

Survey on Profile Creation and Privacy in Personalized Web Search

Seetha Lekshmi B., Annie George

Computer Science and Engineering, Sree Buddha College Of Engineering For Women ,

Elavumthitta, Pathanamthitta

seethathoppil@gmail.com, annieasg@gmail.com

Abstract : *The web search engine in an important source for looking information on the web. Amount of information in search engine increases day by day. But in sometimes result obtained through search engine is irrelevant because it does not meet user's intensions. Personalized Web Search (PWS) overcome this problem by providing better search result. PWS is an effective way to improve the search engine quality by providing customization in search result. This paper focus on the survey of profile creation and privacy protection in PWS established by different authors.*

Keywords : Personalized Web Search, implicit profile, explicit profile.

I. Introduction

Data mining is a process designed for extracting information from a large dataset. It is a computational process of extracting patterns from huge dataset such as big data etc. So data mining is a knowledge discovery from databases. One of the main dataset used for mining information's is World Wide Web (WWW). The main function of internet is collection and sharing of information. Therefore WWW is a main source for extracting information. Due to the wide usage of internet users depend on different website for getting information. Extraction of information from web pages is through different search engines. Google, Yahoo etc. are example for such search engine. But search engines works based on "one information fits all" constraints. The results given by the search engine does not consider the user's intension. Different users have different interest, but search engine doesn't consider the intension of user and it retrieves same result for same query given by different users. For example if the query is java, then java is an island, a programming language etc. So to avoid this problem personalization is added into the search engine. Thereby quality of search engine is improved by giving better results.

PWS is a search engine used for obtaining better results. It considers the user's query and analyzes the user's information, based on that result provided. So PWS improve the search quality by providing useful information. Factors affecting on personalization web search are divided into two: (1) spatial factors such as queries, visited pages, browsing history etc. and (2) temporal factors such as time taken for page view, query usage etc. [iii]. PWS are of two types: Click log based and profile based [i]. Click log based method directly give result based on the user's query history. In Profile based method, result is based on user's interested area generated from user profile. So search engine quality is improved. Here information is collected implicitly from user's query history,

documents, bookmarks etc. It is very effective for all type of queries but in some condition it is not stable.

Both PWS technique has advantages and disadvantages. In comparing them profile based technique is better because it improve the search engine quality by providing good results.

II. Literature Survey

Currently user provides the feedback and builds the user profile by selecting their interested area. Based on the feedback, profile will created explicitly. But user will not spend more time to specify their interested area. To overcome this Jaime Teevan, Susan T. Dumais, Eric Horvitz [iv] propose a method for creating the profile implicitly. Implicit profile construction based on user query history, web pages visited, emails etc. So user need not maintain their interested area and web page ranking is giving by using BM25. It is a well-known probalistic weighting scheme. This method is easy to implement in client machine.

Mirco Speretta, Susan Gauch [v] implements a wrapper for Google. Wrapper is used for collecting information from user profile and creates the profile implicitly. Here profile shown as a weighted concept taxonomy. Thereby it avoids the explicit profile creation and use proxy server for capturing the information. Wrapper is used for creating user profile and performs classification based on key concepts. Advantage of using wrapper is collection of user interest in a non-invasive manner, search based on the data in search engine and system performance based on user activity rather than feedback given by user. But profile hierarchical is created statically.

Amount of information in internet increases rapidly so searching is very difficult. After giving a query search engine return tremendous amount of resultant page, from this only few of them are relevant to the user. This is due to short query and user doesn't specify their intension. A. Pretschner and S. Gauch [vi] suggest that relevant web pages are developed by re-ranking and filtering instead of feedback. Re-ranking means modifying the rank return by web pages and filtering means excluding some documents based on ranking function.

Requirement for a personalized web search is building the user profile based on user interest. Profile creation is in a hierarchical format. Two representation are commonly used for user profile, 1) use the frequently occurred word in user document, 2) using an ontology called DMOZ [vii]. Krishnan Ramanathan, Julien Giraudi, Ajay Gupta [vii] propose a method for creating user profile by using Wikipedia as vocabulary for specifying interested area. It consists of 3 steps. (1) Mapping web page to Wikipedia concept, (2) Hierarchical

profile created from Wikipedia concept, (3) Tagging the profile concepts in two ways, 1st tag describes transactional and recreational concepts and 2nd tag measures users' interest in the concepts.

Zhicheng Dou, Ruihua Song and JiRong Wen [ii] propose a similarity checking between user interest and web search. Here user profile created as a hierarchical tree and represented as a weighted topic. So user doesn't need to give feedback, instead of that it is automatically learned from browsing history. Thus profile created implicitly.

Now profile creation in PWS is done by implicitly based on user interest, preference, user document, browsing information. But this implicit profile creation leads to privacy problem in PWS, i.e. takes the private information about the user. So user reluctant to disclose their private data into a search engine. This leads to consideration of privacy in PWS.

Xuehua Shen, Bin Tan, ChengXiang Zhai [viii] concerns about privacy in PWS. There are different types of users, some may not will to disclose their personal data but some share their information. So level of privacy is tuned based on the user preference. Here explain four level of privacy protection. In level 1 is Pseudo Identity which is a mapping of user ID to pseudo identity. It gives low level of protection. Level 2 is Group Identity, where group of users share same user identity. Here need a proxy for group of users. By comparing level 1 and 2, level 2 give higher protection than level 1. Level 3 is No Identity, so privacy protection is through anonymous network. It is better than level 2. Level 4 is No Personal Information, it does not given ID and user description. Search engine return only normal search result to the user. So level 4 give better protection than other levels.

S. Sendhilkumar and T. V. Geetha [iii] explains about the semantic search for improving search result, because current web search not consider relevant page unvisited by the user. So for avoiding that problem result set provide a link to the relevant page. This is done by identifying relevant page and recommended it to the user if page is unvisited. Next is identifying shortest path and highlight it to the user. Thus search result will improve.

III. Conclusion

Personalized web search is an effective method for improving the quality of search engine. Due to the enormous increase of information in internet, search engine doesn't provide good result in some situation. So personalization is applied to the search engine for getting better result. In this paper a survey is done on how user profile is created from explicitly to implicitly. Also it focuses on privacy consideration in PWS.

References

- i. Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection in Personalized Web Search"
- ii. Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," *Proc. Int'l Conf. World Wide Web (WWW)*, pp. 581-590, 2007..
- iii. S. Sendhilkumar and T. V. Geetha, "Concept based Personalized Web Search"
- iv. Jaime Teevan, Susan T. Dumais, Eric Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities"
- v. M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," *Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI)*, 2005.
- vi. A. Pretschner and S. Gauch, "Ontology-Based Personalized Search and Browsing," *Proc. IEEE 11th Int'l Conf. Tools with Artificial Intelligence (ICTAI '99)*, 1999.
- vii. Krishnan Ramanathan, Julien Giraudi, Ajay Gupta, "Creating hierarchical user profiles using Wikipedia"
- viii. Xuehua Shen, Bin Tan, ChengXiang Zhai, "Privacy Protection in Personalized Search"
- ix. Xuehua Shen, Bin Tan, ChengXiang Zhai, "Implicit User Modeling for Personalized Search"
- x. Shraddha Tare, Prof. Sachin Chavan, "An Advanced Approach for Privacy Protection in Web Based Personalized User Search"