IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-12 E-ISSN: 2347-2693

A Point of Two Mode-Session Logs Based Web User Interest Prediction System From Web Search Engine

R. Velmurugan[1*], S.P. Victor[2]

[1*] Kristu Jayanti College (Autonomous), Bangalore University, Bangalore, India
[2] St. Xavier's College (Autonomous), Manonmaniam Sundaranar University, Tirunelveli, India

[*]Corresponding Author: velmurugan13777@gmail.com, Tel.: +91-98458-00376

Available online at: www.ijcseonline.org

Received: 14/Nov/2017, Revised: 30/Nov/2017, Accepted: 15/Dec/2017, Published: 31/Dec/2017

Abstract— Information prediction from web search is a hands-off process based on user interest. By differently user may think based on the knowledge or relevant search things, but web mining tasks are complex to understanding the user behaviour and knowledge to retrieve the right things to the user. To propose a new intent interest prediction for improving the efficiency, we propose a session timing preference based two mode user interest prediction methods called semantic log pre-fetch clustering (SLCPC) algorithm and personalized feed ranking (FPR) algorithm to deduce a set of related categories for each user query based on the retrieval history of the user session, i.e. different contents have to be placed for different users according to the user relevant search profiles spending time on interested logs. Two modes analyse behaviours like time of visit, navigation URL, weblogs, and user actions on the webpage. SLPC predicts web log data of the user and also identifies the implicit behaviours performed by the user. The identified information is used to identify the user interest and seeds are generated by logging page. So the user doesn't wait for long more time to search our interesting factor frequently. The web users are clusters based on the feed ranking interest which is used continues searches. Our proposed system improves the accuracy of personalized user search compared to an existing approach.

Keywords— Personalization, Prediction mining, Web search, Clustering, Session logs

I. INTRODUCTION

General web search is performed predominantly through text queries to search engines. Because of the enormous size of the web, text alone is usually not selective enough to limit the number of query results to a manageable size. The user interest be calculated dissimilar reviews between users ranks, ratings, it user visited pages in search engine to take advantage of the linkage structure of the web to compute the importance of user interest in rankings can be used to influence the ranking of search results. To encompass different notions of search importance for different users and queries may vary. The basic Page Rank algorithm can be modified to create "personalized views" of the web, redefining importance according to user preference. For example, a person can also want to specify his bookmarks as a set of desired pages, so that any question results which are crucial with recognizing to his bookmarked pages could be ranked higher. At the same time as experimentation with the use of personalized Page Rank has proven its utility and promise, the dimensions of the internet make its practical awareness extremely tough. This can be evaluated by PageRank algorithm and its extension for personalization search by the user. The essential motivation underlying PageRank is the recursive notion that crucial pages are that related-to by means of many important pages.

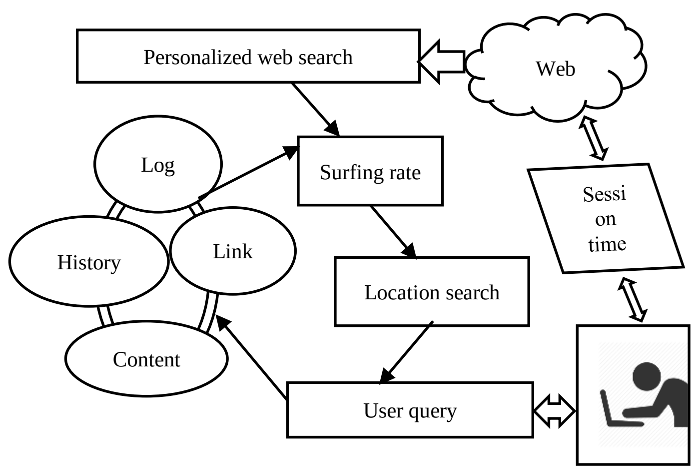


Figure 1. Structure of Personalized Web Search

A page with only two in-links, for example, may seem improbable to be an important page, but it may be significant if the two referencing pages are the search engine, which themselves are important pages since they have numerous in-links. One way to formalize this recursive idea is to use the "random surfer" model presented as the search engine.

Our important contributions to develop user behaviour based interest surfing with higher performance with interest pointed pages from seeking engine. Consider that trillions of random surfers are browsing the net: if at a sure time step a surfer is asking at web page p, at the subsequent time step he appears at a random out-neighbour of p. As time goes on, the predicted percentage of surfers at each page p converges (beneath sure situations) to a restrict r(p) that is unbiased of the distribution of starting factors based on the retrieval process. Intuitively, this restriction is the page rank of p and is taken to be an importance rating for p, since it displays the number of people expected to be searching at p at any one time interest rate to predict information. This implementation precise in detailed process of semantic prefect log process problems analysed in section 2 as literature review, follows to provide our new intents of implementation in section 3and our resultant test case proves the performance in section 4.

II. LITERATURE SURVEY

Most prevailing user outlining approaches are founded on substances that users are attentive in (i.e., positive preferences), but not the objects that users dislike (i.e., negative preferences). In review mainly focus on search engine personalization and develop several concept-based user profiling methods that are based on both positive and negative preferences [1]. Most clusters are intended as the proposed methods besides our before and projected modified query clustering method.

A clustering approach is that captures the user's conceptual preferences with the intention to provide personalized query pointers. We achieve this intention with two new techniques. Online strategies that extract concepts from the net-snippets of the search result returned from a question and used the principles to identify associated queries for that query [2]. 2d, we propose a new two-segment personalized agglomerative clustering algorithm that is capable of generating personalized question clusters. To the nice of the authors' knowledge, no preceding work has addressed personalization for question guidelines especially, we build consumer profiles primarily based on activity at the search webpage itself and look at the usage of those profiles to provide personalized seek effects. by implementing a wrapper around the Google search engine [3], we had been able to gather information about individual consumer seek activities. Specifically, we gathered the queries for which as a minimum one search result changed into tested and the snippets (titles and summaries) for each tested result. User profiles were created by means of classifying the accrued statistics (queries or snippets) into ideas in a reference idea hierarchy. Users are uncomfortable with exposing personal desire information to search engines [6]. However, privacy isn't absolute, and often can be compromised if there's a benefit in service or profitability to the user.

Classical net utilization mining does not take semantic know-how and content material into sample generations. Latest researches display that ontology, as past historical understanding; can improve sample's excellent [7].

This user choice aims to lay out a hybrid recommendation device based on integrating semantic information with net utilization mining and net page clustering based totally on semantic similarity. The internet directory is considered as a thematic hierarchy and personalization is found out via constructing person community models by usage information. We enhance the clustering and probabilistic techniques supplied in prediction rates and additionally intents a new optimization with new set of rules that mixes those strategies [9]. The resulting community fashions take the form of community internet directories. The proposed personalization method is evaluated each on a specialized synthetic and a popular-motive web directory, indicating its ability cost to the internet consumer. To consider a Web event as a system composed of different keywords, and the uncertainty of this keyword system is related to the uncertainty of the particular Web event [10]. Based on keyword association linked network Web event representation and Shannon entropy, we identify the different levels of semantic uncertainty and construct a semantic pyramid (SP) to express the uncertainty hierarchy of a Web event.

It makes tough to discover exact search result in step with person choices. In this paper, we proposed a way for personalized net seek [13]. Customized netseeks any action taken to optimize the quest result according to user's character choices.

Session hobby concept is defined as a pair of quantity and purpose where the quantity covers a set of documents decided based on user seeking out comes, and the cause covers a set of keyword features extracted from the chosen files [18]. And, with the search case intention to make a concept network grow is to calculate the similarity between a new concept and existing principles, and to this end.

III. IMPLEMENTATION OF PROPOSED SOLUTION

Personalization of web search is to carry out retrieval for each user incorporating his/her interests. For a given query, a personalized Web search can provide different search results for different users or organize search results differently for each user, based on their interests, preferences, and information needs. There are many personalized web search algorithms for analyzing the user interests and producing the outcome quickly; User profiling, Hyperlink Analysis, Content Analysis and collaborative web search are some of the instances for that kind of algorithms. In this paper, we are analyzing various issues of personalized web search.

We provide a formal framework to investigate the problem of learning a user's interest based on his/her past click history. As part of this framework, we propose a simple yet reasonable model on how we can succinctly represent a user's interest and how the interest affects his/her web click behavior. Based on the formal user model, we develop a method to estimate his/her hidden interest automatically based on his/her observable past click behavior. We provide theoretical and experimental justification of our estimation method. Finally, we describe a ranking mechanism that considers a user's hidden interest in ranking pages for a query based on the user interest. We conduct a user survey to evaluate how much the search quality improves through this personalization.

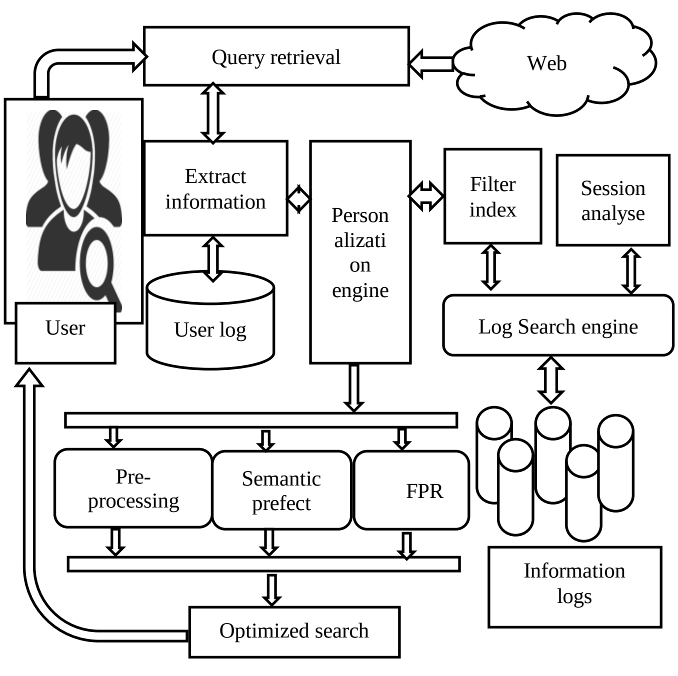


Figure 2. Architecture Diagram for Proposed Solution

Our proposed approach is a form of client-side personalization based on an interest-to-taxonomy mapping framework and result categorization. It piggybacks on a standard search engine such as search tool and categorizes and displays search results on the basis of known user interests. As a novel feature of our approach, the mapping framework automatically maps the known user interests onto a set of categories in a Web directory.

3.1 PRE-PROCESSING

It's a process of the procedure of splitting a string of written language into its phrases. Textual content statistics includes a block of characters called tokens. It's much used to do away with noisy, inconsistent and incomplete facts for doing the classification; textual content pre-processing and function extraction is an initial phase. So the files are being separated as tokens and were used for in additional processing. Elimination of stop words are the phrases which are needed to be filtered i.e., can be before or after natural language processing. Represent words are phrases which contain little informational. Various equipment mainly keeps away from to put off those stop phrases a good way to aid phrase search. Numerous collections of words may be chosen as preventing words for any purpose.

Step 1: load web dataset $D_{s.}$

Step 2: Identify No of items N.

Step 3: Initialize Occurrence words OM.

Step 4: count number of lexical words T in Ds.

Step 5: for each item Ii in Ds

Noc=Count Number of occurrence in all records T.

OM[Ii]=Noc.

Store Noc in Occurrence words.

End.

Step 6: End dataset

Some search engines remove most of the commonplace phrases which include lexical phrases inclusive of "need" from a textual content on the way to improve performance. Seek engine or natural language processing may additionally comprise a ramification of stop phrases. it includes English prevent phrases such as “and”, “the”, “a”, “it”, “you”, “may also”, “that”, “i”, “an”, “of” etc. that are taken into consideration as ‘useful words’ as they don't have meaning.

3.2 USER SESSION TIME QUERY EVALUATION

To evaluate the behavior objective of user session time is to extract and identify the main learning goals and higher-level learning tasks from a session log. Typically, a learner focuses on one or two main activities in a session. Using a waiting occurrence approach we can identify the main learning objectives by looking at the predominant types of interaction with the system.

Semantic pre-fetch link prediction

Step 1: Read total No of items N links.

Step 2: Initialize Transition matrix TR search link with size

TRm×n.

Step3: Read occurrence matrix OM.

Step4: for each item Ii in item set I

for each item Ij in item set I

Identify number of occurrence of Ii from OM link.

Q= OM(Ii);

Identify number of occurrence of Ij from OM link.

R= OM(Ij);

Compute probability of transition from Q to R as follows.

Pb = Q/R.

Assign Pb in the transition matrix TR.

TR(I,j)=pb.

End

End

Step5: End

User spends substantial session time on a particular activity or submits a high number of requests for these resources and activities, and then this activity is a manifestation of particular session the user interesting state. The requests of pages of the individual classes are counted.

3.3 FREQUENCY ESTIMATION OF THE WEB PAGE

More frequent terms are more significant than less frequent terms. A document that contains a matching term a number of times will be more related to a user's interest than a document that has the matching term only once. We estimate the probability, P(Eti), of a matching term ti at frequency Eti in a web page to measure the significance of the term. In general, frequent terms have a lower probability of occurring term which occurs in many documents is not a good discriminator, and has less significance than one which occurs in few documents the same reasoning for Inverse Document Frequency (IDF).

$$P(Eti)=\frac{\text{number of web pages Eti that contain the term ti}}{\text{total number of returned web pages}} - (\text{number of distinct terms}) \quad (1)$$

We measure term specificity by estimating the likelihood P(Eti) of the matching term, ti, appearing in the documents.

3.4 PERSONALIZED FEED RANKING (FPR)

We calculate total image term scores and combine them with the public score to get another rank order. After that, we compute the probability of relevant/retrieved web pages belonging to certain based on term scores and image term scores separately. The probability of relevant/retrieval, P(relevance)/P(retrieval), is estimated as:

Personalized Relevance score,

$$p = \frac{\text{number of relevant web pages in a bin}}{\text{total number of relevant web pages in query results}} \quad (2)$$

Retrieval score,

$$Rp = \frac{\text{number of web pages at top 10 ranks in a bin}}{\text{top rank items}} \quad (3)$$

Step1: start

Step2: read webpage Ts.

Step3: identify attribute set As.

Step4: compute number of transaction TN.

Step5: compute combination of possible patterns Ps = ((TN X §)/(K X(TN – K) X §)).

k- Specific pattern.

Step6: for each pattern Pi from Ps

Compute count = ø(Pi£(Ps)).

Ø- Number of pattern pi contained in ps.

Compute support = count/TN.

If support > ST then

Add Pi to Sanitization set Ss.

End

End

Step7: return Ss

Step8: Stop

To generate relevance score p(retrieval)−p(relevance) obtained from the net pages against the median web page accrue in a bin based on term compute cluster ratings and based on web pages p(retrieval) is higher than p(relevance) and for the pages relevant from retrieval. Primarily based on excessive scoring internet pages still have a bias to be ranked higher and shorter web pages to be ranked lower. While initial, our survey result indicates vast improvement within seeks satisfactorily. We have a look at about 25% development over the exceptional existing technique demonstrating the potential of our technique in personalizing web seek.

3.5 PRE-FETCH PERSONALIZATION CLUSTER WEB LINKS

To make use of the pre-fetch personalization level of context to personalize seeking user consequences by re-ranking the consequences returned from a seek engine for a given query assuming a semantic consumer profile with evaluating the semantic scores exists, and we've got a set of seeking consequences, algorithm is applied to re-rank the quest effects based on the user interest scores, user preference score, and the semantic cluster score.

Input: cluster listing, semantic database, web cluster database

Output: ranked output: ranked personalized list steps:

1. Get the input query.
2. Retrieve the ontology tree node, on, matching the question.
3. Get all leaf nodes, upload weight to them.
4. Get person query
5. Retrieve all information in which question fits the web cluster database keyword.
6. Upload weights to those clusters
7. Get consumer personalization –semantic ranking ratio.
8. Multiply the semantic listing with the ontology ratio.
9. Multiply the customized listing personalization ratio.
10. Merge the semantic and personalization listing
11. Healthy the list with the semantic cluster listing obtained from set of rules.
12. Discard non-matching clusters.
13. Re-rank the top order list.

As a result, suitable weights are added to the cluster depending on given personalization level. Therefore the use of dynamic consumer profile and ontology cluster list to attain a ranked cluster listing which satisfies the user interest.

3.6 PRACTICAL MEASURES

Our evaluation used precision, mean feed rank, normalized cumulative grade point, and average score. We computed these measures separately for combined log futures. This measure required that the top predicted category user behaviors for a context value matched its top actual search in the specified future duration. If so, the user interest model would have been given with a score 1 and 0 otherwise. The overall scores context of user search was then averaged to provide a final score for each set.

This measure finally compared top predicted category label for a context analysis with educational web labels (i.e., data mining, networking, image processing related domains), in the specified future duration. If there was a match, the user interest model would be given a rank score 1, and 0 otherwise. Scores were averaged to compute final scores as before division by sum of the category. Assumes that at most one label prediction would be used in a real system, but correctly predicting any of three dominant interest as well calculated results. The Final Results are based on top-rated scores as categorized.

IV. RESULTS AND DISCUSSION

The results are carried out to test by UCI web link repository with search engine framework. The proposed personalized webs search be implemented pre-fetch clustering (SLCPC) algorithm and personalized feed ranking (FPR) algorithm to deduce a set of related categories for each user query based on the retrieval history of the user session. The proposed method has produced efficient results on clustering and improved the performance also. Parameters are tabulated given below.

Table 1. Details of Dataset

Parameter	Value
Number of services	10
User logs	15
Datasets used	Web resources

The Table 1, shows the details of data set being used to evaluate the performance of the proposed multi-attribute opinion rate support measure based approach. The figure given below shows the evaluation of user interests using java net beans framework with objective weblog datasets.

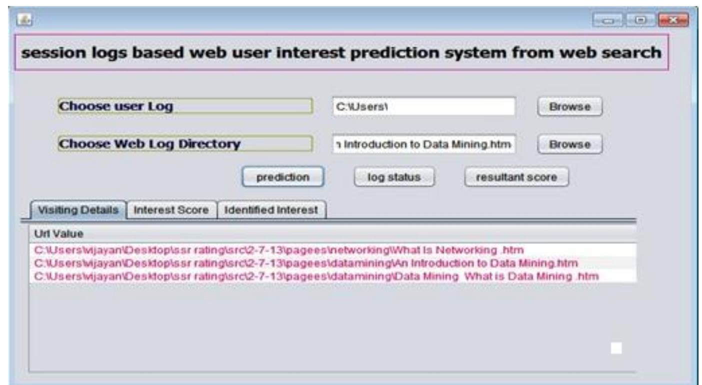


Figure 3. User Log Visiting Details

We performed a comparison of the predictive accuracy of user interest models generated based on only and the five sources of contextual evidence like. Above figure shows the results of user visit logs for each of the interest models, at each future time duration being used by the users continuously. Evaluation measures were computed over each experimental set, and the results averaged as interest score.

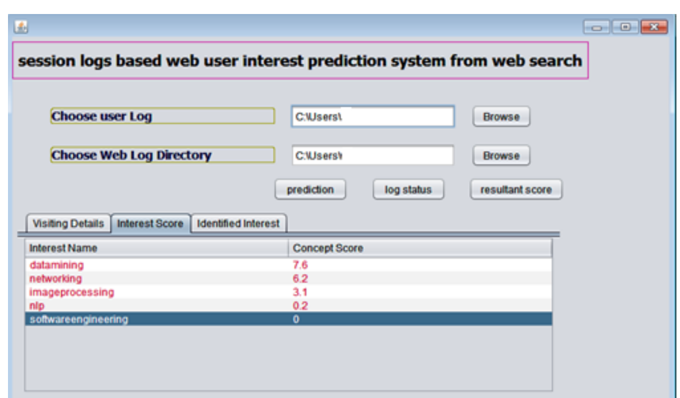


Figure 4. Top Feed Rank Interest Score

The list of predicted logs that generated by the user interest score is shown in above figure 4 is then compared to the ideal vector of relevant top score other searches, and a discounted cumulative gain score is computed using a standard discount log factor of the Computed score. Our modification of the standard computation was to restrict the depth of the comparisons between the two label vectors to the minimum length as least score. The score was then normalized by dividing it by the maximum possible search domain value that could be obtained to this depth. The scores overall context trails were averaged to provide a final score for each set.

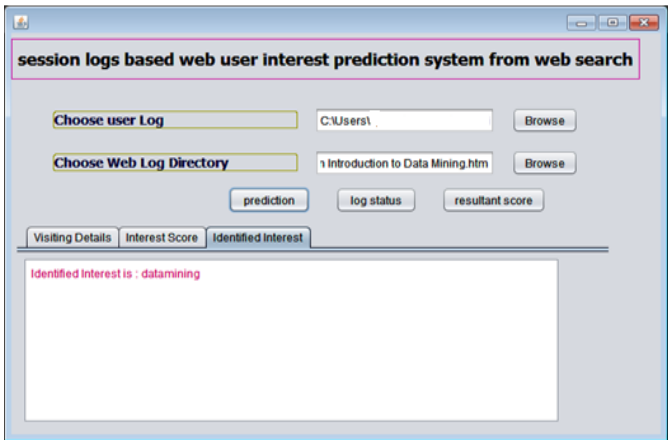


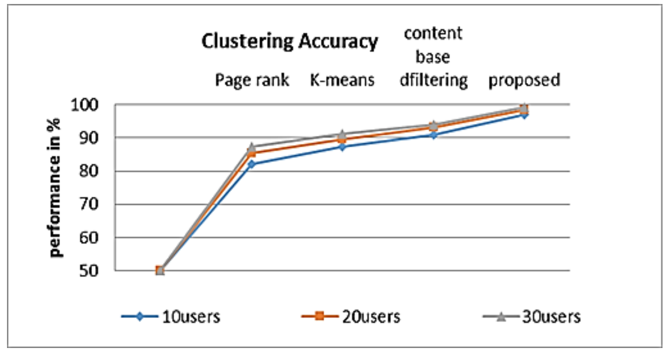
Figure 5. Resultant of User Interest Identified

The above figure shows the results that the interaction context predicts user interests most accurately in the time immediately by user way of behaviour from log history. This is likely because does not represent the beginning or the end of the current task, and the interaction that occurs before and after is task-related. The findings show that the interests of the user within previous search log by most accurately predicted by the task context, suggesting that the active work task may be lengthy. The findings also show that the long-term interests of the user are most accurately predicted by the historical context of the user, but also the social context comprising the interests of other users who also visit.

The performance of SLPC is evaluated through clustering accuracy (cs), precision rate, and recall rate and time complexity.

Clustering accuracy,

$$cs = \sum_{k=0}^{k=n} \times \frac{\text{total number of cluster group dataset(Cds)predictedlinks}}{\text{Total related datsetsTr)}} \quad (4)$$



Graph 1: Comparison on Clustering Accuracy

The Graph 1, shows the comparison of clustering accuracy and shows that the proposed method has produced higher clustering accuracy than other methods.

Table 2. Comparison of Clustering Accuracy

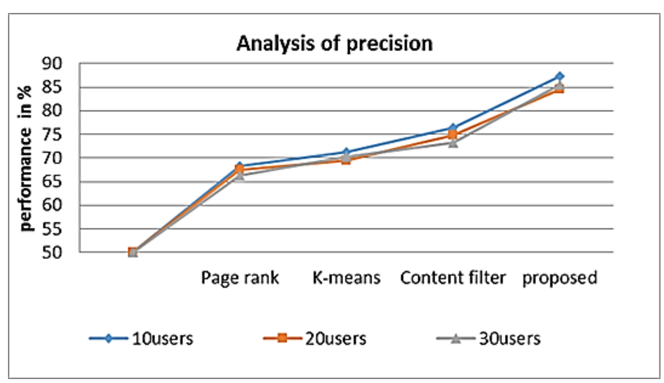
Methods/ number of records	Page rank	K-means	Content -based filtering	Proposed SLCPC
10 users	82.2	87.3	91.1	96.1
20 users	85.4	89.5	93.2	97.5
30 users	87.4	91.3	94.1	98.1

The Table 2, shows the comparison of clustering accuracy produced 10 users as 96.1%, 20 users as 97.5% and 30users as 98.1 % shows that the proposed approach has produced higher clustering accuracy.

Analysis of precision rate:

Precision, Pr is defined as the proportion of a total number of relevant URL links and the total number of retrieved URL links, where R is the relevant URL links calculated manually and A is the total number of retrieved URL links.

$$\text{Precision, Pr} = \frac{\text{Relavant links (R)}}{\text{Total number of retrived Links (A)}} \text{ x } 100 \quad (5)$$



Graph 2: Comparisons of Precision Rate

The Graph 2, shows the comparison of precision rate produced by different methods and the proposed method has produced higher performance rate than other methods.

Table 3. Comparisons of False Classification

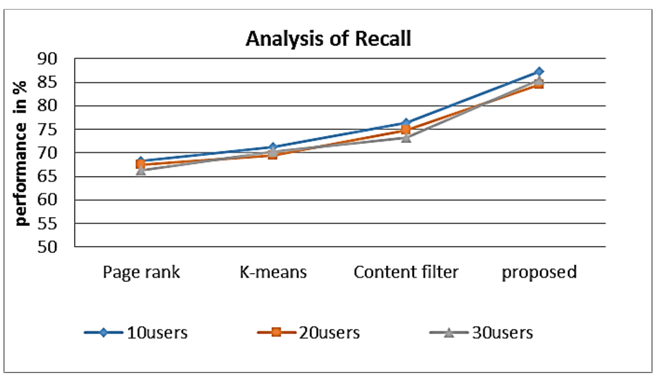
Methods/ number of users	Page rank	K-means	Content-based filter	Proposed SLCPC
10 users	68.2	71.2	76.3	87.3
20users	76.4	69.4	74.8	84.6
30users	66.2	70.2	73.2	85.5

The Table 3, shows the comparison of precision ratio produced 10 users as 87.3%, 20users as 84.6% and 30 users as 85.5 % shows that the proposed approach produces higher performance ratio.

Analysis of Recall:

Recall, Rc is defined as the proportion of a total number of retrieved URL links within the relevant URL links and the total relevant URL links with paged ranking.

$$\text{Recall, Rc}= \frac{\text{total retrived from relavant links (RA)}}{\text{relavant links(R)}} \text{ x } 100 \quad (6)$$



Graph 3: Comparisons of Recall

The Graph 3, shows the comparison of false recall ratio produced by different methods and the proposed method has produced higher performance other methods.

Table 5. Comparison of Recall

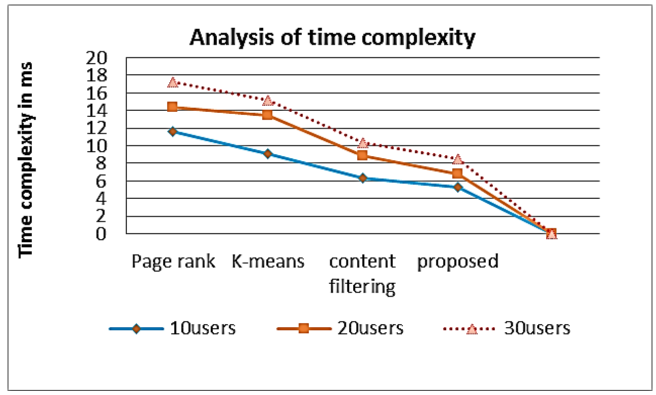
Methods/ number of records	Page rank	K-means	Content-based filter	Proposed SLCPC
10 Users	68.2	71.2	76.3	87.3
20 Users	67.4	69.4	74.8	84.6
30 Users	66.2	70.2	73.2	85.5

The above table 5 shows the comparison of recall page rank value that produces higher performance compared to other methods.

Analysis of Time Complexity:

Time complexity,

$$Tc = \sum_{k=0}^{k=n} \times \frac{\text{clustering Accuracy(cs)+false classification(Fcr)}}{\text{Time taken(Ts)}} \quad (7)$$



Graph 4: Comparisons on Time Complexity

The Graph 4, shows the comparison of time complexity produced by different methods and shows that the proposed approach has produced less time complexity than other methods.

Table 4. Comparisons of Time Complexity

Methods /number of records	Page rank	K-means	Content-based filter	Proposed SLCPC
10 Users	11.6	9.1	6.3	5.3
20 Users	14.4	13.4	8.8	6.6
30 Users	17.2	15.2	10.3	8.5

The Table 4, shows the comparison of time complexity proposed prefect clustering produced 10users as 5.3(ms), 20users as 6.6(ms) and 30users as 8.5(ms) shows that the proposed approach has produced less time complexity.

V. CONCLUSION

Mining information from huge resource becomes crucial one for user interest prediction. This implementation proves the semantic rule prediction problem using an efficient semantic log pre-fetch clustering (SLCPC) algorithm and personalized feed ranking (FPR) been presented to process the personalized web search. Initially, pre-process the web resource based on the user interest to identify and reduce the dimensions of session time and search logs, opinions. Then the method initializes the cluster with a set of points, and for each input data point of web links, finally the method computes the prefect user log support measure. Based on the measure of session time relevant link has been selected and indexed. The same is used to perform intelligence generation and produces efficient results on clustering and intelligence generation. Also, the method reduces the time complexity as well. Our proposed system improves the search performance based on the user interest to deals time complexity and purpose to search.

REFERENCES

[1] Kenneth Wai-Ting Leung and Dik Lun Lee. "*Deriving Concept-Based User Profiles from Search Engine Logs*," IEEE Trans. Knowledge and Data Eng., vol. 22, Issue. 7,pp.969 – 982,2010.

[2] K.W.-T. Leung, W. Ng, and D.L. Lee, "*Personalized Concept-Based Clustering of Search Engine Queries*," IEEE Trans. Knowledge and Data Eng., vol. 20, Issue. 11, pp.1505-1518,2008.

[3] M.speretta and S.Gauch, "*Personalized Search Based on User Search Histories*," IEEE Trans.Web Intelligence, vol 10,Issue 7,pp.192.212,2005.

[4] R.Baeze-yates, C.Hurtado, and M.Mendoza, "*Query Recommendation Using Query Logs in Search Engines*," Proc.Int'l Workshop Current Trends in Database Technology, pp.588-596, 2004.

[5] Y.Xu,K. Wang, B.Zhang, and Z.Chen, "*Privacy-Enhancing Personalized Web search*," Proc. World Wide Web(WWW) Conf., 2007.

[6] Zheng Lu,By, HongyuanZha, Xiaokang Yang, Weiyao Lin, And ZhaohuiZheng "*A New Algorithm For Inferring User Search Goals With Feedback Sessions*" , IEEE Trans. Knowledge and Data Eng., vol. 25, Issue. 3, pp.502-513,2013.

[7] Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu. "*Web-Page Recommendation Based on Web Usage and Domain Knowledge*", IEEE Trans. Knowledge and Data Eng., vol. 26, Issue. 10, pp.2574-2587, 2014.

[8] AthanasiosPapagelis and Christos Zaroliagis "*A Collaborative Decentralized Approach to Web Search*",IEEE Transactions on Systems, Man, and Cybernetics ,vol 45,issue 5,pp 1271- 1290, 2012.

[9] Dimitriosierrakos and George Palioura, "*Personalizing Web Directories with the Aid of Web Usage Data*" IEEE Trans. Knowledge and Data Eng., vol. 22, Issue. 9.pp 1331-1343, 2010.

[10] JunyuXuan, XiangfengLuo, Guangquan Zhang, Jie Lu, and ZhengXupp "*Uncertainty Analysis for the Keyword System of Web Events*, IEEE Trans. Knowledge and Data Eng., vol. 42, Issue. 10 ,829-842,2016.

[11] Rakish Kumar and AditiSharan, "*Personalized web search using browsing history and domain knowledge*", IEEE Trans. Intelligent computing, vol. 12, Issue. 2 pp. 2161-2174, 2014.

[12] Anoj Kumar and Mohd. Ashraf, "*Efficient Technique for personalized web search using users browsing history*", IEEE Trans. Intelligent computing ,vol 3, Issue 4. 2015 .

[13] Y.Mao, H Shen, "*Web of Credit: Adaptive Personalized Trust Network Inference From Online Rating Data*", IEEE Trans. Network and Intelligent computing ,vol 2, Issue 4.pp.234-2452, 2015.

[14] KamleshMakvana, Pinal Shah and Shah, "*A Novel Approach to Personalize Web Search through User Profiling and Query Reformulation*", IEEE Trans. Knowledge and Data Eng., vol. 26, Issue. 7.pp 1121-1139 2014.

[15] Jayaraj Jayabharathy , Selvadurai Kanmani " Correlated concept based dynamic document clustering algorithms for newsgroups and scientific literature " springer decision analytics, 2014.

[16] R. Hu, W. Dou, X. F. Liu, and J. Liu, "*Personalized searching for web service using user interests*," in Dependable, Autonomic and Secure Computing (DASC),2011 IEEE Ninth International Conference on. IEEE,India, 2011.

[17] Enrico Sartori, Yannis Velegrakis, and Francesco Guerra," Entity-Based Keyword Search in Web Documents", Proceeding Transactions on Computational Collective Intelligence, Springer, pp. 21–49, 2016.

[18] H.-j. Kim, S. Lee, B. Lee, and S. Kang, "*Building concept network-based user profile for personalized web search*," in Computer and Information Science (ICIS), IEEE/ACIS 9th International Conference on. IEEE,India, pp. 567–572. 2010

[19] EhsanElhamifar, and Rene Vidal, "*Sparse Subspace Clustering: Algorithm, Theory, and Applications*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, Issue. 11, pp. 2765-2781, 2013.

[20] Caimei Lu, Xiaohua Hu and Jung-ran Park, "*Exploiting the Social Tagging Network for Web Clustering*", IEEETransaction on Systems, Man and Cybernetics, vol.41, Issue. 5, pp. 840-852, 2011.

Authors Profile

Mr Velmurugan R pursued Bachelor of Science from Universsity of Madras, Chennai, India and Master of Science from Bharathidasan University, Tiruchirappalli, India in year 1999. He is currently pursuing Ph.D. in Manonmaniam Sundaranar University, Tirunelveli, India and currently working as Associate Professor in Department of Computer Science (PG), Kristu Jayanti College (Autonomous), Bangalore, India since 2003. He has published 2 research papers in reputed International Journals which are available online. His main research work focuses on Web Mining, Prediction mining, Web search, Clustering, Session logs based education. He has 18 years of teaching experience and 7 years of Research Experience.

Dr. Victor. S.P., pursued Bachelor of Science from St. Joseph's College, Bharathidasan University, Trichy, Tamil Nadu in the year 1986, Master of Computer Applications from St. Joseph's College, Bharathidasan University, Trichy, Tamil Nadu in the year1988, Masters of Engineering in Computer Science and Engineering from Anna University, Tirunelveli, Tamil Nadu in the year 2012 and Ph.D in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu in the year 2005. He is currently work as Dean of Science and Associate Professor of Computer Science, St. Xavier's College, Palayamkottai, Tirunelveli, Tamil Nadu with over 29 years of teaching experience. He is Expert Committee Member of DOEACC, Department of Electronics, Government of India, New Delhi. He has published 02 National and 86 International articles in reputed journals. Guiding students in the areas like Cloud Computing, Data Mining, Web Mining, Graph Mining and Image Processing.