A Re-ranking Approach Personalized Web Search Results by Using Privacy Protection

Vijayalakshmi Kakulapati and Sunitha Devi Bigul

Abstract Various search services quality on the Internet can be improved by personalized web search. Users face sort of dissatisfaction when the results fetched by search engines are not related to the query they have asked for. This irrelevance result is retrieved huge based on the enormous variety of consumers' perspective and backgrounds, as well as the ambiguity of the contents. However, evidences show that the user's private information which they search has become public due to the proliferation of Personalized Web Search. The proposed framework RPS implement re-ranking technique, which adaptively make simpler user profiles by queries while respecting the consumer particular constraints of privacy. The great challenge in personalized web search is Privacy protection. To increase the efficiency and accuracy of web search privacy we use Greedy IL algorithm, i.e. GreedyDP and GreedyIL, for runtime generalization. Experiment assessment results show that the privacy-preserving personalized framework and re-ranking approach is highly effective and accurate enough for user profiling privacy personalization on the web search.

Keywords Web search · Privacy · User profile · Ranking · Query

1 Introduction

Now days, to acquire any useful data about anything on the internet, the very important gateway which help in achieving this is a web search engine. Sometimes these search engines retrieve results for users with moot info that don't fulfill user

V. Kakulapati (✉)
Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Hyderabad, India
e-mail: vldms@yahoo.com

S.D. Bigul
Department of Computer Science and Engineering, CMRIT, Hyderabad, India

S.C. Satapathy et al. (eds.), *Information Systems Design and Intelligent Applications*, Advances in Intelligent Systems and Computing 434,
DOI 10.1007/978-81-322-2752-6_7

desires. This connectedness is predicated on the sizable amount of user contexts further maximum amount open retrieval info. The customized net search could be a common variety of look for techniques which give higher retrieved results and meet user desires. Gathering and analyzing user info offers user intention behind the sitting question. The customized net search is of two varieties, one is the click through knowledge and another is user identification technique. Within the click through knowledge users retrieve the tendency to clicked web content within the question record. Though this strategy has been with the efficiency question result [1], a Click through knowledge will work solely on continual queries. This is often the most disadvantage of this strategy.

Profile primarily based techniques are successful for a large vary of inquiries, yet are seen to be unsteady beneath some conditions. Profile primarily based strategies improve search results with difficult user models created from user identification methodology. There are a unit favorable circumstances and inconveniences for all styles of PWS techniques, the profile primarily based PWS is additional successful in enhancing the character of the internet hunt. By utilizing personal and behavior data of user profile, that area unit usually gathered from the internet logs [2, 3], consumer query history [4–6], bookmarks [7], click through knowledge [1, 8, 9], user documents [2, 10]. Unfortunately, individual personal data will simply relate a user's non-public data. Security problems build uneasiness in clients yet as reduce excitement in giving custom-built pursuit. For look after the user has to privacy in personalized internet search folks to believe two things in the search method. One is, they need to enhance the search quality victimization personal question of the user. Another issue is, they need to cover non-public data accessible within the user's profile so as to stay privacy risk in check. However, a few of previous history [10, 11] demonstrates that people are ready for agreement security on the off likelihood that they improve question things by provision user profile. In a perfect case, important measures of knowledge are often obtained at the value of solely tiny a part of user profile known as generalized profile. During this means privacy of the user are often protected with none negotiation. There's a balance earned between search quality and generalization gives privacy protection.

Privacy preserving PWS existing works don't seem to be satisfactory. The problems with the existing system are as follows:

1. The previous profile based mostly PWS do no generalize profile at runtime. A user profile is generalized one time which is in offline mode and queries accustomed modify square measure from a similar user. One issue reportable is [1] that typically profile based mostly personalization technique doesn't support unexpected queries. On-line identification is that the higher method, however, no earlier works have supported this.
2. The previous system doesn't contemplate modification in privacy needs. During this user privacy aren't protected properly.
3. While making customized search results several existing personalization techniques need continual user interactions.

All the above drawback of the system is resolved in our RPS (Re-ranking Privacy preserving search) structure. Construction works with the belief that the queries are with none sensitive data, progressing to shield not solely the privacy of individual users, however additionally retentive their quality for PWS. Framework usually uses two phases. One is the offline phase and other one is the online phase.

1. A consumer poses a keyword Q on the search engine, a generalized consumer profile G is creating by proxy gratifying confidentiality.
2. The keyword Q and the consumer profile G are then sent together to Personalized Web Server log file.
3. The Search results tailored according to profile is sent back P to proxy server log file.
4. At last, the web logs either present unrefined results P to the consumer or ordered them P' for the entire consumer profile.

Our main contributions are as listed below:

1. A Re-ranking privacy-preserving personalized architecture RPS, generalizes outlines the query based on the consumer privacy requirements.
2. The two metrics, web search personalization utility and privacy possibility are taken into consideration and originate the difficulty of the privacy-preserving personalized search.
3. Two simple and efficient algorithms GreedyDP and GreedyIL are developed to facilitate dynamic profiling.
4. Re-ranking algorithm applied on the generalized personalized web search.
5. Client can decide to personalize a query in RPS before each runtime profiling.

2 Related Work

2.1 Profile-Based Personalization

For the improved search results we tend to use profile primarily based personalization. To facilitate totally different personalization methods several profile representations are out there. Most of the class-conscious representations are made with weighted topic hierarchy. Our framework doesn't concentrate on the implementation of user profiles; it will with efficiency implement any class-conscious illustration supported information taxonomy.

In order to scale back human participation in performance mensuration, researchers have projected alternative metrics of customized internet search like an average preciseness [10, 12], level rating [13], and normal Rank [5, 9]. In this paper, we tend to the use typical preciseness measure projected by Dou et al. [1] that measures usefulness of personalization in cps. We tend to propose two

prognostic metrics, specifically measure of service and measure of confidentiality on a profile while not demanding consumer response.

2.2 Privacy Protection System in Personalized Web Search

Privacy protection issues are classified into two categories [14] for PWS. The first category contains of those treat privacy because the detection of a private. The second category contains of those contemplate the kindliness of the info, notably the consumer profiles, representation to the Personalized Web Search server. Distinctive work within the study of protective consumer identifications attempt to resolve the confidentiality downside completely dissimilar levels, together with the simulated uniqueness, the cluster distinctiveness, no uniqueness, and no individual information. Resolution of the major stage is confirmed breakable. The next two levels are unreasonable as a result of the high price in message and cryptography. The prevailing attempts specialize in the subsequent level. Each [2, 8] give on-line namelessness on consumer profiles by make a bunch of profile for k consumers.

Exploitation this advance, the association between the question and the consumer is broken. Mix up queries among a bunch of consumers of United Nations agency concern them [9] to plan as a worthless consumer profile protocol [3]. As a consequence, any individual cannot profile an explicit entity. All of these efforts assume the continuation of an expectation of third-party anonymizer, that isn't promptly out there over the web at giant. Viejo and Castell_a-Roca [4] use inheritance social networks rather than the moderator to supply an imprecise consumer profile to the online computer program. Within the theme, each consumer acts as a groundwork activity of his or her neighbors. Consumers will attempt to propose the question on behalf of the United Nations agency issued it, or promote it to different neighbors.

2.3 User Profile Generalization

Removing topics with low sensitivity is reserve. Hence, merely forbidding sensitive topics don't defend the consumer's confidentiality wants. To resolve this drawback with forbidding, we have a tendency to propose a brand new technique. This method identifies and removes set of topics from user profile specified the privacy risk is in check. This method is named generalization, and also the output of this method could be a generalized profile. Generalization is assessed into offline generalization and on-line generalization. Offline generalization is performed while not involving consumer queries. But it's unreasonable to perform offline simplification as a result of the output during this method might contain topic branches tangential to a question. Online generalization [15] avoids reserve privacy revealing

and additionally removes topics tangential to the present question. Over generalization causes ambiguity in personalization, resulting in poor search results.

The dilemma of confidentiality maintaining generalization in the cycle is outlined supported utility and risk. Utility calculates the personalization service of the comprehensive profile, whereas risk measures the privacy possibility of exposing the profile.

2.4 The Re-ranking Approach

In this architecture, we present an absolutely unique method for fabricating ontological consumer profiles by allocating significant scores to existing suggestions in domain ontology. All of these profiles are continued and revised as explained interests of predecessor reference realm ontology. In this regard, we propose an extending commencement algorithmic program for maintaining interest scores within the consumer profile holding the user's current performance. The RPS experimental results show that supported the significant scores and the semantic proof for associate degree, ontological consumer profile with success provides the consumer with a customized read of the search results by the delivery results nearer to the uppermost after they are appropriate to the consumer.

2.5 Generalization Metrics

(1) **Utility Metric** This metric predicts the search quality of the query on a generalized profile. We have a tendency to remodel the utility prediction downside to the analysis of characteristic power of a given question on a generalized profile. Similar suggestion has been created in [11] to form of utility; however, this measure cannot be utilized for downside settings, as we've got a profile along with hierarchical data structure rather than flat one.
(2) **Privacy Metric** When a generalized profile is exposed the entire kindliness contained in normalized kind is outlined as privacy possibility. If the unique profile is uncovered the chance of exposing all insightful topics is the peaks.

2.6 Profile Generalization Algorithms

(1) **Brute Force Algorithm** *The m*ost favorable generalization is created by generating all rooted sub trees of our seed profile by using Brute Force algorithm and the associate tree and the best service is taken as the consequence.

(2) **GreedyDP Algorithm** We apply this algorithm on a generalized profile. We remove the leaf topic of this profile to generate optimal profile. Algorithm works [16] in a bottom up the manner. With the repeated iterations we generate profile with maximum distinguishing power and satisfying δ risk constraint. And this is the final output of GreedyDP algorithm.
(3) **GreedyIL Algorithm** GreedyIL algorithm [16] reduces the information loss. When δ risk is satisfied stop the iterative process and this reduces the computational cost. Then it simplifies the computation of information loss. It reduces the need of information loss recomputation.

3 Framework for Privacy Preserving and Personalization

The Framework describes our proposed key components of the framework and re-ranking approach for personalization.

3.1 The Proposed Framework

Our framework (Fig. 1) implements a re-ranking process and enables an effective personalization using the user query log and click through data. The framework consists of five components: Request Handler, Query Processor, Result Handler, Event Handler and Response.

3.2 Greedy DP Algorithm

In this planned the model of RPS, hand and glove with a greedy algorithmic rule Greedy DP named as Greedy Utility to sustain online recognition supported on prognosticative measures of personalization effectiveness and confidentiality problem. Greedy algorithmic rule Greedy DP works during a bottom up the manner. The most downside of Greedy DP is that it needs computation of all candidate profiles generated from tries of prune-leaf manner. Formally, we denote by $gi - t\ gi + 1$ the procedure of trimming leaf t from G_i to obtain Gi + 1. Visibly, the most favorable profile G * can be created with a finite-length transitive closure of trim-leaf. Greedy DP algorithm employed in a bottom up manner. This algorithm starts from G0, in every ith iteration, Greedy DP choose a leaf topic t€TGi(q) for trimming, striving to maximize the effectiveness of the output of the recent iteration, namely Gi + 1. In these iterations, maintain a preeminent profile-so-far, which gives the Gi + 1 having the highest perceptive power while satisfying δ-problem restriction. The iterative procedure concludes when the profile is indiscriminate to a

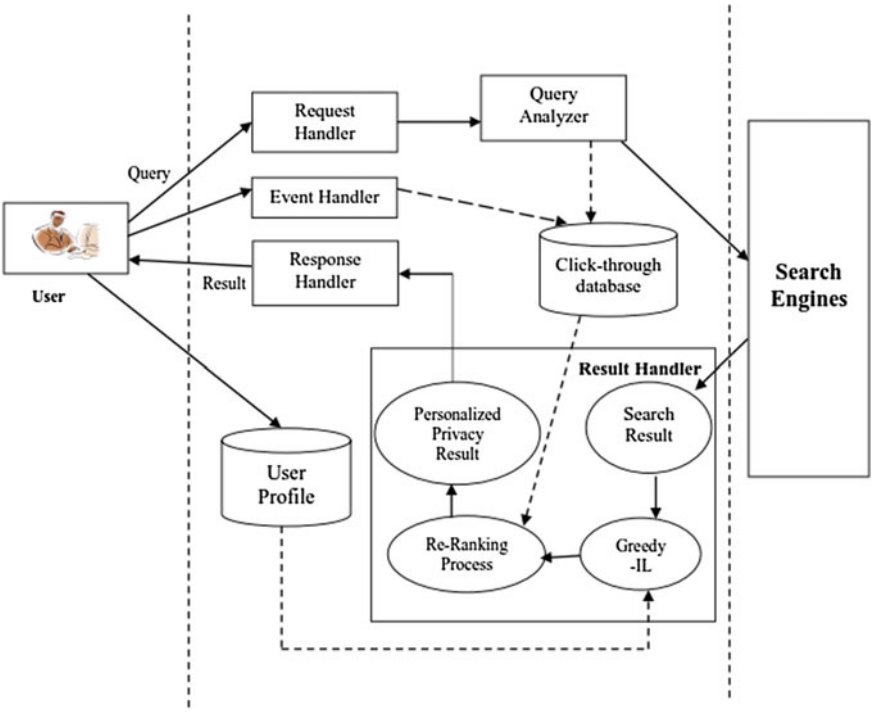

Fig. 1 Framework for re-ranking approach for privacy and personalized web search

root-topic. The final result (G*) of the algorithm will be the best-profile. The major obstacle of Greedy DP is that it requires recomputation of all candidate profiles (together with their perceptive power and privacy issue) produced from tries of trim-leaf on all t€TGi (q). This will lead considerable memory requirements and computational cost.

3.3 Greedy IL Algorithm

In this planned a brand new profile generalization formula referred to as Greedy IL. The Greedy IL formula improves the potency of the generalization using heuristics supported varied conclusions. One of the results is that any trim-leaf operation shrinks the perceptive authority of the profile. In different statements, the refugee demonstrates monotonicity by trim-leaf. Greedy IL any diminishes this live with Heuristic. The less iterations the algorithm desires the bigger the isolation threshold.

3.4 The Re-ranking Approach

In this method, the ranking rule is about in stepping with the category attain of the item capable the quantity of selections of the constant user profile in past. From the primary ranking tend to plan replacement ranking to redefine the consumer preference record. We tend to utilize Probalistic Similarity measure and cosine function Similarity measure for Item cosine and Ranking for base Search.

1. ***Algorithm: Ranking (*****Privacy Personalized Result** Set ***PPRS)***

```
Input: Privacy Personalized Result Set PPRS.
Output: Arranged Result List with Ranking r.
        do
                if (PPRS i >PPRS j) then
                        Swap (Ii,Ij)
                else
                        Return PPRS I with ranking Order
        Until (no more Items in PPRS)
```

2. Algorithm: Re-ranking (Ranked **Privacy Personalized Result** Set RPPR ***S)***

```
Input: Ranked Privacy Personalized Result Set RPPRS.
Output: Ordered Result List with Re-Ranking r.
        CTD<--GetClick_ThroughData (q, r, s);
        do
                if  (CTD=True && RPPRS i > RPPRS j) then
                        Swap (Ii, Ij)
                else
                        Return RPPRS I with Re-ranking Order
        Until (no more Items in RPPRS)
```

4 Experimental Evaluation

Required datasets for experiment evaluation is collected through java based application. We created 50 user profiles (UP) with different interests and then perform query search for each user using different queries and create more than 1000 click-through database records for evaluation. We use the Yahoo search engine to retrieve the search results. To measure the effectiveness of the proposed framework approach we measure the personalized precision rate and recall of obtain results for existing (i.e., based on click-through data only) and proposed using profile-based with Greedy-IL. The measure of results is considered based on the number of similar and relevant results are re-ranked based on both existing and proposed approach.

Table 1 Precision and recall without user profile

Without user profile (WO-UP)					
Query	Number of total search result	Relevant results to the query	Number of similar and relevant results	Precision	Recall
Movies	50	6	4	0.08	0.667
Sports	50	9	5	0.1	0.556
Music	50	11	5	0.1	0.455
Electronic	50	10	5	0.1	0.5
Travels	50	11	6	0.12	0.545

Table 2 Precision and recall without user profile

With user profile (W-UP)					
Query	Number of total search result	Relevant results to the query	Number of similar and relevant results	Precision	Recall
Movies	50	31	8	0.16	0.258
Sports	50	18	7	0.14	0.389
Music	50	38	8	0.16	0.211
Electronic	50	38	7	0.14	0.184
Travels	50	35	8	0.16	0.229

In the first run we evaluate the framework without user-profile and Greedy-IL. We run the query in five different domains as *Movies, Sports, Music, Electronic and Travels* with click-through data and user-profiles. We observe that most of the results which are similar and relevance to the query keywords, but not in relevance to the user profile interest in case of existing click-through based, whereas high relevancy is observed in case user-profile with Greedy-IL as shown below in Tables 1 and 2.

4.1 Personalized Precision Measure

It is a measure of correctly predicted results by the system among all the predicted results. It is defining as the number of relevant results retrieved by a search divided by the total number of results retrieved by that search.

$$Personalize\ Precision\ (PP) = \frac{|number\ of\ simillar\ and\ relevant\ results|}{|No.\ of\ total\ Search\ Result|} \times 100$$

4.2 Personalize Recall Measure

Recall is a measurement of correctly predicted results by the system among the positive results. Recall is defined as the number of relevant results retrieved by a search divided by the total number of existing relevant results.

$$Recall\ Ratio = \frac{|The\ number\ of\ simillar\ and\ relevant\ results|}{|Relevant\ results\ to\ query|} \times 100$$

Figures 2 and 3 illustrates personalized precision and recall performance at different query categories of search results with the help of click-through data and user profile with Greedy-IL. The result shows an improvement in the personalized precision rate with different query categories in case of with-UP and the low recall rate in compare to without-UP. It's suggested more appropriate to meet the satisfactory level of motivation of the proposal. Improvisation is due to an online

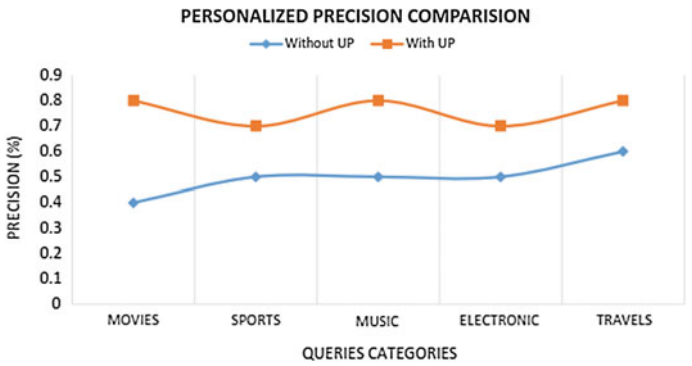

Fig. 2 Personalized precision comparison between with and without user-profile

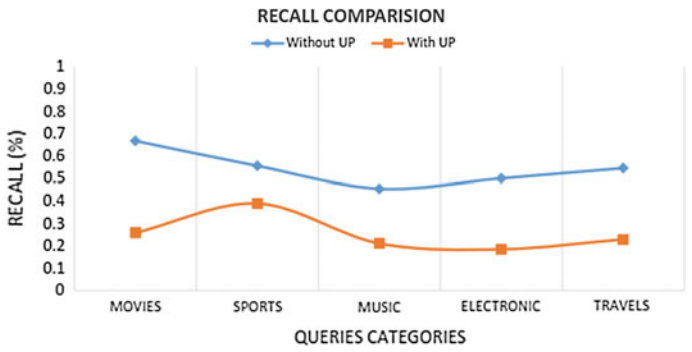

Fig. 3 Recall comparison between with and without user-profile

prediction method used to make a decision regarding relevancy using user profile for high query relevancy and make more beneficial for accurate in information retrieval. The results revealed that using user-profiles could achieve accurate and quality search results while preserving the user's tailored privacy requirements. The final results conforms the effectiveness and efficiency of privacy personalized search.

5 Conclusion

In this work, we proposed a re-ranking privacy protection framework called RPS for privatized website searching. RPS could be used by any Personalized Web Search that holds user profiles in a definite hierarchical categorization. The framework allowed consumers to specify customized privacy related requirements via the hierarchically categorized profiles. Along with this, RPS also had acted to generalize user profiles to ensure the personal privacy without undermining the search quality. User profiles are generalized using greedy IL. The result induces us to keep up a precedence queue of trim-leaf actors in a digressive arrangement of the data mislaying caused by the operators. The queue q permits quick retrieval of the most effective candidate operator. Filtering results is re ranked by using ranking rule supported RPS and results are shown to the user.

6 Future Enhancement

In our further work we will also make an attempt to resist opponents with a wider relationship among topics or capability to capture a more relevant personalized privacy preserving the result of queries posed by users based on the privacy of user profiles. We will also look into better-brushed methods to build the consumer profile, and more effective measures forecast the performance of RPS.

References

1. Silvia Quarteroni, Suresh Manandhar "User Modeling for Personalized Question Answering" AI*IA 2007: Artificial Intelligence and Human-Oriented Computing Lecture Notes in Computer Science Volume 4733, 2007, pp 386–397.
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. 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.

4. http://www.citeulike.org/user/abellogin/article/776870M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
5. 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.
6. 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.
7. A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services," J. Artificial Intelligence Research, vol. 39, pp. 633–662, 2010.
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. Archana Ukande, Nitin Shivale, "Supporting Privacy Protection in Personalized Web Search with Secured User Profile" International Journal of Science and Research (IJSR), Volume 3 Issue 12, December 2014.
10. X. Shen, S. Dumais, and E. Horvitz.: Analysis of topic dynamics in Web search". In *Proceedings of the International Conference on World Wide Web*, pages 1102–1103, 2005.
11. P. Sudhaselvanaya et al "Confidential User Query Profile Construction for Personalized Web Search", International Journal On Engineering Technology and Sciences–ISSN (P): 2349–3968, Volume 1 Issue 6, October 2014.
12. Ahu Sieg et al., "Web Search Personalization with Ontological User Profiles", CIKM'07, November 6–8, 2007, Lisboa, Portugal. Copyright 2007 ACM 978–1- 59593-803-9/07/0011.
13. V. Ramya, S. Gowthami, "Enhance privacy search in web search engine using greedy algorithm" International Journal of Scientific Research Engineering & Technology, ISSN 2278–0882 Volume 3, Issue 8, November 2014.
14. 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.
15. Yusuke Hosoi et al, "Generalization User Profiles to Context Profiles and Its Application to Context-aware Document Clustering", ISBN: 978-960-474-361-2, Recent Advances in Computer Engineering, Communications and Information Technology.
16. Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection in PersonalizedWeb Search", IEEE Transactions on Knowledge and Data Engineering Vol. 26 No. 2, 2014 doi no. 10.1109/TKDE.2012.201.