

# Privacy Preserving in Personalized Web Search

J. Haritha & P. Sriveni

Assistant Professor, Department of CSE, Balaji Institute of Technology & Science, Warangal District, Telangana, India.

## Abstract:

*In this work, the value of privacy in web searches considering personalization. To Beginning will be a brief introduction to anonymization and Personalization given. This contains different ones Species and their requirements. After that, on the Main topic "What is the value of privacy at Web Search?" This will be an experiment and used a survey. For the experiment should Subjects perform various web searches? Were whose behavior and privacy attitudes are observed and evaluated. In the further course of the work the procedure becomes of the experiment and its results. Following the structure of the survey will be explained and explained Evaluation of the results of the experiment.*

## Keywords

Web search, search engine, anonymization, personalization, Profiles, privacy, privacy, types of personalization, Value of Privacy, Personalized Search

## 1 INTRODUCTION

"How important is your privacy to today's society?" or "How important is my privacy?". By giving personal information and own Interests are greatly simplified when dealing with websites. However, restricting the access of these data again, or they even deleting is far more difficult than creating it before. An example on this topic is "The Right to Forget", as it is called in the media. This is Google after the Lost Precedent [4] at the European Court of Justice the possibility of deletion within the EU to give your own query results. True, that is basic idea comprehensible, but you can at the moment assume that, in my opinion, this right for a long time only a few will be used. On the one hand this is because the process is expensive to perform by completing a form and every single request On the other hand, this right applies only to Google domains in the European Union. [5] The Main website google.com is not affected by this not operated within the EU. Next comes that a copy of the identity card of the private individual must be uploaded, which is a unique Enable identification by Google. This is done by

criticized several pages. An example can be found below [14]. Thus, the following considerations come up: "What is us actually worth the privacy on web searches? "and" how does personalization look like today? "

## **2.ANONYMIZATION AND PERSONALIZATION**

After Federal Data Protection Act [13] is Anonymization Are defined as "[...] changing personal data in such a way that the details of personal or factual circumstances no longer or only with a disproportionate amount of time, money and costs Worker of a particular or determinable natural Person can be assigned. ". Thus summarized, changing data that they are no longer a person can be assigned. In contrast, is Personalization defined as the collection and processing of Data of a person, it is the "raising [...] the procurement data about the person (person) concerned. "and the processing "[...] saving, changing, transmitting, locking and Delete personal data. "

### **2.1 Possibilities of anonymization**

Information gathering has always been for people important, but not every topic you talk about on the internet, you want to have logged in his public profile with. Especially if these logs are a person whose environment with

friends, relatives, their buying behavior and interests to include in the smallest detail. How much with such Records can begin and how to recover these For example, on the Google Trend Function [17] listing the most searched terms and trend trends are evaluated. That's why also like the possibility of anonymisation resorted. There are different types of his identity on the Internet or at Disguise web searches. For one thing, there is the possibility false information (eg age, sex, place of residence) at the To register. This is the simplest for the user Possibility to distract yourself. On the other hand, this does not help at the problem of personalization, because so far still one, if only virtual, person exists. Also at the use of different "pseudo" profiles the possibility that by means of behavioral pattern analysis a Personalization is performed. That's why often other external services involved (eg Tor [6], JonDonym [7] and various proxy or VPN providers) that not only their own access location, but also the identity of a each person through a public, multi-used, Replace identity. Thus, for example, in the case of Web searches, no personalization based on interests Be there different users over the same IP address with their respective interests on the web search engines access at the same time. When using external services

Nevertheless, there is still the danger of data storage. Because these services must, if only temporarily, the store correct data of persons to be disguised, So network traffic to the right people can be forwarded. Another option around the

Identify people based on this data storage To avoid this would be the use of public Internet access (eg the use of public hotspots on Railway stations or airports). You should also note once you have personal information, even over an anonymized Connection, shipped, you can quickly unmasked become. One last option that is used and in This work is called using one Illegally obtained identity (eg the benefit of a private Wlan connection of a foreign person). So will although its own identity is obscured, but this approach can criminal consequences. Since this is not the subject of this work, it will not continue received. If there is further interest, [15] and [16].

## 2.2 Types of personalization

Just as there are different types of anonymization, There are also different types of personalization. The following types are used as a basis for web searches used. The first type is called "Relevance Feedback and Query Modification ". Although the information

serves not directly personalization, but in the data collection process, it is an important basis. The real goal Here is a better search result to find, based on the entered information and specific results weightings is created. This is the original Search query internally by other results and data (eg

Site relevance). You can do that with a Compare personalization of a single search query. The second type is based on the analysis of the content of the web pages. It compares content from different websites and on the basis of relevance and previous habits or interests of the user adapted. The user's profile has already been used in this method created. In contrast to the previously mentioned "Relevance- System "is here considered a larger space than a single Search query. As soon as a user makes a query, will using the "relevance system" a sort of Returned results. This will be again rearranged by the user's previously created profile. The previous species are relatively static and work on one certain specification. Since it is very expensive, a good profile To create this, the analytic system should dynamically itself can perform. This is now the task of "Recommender Systems ". This is done using algorithms called Database use long-term information, the analysis of Content of the website further adapted to the user. There this entire construct

is based on the relevance system and this with various matching requests extended to structure the result, the Recommender System can estimate which queries the user searches for Would give place. In contrast to the three aforementioned systems, there are also a system based on the analysis of the links, and thus called "link analysis". In this approach, will used the topology of the Internet for relevance determination. So does not require the content itself to be checked for relevance, but the links of the web pages are analyzed. At the so-called topic-sensitive Page Rank algorithm will be different based on the links Websites and the topic itself, different categories and created a rating for websites. The personalized topic-sensitive-Page Rank algorithm is an extension of the general topic-sensitive PageRank. Here are users Profiles or their website preferences in the Review of websites included. Finally, another system should be mentioned, which not just fixed to one user. When so-called "social Search Engine "is the behavior of different User groups included in the rating for websites. This approach emerged from the observation of society, because, for example, book or movie purchases mostly due to Recommendations of others are performed. The "Social Search engine "becomes relative to the mentioned types more often used in the "Recommender

Systems" by Users who share a similar profile are considered together. The page rating is thus under the slogan "Das Collective has executed (mostly) right ". Finally, it can be said that today's search engines based on a mixture of the named species. Also It is noticeable that for a good search result a good profile of the User is required, which by the use of dynamic systems continues to improve and thus requires a continuous analysis of the user. In [2] you will find a deeper insight into this topic

### **2.3 Requirements of personalization**

To carry out a personalization of users, it is enough do not miss out, various data only in a large list too to save. You have to ask yourself various questions that specifically for the species and its destination. Therefore each application area has different requirements are present, only questions are listed in the case a search engine personalization to be answered. in the Following is mainly on the subject of personalization received by Websuchen and explains how this optimal is used. It starts with the question "How to get that Data? ", Respectively" Which method should one use Use storage? "Here, between direct and indirect data collection. As for the user is uncomfortable, with each system use a direct one To give feedback, the indirect form is usually used. Here the behavior of user and

system is observed and partly still with the possibility of optional direct Feedback added. For the web search offers a Indirect data collection very good, as a seamless Communication between user and system exists and Search offer a good basis for personalization. The next question is "How or where should you already created or yet to be created profiles are saved?" Again, there there are two options, the server-side and the client-side Storage. The benefits of client-side data storage lie in the increased privacy and security of the user. So private data remains on the user system and can can also be deleted here at any time. The server-side But data storage would be the evaluation of Better support user behavior patterns, as well plays the client system used (eg different types of Operating systems) does not matter. On the other hand creates a Server-side use a higher load on the server. Thus should be chosen a solution that both sides, client like also server, optimally exploited. It is important to note how often An exchange of information takes place between the two sides. A continuous connection is recommended because Information synchronized immediately on both sides and can thus be kept up to date. Another important question is "Should the system be over the To adjust the duration of

usage to the user? "At Web Search is asking this question with a definite "yes" answer, since Web searching is mostly based on topics, which are interesting for a person during this period. That I Interests but can change quickly, the possibility of the Be given adaptability. It is recommended to the user to allow itself an adaptation of the system. This form

### **3. EXPERIMENT AND SURVEY**

In the following chapters will be an experiment and a Survey with their results on the topic "The value of privacy in web searches" presented and compared.

#### **3.1 Preparation and Procedure of the Experiment**

##### *3.1.1 Implementation*

The goal of the experiment was to find out how much the Users of a search engine is worth the privacy. To were at different times at University College London (UCL) invited 189 test subjects. These were previously addressed within the university area, whether they on a test of a new search engine, which "Find Fever" called would participate. The intention of the behavior and to observe search settings was not explicit called, but in the course of the experiment was asked which Attitude the respective person has on privacy. After the test persons register and register they should answer

different questions, taking them could freely use the internet, but all

Search engine queries linked to the "Find Fever" page were. There were different topics that were discussed Part had a general character, on the other part queried more sensitive information.

**Table 1, credits for specific settings spend and win back for others**

Attitudes	Credits
Do not save my search history	-2
Remove ads on the results page.	-2
Do not share my searches with anyone.	-2
Improve the search result.	-1
Take the search query in the results out.	-1
Pair the search results with mine Location.	free
Use the secure search feature.	free
Save what results I click.	+2
Free Post my search on Twitter.	+2

### 3.1.2 Established hypotheses

Before the start of the experiment were other experiments considered. The experiments are described in [8] and [9]. It was found that privacy is already an important Part is in online stores. So, customers are ready to spend more money on To spend goods, if a more sensitive handling with their Data is given. Still there are gradations, how much more a product may cost

and in what way in the privacy is invaded. Thus, the results showed that as soon as the discount was tempting enough to privacy engaging web pages was changed. Furthermore was found that, for example, hardly spent more money when it comes to sending advertising to the Mail addresses of the users went. See also [10]. Also found out that customers, as soon as it is more sensitive Products or searches, their privacy want to protect more pronounced than in everyday life. How the Person behave exactly, was depending on country and person Although different, but there were tendencies noted become. From this foreknowledge in [1] these five hypotheses which it was to confirm or disprove: H1 - The price of privacy-enhancing features and the proportion of users enabling them are negatively associated. H2 - The more sensitive the search task, the more likely users want to enable privacy-enhancing features. H3 - The more sensitive the search task, the less likely users wants to enable privacy-invasive features. H4 - Users who are more concerned about privacy privacy-enhancing features more often. H5 - Users who consider privacy-enhancing features more important will enable them more often

### 3.3 Survey

To compare the experiment presented was one, adapted to the experiment, created an internet



survey, which itself mainly on the theses and attitudes of the Experiments oriented. This poll was sent to 19 people, distributed with different ages and circumstances, which they have been completed in full and returned. In the Survey, various questions were used to the Search settings. The question became partially so adapted that comparability to the experiment revealed. Nevertheless one could make own comments to the ask questions that are often used and improved the overall picture of the results. At the The beginning of the survey was based on the search engine used, use the "save history" feature and the requested setting options. You could indicate whether the options given in Table 3 be used or not. This question was in two parts split. In the first part it was assumed that the Use of the options available free of charge is provided. In the second part, however, one should for the Activation of the options to pay. On Example of a question is: "Would you like a no-ads- Use function? ", This had the answer option " Yes, I would use no-ads. ", or " no, or it I do not care.". The next part of the survey dealt with the topic of side financing, which options the Respondents would accept. The possible topics were: "Advertising on the results page", "Sponsored Links", "The Search would be posted on Twitter "and" The Search results that you click

will be saved."The conclusion was still after the personal attitude privacy and the frequency of use of Anonymization services and VPN connections are required. [3]

#### 4. SUMMARY

The term data protection or protection of personal data in the Internet in the media since a long time an integral part. One would therefore assume that a minimum level of anonymity and thus the obfuscating own personal data by default when using the Internet and, above all, would have to be present when searching the World Wide Web. Both after the experiment, as well as in the survey, presented out that the users of search engines, according to his own , Is looking very much on data protection. But once costs or incurred certain costs, the protection occurs personal data is often in the background. is also noticed that anonymous search engines rather unknown and are well-known companies such as Google and Microsoft (Bing), currently do not offer any possibility for the anonymization to pay its own data. Thus, there is for normal Internet users only option through the disclosure of his to obtain an accurate search result data being represented.

## 5. REFERENCES

- [1] Sören, Preibusch. 2013. The value of privacy in Web search. Microsoft Research Cambridge UK. The twelfth workshop on the economics of information security in WEIS, 2013.  
<http://weis2013.econinfosec.org/papers/PreibuschWEIS2013.pdf>. Last visited: 30.07.2014
- [2] K. Keenoy and M. Levene. Personalization of Web search. In B. Mobasher and SS Anand, editors, Intelligent Techniques for Web Personalization, Volume 3169 of Lecture Notes in Computer Science, pages 201-228. Springer, of 2005.  
[http://link.springer.com/chapter/10.1007%2F11577935\\_11](http://link.springer.com/chapter/10.1007%2F11577935_11).
- [3] Own survey in May 2014.
- [4] Judgment of the European Court of Justice in Case C- 131/12 of 13 May, 2014.  
<http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&pageindex=0&doclang=EN&mode=req&dir=&Occ=first&part=1&cid=258738>, last called: 06/10/2014
- [5] Googleformforremovalofsearchrequests. ,2014.  
[https://support.google.com/legal/contact/lr\\_eudpa?product=websearch&hl=en](https://support.google.com/legal/contact/lr_eudpa?product=websearch&hl=en). Last visited: 30.07.2014
- [6] Tor Project. July, 2014.  
<https://www.torproject.org/about/overview.html.en>. Last called: 07/30/2014
- [7] JonDonym. July 2014. <http://www.anonym-surfen.de/prinzip.html>. Last visited: 30.07.2014
- [8] Janice Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," in WEIS Of 2007.
- [9] Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch, "Unwillingness to pay for privacy: A field experiment," Economics Letters. vol. 117, no 1, pages 25-27, 2012th
- [10] Nicola Jentzsch, Sören Preibusch, and Andreas Harasser"Study on monetising privacy. An economic model for pricing personal information, "European Network and information Security Agency (ENISA), the 2012th
- [11] Privacy Policy Google. As of March 31, 2014.Heel: we collect information.  
<https://www.google.de/intl/de/policies/privacy/>. Lastcalled: 07/30/2014
- [12]: The Hotspot case. Issue 1/2012 c't. <http://heise.de/-1,394,646th> Last visited: 07.30.2014
- [13] Federal Privacy Act §3. Booth, 2014.  
<http://www.gesetze-im->



[internet.de/bdsg\\_1990/\\_3.html](http://internet.de/bdsg_1990/_3.html). Last visited:  
31.07.2014

[14] Article: Deleted and good? Alexander  
Dröbler.

<http://www.tagesschau.de/wirtschaft/google-urteil-100.html>. Last visited: 31.07.2014

[15] TUM lecture "IT security". Prof.  
Dr. Eckert, Dipl.-Inf. Thomas  
Kittel. Wintersemester 13/14

[16] TUM lecture "Network Security". Prof.  
Dr. Ing. Georg Carle, Dr. Heiko Niedermayer,  
Ralph Wood. winter semester 13/14

[17] Google Trends. June 2014.  
<https://www.google.de/trends/>. Last  
aufgerufen: 01/08/2014



**J. Haritha** is an Assistant professor in  
Department of Computer Science and  
Engineering at Balaji Institute of Technology &  
Science, Warangal and has 1+ years of  
experience in Academics. Her research  
activities are in the area of Mobile Computing,  
Data Mining and Cloud Computing etc.



**P. Sriveni** is an Assistant professor in  
Department of Computer Science and  
Engineering at Balaji Institute of Technology &  
Science, Warangal and has 1+ years of  
experience in Academics. Her research  
activities are in the area of Mobile Computing,  
Data Mining and Cloud Computing etc.