

Effects of Personalisation on Web Search

Author: Vikas Verma¹; Sunil Dhawan²;

Department of Computer Applications, Chandigarh Business School, Landran ¹

Research Scholar, Department of Computer Science, NIMS University, Jaipur²

ABSTRACT

Web searches or using search engines is one of the common tasks that has become an integral and important part of our daily routine. Personalisation of web has taken a giant leap in the past few years and today every search engine from google to yahoo, Bing or any customary web search engine uses some kind of personalisation. In this paper we have measured and analysed the effects of web personalisation on users searching for relevant information. Personalisation is both good and bad for users. It helps in providing the relevant information but sometimes users are simply not able to access information which is relevant to their context due to the filter bubble effect. We investigate the causes of various effects of personalisation. Our results or finding are a step towards understanding the effects and overall impact of personalisation on web searches.

General Terms

Search, Personalisation, Web Personalisation, Web Search Personalisation effects

Keywords

Effect / Causes of Web Personalisation, Impact Personalisation.

1. INTRODUCTION

Web Search industry has grown even its own expectations in the past decade or so. Starting with few keywords based searches today it has shifted to the realm of artificial intelligence based search where the computer tends to know what you want to search. Web Personalisation is one of the techniques that is used by companies like google for providing a good search experience for its users where the whole search system is centered on the user habits or what the user normally searches in the browser. One of the most used identities for personalisation are ip address and accounts of the users. The most common way of personalisation is to use these techniques to get the login information about the user and one can use information stored in cookies by analysing the cookie array. But the scale at which we are considering to analyse the personalisation is huge. The relevance or effects can be seen as very important as 17 billion requests is what google gets as queries just from U.S users [1]. The web search is not limited to a set of topics. It includes a process which provides relevant information relating to news, cricket, buying of products, ranking of players, items etc. The search results today have very

significant impacts. It can completely or partially effect businesses just by ranking certain results in a certain way. The impact can also be seen in foreign affairs of a country e.g. Google's conflict with Chinese [2]. This impact or effect is multiplied when we consider personalisation with page ranking system of search engines.

Different users may get different result based on identity or location. E.g. A person searching for course in PHP may get different result in NEW DELHI from a person who is using the same keywords for his search in MUMBAI. Personalisation provides many benefits to users, including disambiguation and retrieval of locally relevant results. Results for a PHP course in Bangalore may be irrelevant for a person residing in MUMBAI.

But all this personalisation has another dimension to it. It is known as the *Filter Bubble Effect* [3]. The users are at the mercy of the personalisation algorithm. The required or the correct results may be different from what is being provided by the search engine. This effect is exacerbated by the dual issues that most users do not know that search results are personalised or modified, yet users place blind faith in the results ranked or provided by search engine results [4].

In this paper we will find out the effect and analyse their impacts on people and business because ultimately it's people's money and life at stake.

2. Literature Review

For correctly understanding the impacts and effects of the personalization, one has to first understand what it is and for this purpose a literature review was done. The techniques used for personalising search concluded that mining click histories leads to the most accurate results and provides a comprehensive overview of the methods and techniques [5]. The features besides click history have been used to power personalised search. Two studies have shown that user demographics can be reliably inferred from growing histories, which can be useful for personalising content [6, 7]. The paper on investigated privacy-preserving personalised search [8] provides and deals with the filter bubble effect. Two studies refer to personalisation based on the advertisements [9,10] on web especially ad personalisation in Google's own browser Google Chrome which provides the personalised ad services based on user's browsing data and it combines the

data from multiple devices to this. Another study refers to different e-com portals where price personalisation is conducted [11].

In literature review we have also studied personalisation's in different search engines using comparison studies. Two studies have performed user studies to compare search engines [12, 13]. These studies provide us in-depth knowledge regarding competing search engines but no study effectively examines the impact of personalisation.

3. Personalisation in Web Search

Personalisation in web search is synonymous with the google and yahoo. Why? It is because of the fact that most of the users use these services to find out the relevant links for information. Even if the know the direct URL of the needed resource people use these search engines to do the job. In order to know and analyse the effects of personalisation we need to find out how it started and other particulars of history relating to them.

3.1. History of Personalisation at Google

Google started with Personalisation or Personalised search in 2004 and merged with Google search in 2005[12, 13]. Google was so impressed with the results that in 2009 it started doing personalisation for users who weren't google account holders. No one can surely say how google does it personalised search but the information from google blog states that Google Search uses user's language, geolocation, history of search, queries and google+ posts to do personalisation[14]. Users get results on the basis or the order of results depend upon the combination of temporal order of searches and whether user click on the results. Use of social data can be there but there is no concrete evidence that how google uses it.

3.2. Use of Google Accounts

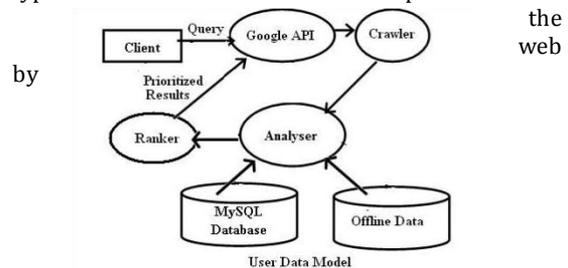
Google started as a search engine and a website which use to provide results to the users queries based on page ranking algorithms. Later on they transformed this into a service. Today Google's portfolio has a number of services like google maps, search, news, mail, calendar etc. Google accounts help google build a single point account management architecture. Google Accounts are the single pony of login for all the google services. You login into one service, you will automatically be logged into all the set of services. A tracking cookie enables all of Google's services to uniquely identify each looted in user. Google uses the data or information of one service to personalise experience in other services commonly known as service information sharing [15].

3.3 User Tracking and Advertising

Google tacks its users as they browse the Web due to their large advertising networks. A study shows that how google uses cookies from tools like DoubleClick, Google Analytics, and YouTube to track users browsing habits [16].

3.3 Google Chrome

Google Chrome uses an integrated approach for storing the cookie and personalisation information. Chrome uses SQLite database and history contains a table of unique URL, and a table for each unique visit called visit calledvisits [17].There are numerous visit types which can be used to trace the path of user on the web



referring URLs and visit types [18]. For the purpose of this research for analysing the effects of personalisation at browser level we use Google Chrome because it's a cloud enabled browser and can take into account the browsing history of many devices which can be used by account history of the users. It is the most advanced browser and the web activity is analysed by fetching the data from database of SQLite using the default APIs of Google Chrome. User History, Bookmarks, Downloads, Cookies and Autosuggestion can be used for extracting or analysing the online behaviour. A user model is followed by us to extract or analyse the behaviour that is given below.

4. Effects of Personalisation

The effects of personalisation can be seen as a one of the attributes that can be used to know or ascertain the need or value of application of personalisation. This research paper deals with Google's search engine and we have worked to know how it affects the end users. Another question that we have tried to answer is to what extent does search personalisation actually affect search results? What are the effects of personalisation on web survey and response rates? Is the data quality affected by personalisation algorithms?

We might have used google search engine as it is the most popular search engine but this methodology can easily be applied to any other search engine as well.

4.1 Measurement Methodology

In order to answer the questions put in the previous section we formulated a methodology for the measurement of effects. We divided our tests into two parts.

1. Queries using google accounts
2. Queries without using google accounts

This was necessary because there is a major difference between a user who is logged in Means you have the browsing history and other data factors available to you rather than a first time user who is just searching for results and we don't know anything about that user. We created 10 google accounts that vary by one or more features. E.g. one

can be male another can be female, age difference etc. Then we executed similar queries from these accounts for 10 days. 1 time each day. Then we compared the results. Following are the effects that we measured and analysed.

4.2 Measured Effects

- * **The Carry over Effect:** This means the effect of one search on the coming or subsequent searches. In other words if a person has searched for India and then searched for Marble. This effects it as the results provided by search engine tends to provide the marble with reference to India. A study conducted shows that sequential searches by a user can be used to refine search results and as no surprise google does the same [19, 20]. The carry over effect can be clearly observed. For continuous 10 days the experiments conducted show that the carry over effect remains for 10 minutes. If some person searches subsequent queries after a gap of 10 minutes the carry over effect is no longer present there. So in order to negate this effect the user must search after a gap of 10~15 minutes. This effect is applicable to both users with accounts and users who are not account holders or not lodged in.
- * **Search Index Effect:** This is another effect that can be used to see why we get different results at different time. Web search services such as google constantly update their search indices. Thus the relevant search query information changes or is constantly evolving so we are able to get different results at different times.
- * **Geo Location Effect:** To personalise search companies like google use the user's IP address to provide localised results. It help them deliver the content based on location from where the user is queried the results. E.g. somebody wants to shop for grocery items. Google uses the ip address of the user to determine the location of the user and using that location google delivers the result containing the addresses of grocery shops in or near that location. This effect can be clearly observed when you do the same query from different ip addresses or location. We used a proxy server to conduct this test and it is very clear that ip addresses are used for personalisation.
- * **Localised Datacenter Effect:** Web search services like google, bing, yahoo have datacenters that are spread across different locations. The anomalies in the search indexes at different locations are the primary reason why search results can be found different if the results are provided by two different datacenters of one company. Thus the personalisation effect can also vary in accordance with the datacenter being used.
- * **Anonymity Effect:** Anonymity is one of the biggest concerns which comes as a byproduct of personalisation. Personalisation uses information provided unknowingly by users and without user's intent to share the information or data is used to provide a personalised experience. Anonymity means that you are using an application or service

as an anonymous user that there is no trace of the user's identity. Personalisation techniques such as sequential data or query analysis, ip address or geo location based methods don't ask for the information. This information is normally concealed and collected from the web browser or user without the acknowledgement of the user. Thus even when you have not logged into the website, you are still being provided a personalised experience by the website using the concealed data. Thus a person cannot have the full anonymity based experience of web search.

- * **Privacy Effect:** Another matter of concern that comes with the application of personalisation it hampers or it is even build on using the private data for personalisation. Personalisation means we are personalising the experience for a particular user and we are not doing it for the complete system. Thus it involves knowing many things or attributes about a user. Today companies have increasing amounts of data about users gathered through various techniques. Personalisation involves or uses this data for online interactions. It drastically improves the relationship with the customer and increases desirable behaviours and increased purchase intent. But personalisation has many negative effects. The increase in perceived information control reduces the privacy concern by providing trusted info to the user.

5. Levels of Effects: Using the measured data we were able to find relevant information that how many effects occur no. of times and thus using the occurrences or the no. of times the information relating to a particular effect is used we devised 5 levels. Where level 5 is the most significant effect that can be seen as the most widely used one and level 1 being the least used. This is using the test cases only and we cannot say it is applicable for all the processes in all the web search engines. Some might employ different criteria to use this information. Using the information gathered over a period of 10 days we calculated and analysed that the carry over effect is the one which is the most important and most widely used. It is being used for presenting the results in a way that if somebody searches for a particular term now and within a predefined time if another term or query is searched that must have some relation with the previous query. Search Index Effect is the one that effects the personalised data in a big way. Search Index use updated data and this data is being updated every now and then. So it changes everything because page and access hits change over a period time. Geo location effect and localised datacenter effect have low relevance as these are used in a passive manner and cannot be used as primary data for personalisation. Anonymity and Privacy effects have very high role-play in the overall personalisation system. The anonymity and privacy concerns can really effect the personalisation as a whole in a search system because it has direct relationship with the

amount of data being used for analysis. The higher the amount of private data being used lower would be the privacy level and similarly higher will be personalisation. On the similar terms we have anonymity. If non self-disclosed data is used as a measure for providing personalisation it cuts down on the anonymity factor. In table 1 we have summarised the effects that were found and these are levelled according to their use and relevance.

Table 1. Table contains the different effects of personalisation

SrNo.	Effect	Level
1	Carry over Effect	5
2	Search Index Effect	5
3	Geo Location Effect	2
4	Localised Datacenter Effect	1
5	Anonymity Effect	4
6	Privacy Effect	4

6. Conclusion

The aim of this research was to explore the potential effects of personalisation on web search. We find out there are different levels of effects that can be used as base to analyse their use in different conditions and search engines. Anonymity and privacy concerns can be reduced using trust as a major factor as it allows the user to feel at home. The privacy concerns can be dealt by providing trustworthy information to the customers. As long as the user has faith in the search system all the effects can be negated and data can be used. But it all depends how the company or search system is using that data and for what purpose? If it is for influencing the buyer in buying something that is not actually right then the customer will eventually distrust your search system. Thus it becomes necessary for the organisation to build a bridge between privacy and trust so that the effect mentioned here can be positively used and can act as a mechanism for building trust.

7. Future Scope

In future we would work over relationship of privacy and behaviour intents. The study will examine the advanced and long term effects of using personalisation methods on referenced links by

search engines. We would also use the relevant metrics for measuring their effect.

8. References

1. comScore August 2012 U.S. Search Results.<http://bit.ly/ThGnOc>.
2. M. Wines. Google to Alert Users to Chinese Censorship.The New York Times, 2012. <http://nyti.ms/IRhGZS>.
3. H. Green. Breaking Out of Your Internet Filter Bubble.Forbes,2011.
4. B. Pan, H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, and L. Granka. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. J. Comp. Med. Comm.
5. Z. Dou, R. Song, and J.-R. Wen. A Large-scale Evaluation and Analysis of Personalized Search Strategies. WWW, 2007.12(3), 2007
6. S. Goel, J. M. Hofman, and M. I. Sircer. Who Does What on the Web: A Large-scale Study of Browsing Behavior. ICWSM, 2012.
7. J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen. Demographic Prediction Based on User's Browsing Behavior. WWW, 2007.
8. Y. Xu, B. Zhang, Z. Chen, and K. Wang. Privacy-Enhancing Personalized Web Search. WWW, 2007.
9. S. Guha, B. Cheng, and P. Francis. Challenges in Measuring Online Advertising Systems. IMC, 2010.
10. C. E. Wills and C. Tatar. Understanding What They Do with What They Know. WPES, 2012.
11. J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting Price and Search Discrimination on the Internet. HotNets, 2012.
12. Personalized Search Graduates from Google Labs. News From Google Blog, 2005. <http://bit.ly/Tndpgf>.
13. M. Hines. Google Takes Searching Personally. CNet, 2004. <http://cnet.co/V37pZD>.
14. A. Singhal. Some Thoughts on Personalisation. Google Inside Search Blog, 2011.
15. A. Witten. Google's New Privacy Policy. Google Official Blog, 2012. <http://bit.ly/wVr4mF>.
16. F. Roesner, T. Kohno, and D. Wetherall. Detecting and Defending Against Third-Party Tracking on the Web. NSDI, 2012.
17. Sandeep Kumar Khanikar, Web Forensics, The Department of Computing Sciences Texas A&M University-Corpus Christi, Corpus Christi, Texas, Fall 2010. <http://scitamucc.edu/~cams/projects/345.pdf>
18. How Rick Santorum's 'Google Problem' Has Endured. <http://n.pr/wefdnc>.
19. Z. Cheng, B. Gao, and T.-Y. Liu. Actively Predicting Diverse Search Intent from User Browsing Behaviors. WWW, 2010.
20. Y. Shen, J. Yan, S. Yan, L. Ji, N. Liu, and Z. Chen. Sparse Hidden-Dynamics Conditional Random Fields for User Intent Understanding. WWW, 2011.