# Multi Keywords with Ranked Search Over Encrypted Data in Clouds

Muppala Swathi[1], P. Vijayaraghavulu[2]
[1]*PG Student, Dept of CSE, Sri Annamacharya Institute of Technology and Science, Rajampet, Kadapa*
[2]*Asst.Professor, Dept of CSE, Sri Annamacharya Institute of Technology and Science, Rajampet, Kadapa*

***Abstract*** **- With increasing number of websites the Web users are increased with the massive amount of data available on the internet which is provided by the Web Search Engine (WSE). The aim of the WSE is to provide the relevant search result to the user with the behavior of the user click where they performed. WSE provide the relevant result on behalf of the user frequent click-based method. From this method no assurance to the user privacy and also no securities were providing to their data. Hence users were afraid for their private information during search has become a major barrier. They were many techniques were proposed by researchers most of that based on the server side, it has provided less security. For minimizing the privacy risk here, we propose the client-side based technique with the combination of Bloom filters method to prevent the user data that we applied in Knowledge mining area.**

***Index Terms*** **- Web Search Engine, personalized search, user query logs, content search and privacy preserving.**

## I.INTRODUCTION

Web search engines are very important in web life. Web search engines are built for all users and not specified for any individual user. Generic web search engines cannot identify the different needs of different users, if user enter improper keyword or ambiguous keywords and lack of user's ability to express what they want are some challenges faced by generic web search engines. To address this issue, we should personalize these results. As it is becoming an important aspect, to provide such environments, different techniques and approaches have developed. But at the same time security of personalized web searches has also gained significance, in which the user's personal or private information cannot be disclosed through web searches.

User's hesitation to disclose their private information during search has become major issue on personalization technologies. For example, system that are personalize some advertisements according to physical location of user or their search history, introduces new privacy challenges that may discourage the wide adoption of personalization technologies. Personalized web search is proving its effectiveness but also raising matter of privacy and securing personal information. Many personalization methods have been exposed and been in practice. But it is not sure that those methods will make sure their efficiency in dissimilar queries for different users. The solutions to PWS can generally be categorized into two types, namely click-log-based methods, and profile-based ones. The click-log based methods are straightforward; they simply impose bias to clicked pages in the user's query history. Although this strategy has been demonstrated to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be potentially effective for almost all sorts of queries but are reported to be unstable under some circumstances. The two contradicting effects [4] during the search process to be considered. Improve the search quality with the personalization utility of the user profile and the need to hide the privacy contents existing in the user profile to place the privacy risk under control. This survey investigates the several privacy preserving techniques and provides idea about the new efficient method in the future. The main goal of this work is to assure the privacy guarantee to the user who is involved in the personalized web search.

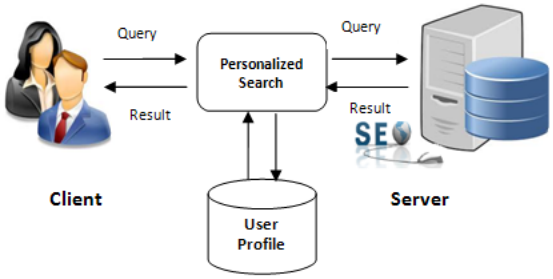



Fig 1: Personalized Search Engine Architecture

By these methods personal data were easily reveal. While many search engines take advantage of information about people in common, or regarding particular groups of people, personalized search based on a user profile that is unique to the individual person. Research systems that personalize search outcomes model their users in different ways. The Personalized Web Search provides a unique opportunity to consolidate and scrutinize the work from industrial labs on personalizing web search using user logged search behavior context. It presents a fully anonymized dataset, which has anonymized user id, queries based on the keywords, their terms of query, providing URLs, domain of URL and the user clicks. This dispute and the shared dataset will enable a whole new set of researchers to study the problem of personalizing web search experience. It decreases the likelihood of finding new information by biasing search results towards what the user has already found. By using these methods privacy of the user might be loss because of clicking the relevant search, frequently visited sites and providing their personal information like their name, address, etc. in this case their privacy might be leak. For this privacy issue, many existing work proposed a potential privacy problems in which a user may not be aware that their search results are personalized for them [6, 7].

## II. RELATED WORK

There are mainly two types of personalized web search they are Click-log-based and Profile-based personalized web search. Before 2000, there was hardly any work aimed to provide a solution for searching on encrypted data. In 2000, D. Song, D. Wagner and A. Perrig proposed the different techniques for searching operation over encrypted data [4].

These techniques for remote searching on encrypted data were provided with security proofs and have a number of crucial advantages. All these techniques were based on Boolean keyword search. Boolean keyword search is not suitable for cloud storage since it sends all matching files to the clients, and therefore incur a larger amount of network traffic and a heavier post-processing overhead for the mobile devices. TF-IDF is a statistic which reflects how important a word is to a document in a collection [5]. Y. Chang and M. Mitzenmacher provided keyword search scheme, but it does not send back the most relevant files [6]. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu proposed a one-to-one mapping OPE which will lead to Statistics Information Leak Control [3]. A. Swaminathan, Y. Mao, G. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard proposed a confidentiality-preserving rank-ordered search [7]. This scheme displays low performances as the relevance scores are computed on the client side, increasing its workload. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou presented a secure ranked keyword search over encrypted cloud data [8]. However, in their work the terms are closely related to the files which could lead to potential information leak. In 2015, Jian Li, Ruhui Ma, Haibing Guan proposed TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud [9]. TEES architecture was introduced to create a traffic and energy efficient encrypted keyword search tool over mobile cloud storages. It offloaded the relevance scores calculation to the cloud server reducing the burden on the mobile clients. It also shortened the retrieval process so that the data user can receive the most relevant files within only one communication. However, TEES architecture uses Single keyword search thus yielding far too coarse results. To improve the search result accuracy as well as to enhance the user searching experience, it is necessary to support multiple keyword searches to narrow down the results. Multi-keyword is potentially the future mainstream encrypted search scheme with higher searching accuracy. Table 1 compares all the previous techniques and mentions their shortcomings.

### A. Click-Log-Based Method

Here, personalization is carried out on the basis of clicks made by user. The data recorded through clicks in query logs, simulates user experience. The web pages frequently clicked by user in past for a particular query is recorded in the history and score is computed for particular web page and based on that web search

results are provided. This method will perform consistent and considerably well when it is works on frequent queries. When a never asked query is entered by user; it will not provide any precise search results, which is the main drawback of this method.

B. Profile Based Personalization

The basic idea of these works is to tailor the search results by referring to a user profile, implicitly or explicitly which reveals an individual information goal. Many profile representations are available in the literature to facilitate different personalization techniques.

- Lists / vectors or bag of words: Earlier techniques utilize term lists/vectors or bag of words to represent their profile. It is the simple representation in information retrieval system. Here a text is represented as the bag of its words, disregarding grammar and even word order [3]. But it keeps multiplicity of those words. In each vector the second entry will be the count of that word.
- Hierarchical representation: Most recent works build user profiles in hierarchical structures. The reason is their stronger descriptive ability, better scalability, and higher access efficiency. Majority of the hierarchical representations are constructed with existing weighted topic hierarchy/graph, such as ODP, Wikipedia, and DMOZ and so on. Using the term-frequency analysis on the user data, the hierarchical profile can be built automatically also.

## III. PROPOSED WORK

There are two classes of privacy protection problems for PWS in general. One class includes those works, treat privacy as the identification of an individual. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server.

A. Identification of An Individual

Typical works in the literature of protecting user dentifications (class one) try to solve the privacy problem on different levels, including the pseudo-identity, the group identity, no identity, and no personal information [13]. Solution to the first level is proved fragile. The third and fourth levels are impractical due to high cost in communication and cryptography. So, the existing efforts focus on the second level.

- Online anonymity: It works based on user profiles by generating a group profile of k users. Using this approach, the linkage between the query and a single user is broken.
- Useless user profile (UUP): This protocol is proposed to shuffle queries among a group of users who issue them. As a result, any entity cannot profile a certain individual. These works assume the existence of a trustworthy third-party anonymizer, which is not readily available over the Internet all the time in large number.
- Legacy social networks: Instead of the third party to provide a distorted user profile to the web search engine, here every user act as a search agency of his/her neighbors. They can decide to submit the query on behalf of who issued it or forward it to other neighbors.

B. Sensitivity of Data

The solutions in class two do not require third-party assistance or collaborations between social network entries. In these solutions, users only trust themselves and cannot tolerate the exposure of their complete profiles to an anonymity server.

- Statistical Techniques: To learn a probabilistic model, and then use this model to generate the near-optimal partial profile. One main limitation in this work is that it builds the user profile as a finite set of attributes, and the probabilistic model is trained through predefined frequent queries. These assumptions are impractical in the context of PWS.
- Generalized Profiles: Proposed a privacy protection solution for PWS based on hierarchical profiles. Using a userspecified threshold, a generalized profile is obtained in effect as a rooted sub tree of the complete profile.

C. Issues

The shortcomings of current solutions in class one is the high cost introduced due to the collaboration and communication. The statistical methods build the user profile as a finite set of attributes, and the probabilistic model is trained through predefined frequent queries in class two. These assumptions are impractical in the

context of PWS and the generalized profile does not address the query utility, which is crucial for the service quality of PWS.

## IV METHODOLOGY

Indeed, the privacy concern is one of the major barriers in deploying serious personalized search applications, and how to attain personalized search though preserving users' privacy. Here we propose a client-side personalization which deals with the preserving privacy and envision possible future strategies to fully protect user privacy. For privacy, we introduce our approach to digitalized multimedia content based on user profile information. For this, two main methods were developed:

Automatic creation of user profiles based on our profile generator mechanism and on the other hand recommendation system based on the content to estimates the user interest based on our client side meta data.

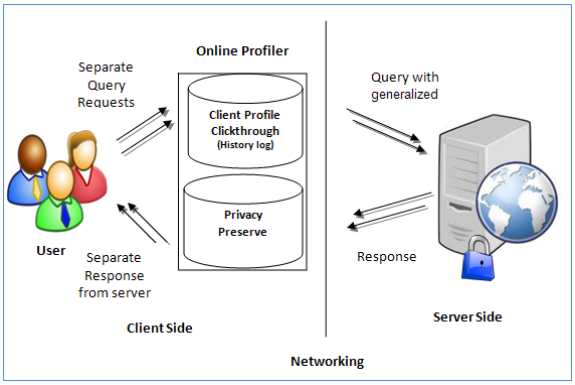



Fig 2: Proposed Architecture

Above figure shows our proposed architecture which is builds in the client-side mechanism and here we protect the data from the server, so only we provide a privacy to the client user.

Every query from the client user were provided by the separate requests to the server, this hides the frequent click through logs or content-based mechanism, from this user can protect the data from the server. In the same case our mechanism maintains the online profiler about the user hence it hides the click logs and provides a safeguard to the user data. After that, online profiler query was processed in the manner of generalization process, it is used to meet the specific prerequisites to handle the user profile and it is based on the preprocessing the user profiles. Our architecture, not only the user's search performance but also their background activities (e.g., viewed before) and personal information (e.g., emails, browser bookmarks) could be included into the user profile, permitting for the structure of a much richer user model for personalization.

The sensitive contextual information is usually not a main aspect since it is strictly stored and used on the client side. A user's personal information including user queries and click logs history resides on the user's personal computer and is exploited to better suppose the user' information requires and provide a relevant search result.

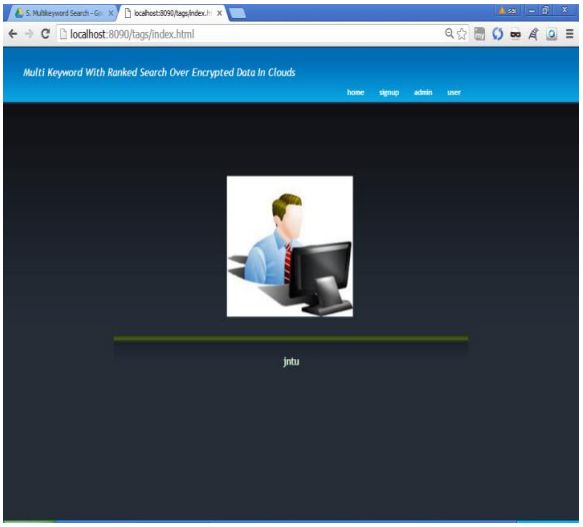
Our proposed algorithm uses the bloom filters method based on the discriminating power and information loss protection to inherit the relations. Here it uses the inherited method to generalize the query.

It allows performing the customization process to protect the data and use the User customizable Privacy-preserving Search framework addressed the privacy problems. This aims at protecting the privacy in individual user profiles.

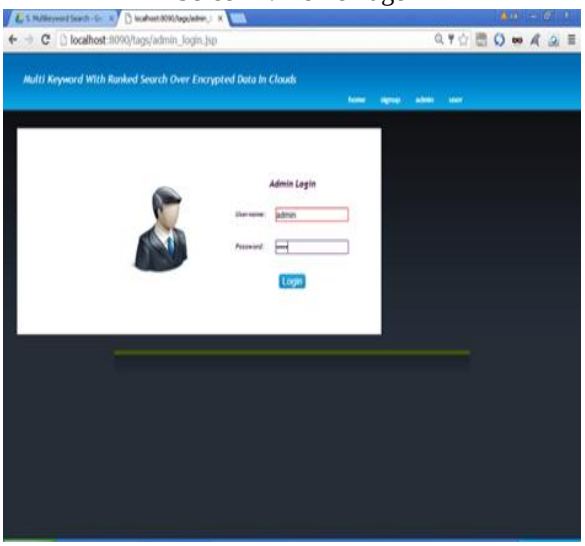
Web users were increases because of available of information's from the web browser based on the search engine. With the increasing number of user service engine must provide the relevant search result based on their behavior or based on the user performance. Providing relevant result to the user is based on their click logs, query histories, bookmarks, by this privacy of the user might be loss. For providing relevant search by using these approaches the privacy of the user may loss. Most existing system provides a major barrier to the private information during user search. That approaches do not protect privacy issues and rising information loss for the user data. For this issue, this paper proposes client-based architecture based on the bloom filters algorithm to prevent the user data and provide the relevant search result to the user in future it can include this work in mobile application.

## IV. RESULT AND DISCUSSIONS

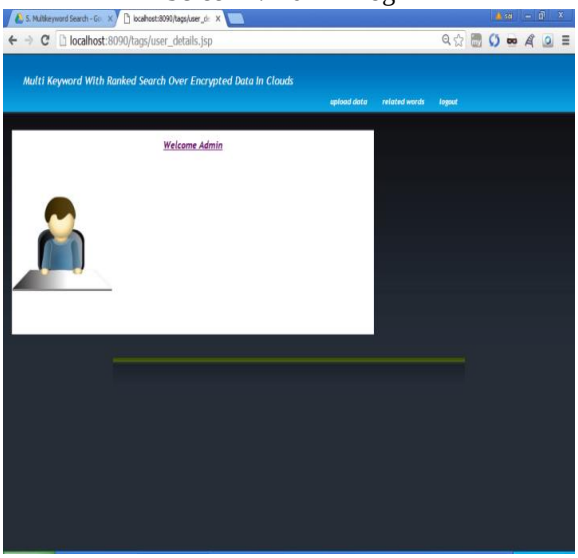
Document Summarization This graph shows efficiency of proposed system is better than existing system. The result shows that retrieval rations in millisecond. Here compare three different algorithms from which Sum all text algorithm required minimum retrieval ratio.



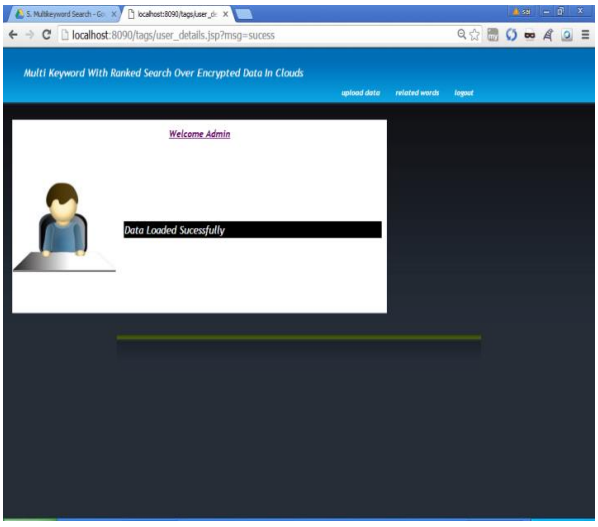

Screen 1: Home Page



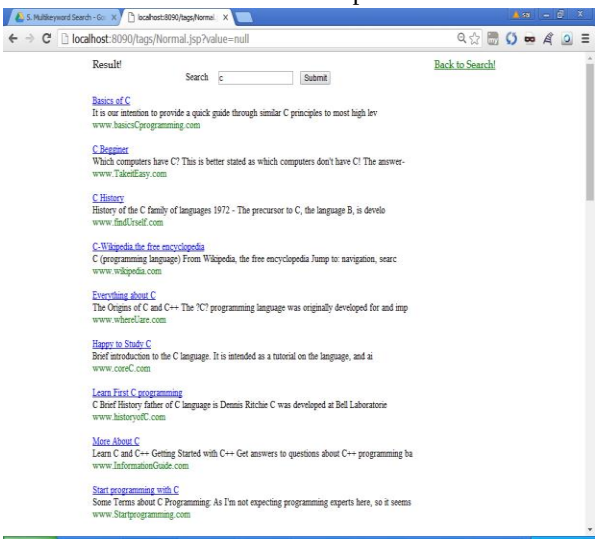

Screen 2: Admin Login



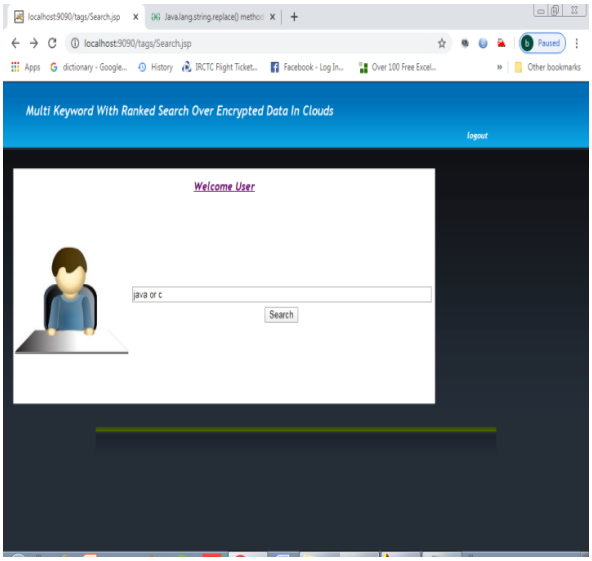

Screen 3: Server Menu



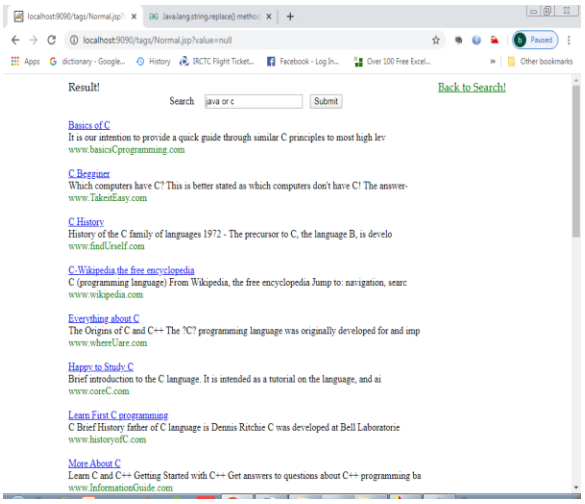

Screen 4: Data Upload




Screen 5: User Search




Screen 6: Multi Keyword Search




Screen 7: Search Result

## IV CONCLUSION

This paper provides a review on personalized web search and the related security concepts. The PWS techniques are developed remarkably in the last decades. A variety of techniques have emerged to increase search effectiveness and to protect privacy using multiple algorithms. Different methods conclude that privacy preservation is not handled well. UPS framework, which is proposed to provide privacy for each user, uses the online profiler to take online decision on whether to personalize a query or not. This framework can significantly reduce the risk of attack and performs better as compared to others. The main goal of this work is to assure the privacy guarantee to the user who is involved in the personalized web search.

## REFERENCE

[1] D. Huang, “Mobile cloud computing,” IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter, 2011.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” the Journal of machine Learning research, vol. 3, 2003, pp. 993–1022.

[3] R. Agrawal, J. Kiernan, R. Srikant and Y. Xu, “Order preserving encryption for numeric data,” in Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM, 2004, pp. 563-574.

[4] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.

[5] A. Aizawa, “An Information-theoretc perspective of tf-idf measures,” Information Processing and Management, 2003, vol. 39, pp. 45-65.

[6] Y. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in Applied Cryptography and Network Security. Springer, 2005, pp. 391-421.

[7] A. Swaminathan, Y. Mao, G. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, “Confidentiality-preserving rank-ordered search,” in Proceedings of the 2007 ACM workshop on Storage security and survivability. ACM 2007, pp. 7-12.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou,” Secure ranked keyword search over encrypted cloud data,” in Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on. IEEE, 2010, pp. 253-262.

[9] Jian Li, Haibing Guan, Ruhui Ma, “TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud,” in IEEE Transactions on Cloud Computing, 2015.

[10] A. Schulman, T. Schmid, P. Dutta, and N. Spring, “Demo: Phone power monitoring with battor.” MobiCom, 2011.

[11] Shen, Xuehua, Bin Tan, and Cheng Xiang Zhai. "Implicit user modeling for personalized search." Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005.

[12] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, “Accurately Interpreting Clickthrough Data as Implicit Feedback,” Proc. 28th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR ’05), pp. 154-161, 2005.

[13] Shen, Xuehua, Bin Tan, and Cheng Xiang Zhai. "Context-sensitive information retrieval using implicit feedback." Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005.

[14] Xu, Yabo, et al. "Online anonymity for personalized web services." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.

[15] A. Viejo and J. Castell_a-Roca, “Using Social Networks to Distort Users’ Profiles Generated by Web Search Engines,” Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.
[16] Xu, Yabo, et al. "Privacy-enhancing personalized web search." Proceedings of the 16th international conference on World Wide Web. ACM, 2007
[17] Xiao, Xiaokui, and Yufei Tao. "Personalized privacy preservation", Proceedings of the 2006 ACM SIGMOD international conference on Management of data. ACM, -2006.
[18] Shou, Lidan, et al. "Supporting Privacy Protection in Personalized Web Search." (2012): 1-1.
[19] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, “Ups: Efficient Privacy Protection in Personalized Web Search,” Proc. 34th Int’l ACM SIGIR Conf. Research and Development in Information, pp. 615- 624, 2011.