ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2019.0264

A WEB PERSONALIZATION BASED ON THE SEQUENTIAL PATTERN MINING FOR IMPROVED WEB ACCESS

A. Vaishnavi and N. Balakumar
Department of Computer Applications, Pioneer College of Arts and Science, India

Abstract

The development of information technology, the web has created a big challenge for directing the client to the website pages according to their need. Web page personalization process user's query and retrieve the search results that corresponds to their interest. Accordingly, the option is to capture the intuition of the client and provide them a list of recommendation. The tedious work is, to find the user's intuition. The web master of an institution ought to utilize methods of web mining to fetch the user's intuition. The web usage mining is one the technique to find the users intuition. Web usage mining can provide patterns of usage to the organizations in order to obtain user profiles and therefore they can make easier the website browsing or present specific pages. The recommendation is one of the applications in web usage mining. Recommender systems area unit one of the most common and easily apprehensible applications. There square measure 2 major ways in which most of advice engines work. They can either rely on the properties of the things that every user likes, discovering what else the user might like. In this paper, we tend to propose a recommendation approach that recommends a number of web pages based on user's interest upon client's history, from the web log. In this approach, it brings the most accuracy of the web pages to be displayed for the user.

Keywords:

Web Usage Mining, Recommendation, Web Personalization, Web Log, Sequential Pattern Mining, Web Mining

1. INTRODUCTION

An enormous measure of data accessible on the World Wide Web [1]. There is a huge influence of method on our present life. Electronic devices, computers and multimedia are things and deal with every day. Particularly Internet is becoming most significant for closely everyone as it is one of the newest and most forward-looking media and confidently - the medium of the future. Therefore we assumed that it would be essential to the think about few good and bad features of how this medium influences us, what effects it has on our behaviour social and what the future will look like. The Internet has changed our life extremely; there is no doubt about that. There are numerous Internet benefits that present you the new medium significance. Web mining is the data mining methods application to the Web data. Web mining is quite useful to resolve the problem of finding out that how users are using web sites. It involves mining or analyzing logs and the steps that have to be follow to get significant data from web logs data collection, preprocessing, data enrichment and pattern analysis and discovery. Data mining techniques application to the WWW, referred to as web mining, has been focus of various present research projects and various papers. However, there is no recognized vocabulary, leading to confusion when comparing research efforts. The term Web mining has been used in two different distinct ways. The first one is as web content mining, the information discovery process from sources across the WWW. The second one is known as web usage mining, is the mining procedure for access patterns and user browsing. In this paper we describe Web mining and present summary of the numerous research problems, methods, and development efforts. The internet is flooded with a huge valuable and useless knowledge. It is most hard to describe valuable knowledge for a specific user which is variable from time to time. The valuable knowledge of one specific time may not be valuable on various times or a several condition. The web itself is concerning currently with newer technologies. Since internet is using freewares that receives structured, non-structured, ordered, non-ordered format to represent knowledge in the web, finding not only the relevant data but to plan them according to the user interest is also a key challenge presently and is called as web personalization. Web personalization is the tailoring methodology a website or content [2] of website to the necessities of all specific user or users set, taking data advantage attained by the exploration of the customer's navigational performance [3].Personalization means to provide facilities to all particular user in a tailored manner. The propagation of information on the internet has ended the personalization system an obligation. The personalization technique must have the capability to resolve the additional data problems and let the clienteles rehearse in any event effort to discover the information they need [4]. Web page personalization includes pleasing between people by finding their inclination and utilizing these discovered inclinations to find the most applicable substance to every user [5]. Main parts of the web personalization system are first, the classification and pre-processing of web information, besides [6] the extraction of relationships between and across various types of such learning lastly, the assurance of the activities that should be advised by a such a web personalization.

The organisation of the paper: Analysis of Existing literature about web page personalization based on user query and optimization techniques are presented in section 2. The purpose of web page personalization is described in section 3. Section 4 gives a detailed description of the proposed methodology along with the sequential pattern mining techniques. Section 5 explains the results the clustering and optimization work and the conclusion part is presented in section 6 along with perspective for this work.

2. LITERATURE SURVEY

Albanese [7] proposed an interesting solution based on pattern recognition techniques, in order to classify a web user, based on its interaction with the web site. Several experiments have been performed and their results have been discussed.

Mobasher et al. [8] discussed that the web is providing a direct communication medium between the vendors of products and services, and their clients. Coupled with the ability to gather

elaborate knowledge at the granularity of individual mouse clicks, this provides a tremendous opportunity for personalizing the web experience for clients. To provided several techniques in which the user preference is automatically learned from web usage data, by using data mining techniques.

Ratnakumar [9] discussed the requirements of web usage mining and the introduction of web 2.0 technology. Improving quality and extension of our models will be the following steps in our project.

Mehtaa et al. [10] mainly focused on the web mining types- Web content mining, web structure mining and web usage mining. Here after that, we have introduced the web mining techniques in the area of the Web personalization.

Ramesh et al. [11] discussed patterns generated by conventional web usage Mining methods do not provide explicit insight into the user's underlying interest and preferences. For this reason there is a need to incorporate semantic data in web usage model to understand web user's navigational behavior at conceptual level.

Patil et al. [12] propose a new method to offer a better web page recommendation system by using semantic enhancement and ontology, which are called as knowledge representation models for recommendation Purpose.

One representation is for domain information second model is for ontology. This classification is a conceptual prediction representation which is used to integrate the web usage and domain knowledge. It is used to form a weighted semantic net of frequently viewed pages or terms. This system is planned to predict next web page requirements of users through querying the information.

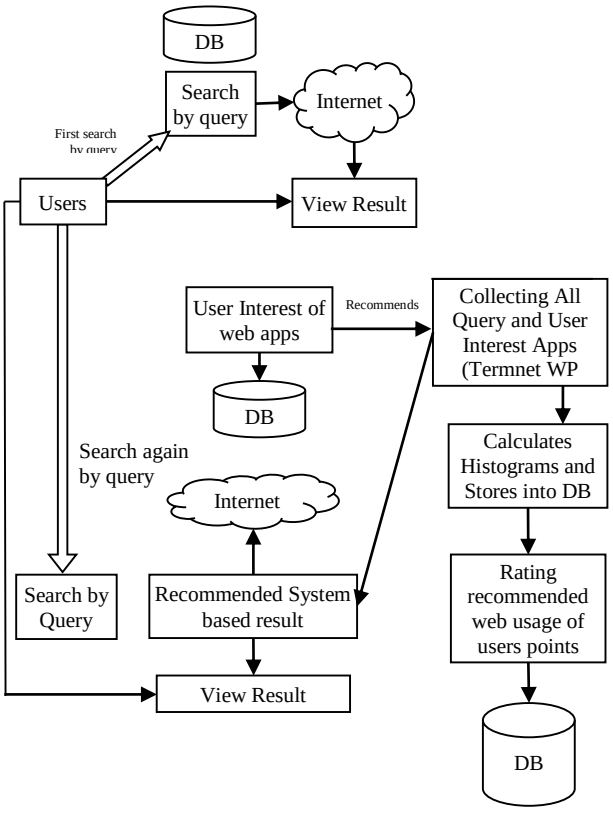

Fig.1. Diagrammatic representation of the proposed work

3. IMPORTANCE OF WEB PERSONALAIZATION AND RECOMMENDATION SYSTEM

Web personalization is strategy, marketing tool, and an a Personalization requires implicitly explicitly collecting visitor information and leveraging that knowledge in your content delivery Web personalization is considered to be the most automated method for customizing content to match users' needs. The recommendation approach uses Web usage analysis powered by an explicit representation of domain knowledge. The recommender system tries to automatically recommend hyperlinks that are deemed to be relevant to the user's interests, in order to facilitate access to the needed data on a large website.

4. METHODOLOGY

The aim of web personalization is to provide the personalized experience of different web pages based on known and relevant information of user query. Web-Page Recommendation Systems are becoming more useful and famous. Network is getting wider and wider. Websites are growing in numbers. So it is demanding task of the webmasters to organize the contents of the particular website in the network to gather the needs of the users. A routine personalization and recommender system technologies have become critical tools, because they help engage visitors at a deeper and more intimate level by tailoring the site's interaction with a visitor to her needs and interests. Next, to that, web personalization and recommender system is generated which contains the keyword with its explanation. Based on the user query, we combine the trained dataset and patterns generated database and find the similarity valued data is attained by the inspired optimization algorithm called sequential pattern mining.

4.1 DATA COLLECTION

The first step is to collect the data from the Syskill and Webert Web Page Ratings which contains the information related to web pages and recommendations to their ratings of an individual user. The four separate subjects on the web pages are Bands- recording artists; Goats; Sheep; and Bio-Medical.

4.2 DOMAIN KNOWLEDGE AND SEMANTIC INFORMATION

The classical advancement of the web usage mining process is extended with new steps, by desegregation domain knowledge in the form of ontology. In this work content, it is assumed that domain knowledge is available in the form of domain ontology provided by the ontology engineer during the design of the web site.

A core ontology is represented as a,

$$5\text{-tuple: } = \{C, R, Hc, rel, A\}$$

where, C is a set of concepts which represent the entities in the ontology domain; R is a set of relations, defined among the concepts; Hc is a taxonomy or concept hierarchy, which defines the is-a relation among concepts; rel element corresponds to a function, rel: $R_C \times C$ that specifies the relations on R. A is a set of axioms usually expressed in a logical language. For example the

set *C* contains product, purchase, supplier, and warehouse as some of the concepts of the domain ontology considered in our model.

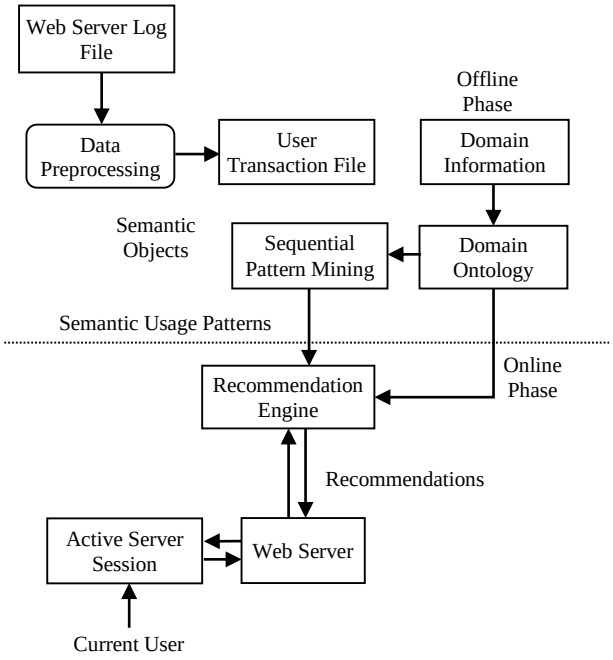

Fig.2. Proposed Recommendation Model incorporating Semantic information in Web Usage Mining process

4.3 LEARNING MODULE

Learning requires a set of positive models of user question (for example, web pages one is interested in) and negative examples (for example, web pages one is not interested in). In this paper, we learn a concept that recognizes pages by rating it by the user inclinations. A learner model needs to perpetually search the relevant data on the domain skills i.e. learn the data from the subjects bands, goats, sheep and biomedical. It will responsible for creating as well as updating the user profile by web page rating.

4.4 SEQUENTIAL PATTERN MINING

Sequential Pattern Mining technique is applied over the semantic space to discover the frequent sequential patterns. The frequent navigational patterns are extracted in the form of Ontology instances instead of Web page views and the resultant semantic patterns are used for generating Web page recommendations to the user. The Semantic Web is based on a vision of Tim Berners-Lee, the inventor of the WWW. The Semantic Web enriches the WWW by machine - processable information which supports the user in accomplishing his tasks more easily [13]. Berendt et al. [3], was the first to explore Semantic Web Usage Mining. The authors have elaborated different ways of how the fields of Semantic Web and Web Mining can cooperate. The first part of the work is on extracting semantics from Web page. The second part is on the improvement of Web Usage Mining by using semantics structures in the form of ontology. Subsequently the authors sketched out the benefits of combining Semantic Web and Web Mining. They further elaborated a process for learning ontologies by mining the Web. They emphasized that by constructing a pattern space over ontology, navigation primitives and Web Mining methods one can find patterns which follows a semantic approach [14]. A variation of Apriori algorithm, enhanced by semantic information for generating frequent sequences [15] and Apriori_Join() for candidate generation is described. Onto SPM generates semantically rich frequent sequences. Apriori_Join procedure uses Semantic distance to prune the candidate sequences, such that if the semantic distance between two (k-1) sequences is more than an allowed maximum semantic distance, then the candidate k-sequence is pruned from the search space without the need for support counting. OntoSPM algorithm and Apriori_Join procedure is given below. The definitions and terms used in this algorithm are from [16].

Algorithm 1 Ontology Based Sequential Pattern Mining

OntoSPM (*S, M, _, min_sup*)

Input: Sequence database *S*, /* sequence of semantic objects */ Semantic distance Matrix *M*, Maximum semantic distance _, /*_: Maximum Semantic distance to prune the search space */, Minimum support *min_sup*

Output: Semantic-rich Frequent Sequences

Algorithm:

Step 1: Scan database *S* to find the set of frequent 1-sequence $L_1 = \{s_1, s_2, \ldots s_n\}$, $k = 1$, $C_1 = L_1$

Step 2: for (k=2; L_k - 1_Ø; k++)

Step 3: for $L_{k-1} \bowtie L_{k-1}$ do /*L_{k-1} join L_{k-1}*/

Step 4: There exist s_i, s_j such that s_i, s_j _ L_{k-1}

Step 5: C_k_$C_k \cup$ Apriori_Join(L_{k-1},*M*, _)

Step 6: end for

Step 7: $L_k = \{c\_C_k \mid c.support\_min\_sup\}$

Step 8: end for

Step 9: return $L = U_k L_k$

Step 10: End

Function Apriori_join () implementation is a variation of the join procedure of the sequential pattern mining algorithm (GSP).

Function *Apriori_join* (L_{k-1}, *M*, _)

Step 1: C_1=Ø

Step 2: for all *P*, *Q*_ L_{k-1} with $P = \{i_1, i_2, \ldots, i_{k-2}, i_{k-1}\}$ and $Q = \{i_1, i_2, \ldots, i_{k-2}, i'_{k-1}\}$and $D(i_{k-1}, i'_{k-1})$_ _ /* $D(i_{k-1}, i'_{k-1})$: defines the semantic distance between i_{k-1}, i'_{k-1}*/, /* _: Maximum Semantic Distance - Maximum allowed semantic distance between any two semantic objects. _ is a user defined value and can be determined as specified in [18]. $D(i'_{k-1}, i'_{k-1})$ is derived from *M* */

Step 3: c = $\{i_1, i_2, \ldots, i_{k-1}, i'_{k-1}\}$

Step 4: C_k _$C_k \cup \{c\}$

Step 5: return C_k.

Step 6: end for

4.5 SIMILARITY MEASURES

4.5.1 Online Recommendation Phase:

The aim of a recommender system is to determine which web pages are more likely to be accessed by the user in the future. In this phase active user's navigation history is compared with the discovered sequential Association rules in order to recommend a

brand new page or pages to the user in real time. Generally not all the items in the active session path square measure taken under consideration whereas creating a commendation. A very earlier page that the user visited is less likely to affect the next page since users generally make the decision about what to click by the most recent pages. Therefore the concept of window count is introduced. Window count parameter '*n*' defines the maximum number of previous page visits to be used while recommending a new page. Semantic distance between objects [16] is taken into consideration to solve the ambiguity problem.

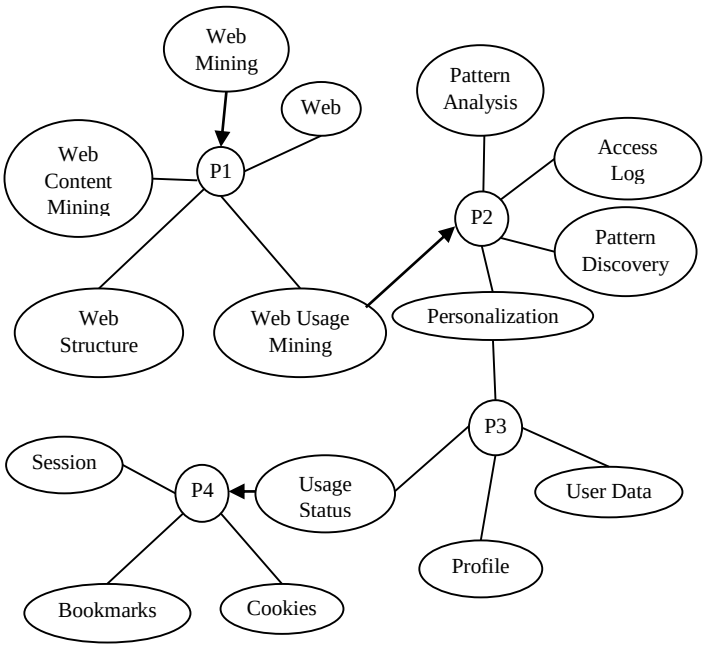

Fig.3. Tracking the user interest through preference network

5. RESULT AND DISCUSSION

The proposed work integrates main knowledge in the form of ontology in all the phases of web Usage Mining process. The generated patterns area unit in terms of ontology instances instead of web page addresses. Such patterns can extract the semantic relatedness between the visited web pages. The discovered semantic rich sequential association rules form the core knowledge of the recommendation engine of the proposed model. Compared with the conventional web usage based mostly recommendation system, our proposed model shows promising results.

5.1 PERFORMANCE MEASURE

Precision is defined as the proportion of the number of relevant recommendations to the number of all recommendations.

$$P = \frac{n(Ret_{Rel})}{n(Ret)} \quad (1)$$

where, $n(Ret_{Rel})$ - Number of relevant pages retrieved and $n(Ret)$ - Number of pages retrieved

Recall gives information about the accuracy of the proposed system. The recall is defined as the ratio of a number of relevant pages retrieved to the total number of relevant pages in the database.

$$R = \frac{n(\mathrm{Re}t_{Rel})}{n(Db_{Rel})} \quad (2)$$

where, Db_{Rel} is the total number of relevant pages in the data base

F-measure: F-measure is a measure of a harmonic mean of both precision and recall. It is described as in Eq.(3).

$$F-measure = 2.\frac{precision..recall}{precision + recall} \quad (3)$$

5.2 PERFORMANCE ANALYSIS

On comparing the algorithms, the proposed model Semantic web mining with sequential pattern and recommendation system contains high precision for all the clusters.

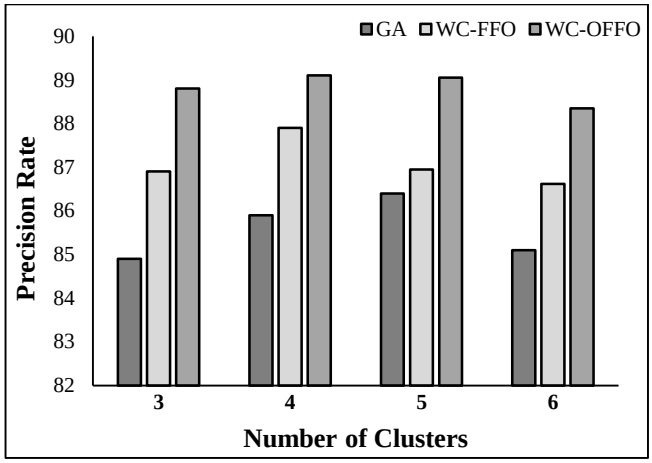

Fig.4. Precision

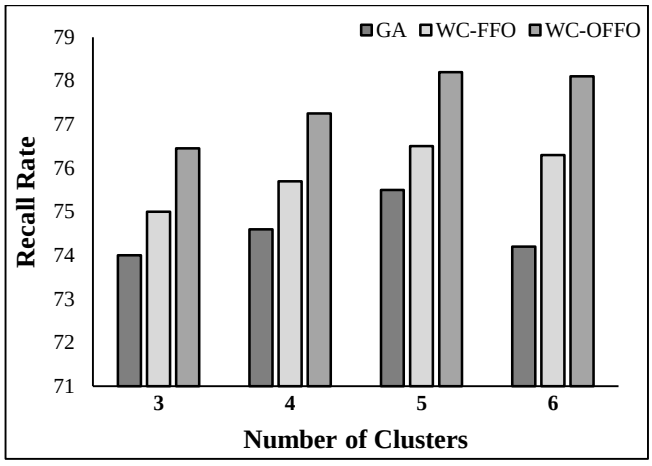

Fig.5. Recall

By using these algorithms, the optimal or maximum similarity measures between user query data among the four subjects such as band, goats, sheep and bio-medical are calculated. The highest recall of 78.09 is attained in the 5th cluster for the proposed WC-OFFO algorithm and the maximum f-measure of 83.25 is achieved in the 5th cluster. By analyzing these measures (recall and f-measure), the efficiency of optimization algorithms is evaluated.

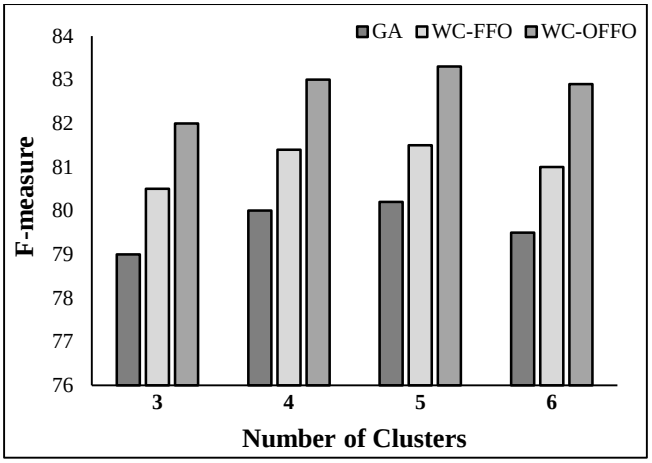

Fig.6. F-measure

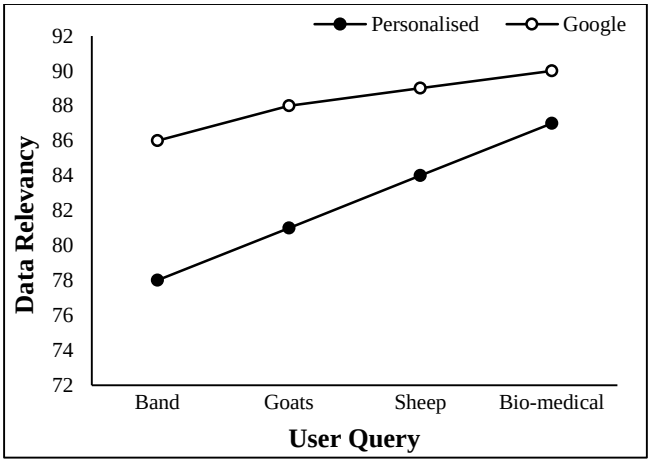

Fig.7. Comparative Analysis

6. CONCLUSION

In this paper we survey the research area of Web mining, focusing on the category of Web mining. We also discussed the different algorithms used in web mining. Web mining deals with retrieving the data from web with best output. The web mining also deals with many algorithms that lead to fetch the data from any website. In general web usage mining is that retrieves the data from website for online user in effective manner. Since this is a vast area, and there a lot of work to do, we hope this paper could be a useful starting point for identifying opportunities for further research. Patterns generated by conventional web Usage mining methods do not provide explicit insight into the user's underlying interest and references. Hence there is a need to incorporate semantic information in web usage model to understand web user's navigational behavior at conceptual level. This motivated U.S.A. to propose the semantically enriched web usage model. The data with maximum similarity measures were clustered by choosing the Cluster Head. Web page personalization is done by comparing the similarity measures of trained dataset. The proposed work integrates domain knowledge in the form of ontology in all the phases of web Usage Mining process. The generated patterns are in terms of ontology instances instead of web page addresses. Such patterns will extract the linguistics relatedness between the visited sites. The discovered semantic rich consecutive association rules type the core information of the advice engine of the planned model. Compared with the conventional web usage primarily based recommendation system, our projected model shows promising results.

REFERENCES

[1] X. Wang, L.T. Wang, X. Xie, J. Jin and M.J. Deen, "A Cloud-Edge Computing Framework for Cyber-Physical-Social Services", *IEEE Communications Magazine*, Vol. 55, No. 11, pp. 80-85, 2017.

[2] Z. Maamar, G. Alkhatib and S.K. Mostefaoui, "Context-based Personalization of Web Services Composition and Provisioning", *Proceedings of 30th Euromicro Conference*, pp. 1-8, 2004.

[3] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization", *ACM Transactions on Internet Technology*, Vol. 3, No. 1, pp. 1-27, 2003.

[4] S. Khonsha and M.H. Sadreddini, "New Hybrid Web Personalization Framework", *Proceedings of IEEE 3rd International Conference on Communication Software and Networks*, pp. 233-238, 2011.

[5] V. Viswanathan and K. Ilango, "Ranking Semantic Relationships between Two Entities using Personalization in Context Specification", *Information Sciences*, Vol. 207, pp. 35-49, 2012.

[6] M. Sah and V. Wade, "Personalised Concept-Based Search on the Linked Open Data", *Journal of Web Semantics*, Vol. 36, pp. 32-57, 2016.

[7] M. Albanese, A. Picariello, C. Sansone and L. Sansone, "A Web Personalization System based on Web Usage Mining Techniques", *Proceedings of 13th International World Wide Web Conference on Alternate Track Papers and Posters*, pp. 288-289, 2004.

[8] Bamshad Mobasher, Robert Cooley and Jaideep Srivastava, "Automatic Personalization Based on Web Usage Mining", *Communications of the ACM*, Vol. 43, No. 8, pp. 142-151, 2000.

[9] A. Jebaraj Ratnakumar, "An Implementation of Web Personalization using Web Mining Techniques", *Journal of Theoretical and Applied Information Technology*, Vol. 8, No. 1, pp. 67-73, 2010.

[10] Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki, "Web Personalization Using Web Mining: Concept and Research Issue", *International Journal of Information and Education Technology*, Vol. 2, No. 5, pp. 1-7, 2012.

[11] C. Ramesh, K.V. Chalapati Rao and A. Goverdhan, "A Semantically Enriched Web Usage Based Recommendation Model", *International Journal of Computer Science and Information Technology*, Vol. 3, No. 5, pp. 23-32, 2011.

[12] Arundhati Patil, "Personalized Web Page Recommendation Using Ontology", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 3, No. 7, pp. 4431-4436, 2013.

[13] G. Stumme, B. Berendt and A. Hotho, "Usage Mining for and on the Semantic Web", *Proceedings of International Conference on Data Mining: Next Generation Challenges and Future Directions*, pp. 461-480, 2004.

[14] B. Berendt, A. Hotho and G. Stumme, "Towards Semantic Web Mining", *Proceedings of International Conference on Semantic Web*, pp. 264-278, 2002.

[15] R. Agrawal and R. Srikant, "Mining Sequential Patterns", *Proceedings of 11th International Conference on Data Engineering*, pp. 3-14, 1995.

[16] Nizar Mabroukeh and C.I. Ezeife, "Using Domain Ontology for Semantic Web Usage Mining and Next Page Prediction", *Proceedings of 18th ACM Conference on Information and Knowledge Management*, pp. 1677-1680, 2009.