Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Advanced Authentication Procedures for User Profiles Based on Biased Ranking

G. Srinivasa Rao
Computer Science & Engineering, DVR & Dr HS MIC College of Technology
Andhra Pradesh, India

Abstract: Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. Users are increasingly pursuing complex task-oriented goals on the Web, such as making travel arrangements, managing finances or planning purchases. To this end, they usually break down the tasks into a few co-dependent steps and issue multiple queries around these steps repeatedly over long periods of time. To better support users in their long-term information quests on the Web, search engines keep track of their queries and clicks while searching online. We propose Biased ranking application development based on their personalized web search of each user present in the data base application procedure. We provide secure privacy to search profiles of each users using hashing secure algorithms. Our experimental results show efficient security operations of each user based on processing of personalized web search.

Key words: Privacy Protection, Web search, SHA, User Profile construction.

I. INTRODUCTION

The webs on the internet look for motor has lengthy become the most essential website for common people looking for useful details on the web. However, customers might encounter failing when Google return unrelated outcomes that do not meet their real objectives. Such irrelevance is largelydue to the tremendous variety of users' situations and background scenes, as well as the indecisiveness of text messages. Customized web look for (PWS) is a general type of look for techniques seeking at providing better look for motor outcomes, which are designed for individual customer needs. As the cost, customer details have to be gathered and examined to determine the userintention behind the released question.The solutions to PWS can generally be classified into two types, namely click-log-based techniques and profile-based ones. The click-log centered techniques are straightforward—they simply encourage prejudice to visited webpages in the user's question record. Although this strategy has been confirmed to execute continually and considerably well, it can only work on recurring concerns from the same customer, which is a strong restriction limiting its usefulness.

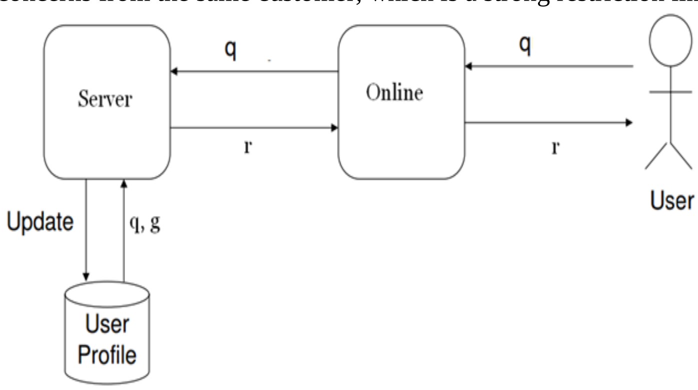

Figure 1: User data structure for web search.

One essential step towards enabling services andfeatures that can help customers during their complicated searchquests on the internet is theability to recognize and team relatedqueries together. Recently, some of the major searchengines have presented a new "Search History" feature,which allows customers to monitor their on the internet queries byrecording their concerns and mouse clicks. For example, Determine 1illustrates a portion of a user's record as it is shown bythe Google on the internet look for motor on Feb of 2010. This historyincludes a series of four concerns shown in reversechronological order together with their correspondingclicks. In addition to watching their look for record, userscan operate it by personally modifying and organizingrelated concerns and mouse clicks into categories, or by sharingthem with their friends. While these functions are helpful,the guide initiatives involved can be troublesome and willbe untenable as the look for record gets longer eventually.

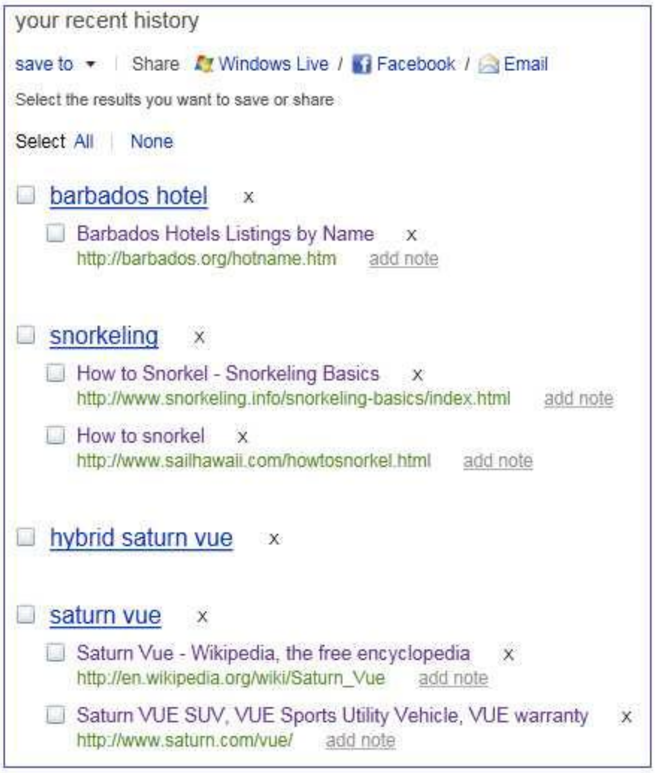

Figure 2: User Profile construction based on query search.

In fact, determining categories of appropriate concerns has applicationsbeyond helping the customers to appear sensible and keeptrack ofconcerns and mouse clicks in their look for record. Firstand major, question collection allows the look for engineto better understand a user's period and possibly tailorthat user's look for encounter according to her needs.Once question categories have been recognized, look for enginescan have a good reflection of the look for contextbehind the present question using concerns and mouse clicks in thecorresponding question team. This will help to improvethe quality of key elements of Google suchas question suggestions, result position, question modifications,session, and collaborative look for.

For example, ifa on the internet look for motor knows that a present question "financialstatement" connected to a {"bank of America", "financialstatement"} question team, it can boost the position of thepage that provides details about how to get a Bankof America declaration instead of the Wikipedia article on"financial statement", or the webpages appropriate to financialstatements from other financial institutions.Query collection can also assist other customers by promotingtask-level collaborative look for. For example, given aset of question categories designed by expert customers, we can selectthe ones that are in accordance with the present user'squery activity and suggest them to her. Explicitcollaborative look for can also be conducted by allowingusers in a reliable community to find, share and combine appropriate question categories to execute larger, long-term projects on the Web.

In this paper, we study the problem of planning a user's look for record into a set of question categories in an computerized and powerful fashion. Each question team is a collection of concerns by the same customer that are appropriate to each other around a common informative need.

The nature of a site page is controlled by a mix of numerous particular elements. To begin with, it needs to contain unique, dependable, and forward substance of honest to goodness esteem. It ought to likewise give metadata that precisely depicts the substance of a page, and contain joins that can go-to people to other related assets. At last, website page design ought to be reliable and take after the standards of client driven web outline, by permitting perusers to easily explore to the pertinent data on the page . As record quality is affected to some degree by these variables, the nature of a page ought not be seen as a dichotomy, yet rather as a constant range. Toward one side of this quality range are surely understood assets for fantastic web reports, for example, Wikipedia. Wikipedia articles are always observed and redesigned by editors, have a predictable design and for the most part contain connections to other related Wikipedia articles and website pages of hobby. On the flip side of this range are spam pages that utilize systems, for example, content duplication, connection plans, substance shrouding and catchphrase stuffing to falsely expand their web crawler positioning and give no helpful substance (or even fake and destructive substance) to their per users. In our paper we propose another way to deal with quality-one-sided positioning which incorporates making of new aspects of importance what's more, execution of various components, catching the nature of a site page along the proposed measurements. On the premise of a few quality aspects we frame a combined rating, which is called business pertinence. As opposed to we extrapolate business pertinence names to the entire figuring out how to-rank data set. For the topically applicable pursuit results we characterize the bound togetherimportance name as the weighted aggregate of topical and business pertinence scores. Our methodology permits to altogether enhance disconnected from the net and also online measurements contrasting with the default positioning calculation.

II. BACKGROUND APPROACH

We show the techniques completed for every client amid two diverse execution stages, in particular the disconnected from the net and online stages. For the most part, the logged off stage builds the first client profile and after that performs security necessity customization as per client indicated subject affectability. The consequent online stage finds the Optimal -Risk Generalization arrangement in the hunt space dictated by the modified client profile.Previously we are used Greedy DP and Greedy IL algoritham. So we need better security for using SHA Algoritham in proposed.

SHA ALGORITHAM

Input: Information in the form Packets.

Output: Encrypted form / Decrypted form data with buffer size requirements.

Step 1: Appending padding bit of information, divide message into 64 bits with multiples of 512 bits.

Step 2: Append the length (In binary format indicating length of the original message into 64 bit)

Step 3: Prepare processing functions like

f(t;B,C,D) = (B AND C) OR ((NOT B) AND D) (0 <= t <= 19) f(t;B,C,D) = B XOR C XOR D (20 <= t <= 39)

f(t;B,C,D) = (B AND C) OR (B AND D) OR (C AND D) (40 <= t <=59)f(t;B,C,D) = B XOR C XOR D (60 <= t <= 79)

Step 4: Prepare processing constants related to original message:

K(t) = 0x5A827999 (0 <= t <= 19)

K(t) = 0x6ED9EBA1 (20 <= t <= 39)

K(t) = 0x8F1BBCDC (40 <= t <= 59)

K(t) = 0xCA62C1D6 (60 <= t <= 79)

Step 5: Initiate buffers sizes with equivalent constants depending on the number of words:

H0 = 0x67452301

H1 = 0xEFCDAB89

H2 = 0x98BADCFE

H3 = 0x10325476

H4 = 0xC3D2E1F0

Step 6: Processing Message in 512 bit blocks:

K(0), K(1), ..., K(79): 80 Processing Constant Words

H0, H1, H2, H3, H4, H5: 5 Word buffers with initial values.

Pseudo Code for converting buffer size requirement.

For loop on k = 1 to L

(W(0),W(1),...,W(15)) = M[k] */* Divide M[k] into 16 words */*

For t = 16 to 79 do:

W(t) = (W(t-3) XOR W(t-8) XOR W(t-14) XOR W(t-16)) <<< 1

A = H0, B = H1, C = H2, D = H3, E = H4

For t = 0 to 79 do:

TEMP = A<<<5 + f(t;B,C,D) + E + W(t) + K(t) E = D, D = C,

C = B<<<30, B = A, A = TEMP

End of for loop

H0 = H0 + A, H1 = H1 + B, H2 = H2 + C, H3 = H3 + D, H4 = H4 + E

End of for loop

Output:

H0, H1, H2, H3, H4, H5: Word buffers with final message digest

For both SHA one begins by converting the message to a unique representation of the message that is a multiple of 512 bits in length, without loss of information about its exact original length in bits, as follows: append a 1 to the message. Then add as many zeroes as necessary to reach the target length, which is the next possible length that is 64 bits less than a whole multiple of 512 bits. Finally, as a 64-bit binary number, append the original length of the message in bits..

III. PROPOSED WORK

BIASED QUALITY RANKING

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 this section, we summarize our suggested likeness operate simmer to be used in the on the internet question collection procedure. For each question, we maintain a question image, which symbolizes the importance of other concerns to this question. For each question team, we sustain a perspective vector. We then propose a likeness operate simrel for two query categories based on these concepts of perspective vectors and question pictures. Note that our suggested explanations of question reformulation chart, question pictures, and perspective vectors are crucial ingredients, which offer significant unique to the Markov chain procedure for determining importance between concerns and question categories.

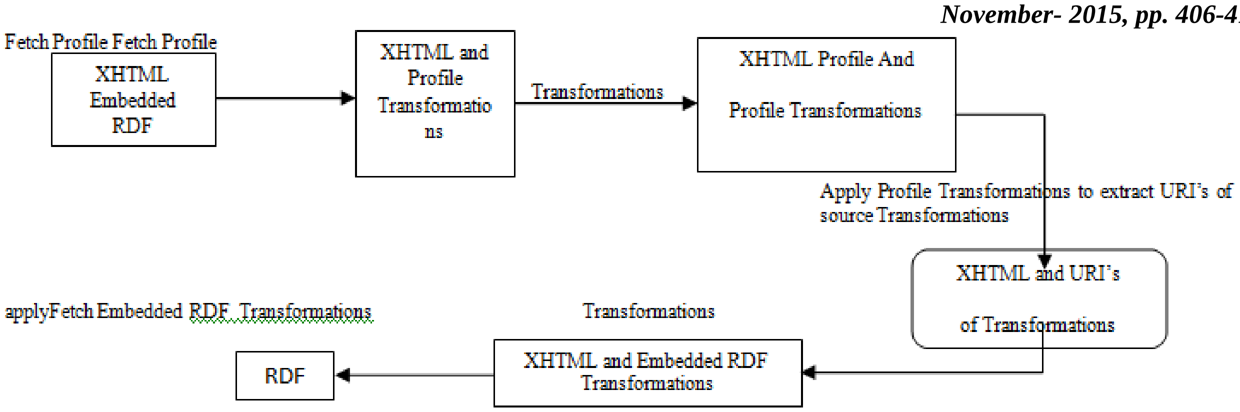

Figure 3: Sequence of RDF documents with profile construction.

Online Query Grouping: The likeness measurement that operates on the imagesof a question and a question team. Some programs such as question recommendation may be assisted by fast on-thefly collection of user concerns. For such programs, we can avoid performing the unique walk calculations offusion importance vector for every new question in real-time, and instead pre-compute and storage area cache these vectors for some concerns in our chart. This works especially well for the popular concerns. In this case, we are basically trading off hard drive storage area for run-time performance. In this segment, we detail the standards of value one-sided positioning, in light of the Markov Random Field model for Information Retrieval (MRF-IR), initially proposed by Metzler and Croft. MRF-IR has reliably shown cutting edge recovery adequacy in an assortment of hunt undertakings, and particularly for pursuit over extensive web accumulations. A few top performing entries at the Text Retrieval Meeting (TREC) in the web hunt tracks (Terabyte Track 2004-2006, Million Query Track 2007-2008) have utilized this model as a part of the most recent five years. Currently,the MRF-IR model is a standout amongst the best freely revealed content based recovery models for web look. On the other hand, tothe best of our insight, there is no distributedexamination on effectively fusing the thought of record quality into the MRF-IR model. As needs be, in this area, we talk about the mix of elements speaking to the archive's nature content into this model.

We are currently prepared to completely indicate the quality-one-sided using so as to position capacity the component capacities characterized in the past segment. Utilizing the three sorts of potential capacities in the consecutive reliance model (characterized over term archive, bigram-report and record just inner circles)

$$score(Q,D) = \lambda\tau f\tau(q,D)$$

$$+\sum_{L \in C} \lambda_L f(D)_L$$

The capacities fT ,fO and fU depend on weighting capacities, which have been effectively utilized by specialists as a part of the past. Capacities fLdepends on the record quality elements.

IV. PERFORMNANCE EVALUATION

In this area, we research the actions and efficiency of our methods on dividing a user's question history into one or more categories of relevant concerns. For example, for the series of concerns "caribbean cruise";"bank of america"; "expedia"; "financial statement", we would expect two outcome partitions: first, {"caribbean cruise", "expedia"} associated with travel-related concerns, and, second, {"bank of america", "financial statement"} associated with money-related concerns.

Data: To this end, we acquired the question reformulation and question just click charts by consolidating a variety of per month look for records from a professional online look for engine. Each per month overview of the question log contributes roughly 24% new nodes and sides in the chart asopposed to exactly previous per month overview,while roughly 92% of the huge of the chart isobtained by consolidating 9 per month pictures. To decrease the effect of disturbance and outliers, we trimmed the question reformulation chart by maintaining only question places that showed up at least two times (q = 2), and the queryclick chart by maintaining only query-click sides that had at least ten mouse clicks (c = 10). This created question and just click charts that were 14% and 16% more compact in comparison to their unique specific charts. Depending on these two charts, we designed the question combination chart.

Figure 4: Varying threshold value with respect to time.

In purchase to create test cases for our methods, we used the look for action (comprising at least two queries) of a set of 200 customers (henceforth called the Rand200 dataset) from our look for log. To produce this set, customers were selected arbitrarily from our records, and two humanlabelers analyzed their concerns and allocated them to either an current team or a new team if the labelers considered that no relevant team was present. A user's concerns were involved in the Rand200 dataset if bothlabelers were in contract to be able to decrease prejudice and subjectivity while collection. The labelers were permitted access to the Web to be ableto figure out if two apparently remote concerns were actually relevant (e.g. "Alexander the great" and "Gordian knot").

Performance Measurement: To assess the quality of the outcome categories, for each user, we start by processing question places in the marked and outcome categories. Two concerns form a couple if they are part of the same team,with only concerns coupling with a special "null" question. To assess the efficiency of our methods against the categories created by the labelers, we will use the Rand Catalog metric, which is a generally employed assess of likeness between two categories. The Rand Catalog likeness between two categories X,Y of n components each is determined as

RandIndex(X, Y) = (a + b)/n2

where a is the variety of places that are in the same set in X and the same set in Y , and b is the variety of places that are in different places in x and in different places in Y .

In our first research, we research how we should merge the question charts arriving from the question reformulations and the mouse clicks within our question log. Since mixing the two charts is taken by the parameter. we analyzed our criteria over the charts that we designed for increasing principles of α.

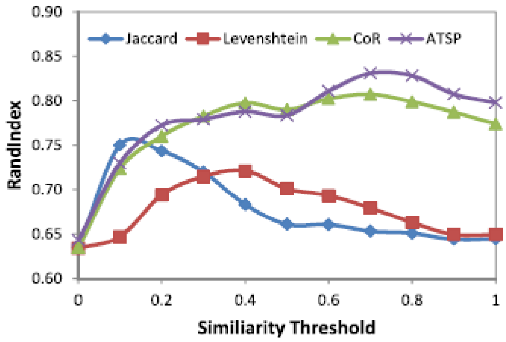

Figure 5: Similarity random index with random clicks.

The horizontally axis symbolizes (i.e., how much weight we give tothe question sides arriving from the question reformulation graph), while the straight axis reveals the efficiency of our criteria in terms of the RandIndex metric. As we can see from the chart, our criteria works best (RandIndex = 0.86) whenis around 0.7, with the two extreme conditions (only sides from mouse clicks, i.e., = 0.0, or only sides from reformulations, i.e., = 1.0) executing lower. It is exciting to note that, in accordance with the shape of the chart, sides arriving from question reformulations are considered to be a little bit more helpful in comparison toedges from mouse clicks. This is because there are 17% less click-based sides than reformulation-based sides, which means that unique walking conducted on the question reformulation chart can recognize better question pictures as there are more available routes to follow in the chart.In conclusion, from the trial results, we notice that using the just click chart in addition to question reformulation chart in a specific question combination chart helps improve efficiency. Additionally, the question fusiongraph works better for concerns with higher utilization details and easily surpasses time-based and keyword and key phrase similarity-based baselines for such concerns. Lastly, keyword and key phrase similarity-based methods help supplement our method well offering for a high and constant efficiency regardless of the utilization details.

V. CONCLUSION

In this document, we display how such details can beused successfully for the process of planning customer search. Histories into question categories.Here propose Biased ranking application development based on their personalized web search of each user present in the data base application procedure. Here provide secure privacy to search profiles of each users using hashing secure algorithms. Our experimental results show efficient security operations of each user based on processing of personalized web search.We proposecombining the two charts into a question fusiongraph. We further display that our strategy that is based on probabilistic unique walking over the question fusiongraph outperforms time-based and keyword and key phrase likeness basedapproaches. We also discover value in mixing ourmethod with keyword and key phrase similarity-based techniques.

REFERENCES

[1] "Supporting Privacy Protection in Personalized Web Search", by LidanShou, He Bai, Ke Chen, and Gang Chen, in IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL:26 NO:2 YEAR 2014.

[2] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.

[3] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
[4] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.
[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.
[6] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.
[7] R. Baeza-Yates and A. Tiberi, "Extracting semantic relations fromquery logs," in *KDD*, 2007.
[8] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*.Morgan Kaufmann, 2000.
[9] W. Barbakh and C. Fyfe, "Online clustering algorithms," *International Journal of Neural Systems*, vol. 18, no. 3, pp. 185–194, 2008.
[10] M. Berry and M. Browne, Eds., *Lecture Notes in Data Mining*.World Scientific Publishing Company, 2006.
[11] V. I. Levenshtein, "Binary codes capable of correcting deletions,insertions and reversals," *Soviet Physics Doklady*, vol. 10, p. 707,1966.
[12] M. Sahami and T. D. Heilman, "A web-based kernel function formeasuring the similarity of short text snippets," in *WWW '06:Proceedings of the 15th international conference on World Wide Web*.New York, NY, USA: ACM, 2006, pp. 377–386.
[13] J.-R. Wen, J.-Y.Nie, and H.-J. Zhang, "Query clustering using userlogs," *ACM Transactions in Information Systems*, vol. 20, no. 1, pp.59–81, 2002.
[14] A. Fuxman, P. Tsaparas, K. Achan, and R. Agrawal, "Using thewisdom of the crowds for keyword generation," in *WWW*, 2008.
[15] K. Avrachenkov, N. Litvak, D. Nemirovsky, and N. Osipova,"Monte carlo methods in PageRank computation: When oneiteration is sufficient," *SIAM Journal on Numerical Analysis*, vol. 45,no. 2, pp. 890–904, 2007.
[16] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.
[17] J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.
[18] 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.
[19] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.
[20] J. Teevan, S.T. Dumais, and D.J. Liebling, "To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 163-170, 2008.
[21] 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.

