When Personalization Becomes Too Personal

Gkarnara Christina
Department of Archives and Library Science
Ionian University
pms09gka@ionio.gr

Abstract

With the advent of the Web thousands of users interact with search engines daily seeking information on a wide variety of topics. The problem is that search engines cannot discriminate the different cognitive and search skills of individual users who query the Web for multiple topics often using similar queries. Web personalization is a promising approach to provide users with the information they expect by customizing the results to each specific user. One way to achieve personalized search results is to ask users about their preferences and information needs. However, users are not willing to reveal private information and often they are concerned about their privacy breach. In this article, we give an overview of the privacy issues in web search that have been previously addressed so as to apprise users on how to achieve a balance between effective personalization and simultaneously achieve a high level of privacy protection.

Keywords - web search, personalization, privacy, pseudo identity, group identity, anonymity, encryption.

1 Introduction

With the advent of the WWW the number of users who interact with search engines increases rapidly as rapidly flourish the documents that are indexed by search engines. Users can find the desirable data in two predominant ways: they can either *search* or *browse* [18]. Via the first way, users submit a keyword to a search engine, which retrieves documents that contain these keywords. This way is the most popular way to seek information and the advantage is that the user finds quickly the information by identifying the pages which contain the information query [17]. The second way is by browsing and is done through an existing ontology of topics on which user navigates by clicking the preferred topical node until he reaches the desirable area of interests [18]. This way of seeking information is efficient when the user isn't connoisseur of the search domain. The Web is an information repository in which the data structure is heterogeneous. Many times users resent by the numerous unrelated results that search engines retrieve thus making the user's navigation a time consuming procedure.

To fill this void, personalization of search results has been introduced as a promising direction in enhancing the accuracy of the retrieved results in terms of user-specific needs. Personalized search targets to design systems which retrieve tailored collections of pages adaptable to each individual user profile. There exist two approaches to construct the user's profile, through *explicit* or through *implicit* feedback [17]. Both ways require collecting and storing user information which illustrate the users interests, preferences, needs, tastes and general users personal data in order to decipher the query intention. Hence, in order to achieve search

personalization, search engines need to collect large amounts of users' personal data. But, users are reluctant to reveal private information or their preferences and often are concerned about their privacy breach. In this article, we survey privacy issues pertaining to web search that have been identified by numerous researchers and we highlight the avenues for future research on maintaining profile privacy in personalized web search. The quest of our survey is to demonstrate how users can achieve a balance between personalized and secured search results.

The article is organized as follows. In Section 2, we highlight the contribution of personalization when searching the Web. In Section 3, we present the two approaches that exist to obtain personalization and we introduce the concerns they raise on privacy issues. In Section 4, we underline how to achieve privacy in web search by presenting existing approaches. In Section 5, we list some challenges that need to be addressed and in Section 6 we conclude the article.

2 The Contribution of Personalization

Search engines have been designed to gratify different users and meet their growing needs and expectations while they seek for information. Despite the advancement of web search, search engines are still not optimally operand. This is due to the following: firstly, queries might be polysemous [21]. When different information seekers submit the same query in a search engine, each of them expects dissimilar retrieved results. For, example, for the query "cookies", some users desire to inform about the mechanism that text files are storing on the client side computer, while others expect information related to the dessert and how to cook the dessert "cookies". Another example, for the query "java", some users desire to inform about the programming language "java", while others expect information related to "coffee" [21]. Therefore, the second problem that emerges is that the search engines for every input information query like above retrieve the same information to all users, is that they follow the model of "one size fits all" [17]. This model makes the user's navigation through the retrieved results until reaching to the desired information a time consuming procedure.

To overcome the above, we can personalize search results according to user-specific needs. Personalized search targets to design systems which retrieve tailored collections of potentially relevant pages classified in such a way that reflects the pages relevance to the query keywords and to be adaptable to each individual users profile [17]. In order to achieve personalization, systems need to construct a user profile which represents the user's interests and preferences. The approaches towards identifying user profiles are discussed in the section 3. One example to illustrate the contribution of personalization is referred in [20]. The system called "Susanna" affords the opportunity to a user to adopt on its own the navigational pattern he desires. Namely, the system gives the opportunity to the user to choose if he wants to retrieve personalized results or not. Assuming there are two users each of them wishes to get personalized results and they both issue the same query, e.g. searching for a book and especially a book written by an author named King. In the case of the first user who doesn't want personalized results, the system retrieves results which contain books of the author Stephan King which they don't belong into a specific thematic hierarchy. In the case of the second user who desires personalized results, before issuing the query, he declares to the system his interests and preferences and the system constructs a user profile. When the user inputs the same query, the system retrieves personalized results and particularly the system retrieves results which

contain books of the author Robert King that belong to a particular thematic hierarchy which is “Computers & Web”. Thus the second user saves time and gets the desirable results in a more effective and efficient manner. Consequently, the contribution of personalization is manifold and the essential prerequisite to achieve personalization is the construction of user search profiles. In the next section, we present the two main approaches towards creating search profiles.

3 Approaching Personalization

This section focuses on personalization approaches and in particular on the two main directions via which user profiles are constructed. Figure 1 illustrates the overall search personalization process for both of the presented ways. The first approach of user modeling is the technique of *explicit profiling* [17]. The user’s profile is constructed via the explicit declaration of preferences, usually through registration forms or questionnaires. Hence, the interests of each individual user can be identified. The system which implements the explicit approach requires the direct involvement of the users. The drawback of this approach is that the disclosure of personal data makes the users disinclined to explicitly specify their needs and they find the registration a time consuming and confusing task. The second approach of user modeling is through *implicit profiling* [17]. To overcome the challenges of the explicit feedback, the implicit profiling does not require the user’s direct involvement. The user’s profile is constructed by tracking and monitoring the user’s querying and results’ navigational behavior. The systems which implement the implicit approach capture the personal needs of each specific user via the interaction among a user and the system. Particularly, by capturing the complete navigational path such as the pages the user visits, the URLs and anchor texts the user follows for particular queries, the time and date of access [17].

Thus, the system emanates the user’s interests in personal topics and constructs the profile. An overview of above can be found in [2,3,4,8,14,23,24].

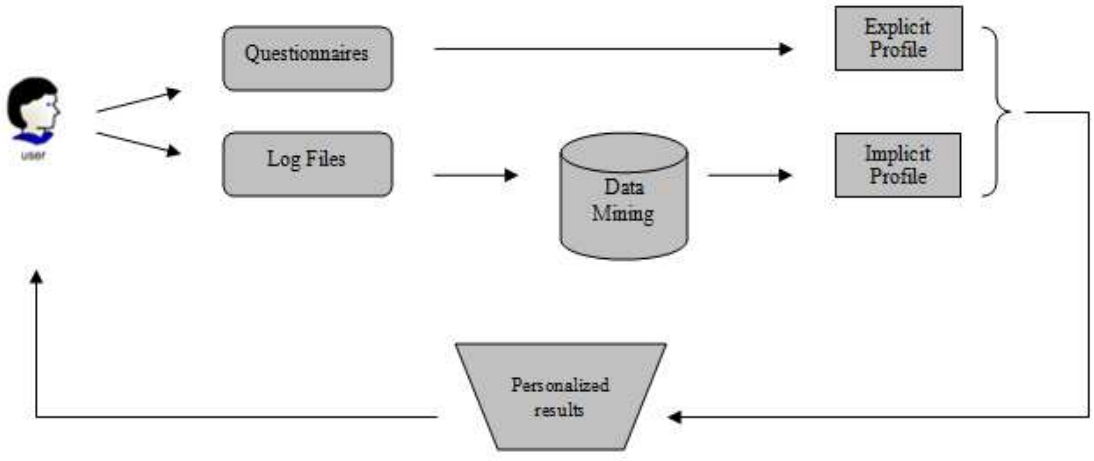

Figure 1. Explicit & Implicit Profiling to learn and capture user’s interests to retrieve personalized results.

Implicit profiling relying upon the recorded users' interaction with the search engine is effective in that it does not interfere directly with the user. Still, it raises concerns about the level of distribution and privacy of the collected data. Such data concern the full record of the user's search logs. A current example of data disclosure is the case of AOL search engine [1,27,15]. In August of 2006, the American Online (AOL) search engine disclosed an extremely large amount of query log. This query log extracted over a period of three months. 20 million search queries which have been submitted by 658,000 users were released in order to help researchers in the Information Retrieval (IR) community. This release denotes expose of private data for a number of AOL users. The mistake of the AOL search engine was that they haven't sanitized the queries; they have just replaced the user IDs but not the search queries. For example, journalists from New York Times (not professionals specialized in data mining) could accomplish to decipher the user corresponding to ID 4417749, even without taking account the existence of social security numbers, driving license numbers and credit card numbers. Through observation the detailed and multiple queries they found that the user with the ID 4417749 was corresponded to a 62 year old woman living in Georgia. Another example [27] is the search queries entered by the user 1515830:

calories in bananas
surgical help for depression
jobs in denver colorado
teaching positions in denver colorado
anti psychotic drugs

As the above studies suggest, it is relatively easy to decipher private information about concerns the user such as health and mental condition, profession, geographical location. We can comprehend through this example how users unintentionally reveal information about themselves when submitting queries in search engines. Beyond the analysis of user's search logs, another way a search engine can identify the user is under their IP address. In order for a user to communicate and interact with search engines, an IP address is indispensable. An IP address is a unique string of numbers assigned to each user's computer. In short, IP is like a user's street address or telephone number. Furthermore, "cookies" enable search engines to decipher a user like a recurrent visitor and glean his searches and store them even using different IP address [27]. From the above we understand that users are treated as easy prey. Most information seekers are concerned about their privacy and they desire safety and security during their navigation. In the next section, we discuss technological approaches that have been researched so as to enhance users' privacy protection and accomplish a balance between effective personalization and privacy protection.

4 How to achieve privacy in personalized web search

In the current section we present a bibliographic overview of the technical solutions that have been examined in order to account for privacy issues with respect to personalization. Privacy Enhancing Technologies have been deployed since early eighties. Recent surveys indicate that although nearly 80% of the information seekers are interested in personalization [16], 83% of them are concerned about the privacy of their personal data, 82% refuse to disclosure personal information to a web site, 27% are not inclined to provide any personal information to a web site, 34% of the information seekers submit false information when asked to register or fill out an

online form and 49% believe that sites share users' personal information with other sites, thus resulting to privacy breach [26]. Users differ from one another, which indicates that every user has different requirements for privacy protection and as such their opinions on privacy enhancement differ widely. There are users who don't desire to share personal information with anyone, while others are eager to disclosure personal information in order to achieve high quality services. The level of privacy protection depends on the individual needs and expectations [13] and as such it should be adaptable so as to satisfy the gamut of user requirements. Regardless of the different approaches that have been investigated towards achieving the above, one thing they all share in common is the process of anonymizing the search logs. User data anonymity can be achieved through many technical solutions such as *pseudo identity, group identity and encryption.* Each of the above approaches is a gradational advancement of privacy protection of the precedent. In the following paragraphs, we outline the most widely used approaches to web data anonymization without impairing personalization.

Pseudonymous Personalization

One personalization system which allows users to remain anonymous during their personalized web browsing is the Janus Personalized Web Anonymizer [9]. Janus functions as a proxy between a user and a web site. For each user's search session, Janus automatically generates a different alias so as to establish an anonymous user account at the website. For example, user John Smith desires to navigate to the New York Times web site. Before John starts up his web browser and connect to the New York Times web site, Janus request from John to be recognized from Janus proxy and thus asks to fill a Janus authentication form. This form requires John to give a username that is recommended to be the email address and a secret- id1 and S1-. After Janus has achieved the required information about John, it allows John to navigate to the New York Times site. If John navigates for the first time to this site, in order to establish an account, sends John a registration form. John can simply fill the registration form with strings *"\U"* for username, *"\P"* for password and "\@" for e-mail address. These strings are recognized by Janus and compute aliases on John's behalf. For every site in which John navigates, different aliases are computed by Janus, which are all distinct from other users (u1, p1), (u2, p2), (u3, p3).

A similar system to Janus is Lucent Personalized Web Assistant (LPWA) [10]. LPWA is a pseudonymous tool that let Web sites offer identification based services without linking to user's actual identities.

Another approach [11] presents an architecture that allows users to create with minimal effort their personas with the aid of Persona Manager (subset of information) and through these authenticated personas provide information to service providers without revealing their identity. Service providers use these personas to enhance accuracy in personalized services. For anonymity and further privacy protection the architecture support multiple personas for each user, so as to not underline identity and not match linking personas to each user. So, when an information seeker navigates to a service provider website the communication is established between the Personal Server Device (PSD) and service provider's system. The user selects a persona to release to service provider and the service provider after authentication of the persona, begin to communicate with the user so as to provide personalized services.

Distributed Personalization

Distributed personalization has been investigated in the domain of collaborative filtering. Systems which support collaborative filtering based on the assertion that like-minded users with similar needs can be divided into groups of similar users. Consequently, personalization is done in users' group level instead of an individual user level. Search engines construct a group users profile and according to the variation of topical interests, dissimilar group of user profiles are constructed which each group profile sharing a distinct user identity [21].

An approach of privacy protection in collaborative web personalization is the meta-search engine I-Spy [22]. I-Spy relies on tracking and capturing via the click through data the user's preferences. It also relies on the *hit-matrix* so as to construct a community profile. Each individual interaction of user doesn't record and thus search history cannot be related to a particular individual user and thereby there is no concern of compromising user's privacy. Hit-matrices decay hit-values so as to keep up with the alternating preferences of community over time so as to achieve personalization while preserving anonymity of individual users. It is important to declare that different hit-matrices exist for different communities of users. A community of users is discernable from other popular and recently accessed communities. In the I-Spy start-page of a community, near the query box there is a tick-box "private" which gives the opportunity to users to exclude their search queries from being shown to other community members.

Another approach is a Firefox web browser plug-in called TrackMeNot [21] which periodically randomizes search queries to search engines. This method's goal is to make an individual's user profile to look like a group of users profile by matching firstly common queries and mix them so as to resemble that belong to a group of users and not to a certain user.

[13] implemented a system called Masks (Managing Anonymity while Sharing Knowledge to Servers). The system adapts pseudonym to group-based personalization. The system consists of a server-side, Masks server, which functions as proxy between users and web sites, and privacy and security agent (PSA) which functions as an intermediary between users and Masks server. Each user has a temporary identification that adopts while interacting with a web site and is associated with the user's interest in a thematic topic through a use of a semantic tree. For example, sites that offer travel information and the group interested in travel shares one mask, thus on one hand users can savour personalized services but on the other the system can't profile each individual user. Users can associate with a majority of groups and masks. The PSA gives the opportunity to the user to configure the masks and undertake other functionalities such as blocking and filtering privacy violations, such as cookies and web bugs.

Anonymous personalization

Anonymity means that neither a user's identity nor his location can be emanated and tracked online [5]. Thereby, construction of a user's profile can't be build even at a group level [21]. There exist anonymous networks such as the web browser Torpark, which is based on Tor onion-routing network, to obscure the communication path of users, using a network of routers that breaks the link between incoming and outgoing traffic [19]. A search engine disables to decipher where the query derives from, but the retrieved results can return to the correct user through Tor Network [21].

Tor [19] consists of over 800 routers and is at disposal to an estimated number of 200.000 users. Here, we present approaches that we pay regard to online anonymity for personalized web services.

One approach is presented in [29]. This approach is based on the assumption that a personalized query consists of two parts <d, q>. The query *q* is unstructured data and contains sensitive information while *d* contains demographic and interests' information which is used to personalize the retrieved results. At this point, a privacy breach is discerned. A detailed navigation *d* may match through links and observed paths of users to a small number of users or to a unique *q*. To fill this void, a user pool is introduced which utilizes the semi-honest model and anonymous communication channel exists between each user and the user pool. Thereby, before the user interacts with the search engine and submits a query *q*, he must first register with the user pool his personal information *d*. The user pool presents the personal information generalized *d´* where *d´* contains less but important information, than *d*. For example, a user in his demographic data reveals age, gender, zip code. If *d* contains "Age=25", then *d´* after generalized personal information becomes "Age in [20,30]". Therefore, the user submits to the web service consequently < d´, q> and it is feasible to achieve personalization without privacy breach and without the user being concerned about disclosing his identify because personal information according to this approach remains anonymous.

Another approach is presented in [1]. This approach can be classified both in the case of anonymity and the encryption that we describe next, because based on two specific solutions so as to achieve balance between privacy and personalization. The first approach of anonymizing query logs is through a cryptographic technique based on secret shares. This technique relies *on-the-fly elimination* with secret sharing and consists of two primary issues. The first issue is to remove retrieved results that uniquely identify a user. But many times a specific query may prove to be not as uncommon as originally presumed and if the data is removed it is difficult to be recovered, thus meaning that valuable data can be potentially lost. The second issue is that a service provider needs to keep a histogram of submitted queries so as to achieve accuracy in the retrieved results. To fill this void, threshold cryptography or a secret sharing application has been proposed. The secret S, illustrates the query and then splits it into a number of shares. Each share is not indicative on its own, but in combination the secret can be decoded if it appears a *t* times before been decoded. The second solution relies on splitting personality so that a service provider can't correlate the user interests with a specific user. For example, if a user is interested in "football" and "cooking", upon personality split, the user will look like two distinct users with two separate preferences. In example of the 62 year old woman living in Georgia from AOL disclosure we mentioned in a previous section, they found that she consists of 165 different personalities. These two mechanisms contribute to anonymize the users and make them seem dissimilar so as to achieve a balance between privacy and personalization.

Encryption

Encryption is the process of transferring information -plaintext- with the aid of an algorithm -cipher- to make the plain text unreadable by anyone, except those possessing specific knowledge, usually referred to a possession of public or private key. Each key is interdependent to each other. If one key is used for encryption, the

other is used for decryption. Encryption is used to protect data which is being transferred via networks [31].

One approach in this respect is introduced in [30] and it is called the de-identification approach. This approach distinguishes the data in identifiable which contains user names, account IDs, card numbers. These data can easily map users. The other classification of data is the piece of data which contains the user preferences and interests which is not as sensitive as the previous. The basic idea of the de-identification approach is to encrypt identifiable data with the domain salt and makes it infeasible to correlate users with their profiles under different sites. The remaining data which consists of preferences doesn't undergo changes so that the personalization can be achieved.

5 Challenges to Web Data Anonymization - Further Proposals

Privacy in personalized web search poses many challenges that need to be addressed. In this section we outline some challenges that have been identified after applying some of the approaches previously discussed. The quest in privacy preservation for personalized applications is to ensure a balance between personal data protection and user-specific services. The design guidelines which we described above account for different levels of privacy protection with the encryption method to be at the highest level.

The first approach for safeguarding and enhancing privacy in personalized search we have addressed is the pseudo identity approach. Pseudonymous personalization enables information seekers to remain anonymous through aliases during the different user search sessions [28]. Namely, when a user reveals his IP address, the system emanates geographic information which denotes the *whois* service [21]. Furthermore, the majority of the online web sites require the user -in order to fill a registration form- to provide a valid e-mail address which has a result to reveal personal information about the user's identity. With a pseudo identity, such personal information is protected. But, the approach of pseudonymous personalization ostensibly seems to be a panacea however, actually offers the lowest level of privacy protection because if grouping a number of queries from the same user, it is possible to identify the user.

The approach of group identity denotes higher degree of privacy than the first level of pseudonymous personalization because on one hand an individual user profile cannot be constructed and on the other the identity of one user is not discernable and lost in a group identity, so it is a hard task to emanate true information of each individual user [21]. Higher privacy protection than the approach of group identity provide the anonymous personalization but some user information is kept at the search engine in order to mine search logs and thus it is possible to emanate a user's identity from distinct searches. The encryption approach offers a fully protected percentage of user's privacy but it is difficult to be achieved because the implementation of encryption and decryption is a costly procedure [21].

Achieving a substantial gain and to bridge the gap between privacy and personalization we have to overcome these challenges and define new protocols. Personalization is a promising approach but the concern of user's privacy is beyond dispute. One movement to enhance privacy is the Platform for Privacy Preferences Project (P3P), created by the Wide World Web Consortium [6]. This Platform empower websites to express their privacy policies in a standard machine-readable

vocabulary so as to provide information to users about the sites' privacy policies and underlying to them when a privacy breach is exist.

Another proposal for enhancing privacy in personalized search arouse by [7] which supports that personalization gives amelioration not to the majority of retrieved results after inputting a query. So according to [25], not all searches need to be personalized. The contribution of personalization is useful when a large click entropy is discerned, meaning that various pages have been clicked for an input query, while the detection of low click entropy indicates that only a few pages have been clicked, thus there is no need to apply personalization. As a result the need to search results is mainly pronounced for informational queries and less so for navigational ones. Furthermore, according to [12] introduce manually selection of privacy according to the users' requirements. The system offer the opportunity to users' designate the level of privacy they prefer and declare which preferences and interests desires to be *public*, *semi-public*, *private and don't share*. This system with the correlation of personalize only queries with large entropy, users achieve personalization and control by themselves the privacy level they prefer. Lastly, taking account the P3P Platform, users have a better image about the privacy policy that a website supports and enhance the navigation to the current website without concerns on behalf of the users for their privacy breach.

6 Conclusion

Personalization is a promising approach to enhance accuracy in retrieved results so as to cater for the user needs and overcome the problem of polysemy by the reason of rapidly flourishing documents that are indexed in search engine databases. The two predominant approaches towards personalization are: explicit and implicit user profiling, with the latter respecting the user interests and search preferences. The acquisition of user personal data is a considerable task and the initiatory stage in order to achieve effective personalized assistance and to avoid mismatching retrieved results. These pieces of personal data are very beneficial to capture the tastes of users and retrieve relevant results but users are concerned about their privacy breach. In this paper we have summarized the most known privacy issues in personalized web search and outlined their strengths and weaknesses in the hope of paving the ground for most advanced data protection services in the next generation search engines.

References

1. Adar E. 2007. “User 4xxxxx9: Anonymizing Query Logs”. In *Query Logs Workshop, at the 16th WWW.*

2. Agichtein E., Brill E., Dumais S. and Ragno R. 2006. “Learning User Interaction Models for Predicting Web Search Result Preferences”. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA.

3. Agichtein E., Zheng Z. 2006. “Indentifying “Best Bet” Web Search Results by Mining Past User Behaviour”. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD, New York, NY, USA.

4. Aktas M., Nacar M. and Menczer F. 2004. “Personalizing PageRank Based on Domain Profiles”. In *Proceedings of the sixth WEBKDD workshop, in conjunction with the 10th ACM SIGKDD conference*, Seattle, Washington, USA.

5. Al-Muhtadi J., Hill R. and Campbell R. 2004. “A Privacy Preserving Overlay for Active Spaces”. UbiComp, Privacy Workshop, Nottingham, England.

6. De Grande R. and Zorzo S. 2006. “Privacy Protection Without Impairing Personalization by Using the Extended System MASKS and the Extended Contextualized P3P Privacy Policies”. In *Proceedings WebMedia '06 Proceedings of the 12th Brazilian Symposium on Multimedia and the web*, ACM, New York, NY, USA.

7. Dou Z., Song R., and Wen J.R. 2007. “A large-scale evaluation and analysis of personalized search strategies.” In *Proceedings of the 16th international conference on World Wide Web*, 581-590, ACM, New York, NY, USA.

8. Eirinaki M., Vazirgiannis M. and Varlamis I. 2003. “Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process.” In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD, New York, NY, USA.

9. Gabber E., Gibbons B. P., Matias Y. and Mayer A. 1997. “How to Make Personalized Web Browsing Simple, Secure, and Anonymous”. In *Financial Cryptography ´97* (Vol. 1318), Berlin- Heidelberg - New York: Springer Verlag.

10. Gabber E., Gibbons B. P., Kristol M. D., Matias Y. and Mayer A. 1999. “Consistent, Yet Anonymous, Web Access with LPWA”. In *magazine Communications of the ACM*, Volume 42, Issue 2, New York, NY, USA.

11. Hitchens M., Kay J., Kummerfeld B. and Brar A. 2005. “Secure Identity Management for Pseudo-Anonymous Service Access”. In *Proceedings of the Security in Pervasive Computing: Second International Conference*, Lecture

Notes in Computer Science Volume 3450/2005, pp. 48-55, Boppard, Germany.

12. Hawkey K. and Inkpen M. K. 2006. “Examining the Content and Privacy of Web Browsing Incidental Information”. In *Proceedings of the 15th international conference on World Wide Web,* ACM, New York, NY, USA.

13. Ishitani L., Almeida V. and Wagner M., Jr. 2003. “Masks: Bringing Anonymity and Personalization Together”. IEEE Security & Privacy Magazine, Volume 1, Issue 3, 18-23. IEEE Educational Activities Department Piscataway, NJ, USA.

14. Kazunari S., Hatano K. and Yoshikawa M. 2004. “Adaptive Web Search Based on User Profile Constructed without Any Effort from Users”. *In Proceedings of the 13th international conference on World Wide Web,* New York, NY, USA.

15. Kiraly C. 2007. “Privacy vs. Profiling on the Internet”. Conference “INTER: A European Cultural Studies Conference in Sweden”, organised by the Advanced Cultural Studies Institute of Sweden (ACSIS) in Norrkoping.

16. Kobsa A. 2007. “Privacy-Enhanced Personalization”. In *Communications of the ACM*, Vol. 50, Issue 8, New York, NY, USA.

17. Micarelli A., Gasparetti F., Sciarrone F. and Gauch S. 2007. “Personalized search on world wide web”. Published in book: The adaptive web, Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-72078-2.

18. Pretschner A., Gauch S. 1999. “Ontology Based Personalized Search”. In *Proceedings of the11th IEEE Intl. Conf. on Tools with Artificial Intelligence,* Chicago, pp. 391-398.

19. Saint-Jean F., Johnson A., Boneh D. and Feigenbaum J. 2007. “Private Web Search”. In *Proceedings of the ACM workshop on Privacy in electronic society.* New York, NY, USA.

20. Semeraro G., Degemmis M., Lops P., Thiel U. and L’ Abbate M. 2003. “A personalized information search process based on dialoguing agents and user profiling”. In *Proceedings of 25th European Conference on IR Research, ECIR*, Lecture Notes in Computer Science,Volume 2633/2003, 548.,Pisa, Italy

21. Shen X., Tan B. and Zhai C. 2007. “Privacy Protection in Personalized Search”. ACM SIGIR Forum, Volume 41 Issue 1, New York, USA.

22. Smyth B. and Balfe E. 2006. “Anonymous personalization in collaborative web search”. Information Retrieval, Springer Link, Vol. 9, No 2, 165-190.

23. Sugiyama K., Hatano K. and Yoshikawa M. 2004. “Adaptive Web Search Based on User Profile Constructed without Any Effort from Users”. In

Proceedings of the 13th international conference on World Wide Web, New York, USA.

24. Tanudjaja F. and Mui L. 2002. “Persona: A Contextualized and Personalized Web Search”. In *Proceedings of 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*-Volume 3, Big Island, Hawaii, pp. 67.

25. Teevan J., Dumais S. and Liebling D. 2008. “To personalize or Not to Personalize: Modeling Queries with Variation in User Intent”. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA.

26. Teltzrow M. and Kobsa A. 2004. “Impacts of user privacy preferences on personalized systems”. Designing Personalized User Experiences in eCommerce, Human-Computer Interaction Series, Springer Link, Volume 5, Section 5, 315-332.

27. Tene O. 2008. “What Google Knows: Privacy and Internet Search Engines”. http://works.bepress.com/cgi/viewcontent.cgi?article=1001&context=omer_tene&sei-redir=1#search="What+Google+Knows:+Privacy+and+Internet+Search+Engines"

28. Wang Y. and Kobsa A. 2008. “Technical Solutions for Privacy-Enhanced Personalization”. In Mourlas, C. and Germanakos, P. (Eds.), *Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies*. PA: IGI Global, Hershey.

29. Xu Y., Wang K., Yang G and Fu W.C. A. 2009.”Online Anonymity for Personalized Web Services”. In *Proceedings of the 18th ACM conference on Information and knowledge management*, New York, NY, USA.

30. Zheng J., Yao J. and Niu J. 2008. “Web User De-Identification in Personalization”. In *Proceedings of the 17th international conference on World Wide Web*, New York, NY, USA.

31. http://en.wikipedia.org/wiki/Encryption