International Journal of Engineering &Technology, 7 (4.24) (2018) 59-66

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Web Search Personalization Based on Probability of Semantic Similarity between User Log and Query with Web Page

Y.Raju[1]*, Dr D. Suresh Babu[2], Dr K. Anuradha[3]

1 Department of Computer Science and Engineering, ISL College of Engineering, Hyderabad, India.
2 Department of Computer Science, Head, Kakatiya Government College, Kakatiya University, India.
3 Department of Computer Science and Engineering, GRIET, Hyderabad, India.
**Corresponding author E-mail: raju.yeligeti@gmail.com*

Abstract

Web search personalization is recognized as a competent solution to address the problem of query-relevant search as per the user interest, while it able to present dissimilar search results based upon the preferences and information requirements of users. The popular search engines provide their search results interpreting the user query only, which mostly have unrelated results due to the keywords ambiguity problem. In order to have satisfied and user interesting result, it is important to personalize the results according to their relevancies. In this paper, we propose a Web search Personalization based on a Probability of Semantic Similarity (WP-PSS) between user log and query with search result webpage. It performs a probability of semantic similarities computation between the user query and search result webpage snippet, and compute the frequency of link associated with the log data. Based on these two computed factors a probability of similarities association is computed to group and re-rank the search results for the personalization. Experiment evaluation over a set of multi-domain web searched data collection shows an accuracy improvisation.

Keywords: *Web Search, Personalization, Semantic Similarity, Weblog, Query, Web page*

1. Introduction

As the World Wide Web has exploded so dramatically over the last 15 years, the information available to users continues to grow. In this context, search engines have become a vital tool for users to find the information they need at the sea of massive information. However, because there is so much information to search, traditional search engine technology is becoming less and less useful. Many studies show that the majority of search terms for search engines are not short and clear, and users can have completely different intentions for the same query [1], [2], [3]. This can be described by the search term "*office*" or "*Jaguar*". In both cases the result return by search engine might be prefer different for a different users, such as a real-estate agent may interested for looking office spaces and software professional may interested for look office software, in similar, a car buyer may interested car models and an animal researcher may interested in the wild cat species. To address such ambiguity problems personalization of web search is most preferable solution [19], [20], [26], [29].

In spite of the benefits of personalized search, there is currently no big utilization of personalized search services due to it faces several challenges in terms of accessing the weblog privacy, association of the web usage accurately and some case the ineffectiveness of the personalization due to the irrelevancies of the query and result association [4]. However, personalization can be improved through regular web searches rather than explicitly relying on specific user interests. Even the "Google" and other web search engines are currently attempting a personalized search [1], [5], [6].

Currently, web search personalization primarily uses user profile and weblog data information to learn the necessary personalization. Most of these tasks are performed by the majority of users [2], [3], [7], [8] to provide a solution based on historical activities in the form of Web log data or explicit user feedback learning. Many research on personalized web searches focuses on new mechanisms to get automatically be trained for the user preferences exclusive of the user's direct effort because they are unwilling to give "explicit feedback" on their interests. User profiles can typically aggregate user's historical information and indicate the user's long-period interest in information necessitates.

In various cases, research has explored whether such long-period user profiles are unproductive. Considering the case which described in the historical activity of the user's weblog data format, the needs of different time zones vary depending on the situation. In this situation, personalization based on the user's long-period interests might not offer satisfactory performance, since similar results may be returned repeatedly. Several works [1], [3], [13], [23] have considered using a user's active perspective to indicate short-period information requirements. The search perspective is either incorporated into the user profile or is made up of a different short-period user model or profile and is utilized to estimates the user's information needs.

Even some web search personalization approaches suggest "PageRank" depending on the re-ranking of the resulting documents using click-through data [28]. Unfortunately, to calculate a re-ranking model, it will need to go through several iterations through weblogs or click-through data to create a re-ranking result. Also, if many users use the search engine, many "personalized PageRank" cannot be calculated and saved offline. But the experiment concludes that web search personalization using existing works [1], [7], [18], [19] results using "PageRank scores" is able to enhance the web search, however, the number of personalized results is limited because of computational requirements. In fact, instead of "personalizing" rankings to a particular person, this query changes the rankings depended on the topic of the query and query perspective.

Based on the above need and limitation observation of the personalization, in this paper, we propose a Web search Personalization based on a Probability of Semantic Similarity (WP-PSS) between user log and query with search result webpage. It contributes a probability of semantic similarities computation between user query and search result webpage snippet to overcome the problem of user’s long-term interests satisfying performance, through grouping the most relevant results, and secondly it contributes to solving the problem of "PageRank" through a runtime computing the frequency of link associated with the log data. The outcome of this computation is effectively be utilized to construct the most preferable personalization search result.

The remaining paper is categorized as follows: Related works are discussed in Section-2. In Section-3, it presents the architecture and personalization methodology of WP-PSS. In Section 4, it presents the experiment mechanism and results in the evaluation, and section-5 presents the conclusion of the paper.

2. Related Works

The previous work on search personalization [13], [17], [18], [19], [20], [21] is usually characterized through the data source utilized to learn about user interest, and the approach where a user is relating to these data. The factors that are mostly considered for web personalization are "user queries", "weblog data" and "click-through data". All these data requires appropriate aggregation to create personalization according to individual users. As users randomly search, they are never interested, they are also part of the weblog, so this kind of logs for personalization can be irrelevant. The user query is the main input in all these acceptance of privatization [30], the only information is being extracted based on these query keywords. Without adequate keywords or unclear keywords, there may be a negative impact on personalization in case of any query.

Since the user's interest preferences are strongly correlated with the query context, it is very important to study the relevance of the semantic relations between the search results and the query keywords [7], [13], [15]. The semantic similarity between entities changes in different areas eventually as recent words are created continuously and fresh senses are assigned to existing words. C. Chen et al. [7] recommend "Location-Aware Personalized News" using in-depth semantic analysis. It improves the relevance between users to estimate the similarity between the user's current location and the topic of the candidate news, and the "Explicit semantic analysis (ESA)" [15] related to the topic of the news article. The news with the most relevant top-*k* is recommended for users. Because of the ability to extract an effective representation of the necessary information using semantic similarity [11], implicit semantic similarity learning has already been applied to many personalized referral applications such as "music", "movie" Successfully applied. "Multiple View Items" are for the recommendation [2], [7], [9], [12].

T. T. Sang Nguyen et al. [9] proposes a "Web page recommendation" that semantically integrates web domain knowledge with web usage knowledge. By integrating semantic information with Web-enabled mining, it can achieve higher performance than existing web-using mining algorithms [14], [15], [16]. Ganesh et al. [25] proposed association measures for the intention of optimizing the Web crawler's visit URL order. Here, for every link URLs, the association measures estimate the semantic content related to the reference domain explicit ontology model. In addition, dimensions in URLs can be categorized by analyzing the link strength between the parent and child web pages in subsequent of the web pages are downloaded. L. Yao [2] also proposed a recommendation system through collaboration filtering and content-based Web services integration. It took into account the ranks of web services and semantic content data in terms of probabilities in terms of search and user preferences.

Many approaches have been proposed in previous [13], [21], [23] using web-based log data to understand user preferences and interests. Likewise, in [19], a user profile is created with a vector of discrete words expressions and is generated by cumulative of user click histories. In reverse the results using "cosine similarity" among the "user profile vector" and the "feature vector" of the retrieved web pages. J. Teevan et al. [21] and P. A. Chirita et al. [22] leveraging a prosperous model of user interest construct on "search-related information" and the former information regarding the user. This consist of "documents" and "emails" that it has to interpret and constructed. In [29], keywords are correlated with different types of user profiles which correspond to the "hierarchical category trees" related to keyword types.

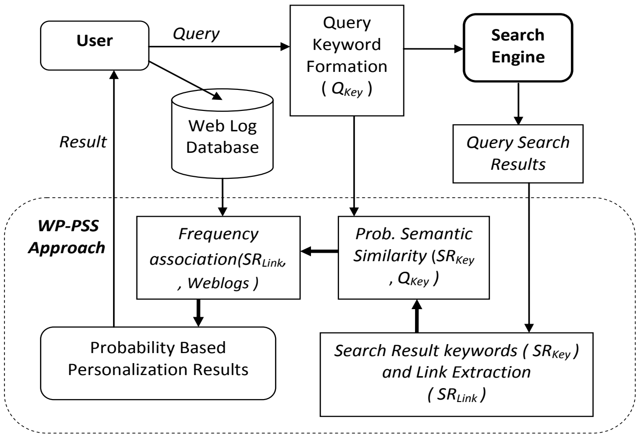
The most commonly used method is to use keyword-based search methods [10] to unearth the relevant web pages and to present suitable ranking schemes. In addition, as demand for user satisfaction has increased, "vertical search engines" have presented specific value information and interrelated services for specific areas, specific individuals, and specific needs (eg, travel searches, online purchases, and search for educational resources) [17]. However, vertical and general search engines still do not get detailed and accurate information. Moreover, ranking enhancements have not been addressed effectively in the Personalization Search, which has happened to a research route for many scientific researchers. Even user "behavior-based techniques" have improved ranking performance [1], [3]. For example, "click models" are precisely considered for personalized searches, where "clicks" of a realistic time period for a particular document suppose the user interest in such results [27], [28], while, it may not be applicable to the former users as such.

Y. Tang et al. [1] proposed a framework for capturing user intentions on personalized Web sites. It executes an "efficient", "configurable", and "intelligent search framework" for personalized Web sites using real-time locations and related feedback technologies. It proposes an implicitly relevant feedback strategy relating to "click-through data analysis" to understands the association among user query situation and search results.

This study infers that the importance of queries, search results, and weblog data can be effective in organizing user personalization needs. However, existing issues related to "computation cost", "weblog privacy", and "accurate extraction and ranking" of results motivate to design a new personalization approach to address the issues of search personalization. An integration with this three entities a query probability of semantic similarity with the extracted web pages, and its link frequency of association with the weblog data are being proposed to meet their real-time personalization information needs.

3. Web Search Personalization

Web search personalized is will be achieved through a probability of webpage content semantic similarity with user query and its links frequency association with the user weblog. A designed architecture for the proposed Web search Personalization based on a Probability of Semantic Similarity (WP-PSS) is shown in Fig.1.



The architecture of WS-PSS is presented in Fig. 1 which consists of three main functions as, 1) Query keyword Formation, 2) Search Result Keywords and Link Extraction, and 3) WP-PSS Mechanism which defines the process of probability of semantic similarity between search result keywords, the computation of frequency association between search result link and weblog, and probability of personalization results generation.

3.1 Query keyword Formation

The user query generally consists of a collection of words along with some regular terms. It needs to clean and the generate the keywords for the result extraction needed from the search engine. The module of query keyword formation implements the method to constructs the unique keywords required as "Q_{Key}". The formation method tokenizes the query into keywords and remove the stop words and submit the "Q_{Key}" to the search engine for the data extraction. The return results are submitted to the WS-PSS mechanism to construct the personalization results. The generated "Q_{Key}" even submitted to the probability of semantic similarity module of WS-PSS mechanism to group the relevant result. In past approaches [10], [17] they stored the user query for the further process which creates the storage overhead, processing runtime overcome this problem.

3.2 Search Result Keywords and Link Extraction

The search results obtained from the search engine based on the keywords "Q_{Key}" presents snippets of information and a link URL. The snippets provide an information of text retrieved through a search engine approximately the query expression of the documents. It presents constructive information concerning the confined context of the query expression. In order to avoid the downloading and processing delay due to the huge return results, we only considered few top-ranking results between 10 - 50 in numbers for efficiently processed. We implemented the "Term Extraction algorithm" as presented in [24], executing over a snippet of every referred web page. It tries to recapitulate the snippet text into a collection of significant keywords. In a similar manner, a link extraction algorithm [14] implemented to retrieve from the webpage.

Based on the user Query, Q input and the generated keywords from "Q_{Key}", it retrieves the keywords from the relevant result documents as D_k from a web search engine. It collects the top 10 results documents for each keyword and an *Extract_Keywords* (D_k) method is processed to extract the keywords vector from the document as "SR_{Key}", and *Extract_Link*(D_k) method to extract the links vector as "SR_{Link}". These search results keywords "SR_{Key}" further utilized for computing the Query Probability of semantic similarity with webpage and "SR_{Link}" for link Frequency of association with a weblog.

3.3 Query Probability of Semantic Similarity with a webpage

To compute the probability of semantic similarity association between the "Q_{Key}", and the keywords of the search document, "SR_{Key}" it initially identifies the most frequent keywords among "Q_{Key}" and then identifies the semantic similarity association between the most frequent with others keywords to generate the required pattern for the classification. The Algorithm-1 describes the procedure of the mechanism.

Algorithm-1: Probability of Semantic Similarity

Input: Q_{Key}*[ ]* → Set of Query keywords.
V[ ] → Sets of searched results.
Z[R] → Sets of searched results keywords.
(where R is the no. of result document retrieved).

Output: *Pr_SS_Value [ ]*
(Probability array of keywords and semantic similarity value)

Method: ***Semantic_Similarity_Association*** **(**Q_{Key} *[ ], Z[R]***)**

```
//-- Semantic Similarity Association for each keywords in QKey --
    for (a=0; a <size of QKey; a++ )
    {
        Ta = QKey [a];
        Acnt = 0;
        SS_Value [ ]; SS_Result [ ];

        //-- For each search result retrieved from, Z[R] --
        for ( x = 0; x < R; x++ ) {
            // -- Getting each result keywords --
            SKey[ ] = Z [x];
            Vx = V [x];

            //-- For each keywords in SKey[ ]
            for ( k = 0; k<size of SKey; k++ )
            {
                E.Key= SKey[k];
                if (Ta == EKey ) {
                  Acnt ++;
                }
                prob_sa_value=0;
                if (Acnt > 0 )     {
                prob_sa_value=((Acnt*100)/(size of SKey));
                }
                //-- Array of Semantic Similarity Association Values --
                SS_Value [x] = prob_sa_value ;
                SS_Result [x] = Vx ;
            }
            Pr_SS_Value [a ] = SS_Value [ ];
            Pr_SS_Result [a ] = SS_Result [ ];
        }
    }
```

The outcome of the Semantic Similarity Association of Keywords generates an array of the probability of similarity, *Pr_SS_Value [ ] and Pr_SS_Result [ ]* to the extracted search result. The value of each search result associated with the keywords helps to group the most relevant results.

To relate the group results of each keyword we related to the weblog data for each result to order the result to meet the user's preferable interest. In the next section, we discuss the process of learning link frequency association to the weblog.

3.3 Link Frequency of Association with weblog

A weblog generally generated by the server implicitly for all kind of users who search and click the result link. The clicked links are recorded into the web login form of URL along with the timestamp, user-id and accessing method. It extracts these link

URL through implementing URL extraction method to built a collection of the link as, "WL_{Links}". The frequency of association between the search results link, "SR_{Link}" and "WL_{Links}" is presented in the Algorithm-2.

Algorithm-2: Link Frequency of Association

```
Input:  WL_Links [ ] → Set of Weblog links.
        SR_Link → Set of Search Result Link.
        R → is the no. of result document retrieved.

Output:  Link_Freq_Value [ ]

Method: Link_Frequency_Association (WL_Links [ ], SR_Link [ ], R)
    for ( x = 0; x < R; x++ )
    {
        // -- Getting each result Link --
        S_Link = SR_Link [x];
        A_cnt = 0;

        //-- For each keywords in S_Key[ ]
        for ( w = 0; w<size of WL_Links; w++ )  {
            L_Link= WL_Links [w];
            if (S_Link = = L_Link )  {
                A_cnt ++;
            }
        }
        link_freq_value=0;
        if (A_cnt > 0 )  {
            link_freq_value=((A_cnt*100)/ (size of WL_Link));
        }
        Link_Freq_Value [x] = link_freq_value;
 }
```

3.4 Personalization using Probability of Semantic Similarity

The construction of personalization is being performed utilizing the computation value of the probability of semantic similarity, *Pr_SS_Value [ ]* and *Link_Freq_Value []*. It utilizes this two computed value to group the most relevant results initially and later re-rank the results in the group based on the link frequency. The methodology is illustrated in the Algorithm-3.

Algorithm-3: Personalization Result

```
Input:  Q_Key[ ] → Set of Query keywords.
        Pr_SS_Value [ ] → Set of Weblog links.
        Link_Freq_Value [ ]  → Set of Search Result Link.
                Z[ ] → Sets of searched results.

Output:  Personalization_Result [ ]

Method:  Prob_Personalization  (Pr_SS_Value  [  ],
Link_Freq_Value [ ], Q_Key[ ], Z[ ] )
   //-- For each keywords in Q_Key [ ]--
   for (a=0; a <size of Q_Key [ ]; a++ )
   {
        T_a = Q_Key [a];

        New_PR_Result [ ];

        SS_Value [ ] = Pr_SS_Value [T_a ];
        SS_Result [ ] = Pr_SS_Result [T_a ];
                p=0;
          for ( k = 0; k<size of SS_ Result [ ] ; k++ )
          {
              SR_k = SS_ Result [k];
              SS_Value = SS_ Value [k];

                //-- Compute the highest relevancy with other
        keywords--
                H_Key  =  GetHighest_SS_Key(  SR_k  ,
        Pr_SS_Value [ ] );
                        if (H_Key = = T_a ) {
                     New_PR_Result [p] = SR_k ;
                                p++;
                        }
                }
          P_Result [T_a ] = New_PR_Result;
                  }
   //-- Re-Ranking of the Grouped Result --
          for ( x = 0; x < P_Result [ ] ; x++ )  {
                P_Result [ ] = P_Result [x];
                P_re-rank  [  ]  =  DoRe-Rank  (  P_Result  [  ],
             Link_Freq_Value [ ] );
                Personalization_Result [ x] = P_re-rank ;
             }
```

Here, the generated *Personalization_Result [ ]* will be most relevant to the query and also will be the highest preferable results according to the search query and user interests. To evaluate this proposal we implement this against few real-time web document retrieve from a different domain. We discuss it more briefly in the next section.

4. Experiment Evaluation

4.1 Datasets and Measures

The WWW includes an enormous number of web pages that represent many semantic relationships. When a user attempts to search for entities in a particular semantic relationship by means of a "keyword-based web search engine", the user has to create a query with a few keywords correlated to entities and its relations. Accordingly to perform the evaluation we collected a set of various domains documents that are used in an informal way to compute the Probability of Semantic Similarity and construct the Personalization results.

For the evaluation, we construct a dataset of 100 web data records collected using Google search engine from each of this query as, "*Tours and Travel Booking*", *"Treatment, Health Care and Hospitals"*, and *"Online purchasing and e-commerce*". We store the data accordingly in the order of rank given by the Google search engine. It consists of few number duplicate results also, to show the effectiveness of this personalization approach we store those duplicate result also. Over the collected web data records we implement the proposed WS-PSS based query mechanism to evaluate the efficiency of the proposal.

To measure the efficiency of the proposal it computes the percentage personalization precision, recall, and accuracy using the equation-1, 2 and 3 as given below.

$$Precision = \left(\frac{\sum No.of\ Correctly\ Personalized\ Results}{Total\ Number\ of\ Personalized\ Results}\right) \times 100 \quad (1)$$

$$Recall = \left(\frac{\sum No.of\ Correctly\ Associated}{Total\ Number\ of\ Relevant\ Results}\right) \times 100 \quad (2)$$

$$Accuracy = \left(\frac{\sum No.of\ Correctly\ Personalized\ Results\ \cap \sum No.of\ Relevant\ Results}{Total\ Number\ of\ Personalized\ Results}\right) \times 100 \quad (3)$$

To compare the performance evaluation measure we compare the proposed results obtains with the popular Google search engine results. The process of evaluation is carried out in three forms. For the first evaluation, it queries the search engine by three keywords and having 500 records of weblog data. Then, it measures the web search personalization performance with varying number of result extraction from 10 to 50 results. In the second evaluation, it measures the personalization ranking performance with varying the size of weblog datasets from 100 to 500 records for the top-5 result extracted, and in the third evaluation, query keywords performance is measured varying the *N* number of keywords.

4.2 Results

In this section, we discussed the performance results of "Web Search Personalization", "Personalization Ranking" and "Query Keywords".

A. Web Search Personalization Performance

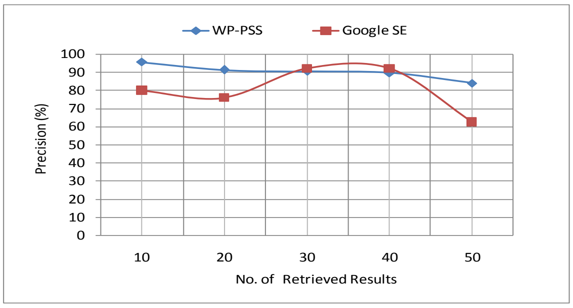


Fig.2. Precision Performance

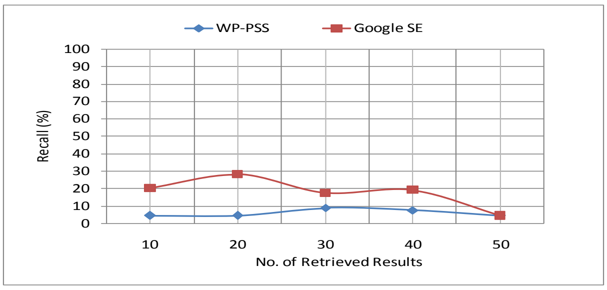


Fig.3. Precision Performance

The precision and recall of web search performance of the proposed WP-PSS are presented in Fig. 2 and 3. It shows that with increasing number of retrieved results the proposed WP-PSS achieve better precision and low recall in compare to the Google search results. The improvisation achieves due to the effectual integration of the user query keyword similarity and weblog link association for the retrieved search results. As with increasing number of retrieved results Google shows low precision because of many duplicates and few irrelevant results retrieval.

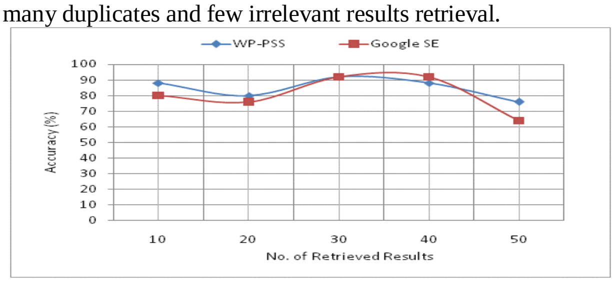


Fig.4. Accuracy Performance

Fig. 4 shows the web search performance accuracy assessment of the WP-PSS and Google results. It shows that in support of an integrated model of query and weblog data for the search personalization improve the accurate and relevant with the increasing number of retrieved results. But, it might have variance depends on the query keyword length and a high number of duplicate data in the result sets. It suggested that the longer the query more precise and accurate the results and personalization also, but in case of the search engine it's results are more associated but it will be attained more recalls.

B. Personalization Ranking Performance

Fig. 5 shows the Personalization Ranking Performance for the top-5 results with varying weblog data from 100 to 500 records. It infers that with increasing number of weblog data it increases their particular rank. When the links of the search results clicked frequently it corresponding weblog also increases, but the rear of the retrieval result list is being clicked low this cause the lower retrieved result as low ranks, but in case of user relevancy preferences a rear of the retrieval result can also improve its rank.

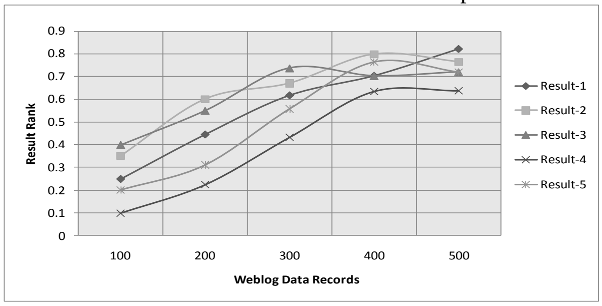


Fig.5. Ranking Performance

C. Query Keywords Performance

Fig. 6, 7 and 8 present the precision, recall and accuracy measure of the query based search personalization performance. The retrieval of the relevance of information completely depends on the query keywords. Here with increasing number of keywords numbers retrieves more relevant results, but at the same time with more number of retrieved results more number of irrelevant and duplicate results also being also retrieved, which affect the personalization precision and accuracy. Because of this with increasing number of keywords it shows the decreasing of precision and accuracy.

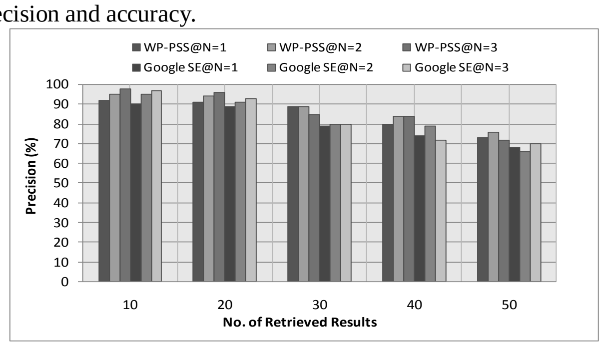


Fig.6. Precision Performance

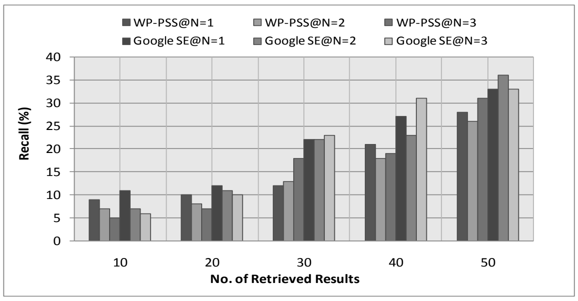


Fig.7. Recall Performance

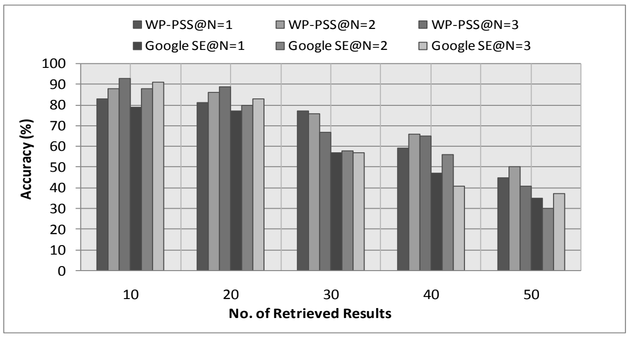


Fig.8. Accuracy Performance

5. Conclusion

In this paper, we proposed a Web Search Personalization based on Probability of Semantic Similarity (WP-PSS) between User Log and Query with the Web page. It investigates the problem and limitation of the web search personalization and suggests the solution to provide precise and accurate results without violating users' privacy. It integrates the two basic implicit data available in the form of weblog and user query to generate the personalization result. It implements a probability Semantic Similarity method to find each keyword association with each individual search result to form a group of results. These groups of results of each individual keywords undergo a link frequency association relating the past weblog data to compute a re-rank model of the grouped results. Based on the obtained probability semantic similarity value and frequency of link association value it generates grouped and re-rank results to present the personalization results. The experimental evaluation of WP-PSS in comparison to the Google search result achieved significant precision and accuracy improvement for the Web Search Personalization and Query Keywords bases Performance. In feature work, it analyses and expands the proposal to associate more user-specific query to present more interesting search results.

References

[1] Y. Tang, H. Wang, K. Guo, Y. Xiao, T. Chi, "Relevant Feedback Based Accurate and Intelligent Retrieval on Capturing User Intention for Personalized Websites", IEEE Access, Vol. 6, pp. 24239 - 24248, 2018.

[2] L. Yao, Q. Z. Sheng, Anne. H.H. Ngu, J. Yu, and A. Segev, "Unified Collaborative and Content-Based Web Service Recommendation", IEEE Transactions on Services Computing, Vol. 8(3), 2015.

[3] S. Hernández, P. Álvarez, J. Fabra, J. Ezpeleta, "Analysis of Users' Behavior in Structured e-Commerce Websites", IEEE Access, Vol. 5, pp. 11941 - 11958, 2017.

[4] S. M. Al-Ghuribi, S. Alshomrani, "A Comprehensive Survey on Web Content Extraction Algorithms and Techniques", International Conference on Information Science and Applications (ICISA) pp. 1 - 5, 2013.

[5] P. Das, R. C. Jisha, G. P. Sajeev, "Adaptive web personalization system using splay tree", International Conference on Advances in Computing, Communications and Informatics (ICACCI) pp. 1582 - 1587, 2017.

[6] J. Liu, N. J. Belkin, "Personalizing information retrieval for multisession tasks: Examining the roles of task stage, task type, and topic knowledge on the interpretation of dwell time as an indicator of document usefulness", Journal of the Association for Information Science and Technology, Vol. 66(1), pp. 58-81, 2015.

[7] C. Chen, X. Meng, Z. Xu, T. Lukasiewicz, "Location-Aware Personalized News Recommendation With Deep Semantic Analysis", IEEE Access, Vol. 5, pp. 1624 - 1638, 2017.

[8] Y. Zuo, J. Zeng, M. Gong, and L. Jiao, "Tag-aware recommender systems based on deep neural networks", Journal of Neurocomputing, vol. 204, pp. 51 - 60, Sep. 2016.

[9] T. T. Sang Nguyen, H. Y. Lu, and J. Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge", IEEE Transactions On Knowledge And Data Engineering, Vol. 26(10), Oct. 2014.

[10] R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multi-keyword ranked query over encrypted data in cloud computing", Future Generation, Computer System, Vol. 30, no. 1, pp. 179 - 190, 2014.

[11] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives", IEEE Trans. Pattern Anal. Machine Intelligent, Vol. 35(8), pp. 1798 - 1828, 2013.

[12] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation", In Proc. Adv. Neural Inf. Process. System, pp. 2643 - 2651, 2013.

[13] G. Venkataraman, A. Ravichandran, "Adaptive Semantic Search: Re-Ranking of Search Results Based on Web page Feature Extraction and Implicitly Learned Knowledge of User Interests", 10th International Conference on Semantics, Knowledge and Grids Pp: 75 - 78, 2014.

[14] T. Moschopoulos, E. Iosif, L. Demetropoulou, A. Potamianos, S. S. Narayanan, "Toward the Automatic Extraction of Policy Networks Using Web Links and Documents", IEEE Transactions on Knowledge and Data Engineering, Vol. 25(10), Pp. 2404 - 2417, 2013.

[15] E. Gabrilovich and S. Markovitch, "Wikipedia-based semantic interpretation for natural language processing", Journal of Artificial Intell. and Reseach, vol. 34(1), pp. 443 - 498, 2009.

[16] S.-X. Bing and Fu-Lei, "Keyword Extraction Algorithm Based on", International Conference on Computational and Information Sciences, Pp. 664 - 665, 2013.

[17] Y. Wu, L. Shou, T. Hu, and G. Chen, "Query triggered crawling strategy: Build a time-sensitive vertical search engine", In Proc. IEEE Int. Conf. Cyberworlds, pp. 422 - 427, 2008.

[18] Z. Dou, R. Song, and J. Wen, "A large-scale evaluation and analysis of personalized search strategies", In Proc. 33rd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 2007.

[19] G. Jeh and J. Widom, "Scaling personalized web search", In Proceedings of the Twelfth Intl. World Wide Web Conf., 2003.

[20] P. A. Chirita, C. Firan and W. Nejdl, "Summarizing local context to personalize global web search", In Proc. Int. Conf. on Information and Knowledge Management, 2006.

[21] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities", In Proc. 31st Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 449–456, 2005.

[22] P. A. Chirita, W. Nejdl, R. Paiu and C. Kohlschutter, "Using ODP metadata to personalize search", In Proc. 31st Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 178–185, 2005.

[23] F. Qiu and J. Cho, "Automatic identification of user interest for the personalized search", In Proc. 15th Int. World Wide Web Conference, pp. 727–736, 2006.

[24] A. T. A. Thuy Vu and M. Zhang, "Term extraction through the unit hood and termhood unification", In Proc. of Int'l Joint Conf on Natural Language Proc., 2008.

[25] S. Ganesh, M. Jayaraj, V. Kalyan, and G. Aghila, "Ontology-based web crawler", In Proc. ITCC: Coding Computing, Las Vegas, NV, pp. 337–341, 2004.

[26] R. Krovetz and W. B. Croft, "Lexical ambiguity and information retrieval", Information Systems, Vol. 10(2), pp. 115–141, 1992.

[27] X. Shen, B. Tan and C. Zhai "Implicit user modeling for the personalized search", In Proc. Int. Conf. on Information and Knowledge Management, pp. 824–831, 2005.

[28] T.Haveliwala,"Topic-Sensitive PageRank", In Proceedings of the Eleventh Intl. World Wide Web Conf., 2002.

[29] F. Liu, C. Yu, and W. Meng, "Personalized web search by mapping user query to categories", In Proc. Int. Conf. on Information and Knowledge Management, 2002, pp. 558–565, 2002.

[30] B. J. Jansen, A. Spink, and T Saracevic, "Real life, real users, and real needs: A study and analysis of user query on the Web", Information Processing and Management, Vol. 36(2), pp. 207 – 227, 2000.