Issues and Challenges in Measuring Security Threats During Personalized Web Search

Hina Ansari

Mahakal Institute of Technology, Ujjain
www.isroset.org

Abstract: The World Wide Web is an important source of data that can come either from the Web content, represented by the billions of pages publicly available, or from the Web usage, represented by the log information daily collected by all the servers around the world. The continuous growth in the size and use of the World Wide Web imposes new methods of design and development of online information services. Most Web structures are large and complicated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. On the other hand, the e-business sector is rapidly evolving and the need for Web marketplaces that anticipate the needs of the customers is more evident than ever. Therefore, the requirement for predicting user needs in order to improve the usability and user retention of a Web site can be addressed by personalizing it. This paper tries to address some issues regarding some of the major challenges faced by Search Engines.

Keywords- PWS, Security threats, ODP Metadata

1. Introduction

For hundreds of years the mankind has organized information in order to make it more accessible to the others. The last media born to globally provide information is the Internet. With the Web, in particular, the name of the Internet has spread all over the World. Due to its impressive size and its high dinamicity, a Web Search Engine is used when we need to search for information on the Web, usually we begin by querying. A Web Search Engine maintains and catalogs the content of Web pages in order to make them easier to find and browse. Even though the various Search Engines are similar, each one of them differentiates from the other by the methods for scouring, storing, and retrieving information from the Web. Usually Search Engines search through Web pages for specified keywords. In response they return a list containing those documents containing the specified keywords. This list is sorted by a relevance criteria which try to put at the very first positions the documents that best match the user's query. The usefulness of a search engine to most people, in fact, is based on the relevance of results it gives back.

Web personalization is defined as any action that adapts the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational behavior and individual interests, in combination with the content and the structure of the Web site. The objective of a Web personalization system is to "provide users with the information they want or need, without expecting from them to ask for it explicitly". With the dramatically quick and explosive growth of information available over the Internet, World Wide Web has become a powerful platform to store, disseminate and retrieve information as well as mine useful knowledge. Due to the properties of the huge, diverse, dynamic and unstructured nature of Web data, Web data search has encountered a lot of challenges, such as scalability, multimedia and temporal issues etc. As a result, Web users are always drowning in an "ocean" of information and facing the problem of information overload when interacting with the web.

2. Motivation

The first web search engines appeared on 1993. Those web search engines did not keep information about the content of the web pages; instead, they only indexed information about the title of the pages. It was in 1994, when web search engines started to index the whole web content, so that the user could search into the content of the web pages, not only in the title. Web search engines are today used by everyone with access to computers, and those people have very different interests. But search engines always return the same result, regardless of who did the search. Search results could be improved if more information about the user was considered.

The wide use of Internet and technology has never made it easier for people to collect, share and exchange data. Meanwhile, as people quickly expand their virtual presence online, the ready availability of personnel data has also put individual privacy under risk in many ways, the ever increasing privacy fear from the public is reflected by surveys and polls[1].To name a few ,from 2004 to 2006,the annual personalization survey from Choice Stream consistently shows that around 70% were not satisfied with current privacy solutions, also predicted that information privacy would be the greatest concern for consumer based e-business. Most recently, Google[2] was ordered by federal judge to hand over the popular online video service YouTube's data to Viacom for copyright issues. The potential threat from exposing the video viewing habits of tens of millions of people provokes another round of public

privacy outcry. The privacy concern has become a major obstacle to information sharing. The implications are twofold [3]: first, organizations such as govt. agencies worry about keeping highly sensitive financial and health data private, and thus are not willing to share it with other organizations, not even monitoring to publish to the public; second, individual users grow ever more wary-and-distrustful-of organizations that handle sensitive data, and thus are not willing to submit their data to organization either. To overcome the obstacle, a large number of techniques has been proposed for protecting individual privacy and sensitive data [4], and they can be generally considered as privacy preserving information sharing and also privacy preserving personalized web searching.

3.   Problem Statement

There are various problems are associated regarding privacy and security of Personalized Web Search so that there is need of better tools for managing and making privacy and security in personalized web search.

To contribute in the improvement of privacy & security in personalized search systems by giving users more control over personal data and personalization features. A privacy tool can be designed that allows users to more easily apply privacy techniques and possible personalization benefits. Since users are highly concerned about their privacy but often do not act accordingly. Thus, better support tools are necessary to inform them about the privacy impacts of their online actions like web search.

4.   Related Work

In information retrieval, much research is focused on personalized search. Relevance feedback and query refinement [5] harnesses a short-term model of a user's interests, and information about a user's intent is collected at query time. Personal information has also been used in the context of Web search to create a personalized version of PageRank [6]. There are still approaches, including many commercially available information filtering systems [7] [8], which require users explicitly specify their interests. However, as [9] pointed out, users are typically unwilling to spend the extra effort on specifying their intentions.

A majority of work focuses on implicitly building user profiles to infer a user's intention [10]. A wide range of implicit user activities have been proposed as sources of enhanced search information. This includes a user's search history, browsing history, click-through data, web community, and rich client side information in the form of desktop indices. Our approach is to compare & study various privacy and security methods based on different users' personnel information collected by search engines.

5.   Web Personalization

It is introduced to improve Web site usage by customizing the contents of a Web site with respect to the users' needs [11]. They provide mechanisms that collect information describing user activity and elaborate this information. Personalization systems usually process the information related to users' sessions, that is, a sequence of pages requested by the same user. For instance the sequence: **Home → Science → Computer → Science → Algorithms** could be a typical session of a person browsing a directory (for example, Yahoo!) with an interest in computer algorithms. Web site personalization can be defined as the process of customizing the content and structure of a Web site to the specific and individual needs of each user taking advantage of the user's navigational behavior. The steps of a Web personalization process include: (a) the collection of Web data, (b) the modeling and categorization of these data (preprocessing phase), (c) the analysis of the collected data, and (d) the determination of the actions that should be performed. The ways that are employed in order to analyze the collected data include content-based filtering, collaborative filtering, rule-based filtering, and Web usage mining. The site is personalized through the highlighting of existing hyperlinks, the dynamic insertion of new hyperlinks that seem to be of interest for the current user, or even the creation of new index pages.

Web personalization is defined as any action that adapts the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational behavior and individual interests, in combination with the content and the structure of the Web site. The objective of personalized search to disambiguate the queries according to users, interest and return relevant results to users. Users tend to issue short queries when searching, resulting in tremendous ambiguity about their informational goals.

6.   Threats to Personalized web search

Personalized Web Search is an advance & better way to improve search quality by making search results for users & people with individual information interest. But sometimes, users are uncomfortable with exposing private preference, need, and their location information to search engines. On the other hand, privacy is not good & up to standard, and it can be compromised if there is a gain in service or profit to the user. Thus, a balance must be made between search quality and privacy-security protection. Therefore to maintain privacy and security during personalized Web Search is very necessary and lack of this will a bigger threat to personalized web search and to its users.

6.1 Way to Protecting Privacy with Maintaining Search Quality

A scalable way for users must be presented so that automatically privacy-security protection tools can be provided during personalized web search. This approach makes a user's interests into a hierarchical organization according to specific interests. For specifying privacy we

proposed to user to choose the content, degree and amount of detail of the profile, personnel, need, preference, location information that is exposed to the search engine. Related studies during analysis showed that the user profiles & related personnel parameters improve search quality when compared. Web search personalization algorithms improve the Web search experience by using an individual's data (e.g., topical categories marked interesting, query history, or term vectors of previously viewed content) to identify the results that are the most relevant to that individual. This can be done in several ways.

7.   Need of Better Security

The Web-search scenario is also a good one in which to have a focused discussion about privacy. It has a few properties that are extremely important from the user's point of view, i.e., with respect to actions the user can take to control what he reveals about himself. First, search services are widely used, and thus there is hope of hiding in the crowd. Second, because of the large number of users, a concrete, widely available tool for enhancing privacy might produce useful feedback. Third, it is a point of connection among most web activities; so the privacy concerns are larger than in more specific web services.

7.1.   Why Security & Privacy Needed

In practice, however, privacy is not absolute [12]. There exist already many examples where people give up some privacy to gain economic benefit. One example is frequent shopping any store. Consumers trade the benefit of extra saving in the grocery stores versus the creation of a detailed profile of their shopping behavior. As another example, consider a cricket fan. He may not be comfortable broadcasting a weekly work-out schedule, but might not mind revealing an interest on cricket if a search engine can help identify any information about his favorite team. Thus, people may compromise some personal information if this yields them some gain in service quality or profitability. Another important observation is that detailed personal information might not be necessary if it is possible to catch a user's interests at more general level. In the above example, the times and locations where the user has played cricket would not be relevant in searching for a favorite cricket team. In fact, such unnecessarily detailed information often becomes noise in the search task. Hence, a proper filtering of a user's private information not only helps protect the user's privacy but also may help improve the search quality. The key is distinguishing between useful information and noise, as well as good balance between search quality and privacy protection [13]. This work targets at fulfilling the conflict needs of personalization and privacy protection, and provides a solution where users decide their own privacy settings based which type of information they want to display/provide against search engine. The above said privacy incorporates the user profile, which offers a convenient way to determine the

limit to which personal information is exposed. This relies on the assumption that more general and frequent terms, which carry smaller self-information, represent information users are more willing to share. There is a kind of privacy infraction or infringement in personalized web services i.e. during personalized web search now a days user requires strict security policy to secure its private information from search engine but at the same time they do not want to compromise with quality of search results. Search-engine queries can reveal a great deal about a user.

8.   Conclusion

As search becomes an increasingly essential part of so many Internet users' daily lives, the breadth and depth of information contained in query logs grows to unparalleled levels. As a body of data that can reveal the interests, preferences, search strategies, and linguistic behaviors of entire populations, query logs are a true bounty for research of all kinds, conducted internally, at the search engine companies, and externally, by academics and others. Achieving the right balance between protecting privacy and promoting the utility of the logs is thus difficult but necessary to ensure that Internet users can continue to rely on Web search without fear of adverse privacy consequences.

References

[1].   A Choicestream Survey on Personalization, 2004-2006, USA

[2].   Google personalized search: http://www.google.com/psearch

[3].   ADAR, E. 2007. Private Information Sharing On Web. In Proceedings of the 16th

[4].   Agichtein, E.,Brill, E., Anddumais, S. 2006. Improving web search by incorporating privacy in user behavior information. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA.

[5].   FOLEY, J. 2007. Are Google searches private? An originalist interpretation of the fourth amendment in online communication cases. Berkeley Techn. Law J. 22, 1, 447–472. ACM Transactions on the Web, Vol. 2, No. 4, Article 19, Publication date: October 2008.

[6].   Glen Jeh and Jennifer Widom. Scaling personalized web search with Relevance Feedback. In Proc. Of the 12th International World Wide Web Conference (WWW), Budapest, Hungary, May 2003.

[7].   ASK.COM. 2007. Ask.com puts you in control of your search privacy with the launch Of 'askeraser'. Www.irconnect.com/ask/pages/news releases.html?D=132847.

[8].   Barbaro, M. And Zeller, T. 2006. A face is exposed for AOL searcher no. 4417749.In The New York Times. Http://www.nytimes.com/2006/08/09/technology/09aol.html?Ex=1312776000.

[9].   H.R. Kim, and Philip K. Chan. Learning implicit user intentions for context in personalization. In Proc. Of International Conference on Intelligent User Interface (IUI),Miami, Florida, January, 2003.

[10]. Howe, D. And Nissenbaum, H. 2008. Trackmenot:Achieving User Intentions in web search. In On the Identity Trail: Privacy, Anonymity and Identity in a Networked Society. Oxford University Press, Oxford, UK.

[11]. M. Speretta, and S. Gauch, Personalizing Web search. In Proc Of International Conference of Knowledge Management(CIKM), Washington D.C., 2004

[12]. R. Agras and D. Phillpe. Privacy preserving Information Searching.In Proc. Of the ACM SIGMOD Conference on Management of Data (SIGMOD), Dallas, Texas, May 2000.

[13]. P. H. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter. Using ODP metadata to personalize search. In the Prof. Of 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Salvador,Brazil, August, 2005.