ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

ENHANCED PRIVACY SEARCH IN WEB AND IMAGE SEARCH ENGINE USING GREEDY ALGORITHM

V.Pushpalatha [1], Dr.S.Ramkumar [2]

P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Asst.Professor, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Abstract: **In web search, there are many techniques used broadly for valuable retrieval of information from the web server. A user query entered into the search engine may return large number of web page results and thus, it becomes extremely important to rank these results in such a manner that it returns accurate and more relevant web results. These tasks of prioritizing the results are performed by ranking algorithms. But the search interest of every user differs with every other user uniquely. In existing system, the search engine that return results for a given user query is not uniquely based on their earlier searching behavior. In this paper, we introduce "task trail" as a new concept to understand user search behaviors. We define task to be an atomic user information need. Web search logs have been studied mainly at session or query level where users may submit several queries within one task and handle several tasks within one session. Although previous studies have addressed the problem of task identification, little is known about the advantage of using task over session and query for search applications.**

Keywords: *Re-ranking, Personalized web Search, Search engine, Profile, User Behavior*

1. INTRODUCTION

The customized privacy requirements on a user profile. Provide framework for privacy preserving profile generalization. Provide protection against a typical model of privacy attack, namely eavesdropping. We can perform the personalized web search. In this module, user enter query to retrieve the results from server. We can analyze whether the query is personalized or general. In this module, online profiler collect information from the user Then convert the profile to generalized profile Generalized profile contains the details about search experience and search history GreedyDP algorithm This algorithm used to improve the search results so we create the Taxonomy repository for quick access. Implement the prune leaf function to remove unwanted search results. GreedyIL algorithm improves the efficiency of the generalization using heuristics based on numerous answers. One significant discovery is that any prune-leaf operation reduces the discriminating power of the profile. In existing system implement the profile-Based Personalization techniques. It leads high cost in communication and cryptography. May chances to reveals the user profiles to others. The existing profile-based PWS do not support runtime profiling. The existing methods do not take into account the customization of privacy requirements. Personalization is the process of presenting the right information to the right user at the correct instant. In order to study on a user, systems must gather personal data, investigate it, and accumulate the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

consequences of the analysis in a user profile. Data can be composed from users in two traditions: unambiguously, for instance ask for comment such as preferences or ratings; or perfectly, for instance detect user behaviors such as the time spent reading an on-line document. Many personalization techniques require iterative user interactions when creating personalized search results. Generalize profiles for each query according to user-specified privacy requirements. We can minimize the information loss in personalized web search system. Providing protection against a typical model of privacy attack, namely eavesdropping. We implement the UPS system, it consist a non trusty search engine server and a number of clients. Each client (user) accessing the search service trusts no one but himself or herself. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. We developed two simple but effective GreedyIL, GreedyDP and generalization algorithms, to support runtime profiling. We provided an inexpensive mechanism for the client to decide whether to personalize a query in UPS.

2. RELATED WORK

Z. Dou, R. Song, and J.-R. Wen et al[1]. Although personalized search has been proposed for many years and many personalization strategies have been investigated, it is still unclear whether personalization is consistently effective on different queries for different users, and under different search contexts. In this paper, we study this problem and get some preliminary conclusions. We present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies using 12-day MSN query logs. By analyzing the results, we reveal that personalized search has significant improvement over common web search on some queries but it also has little effect on other queries. It even harms search accuracy under some situations. Furthermore, we show that straightforward click-based personalization strategies perform consistently and considerably well, while profile-based ones are unstable in our experiments. We also reveal that both long-term and short-term contexts are very important in improving search performance for profile-based personalized search strategies.

J. Teevan, S.T. Dumais, and E. Horvitz, et al [2] . We formulate and study search algorithms that consider a user's prior interactions with a wide variety of content to personalize that user's current Web search. Rather than relying on the unrealistic assumption that people will precisely specify their intent when searching, we pursue techniques that leverage implicit information about the user's interests. This information is used to re-rank Web search results within a relevance feedback framework. We explore rich models of user interests, built from both search-related information, such as previously issued queries and previously visited Web pages, and other information about the user such as documents and email the user has read and created. Our research suggests that rich representations of the user and the corpus are important for personalization, but that it is possible to approximate these representations and provide efficient client-side algorithms for personalizing search. We show that

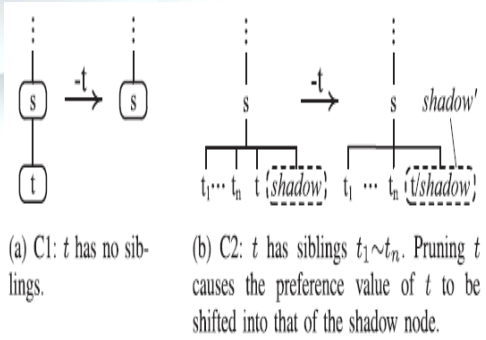
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

such personalization algorithms can significantly improve on current Web search.

M. Spertta and S. Gach, et al[3] . User profiles, descriptions of user interests, can be used by search engines to provide personalized search results. Many approaches to creating user profiles collect user information through proxy servers or desktop bots. Both these techniques require participation of the user to install the proxy server or the bot. In this study, we explore the use of a less-invasive means of gathering user information for personalized search. In particular, we build user profiles based on activity at the search site itself and study the use of these profiles to provide personalized search results. By implementing a wrapper around the Google search engine, we were able to collect information about individual user search activities. In particular, we collected the queries for which at least one search result was examined, and the snippets for each examined result. User profiles were created by classifying the collected information into concepts in a reference concept hierarchy. These profiles were then used to re-rank the search results and the rank-order of the user-examined results before and after re-ranking were compared. Our study found that user profiles based on queries were as effective as those based on snippets. We also found that our personalized re-ranking resulted in a 34% improvement in the rank-order of the user-selected results.

GreedyDP algorithm

In this proposed the prototype of UPS, in concert with a greedy algorithm GreedyDP named as Greedy Utility to support online profiling based on predictive metrics of personalization utility and privacy risk. The main problem of GreedyDP is that it requires recomputation of all candidate profiles generated from attempts of prune-leaf manner. Formally, we denote by $G_i \xrightarrow{-t} G_{i+1}$ the process of pruning leaf t from G_i to obtain G_{i+1}. Obviously, the optimal profile G * can be generated with a finite-length transitive closure of prune-leaf. The first greedy algorithm GreedyDP workings in a bottom up manner. Starting from G_0, in every ith iteration, GreedyDP chooses a leaf topic $t \in T_{G_i}(q)$ for pruning, trying to maximize the utility of the output of the current iteration, namely G_{i+1}. During the iterations, we also keep a best profile- so-far, which indicates the G_{i+1} having the highest discriminating power while satisfying the δ-risk constraint. The iterative process terminates when the profile is generalized to a root-topic. The best-profile- so-far will be the ending result (G*) of the algorithm. The major difficulty of GreedyDP is that it requires recomputation of all candidate profiles (together with their discriminating power and privacy risk) generated from attempts of prune-leaf on all $t \in T_{G_i}(q)$. This causes significant memory requirements and computational cost.



(a) C1: t has no siblings. (b) C2: t has siblings $t_1 \sim t_n$. Pruning t causes the preference value of t to be shifted into that of the shadow node.

Figure 1: Two cases of prune-leaf on a leaf *t*

GreedyIL algorithm

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

In this proposed a new profile generalization algorithm called GreedyIL. The GreedyIL algorithm improves the efficiency of the generalization using heuristics based on several conclusions. One main decision is that any prune-leaf operation reduces the discriminating power of the profile. In extra words, the DP displays monotonicity by prune-leaf. GreedyIL further reduces this measure with Heuristic. The greater the privacy threshold, the less iterations the algorithm needs.

Meta Search Engine:

Meta search engine is a search tool that uses other search engine's data to produce their own results from the Internet. Meta search engines take input from a user and simultaneously send out queries to third party search engines for results. Sufficient data is gathered, formatted by their ranks and presented to the users. Information stored on the World Wide Web is constantly expanding, making it increasingly impossible for a single search engine to index the entire web for resources. Meta search engine is a solution to overcome this limitation. By combining multiple results from different search engines, Meta search engine is able to enhance the user's experience for retrieving information, as less effort is required in order to access more materials. A Meta search engine is efficient as it is capable of generating a large volume of data; however, scores of websites stored on search engines are all different: this can draw in irrelevant documents. Other problems such as spamming also significantly reduce the accuracy of the search. This issue is tackled by the process of fusion which improves the engineering of Meta search engine. Christo Ananth et al. [6] discussed about a system, the effective incentive scheme is proposed to stimulate the forwarding cooperation of nodes in VANETs. In a coalitional game model, every relevant node cooperates in forwarding messages as required by the routing protocol. This scheme is extended with constrained storage space. A lightweight approach is also proposed to stimulate the cooperation. It also reduces the work of users from having individual type searches from different engines to look for resources. Searches can be analyzed in three levels, (a) query level, (b) quest level and (c) session level. In query level it fails to capture the interleaving relationships between different quests. If we analyze the search logs based on session (i.e. session level) the quests will be interleaved. It is difficult to identify what the user is doing because the sessions are chronologically ordered. If we analyze in quest level the topics will be more consistent and relevant to each other. This will help us to understand the intentions behind a user's search. A Query Trail can be defined as sequence of user behavior (a query followed by sequence of browsing behavior). A quest trail represents all user activities within that particular task, such as query reformulations, URL clicks. Session is defined as "a series of queries by a single user made within a small range of time" and the activities done by the user in that time period in a browser is known as session trail.

2.1 Existing System

The existing PWS do not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. The existing methods do not take

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

into account the customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. For example, all the sensitive topics are detected using an absolute metric called surprised based on the information theory, assuming that the interests with less user document support are more sensitive. Many personalization techniques require iterative user interactions when creating personalized search results. They usually refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on.Personalized web search faces several challenges that retard its real-world large-scale applications.

Disadvantages

- Privacy is an issue. Personalized web search, especially server-side implement, requires collecting and aggregating a lot of user information including query and click through history.
- It is really hard to infer user information needs accurately.
- The existing methods do not take into account the customization of privacy requirements.

2.2 Proposed System

Personalized search is a promising way to improve the accuracy of web search, and has been attracting much attention recently. However, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users. Indeed, these concerns have become one of the main barriers for deploying personalized search applications, and how to do privacy-preserving personalization is a great challenge. For a given query, a personalized Web search can provide different search results for different users or organize search results differently for each user, based upon their interests, preferences, and information needs. Personalized web search differs from generic web search, which returns identical research results to all users for identical queries, regardless of varied user interests and information needs. 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. To provide personalized search results to users, personalized web search maintains a user profile for each individual. A user profile stores approximations of user tastes, interests and preferences.

2.2.1 Advantages

- Simplify profiles for each query according to user-specified privacy requirements.
- We can minimize the information loss in personalized web search system.
- Providing protection against a typical model of privacy attack, namely eavesdropping

III.SYSTEM ARCHITECTURE

The amount of information on the World Wide Web is growing rapidly; search engines must be able to retrieve information according to the user's preference. Current web search engines are built to serve all users, independent of the special needs of any individual user. Personalization of web search is

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

to carry out retrieval for each user incorporating his/her interests. Every user has a distinct background and a specific goal when searching for information on the Web. Thus the goal of Web search personalization is to tailor search results to a particular user based on that user's interests and preferences. However, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users. Indeed, these concerns have become one of the main barriers for deploying personalized search applications, and how to do privacy-preserving personalization is a great challenge. Thus, a balance must be struck between search quality and privacy protection. Hence, privacy protection in PWS applications that model user preferences as hierarchical user profiles is proposed using a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements. Along with Personalized Search and Privacy Protection the Custom Search functionality will also be provided so that the users get relevant information.Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided and describe two general approaches to personalizing search results, one involving modifying the user's query and the other re-ranking search results. 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 Risk Profile Generalization. We develop two simple but effective generalization algorithms, Greedy DP and Greedy IL, to support runtime profiling. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). One of the most critical benefits personalized search has is to improve the quality of decisions consumers make. The internet has made the transaction cost of obtaining information significantly lower than ever. However, human's capability of processing information has not expanded much. When facing overwhelming amount of information, consumers need a sophisticated tool to help them make high quality decisions. Two studies examined the effects of personalized screening and ordering tools, and the results show positive correlation between personalized search and the quality of consumers' decisions.

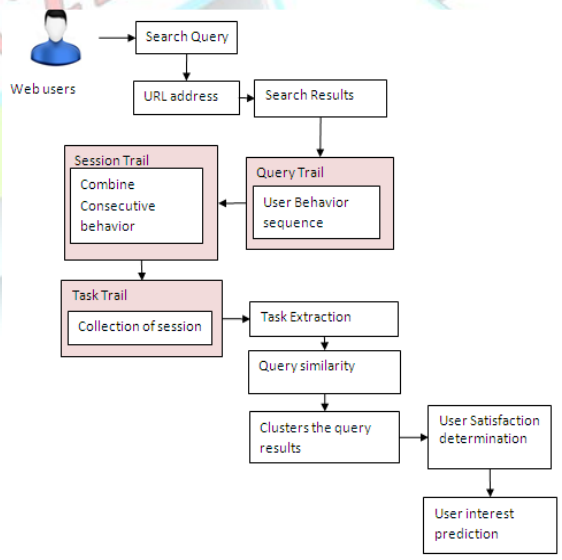


Figure 2: System Architecture

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

3.1 MODULES DESCRIPTION

Personalized Search query:

A web search query is a query that a user enters into a web search engine to satisfy his or her information needs. Web search queries are distinctive in that they are often plain text or hypertext with optional search-directives (such as "and"/"or" with "-" to exclude). They vary greatly from standard query languages, which are governed by strict syntax rules as command languages with keyword or positional parameters. A search query, the actual word or string of words that a search engine user types into the search box, is the real-world application of a keyword – it may be misspelled, out of order or have other words tacked on to it, or conversely it might be identical to the keyword. There are three broad categories that cover most web search queries: informational, navigational, and transactional.

Online profiler:

This module we implement 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. Personal data, i.e. personal documents, browsing history and emails might be helpful to identify a user's implicit intents.However, users have concerns about how their personal information is used. Privacy, as opposed to security or confidentiality, highly depends on the person involved and how that person may benefit from sharing personal information. The question here is whether a solution can be found where users themselves are able to set their own privacy levels for user profiles to improve the search quality. This module is to utilize the user context to personalize search results by re-ranking the results returned from a search engine for a given query. The unified context model for a user is represented as an instance of a reference domain in which concepts are annotated by interest scores derived and updated implicitly based on the user's information access behavior. We call this representation user profile. User profile is a specification of a conceptualization – description of the concepts and relationships that can exist for an agent/user or a community of agents/users.

Greedy DP algorithm:

In this module, we have proposed the prototype of UPS, together with a greedy algorithm Greedy DP (named as Greedy Utility to support online profiling based on predictive metrics of personalization utility and privacy risk. Greedy algorithm Greedy DP works in a bottom up manner. The main problem of Greedy DP is that it requires recomputation of all candidate profiles (together with their discriminating power and privacy risk) generated from attempts of prune-leaf manner. In this module, we can create taxonomy repository in tree structure.

Greedy IL algorithm:

In this module we proposed a new profile generalization algorithm called Greedy IL. The Greedy IL algorithm improves the efficiency of the generalization using heuristics based on several findings. One important finding is that any prune-leaf operation reduces the discriminating power of the profile. In other words, the DP displays monotonicity

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 17, March 2016

by prune-leaf. Greedy IL further reduces this measure with Heuristic. The greater the privacy threshold, the fewer iterations the algorithm requires. This work aims at providing protection against a typical model of privacy attack, namely eavesdropping. In attack model, third party considered as 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.

Search results:

It is important to monitor the personalization utility during the generalization. Using the running example, profiles Ga and Gb might be generalized to smaller rooted sub trees. However, overgeneralization may cause ambiguity in the personalization, and eventually lead to rich search results. Monitoring the utility would be possible only if we perform the generalization at runtime. In this module, we improve the results based on search experience and profession based results. Our experimental results provide the efficient privacy based search results.

IV CONCLUSION

A major problem in web search is that the interactions between the users and search engines are limited by the small form factors of the search engines and various categories. As a result, web users tend to submit shorter, hence, more ambiguous queries compared to their web search counterparts. We proposed PWS to extract and learn a user's history and content preferences based on the user's profiles. Web log segmentation can be done in query level, session level, or task level. Task trail is a sequence of user behaviors occurred within one session, where they collectively define atomic user information need. Following task trail is an effective method to segment the web log and also to determine the user search behavior. Web logs are segmented into sessions by choosing a time threshold. Queries similar to each other are clustered into same task after computing the query similarity. From the extracted tasks, user search behavior can be determined. The final result display unique web results for unique users based on their searched behavior. We also tested a term matching variant that alleviated the need for an exact term match between queries and trails, which led to coverage and diversity gains at the cost of a slight decrease in relevance

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] Christo Ananth, M.Muthamil Jothi, A.Nancy, V.Manjula, R.Muthu Veni, S.Kavya, "Efficient message forwarding in MANETs", International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1,Issue 1, August 2015,pp:6-9

[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. Schu¨ tze, 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.