Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A PERSONALIZED SEARCH ENGINE BASED ON CORRELATION CLUSTERING METHOD

[1] BOURAIR ALATTAR , [2] NORITA MD. NORWAWI

Faculty of Science and Technology, Universiti Sains Islam Malaysia

E-mail: jafar_alattar_80@yahoo.com

ABSTRACT

Nowadays, search engines tend to use latest technologies in enhancing the personalization of web searches, which leads to better understanding of user requirements. This paper aims to address the problem of enhancing the efficiency of search system by mining data logs. Search logs are associated with all the interactions that have been done between the user and the search engine including the query, the resulted pages and the selection. The paper also aims minimize the ambiguity within the query whilst the partitioning-based clustering aims to reduce the search space by dividing the data into groups that contain similar objects. To do so, this paper conduct several experiments to evaluate correlation clustering method. The method of this paper includes a pre-processing phase, which in turn involves tokenization, stop-words removal, and stemming. In addition, we evaluates the impact of the two similarity/distance measures (Cosine similarity and Jaccard coefficient) on the results of the correlation clustering method. Experimental results obtained are quite satisfactory in terms of the Precision, Recall and F-score. .

Keywords: *Personalized Search Engine, Correlation Clustering algorithm, similarity/distance measures.*

1. INTRODUCTION

Personalization in search engine aims to use data logs to observe and store the user's transactions[1-10]. Such logs contain heterogeneous information that can be represented by various aspects such as information that is related to the user's preferences, language, location and interests. Moreover, it also contains information related to the user's behaviour such as his own queries, frequent terms and criteria of results selection. In another aspect, these data logs contain information about the search engine itself including its response and the mechanism of ranking documents. Therefore, these logs tend to be large-scale repository which can significantly influence the effectiveness and efficiency of the search system.

The huge and continually increasing amount of information on the web creates many challenges for the researchers of web search [10-15]. One of the problems of search engines is that when queries are issued by different users, most of them return the same results to the users without considering who have submitted the query. Thus, all users get the same result for the same query, even if they have widely different interests, needs and backgrounds. Current web search engines like Google, Yahoo and many more are built to serve all users, independent of the special needs of any individual user. Majority of the search engines on the web deliver a list of results for the user's query. It has been seen that most of the times the exact search context or interest of the user is ignored. For example, a biologist and a programmer may use the same query "mouse" with different search context, but the search systems would return same results. In addition, it has been seen that most of the times search results present such a big list of outcomes that they have to be categorized into different result pages. There are individual differences with respect to synonymy, polysemy and information needs which create problems for the users. The user might have to browse multiple yet irrelevant results before reaching the expected information. So to fill the gap between the user interest and retrieved search results, personalized web search needs to be evolved[16]. Nevertheless , few of these issues can be alleviated when the users make use of personalization methods while searching [17].

With the growth in the number of Web users, the problem of personalization of web search engine has become very critical and popular [18]. It is highly needed to personalized Web Search effectively as it is an open problem in the information retrieval community [5] .Mining the user logos and search results is considered as one of major approaches toward personalization. Mining data logs is a challenging task regarding to several obstacles. The size of search and browse log data at a search engine is usually very large and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

log data are quite noisy. Unlike traditional text, queries tend to be shorter and yield high degree of ambiguity with possibility to have noisy data such as typos. Furthermore, the meaning of queries may change over time. There is a need to mine data logs in order to enhance the effectiveness (the quality of matching results) and the efficiency (the time and memory) of the search engine. The effectiveness of the personalization of web search based on grouped web usage data depends on the quality of clusters[20]. So efficient clustering algorithm is desired to enhance the effectiveness and efficiency of the Personalization of Web Search [3, 5, 8, 10, 21, 22]. Clustering technique is a method that seeks to organize data into different classes that share identical characteristics. In this technique, intra-class similarities are maximized or minimized. In the clustering technique, it is not common for labels to be attached to data in advance. This research used in this paper presents the results of an experimental study of the correlation clustering method on personalized search engine . The method of this paper includes a pre-processing phase, which in turn involves tokenization, stop-words removal, and stemming. In addition, we evaluates the impact of the two similarity/distance measures (Cosine similarity and Jaccard coefficient) on the results of the correlation clustering method.

This paper is split into four main sections: In Section two, we discussed related works on personalized web search engine document. Then in Section three, we described the How do we implementation out our review. And Section four will be on the experimental findings, and finally, Section five and will be conclusion our work..

2. RELATED WORK

Many methods and approaches have been proposed in terms of web search personalization. For instance, Speretta and Gauch [7] explores the use of a less-invasive means of to gather user information for personalized web search. Particularly build user profiles based on activity at the search site and study the use of these profiles to provide personalized search results. User profiles are created by classifying the collected information into concepts in a reference concept hierarchy and then these profiles are used to re-rank the search results and the rank-order of the user-examined results before and after re-ranking, are compared. Their personalized re-ranking results in a 34% improvement in the rank order of the user selected results.

Lai and Soh [23] compared User profile results and personalized search results. The large amount of information is available on web. When user searches anything, in some cases it provides same results for different type of queries. Therefore, it is difficult for user to get relevant and desirable results because it does not consider user preferences as well as interest. Evaluation of users' specified query search and browsing activities are depends on searching query inputted and clicking rate (i.e. no of the times various user click that link) of each link in the response of query and the amount of time they used particular site. Solution to this is to construct user searching profile and method for document profile construction. The discussion is taken place that conclude how to use this model to combine document and user searching profile to provide desired personalized search results to user.

Kim and Cho [24]implemented a personalized Web search engine by using fuzzy concept network with link structure. Many search engines uses link structure to find out precision. Output of link based search engine is having high quality than text based search engine but they are complex. Personalization is one of the best way to obtain more effective and desirable results. Fuzzy concept with network with link structure is useful to find user query's subjective interest. The proposed approach is used to make results personalized by using link based search techniques. The experimental results shows that the results returns to user are based on user interest and preferences, also appropriate.

Matthijs and Radlinski [11] collect web usage data that is URL of page, duration of page visit, page session date and time ,length of the source HTML using Firefox add on.

Liang [25] proposes effective way to construct user profiles based on user interest and preferences. To construct user profiles three approaches are proposed Support Vector Machines method Rocchio method, k-Nearest Neighbors method and. Experimental results taken from a constructed dataset conclude that the k-nearest method is effective than other.

Leung, et al.[26] proposes a novel web search personalization approach that captures the user's interests and preferences in the form of concepts by mining search results and their click-throughs. They separate concepts into content concepts and location concepts, organize them into ontologies for creation of an ontology-based, multi-facet (OMF) profile to precisely capture the user's content and location interests. The Experimental results prove that OMF

Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

improves the precision significantly as compared to the baseline.

Palleti, et al. [27], by using probabilistic query expansion author developed personalized web search. In this approach, the authors developed a personalized Web search system applied at proxy, which changes to user interests perfectly by generating user profile with the use of collaborative filtering. A user profile consists of probabilistic correlations among query terms and document terms, which are utilized for providing personalized search results. Experimental outcomes prove that this proposed personalized Web search system is very effective and efficient. Smyth [3] introduced community-based approach to provide effective personalizing Web search. At community level, knowledge reflect within search communities by collecting user search query and result chosen by user. All collected data is used to prepare a relevance model that provides promotive community relevant results for all web search. Collaborative web search approach is used here that suggest valuable and sharable knowledge.

With the growth in the number of Web users, the problem of personalization of web search engine has become very critical and popular [18]. It is highly needed to personalized Web Search effectively as it is an open problem in the information retrieval community [5] .Mining the user logos and search results is considered as one of major approaches toward personalization. Mining data logs is a challenging task regarding to several obstacles. The size of search and browse log data at a search engine is usually very large and log data are quite noisy. Unlike traditional text, queries tend to be shorter and yield high degree of ambiguity with possibility to have noisy data such as typos. Furthermore, the meaning of queries may change over time.

Few researches are performed on mining data logs aim for enhancing the quality of matching results and the time and memory that would be consumed [19]. There is a need to mine data logs in order to enhance the effectiveness (the quality of matching results) and the efficiency (the time and memory) of the search engine. The effectiveness of the personalization of web search based on grouped web usage data depends on the quality of clusters([20]). So efficient clustering algorithm is desired to enhance the effectiveness and efficiency of the Personalization of Web Search[3, 5, 8, 10, 21, 22]. Thus it brings interesting clustering challenges in the personalized search framework. Clustering results should change dynamically to detect the changes in user's interest and to reflect the personalized ranking of search results.

3. PROPOSED METHOD

In this study framework includes all the stages of implementing the proposed method. Such stages determine the process of collecting the dataset, preprocessing tasks, feature extraction and the clustering method. The comparative Analysis and evaluation of clustering is carried out by using precision, recall and overall F-measure. Figure 1 shows the framework of the crime document clustering.

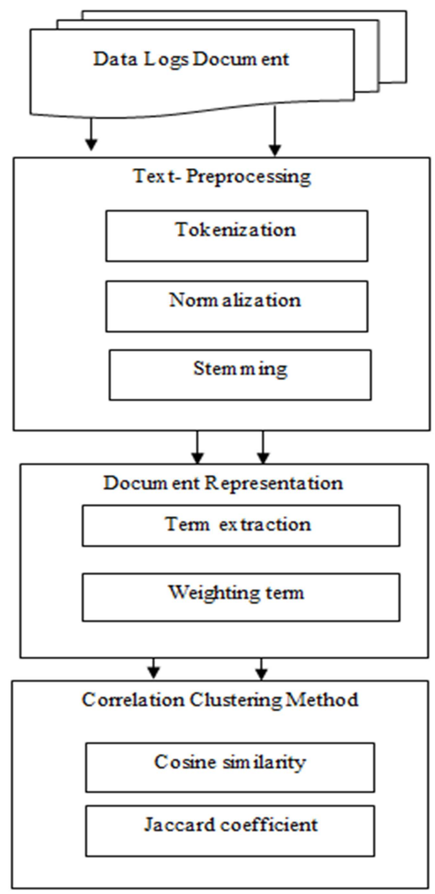

Figure 1: Research Methodology.

3.1 Preprocessing

This phase aims to perform several steps including normalization, tokenization and POS tagging in order to turn the data into a format that enables

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recognizing named entities. The following sub-sections illustrates those steps.

a) Normalization

This step aims to clean the data by removing noisy or unwanted data such as, digits, stop-words and special characters. Such data has been explained as follows:

b) Remove special characters

There are tremendous of special characters which hinders the process of analyzing text. Table 1 shows a sample of such special characters.

Table 1 Sample of special characters

Special characters
! - ~
@ * ^
& +
$ % =

c) Remove stop-words

Generally, documents usually found to contain a lot of unnecessary words in English, such as, pronouns, prepositions, conjunctions and others, which are usually used by authors for the purpose of linguistically enhancing the structures, and in particular, focusing on the syntactic or grammatical function of the language, rather than strengthening the semantic function or the meaning of the content. Therefore, in this particular regard, the process of word removal is very common and of considerable importance to be involved in Document Clustering. This is because, by carrying out word removal. The dimensionality of the terms space will be drastically reduced. stop word as a list of 571 Stop words, and are called so, these are generally regarded stop words because they tend to convey syntactic functions, rather than conveying more than they convey semantic functions, such as, carrying further meaning, which can enhance and strengthen the communicative or informational aspects of the document content.

d) Tokenization

The first step of morphological analyses is the tokenization. The aim of the tokenization is the exploration of the words in a sentence. Textual data is only a block of characters at the beginning. All following processes in information retrieval require the words of the data set. Tokenization is the task of dividing words from text into clusters of consecutive morphemes, one of which typically corresponds to the word stem.

e) Stemming

Stemming aims to eliminate all the derivational inflections form the words. The stemmer that have been used in this research is based on the Buckwalter system which includes three distinct lexicon files (dictionary) listing all prefixes, suffixes, and stems. The algorithm works by dividing a given input word into prefix, stem and suffix and then match those segments with the corresponding lexicons.

3.2 Text Representation

Directly In applying most learning algorithms to text information, in a direct way without representing, it has been proved to be impossible, due to the complex nature of the text information. Therefore, before applying the text using to a machine learning method, it is essential to converting the content of a textual document to a compact representation is necessary. They are Document representation has been found to be efficiently used as a language-independent method, since they are it is independent of the meaning of the language, and perform well in case of noisy text. Term Frequency × Inverse Document Frequency (TF×IDF) weighting is also recognized as a simple method for term weighting.

$$\boldsymbol{W_i = tf_i.\log(\frac{N}{n_i})}$$

Term Frequency × Inverse Document Frequency (TF×IDF) weighting is seen as the most popular method used for term weighting, since it considers this property. By using this approach, assigning the weight of term i in document d to the number of times the term appears in the document is proportional, and it is in inverse proportion to the number of documents in the corpus, in which the term appears.

.

3.3 Similarity And Dissimilarity Measures

Document clustering is the process in which similar documents are grouped to form a coherent cluster. However, complications arise in how to determine if a pair of documents is similar or different. This is not always a straightforward process. In view of the variety of scales, distance measurements (or metrics) between clusters need to be carefully selected. The difference between two patterns is commonly calculated by means of the distance between clusters. The accuracy of clustering depends on a precise definition of the

Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

closeness between a pair of objects, in terms of either the pair wise similarity or distance. This research will focus on well-known measures of distance between patterns. In this regard, this paper focuses on the cosine similarity and Jaccard coefficient as similarity or distance measures.

Cosine

Cosine similarity is one of the most well-known similarity measure which is applied to text documents such as in numerous information retrieval applications and clustering . In measuring the given two documents. $\overrightarrow{t_a}$ and $\overrightarrow{t_b}$, their cosine similarity is:

$$SIM_C(\overrightarrow{t_a}, \overrightarrow{t_b}). \frac{\overrightarrow{t_a} . \overrightarrow{t_b}}{|\overrightarrow{t_a}| * |\overrightarrow{t_b}|}$$

Jaccard

The Jaccard coefficient, which is another similarity measure, also known as the Tanimoto coefficient, is used to measure the similarity in the intersection divided by the union of the objects. For text document, the use of Jaccard coefficient is to make a comparison of the sum weight of shared terms and the sum weight of terms presented in either of the two documents but in condition that they are not the shared terms. The formal definition is as follows:

$$SIM_j(\overrightarrow{t_a}, \overrightarrow{t_b}). \frac{\overrightarrow{t_a} . \overrightarrow{t_b}}{|\overrightarrow{t_a}|^2 * |\overrightarrow{t_b}|^2 - \overrightarrow{t_a} . \overrightarrow{t_b}}.$$

3.4 Clustering Techniques

Correlation clustering

Correlation clustering does not require a bound of the number of clusters that the data is partitioned into; rather it aims to partition the data into the optimal number of clusters based on the similarity between the data points [2]. Basically, it aims to maximize the agreements by maximizing the similarity and minimizing dissimilarity within the same cluster, and minimizing the disagreements by minimizing the similarity and maximizing dissimilarity between clusters. Assume $G = (V, E)$ be a graph on a vertices with edge weights$c_e \geq 0$. Let $e(u, v) \in \{+, -\}$ be the label of the edge (u, v). The positive neighborhood of u is $N^+(u) = \{u\} \cup \{v: e(u, v) = +\}$ and the negative neighborhood of u is $N^-(u) = \{u\} \cup \{v: e(u, v) = -\}$. Let OPT represent the optimal clustering, and for a clustering C, we let C (v) be the set of vertices that are in the same cluster as v. Consider a clustering $C = \{C_1, C_2, \ldots, C_n\}$. A negative labeled edge inside a cluster is considered a negative mistake and a positive labeled edge between clusters is considered a positive mistake [2]. If our goal is to minimize disagreements, we minimize the weight of positive edges between clusters and absolute valued weight of negative edges inside clusters. When maximizing agreements we wish to maximize the weight of positive edges inside clusters plus the absolute valued weights of negative edges between clusters.

4. EVALUATION METRICS

Now, in order to evaluate clustering methods, the common information retrieval metrics precision, recall and f-measure will be used. Precision aims to evaluate the cluster based on the number of correct retrieved candidates out of the total number of retrieved candidates. Hence, we can calculate precision and recall. Precision aims to evaluate the cluster based on the number of correct retrieved candidates out of the total number of retrieved candidates. It can be calculated as follows:

$$Precision\ (cluster\ i) = \frac{\#\ of\ correct\ instance\ in\ cluster\ i}{total\ \#\ of\ cluster\ i\ instances}$$

Whereas, recall aims to evaluate the cluster based on the number of correct retrieved candidates out of the total number of correct instances in the dataset. It can be computed as follows:

$$recall\ (cluster\ i) = \frac{\#\ of\ correct\ instance\ cluster\ i}{total\ \#\ of\ correct\ instances\ in\ the\ dataset}$$

Now it can be possible to calculate the f-measure as follows:

$$f - mesaure = \frac{2 \times Precision\ \times Recall}{Precision + Recall}$$

5. EXPERIMENTAL RESULTS

First experiments aims to evaluate the corrleation clustering with Cosine similarity measure and Jaccard similarity measure in terms of the ability to identify categories of data log's queries. The results of precision, recall and f-measure will be stated with three values of threshold which are 0.3, 0.4 and 0.5. Table 2 shows these results of this phase. As shown in Table 2, the precision, recall and f-measure have been calculated for each threshold values. In fact, the greatest values of precision, recall and f-measure have been obtained by using 0.5 of threshold. Morever, the results obtained using correlation clustering with cosine simailarity outperforms correlation clustering with Jaccard simailarity on three clusters

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2 Results of correlation clustering with three clusters

		Correlation clustering		
		Precision	Recall	Fmeasure
Cosine	0.3	80%	76%	78%
	0.4	82%	80%	81%
	0.5	86%	83%	84%
Jaccard	0.3	78%	75%	76%
	0.4	80%	78%	79%
	0.5	84%	80%	82%

Second experiments aims to evaluate the correlation clustering with Cosine similarity measure and Jaccard similarity measure in terms of the ability to identify clustering of data log's queries with four clusters . The results of precision, recall and f-measure will be stated with three values of threshold which are 0.3, 0.4 and 0.5. Table 3 shows these results of this phase.

Table 3 Results of correlation clustering with four clusters

		Correlation clustering		
		Precision	Recall	Fmeasure
Cosine	0.3	81%	75%	77%
	0.4	85%	78%	82%
	0.5	88%	83%	85%
Jaccard	0.3	82%	78%	79%
	0.4	85%	81%	82%
	0.5	89%	84%	86%

As shown in Table 3, the precision, recall and f-measure have been calculated for each threshold values. In fact, the greatest values of precision, recall and f-measure have been obtained by using 0.5 of threshold. Morever, the results obtained using correlation clustering with cosine similarity is worse than that obtained using correlation clustering with Jaccard simailarity on four clusters.

Second experiments aims to evaluate the correlation clustering with Cosine similarity measure and Jaccard similarity measure in terms of the ability to identify clustering of data log's queries with four clusters . The results of precision, recall and f-measure will be stated with three values of threshold which are 0.3, 0.4 and 0.5. Table 4 shows these results of this phase.

Table 4 Results of correlation clustering with five clusters

		Correlation clustering		
		Precision	Recall	Fmeasure
Cosine	0.3	84%	80%	82%
	0.4	87%	82%	84%
	0.5	90%	85%	87%
Jaccard	0.3	84%	79%	81%
	0.4	87%	82%	84%
	0.5	91%	85%	87%

As shown in Table 4, the precision, recall and f-measure have been calculated for each threshold values. In fact, the greatest values of precision, recall and f-measure have been obtained by using 0.5 of threshold. Morever, the results obtained using correlation clustering with cosine simailarity is almost the same as that obtined using correlation clustering method with Jaccard simailarity on five clusters In general, the two similarity measures have obtained their best results when the number of clusters were 5 where Cosine has achieved 90%, 85% and 87% of precision, recall and f-measure respectively. Whereas, Dice has achieved 91%, 85% and 87% of precision, recall and f-measure respectively. Efficient clustering algorithm is developed to enhance the effectiveness and efficiency of the Personalization of Web Search. Clustering results change dynamically to detect the changes in user's interest and to reflect the personalized ranking of search results

6. CONCLUSION

This paper presents the results of an experimental study of the correlation clustering method in data logs document clustering. In addition, we evaluates the impact of the two similarity/distance measures

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(Cosine similarity and Jaccard coefficient) on the results of the correlation clustering algorithms. the results obtained using correlation clustering with cosine simailarity is almost the same as that obtined using correlation clustering method with Jaccard simailarity on five clusters In general, the two similarity measures have obtained their best results when the number of clusters were 5 where Cosine has achieved 90%, 85% and 87% of precision, recall and f-measure respectively. Whereas, Dice has achieved 91%, 85% and 87% of precision, recall and f-measure respectively. Efficient clustering algorithm is developed in this research to enhance the effectiveness and efficiency of the Personalization of Web Search. In the future work, we paln to extend this work to include other clustering methods such as hierarchical clustering and other similarity measures. To enhance the clustering methods for personalized web search engine , we plan to combine clustering algorithms with meta-heurstic search algorithm.

REFERENCES

[1] P. Brusilovski, A. Kobsa, and W. Nejdl, *The adaptive web: methods and strategies of web personalization* vol. 4321: Springer Science & Business Media, 2007.

[2] B. Smyth and E. Balfe, "Anonymous personalization in collaborative web search," *Information retrieval,* vol. 9, pp. 165-190, 2006.

[3] H. C. Lee and A. Borodin, "Cluster based personalized Search," in *International Workshop on Algorithms and Models for the Web-Graph*, 2009, pp. 167-183.

[4] B. Smyth, "A community-based approach to personalizing web search," *Computer,* vol. 40, pp. 42-50, 2007.

[5] A. B. Rajmane, P. M. Patil, and P. J. Kulkarni, "Personalization of Web Search Using Web Page Clustering Technique," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering,* vol. 9, pp. 2171-2176, 2016.

[6] K. W.-T. Leung, W. Ng, and D. L. Lee, "Personalized concept-based clustering of search engine queries," *IEEE Transactions on knowledge and data engineering,* vol. 20, pp. 1505-1518, 2008.

[7] M. Speretta and S. Gauch, "Personalized search based on user search histories," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, 2005, pp. 622-628.

[8] P. Ferragina and A. Gulli, "A personalized search engine based on Web-snippet hierarchical clustering," *Software: Practice and Experience,* vol. 38, pp. 189-225, 2008.

[9] N. Ramesh and J. Andrews, "Personalized search engine using social networking activity," *Indian Journal of Science and Technology,* vol. 8, p. 301, 2015.

[10] A. Kumar and M. Ashraf, "Personalized web search engine using dynamic user profile and clustering techniques," in *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on*, 2015, pp. 2105-2108.

[11] M. I. Jalgaon, "AN EFFICIENT AND NOVEL APPROACH FOR WEB SEARCH PERSONALIZATION USING WEB USAGE MINING," *Journal of Theoretical and Applied Information Technology,* vol. 73, 2015.

[12] M. Sah and V. Wade, "Evaluation of Personalized Concept-Based Search and Ranked Lists over Linked Open Data," in *UMAP Workshops*, 2014.

[13] M. Sah and V. Wade, "Personalized concept-based search and exploration on the web of data using results categorization," in *Extended Semantic Web Conference*, 2013, pp. 532-547.

[14] K. W.-T. Leung, D. L. Lee, and W.-C. Lee, "Pmse: A personalized mobile search engine," *IEEE transactions on knowledge and data engineering,* vol. 25, pp. 820-834, 2013.

[15] D. Sontag, K. Collins-Thompson, P. N. Bennett, R. W. White, S. Dumais, and B. Billerbeck, "Probabilistic models for personalizing web search," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 433-442.

[16] P. Jay, P. Shah, K. Makvana, and P. Shah, "Review on web search personalization through semantic data," in *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*, 2015, pp. 1-6.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[17] G. Bordogna, A. Campi, G. Psaila, and S. Ronchi, "Disambiguated query suggestions and personalized content-similarity and novelty ranking of clustered results to optimize web searches," *Information Processing & Management,* vol. 48, pp. 419-437, 2012.

[18] P. Kumar, R. S. Bapi, and P. R. Krishna, "SeqPAM: a sequence clustering algorithm for Web personalization," *International Journal of Data Warehousing and Mining,* vol. 3, p. 29, 2007.

[19] D. Jiang, J. Pei, and H. Li, "Mining search and browse logs for web search: A survey," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 4, p. 57, 2013.

[20] K. Vijayalakshmi and S. Jena, "Web Usage Classification and Clustering Approach for Web Search Personalization," in *Proceedings of the Sixth International Conference on Computer and Communication Technology 2015*, 2015, pp. 376-383.

[21] Y. Inoue, T. Imada, S. Doi, L. Chen, T. Utsuro, and Y. Kawada, "Clustering Search Engine Suggests by Modeling Topics of Web Pages collected with Suggests," in *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*, 2016, p. 50.

[22] J. G. Moreno, G. Dias, and G. Cleuziou, "Query log driven web search results clustering," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 777-786.

[23] J. Lai and B. Soh, "Personalized Web search results with profile comparisons," in *Third International Conference on Information Technology and Applications (ICITA'05)*, 2005, pp. 573-576.

[24] K.-J. Kim and S.-B. Cho, "A personalized web search engine using fuzzy concept network with link structure," in *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, 2001, pp. 81-86.

[25] C. Liang, "User profile for personalized web search," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, 2011, pp. 1847-1850.

[26] K. W.-T. Leung, D. L. Lee, and W.-C. Lee, "Personalized web search with location preferences," in *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*, 2010, pp. 701-712.

[27] P. Palleti, H. Karnick, and P. Mitra, "Personalized web search using probabilistic query expansion," in *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*, 2007, pp. 83-86.