	09600095047
new	new	new	new@gmail.com	india	chennai	9600095046
hema	hema	hema	dataalcott@gmail.com	india	Chennai	9600095046

Fig. 11 Feedback with Positive and negative impression

During search result the feedback helps to provide the key for further search in virtue of its sequence

5. Conclusion

This paper presented a client-side privacy protection framework called UPS for personalized web search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. In addition, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality. We proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution. For future work, we will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary in Section 3.3) from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

6. 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 via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.

[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 Web Search Based on User Profile Constructed without any Effort from 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 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] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.

[9] J. Pitkow, H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.

[10] 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.

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

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

[13] J.S. Breese, D. Heckerman, and C.M. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), pp. 43-52, 1998.

[14] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

[15] A. Pretschner and S. Gauch, “Ontology-Based Personalized Search and Browsing,” Proc. IEEE 11th Int’l Conf. Tools with Artificial Intelligence (ICTAI ’99), 1999.

[16] E. Gabrilovich and S. Markovich, “Overcoming the Brittleness Bottleneck Using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge,” Proc. 21st Nat’l Conf. Artificial Intelli- gence (AAAI), 2006.

[17] K. Ramanathan, J. Giraudi, and A. Gupta, “Creating Hierarchical User Profiles Using Wikipedia,” HP Labs, 2008.

[18] K. Jä rvelin and J. Kekä lä inen, “IR Evaluation Methods for Retrieving Highly Relevant Documents,” Proc. 23rd Ann. Int’l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), pp. 41-48, 2000.

[19] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley Longman, 1999.

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

[21] 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.

[22] 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.

[23] J. Castell´ı-Roca, A. Viejo, and J. Herrera-Joancomart´ı, “Preserving User’s Privacy in Web Search Engines,” Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.

[24] A. Viejo and J. Castellā-Roca, “Using Social Networks to Distort Users’ Profiles Generated by Web Search Engines,” Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[25] X. Xiao and Y. Tao, “Personalized Privacy Preservation,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2006.

[26] J. Teevan, S.T. Dumais, and D.J. Liebling, “To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent,” Proc. 31st Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 163-170, 2008.

[27] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, “Ups: Efficient Privacy Protection in Personalized Web Search,” Proc. 34th Int’l ACM SIGIR Conf. Research and Development in Information, pp. 615- 624, 2011.

[28] J. Conrath, “Semantic Similarity based on Corpus Statistics and Lexical Taxonomy,” Proc. Int’l Conf. Research Computational Linguistics (ROCLING X), 1997.

[29] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, “Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies,” Proc. Int’l Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.