I J C T A, 9(2) 2016, pp. 467-476

Enhancing Web Search Performance with Active User Categorization

G. Sumathi[1] and S. Sendhil Kumar[1]

ABSTRACT

Active user categorization system with interest based user clustering is proposed to extract the suitable group of pages for enhancing the web search performance. Existing users are grouped in light of the Swarm Intelligence (SI) methodology of Particle Swarm Optimization (PSO). The keyword based methodology is proposed to consider the user interest comparability for separating the group of existing user's travelled pages. The crawler data is cleaned and grouped in our experimental environment to evaluate accuracy of our categorization system using page correctness rate criteria. The experimental results show that 90% of separated pages are valuable for active user.

Keywords: PSO Clustering, User categorization, ODP taxonomy, Keyword Extraction, Keyword Mapping and Categorization

I. INTRODUCTION

As the World Wide Web is developing at a fast rate, the web-based services and applications turn out to be more prevalent. Users are interested with looking for valuable data on Internet and are exceptionally asking for the accuracy in rapidly response.

To develop suitable models by analyzing user access logs to distinguish category of the active user to encourage the perfecting of suitable pages to enhance the web search performance. Web mining technology is generally utilized on developing prediction models to dive profoundly into the interests of web users by examining their searching behaviors. In this way, the web administrator can adjust the site structure and foresee web user prospective traveling path to give the suitable data to users through examining the user behaviors [1][2]. Numerous particular systems are proposed to build prediction model, for example, as association rules [1] and Markov chain [3]. In this paper, we will propose a keyword based categorization framework including two stages cluster generation phase and active user categorization phase. These stages are utilized to analyze user behaviors and group them in view of the comparable searching travels. The keyword similarities between the active user search query and the user clusters can be mapped to extricate suitable group of looked pages for enhancing the web search performance. Additionally, active user categorization system will be incrementally developed for enhancing the exactness and effective of prediction. The measure of page correctness rate is intended to assess the exactness of our categorization framework.

The manuscript is organized as follows. Section II focuses on the literature review on user's searching behavior identification. Section III discusses methodology adopted in this paper while section IV presents experimental results. Section V provides conclusion.

II. BACKGROUND WORKS

Because of the dramatic growth of the internet, the number of transaction gets to be bigger and bigger. It results in the trouble of examining these entangled information utilizing conventional analysis system.

[1] Department of Information Science and Technology, CEG Campus, Anna University, Chennai, India, *E-mail: sumisundhar@auist.net, ssk_pdy@yahoo.co.in*

Numerous researchers have a tendency to find the potential behaviors utilizing data mining approaches to deal with foresee the active user behavior [4].

The focus of this section of the paper is to study and differentiation distinctive available systems to distinguish user's behavior.

Alexandros Nanopoulous *et al.* developed an effective model for expectation of web user accesses using web pre-fetching. It can be utilized as a part of diminishing user perceived latency problem present in each online application. As web prominence is expanding step by step in enormous folds, there is heavy traffic on the web which brings in delay in response because of congestion. A method is proposed which goes for reducing the delay of client future request process on the web by getting the pages into the cache in a background environment before an explicit request is made for the page. Notwithstanding this, another algorithm Ordered Web Mining(WMO) is created and is thought about against already proposed and existing approaches on web-log mining for web pre-fetching and the outcome demonstrated that WMO algorithm accomplished higher precision levels in expectation with very low overhead in system traffic [5].

In 2003, Mathias Gery et al predict the user's next request by utilizing web usage mining methodologies. In this technique three recognized web mining methodologies that are fit for investigating web logs were produced and examinations were additionally directed with real web log data sets:

(a) Association Rule (AR): In data mining Association Rule learning is a prevalent exploration system for finding interesting relations between variables in substantial database as in [6]. It depicts, examine and provide strong rules for discovery in databases utilizing distinctive measures of interestingness as in. The issue of discovering website pages that are gone to together in a specific arrangement is like discovering a relationship among item sets in transaction databases. When transactions have been recognized each of them could represent a basket and every page in the basket could represent an item [7].

(*b) Frequent Sequences (FS):* This methodology is utilized to find time requested successions of URL's that have been gotten to by past user's.

(*c) Frequent Generalized Sequences (FGS):* A generalized sequence is an arrangement permitting wildcards with a specific end goal to decide the user's navigational pattern flexibly [7].

The result of different examinations on an accumulation of web log datasets assessed that Frequent Sequences (FS) gives preferred exactness over AR and FGS [7].

Enrique Frias-Martnez *et al.* created a customizable behavior model for temporal prediction of web user sequences. Clustering and association rule, capture the consecutive and the temporal nature of website pages as they will be visited and are termed as sequential rules. To accomplish sequentiality the rules are are conveyed in a way such that they can store the click stream of the antecedent and the consequent. For Temporality the distance between the antecedent and the consequent is measured by the quantity of clicks required to travel between various pages. The prediction framework with utilizations the idea of antecedent and consequent is reliable as they gives the data about when the pages will be got to alongside the data of what pages will be accessed [8].

They also proposed a adjustable prediction framework i.e., the prediction framework is versatile to different environment relying upon the characteristics of the server which can incorporate number of pages, architecture of server, number of links per page and so on, in order to capture more precisely the behavior of its user's. The technique utilized measures the distance between the first page and the last page as far as number of clicks required to go from one page to other. Likewise with the quickly growing internet a customizable prediction framework is required. The author has introduced such a framework that represents order of the arrangement of URL's between antecedents and consequent along with the distance between them which makes it conceivable to decide when the pages are developing to be gone to [8].

In 2007, authors exhibited a web user behaviors prediction framework utilizing trend comparability. A trend based application framework is utilized to dissect user's behavior and predict the future path of user taking into account trend similarity. It is not reasonable to predict the searching behavior of new user according to the similar past behavior of different user's. Henceforth, a trend based prediction model is proposed to anticipate the future travelling path by generating ordered searching sequence. The framework proposed works in two stages. One is the construction stage and another is the prediction stage. The construction phase helps to discover useful common browsing patterns for the experts and after that they utilize it to foresee the further browsing sequences. In predicting stage, the searching behavior of a target user is nourished into the framework to be looked at inside of the prediction framework in order to produce pages that can be pre-fetched to enhance the searching performance. Application of replacement algorithm on proxy servers and test results demonstrates that the performance of the proposed model is valuable to pre-fetch user's pages ahead of time subsequently decreasing search time [9].

Authors utilized another classification model for online anticipating user's future movement. Web usage mining is normally actualized for two parts online and offline. The offline structure separates knowledge from the historical log records and after that this knowledge is utilized by the online part. They also proposed advance architecture for enhancing accuracy of classification in online stage. In this framework the classification is done using Longest Common Subsequence (LCS) algorithm. Pre-treatment makes the data refined so excess and oddities can be evacuated in the before stage itself. The semantic knowledge about fundamental domain can be utilized to enhance the nature of the recommendations [10].

Mehrdad Jalali propelled their past work and renamed there architecture as WebPUM. In this they utilized a novel formula for assigning weights to edges of undirected graphs to arrange new user movement. They utilized LCS algorithm to predict user not so distant future movement and conducted two fundamental analyses for navigation pattern mining and prediction of user's next request. What's more they discovered grouping patterns for user searching behavior and nature of the utilized datasets CTI and MSNBC enhanced [11].

Anitha introduced a new web usage mining methodology for next page access prediction. Mix of Markov model in view of sequential pattern mining with clustering demonstrated that prediction accuracy is expanded by 12% when contrasted with traditional Markov model. Clustering was utilized to distinguish comparable access pattern from web logs utilizing pair-wise nearest neighbour and afterward sequential pattern mining is performed on these recognizes patterns to decide next page possible accesses. The compactness of groups is enhanced by setting similarity threshold while framing groups. At the point when in future mining is done on these patterns, prediction exactness will be enhanced when contrasted with the accuracy when mining is done on disparate access patterns. Hence, a sequential mining method called "Markov model" is recommended in blend with pattern discovery. Markov Model gives great prediction accuracy in the event that it is utilized as a part of agreement with sequential mining [12].

Mamoun A.Awad et al applied the Markov Model for user's search behavior prediction. They concentrated on another altered Markov model to alleviate the scalability issue in the number of paths. Notwithstanding this, another two-tier prediction system that makes an Example Classifier (EC), based on the training examples and the generated classifiers is created. Tests demonstrated that such system can enhance the prediction time without compromising the accuracy. Prediction accuracy rate expanded to a higher rate [13].

III. ACTIVE USER CATEGORIZATION FRAMEWORK

In order to reduce the latency of browsing websites, a keyword based categorization method is proposed in our categorization system to download some relevant web pages in advantage on local user's computer according to the existing user's browsing behavior. This system is divided into two components which are picturised in Fig. 1. One is Cluster generation phase, and the other is Active user categorization phase.

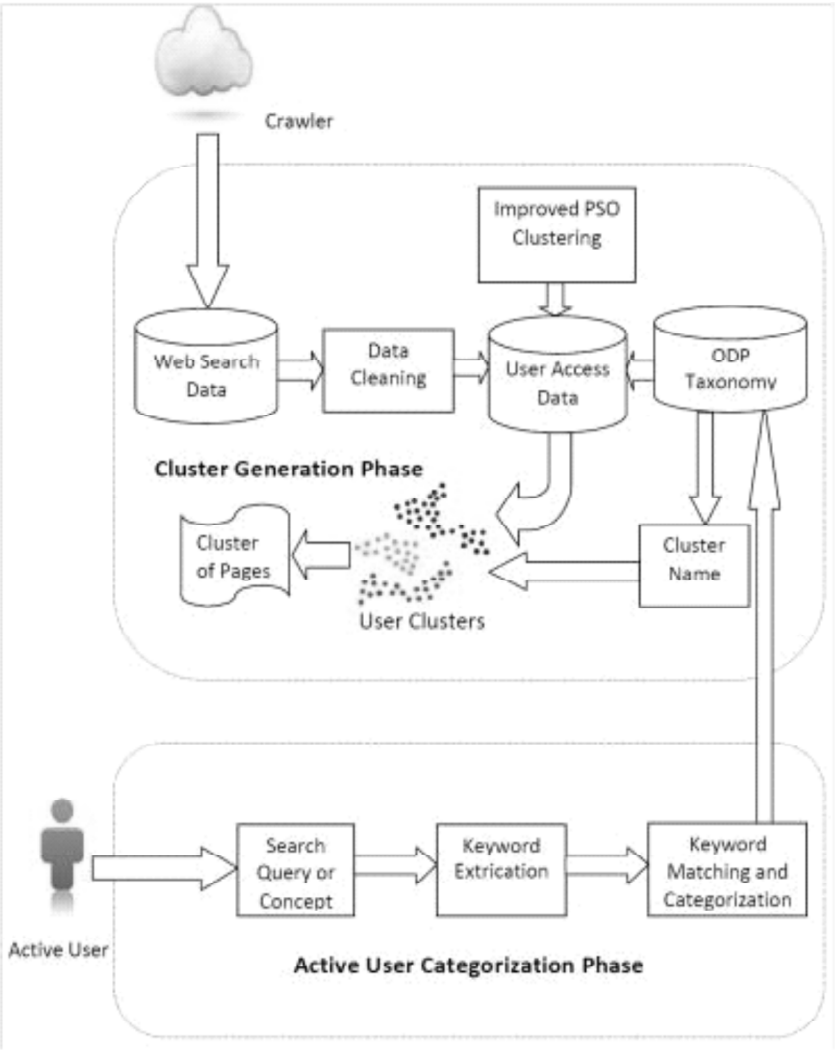


Figure 1: System Architecture

First, the user clusters with searched pages could be generated in the cluster generation phase using PSO algorithm after collecting enough web search data from the crawler. In the active user categorization phase, the search query of active users can be obtained to compare the similarity with existing user clusters formed in cluster generation phase. Then, the pages of matched cluster could be extricated to improve the browsing performance.

(A) Cluster Generation

In cluster generation stage, the user clusters with searched pages could be produced utilizing PSO algorithm and ODP taxonomy. This stage includes data cleaning and improved PSO clustering steps to make group of pages with common category.

(1) Data Cleaning

User search data are the crude data. This information contains the greatest measure of undesirable or insignificant information. This information makes the user clustering strategy wasteful. Cleaning of these search information is the beginning step which is required for the user clustering. Every search log entry of these information comprises of: URL, IP address, timespent, scrollingspeed encountered and average speed.

Cleaning process is associated with the user access data for taking out the insignificant or undesirable information. The following two steps are applied for this process:

(1) The information contains archive names with GIF, JPEG, CSS, etc., can be found and these values are removed.

(2) The links with unsuccessful pages can be traced and these links are removed.

Fom the relevant data, the users are grouped by the help of IP addresses. Since the IP locations of users are same, the users' searching activity was traced to identify the user interest. Thus, the title value of each page visited by all users was extricated from the searched URL. The relevant data for user clustering is picturised in Fig. 2 [14].

	A	B	C	D
1	User	FName	Url	Title
2	Usr_1	log1_1	http://www.annauniv.edu/research/index.php	Home - Anna University
3	Usr_1	log1_2	http://www.annauniv.edu/research/index.php	Centre For Research - Anna University
4	Usr_1	log1_3	http://www.annauniv.edu/research/announcement.php	Centre For Research - Anna University
5	Usr_1	log1_4	http://www.annauniv.edu/research/announcement.php	Centre For Research - Anna University
6	Usr_1	log1_5	http://www.annauniv.edu/research/index.php	Centre For Research - Anna University
7	Usr_1	log1_6	http://www.annauniv.edu/research/supervisor.php	Centre For Research - Anna University
8	Usr_1	log1_7	http://www.annauniv.edu/research/supervisor.php	Centre For Research - Anna University
9	Usr_1	log1_8	http://www.annauniv.edu/centres.php	Centre For Research - Anna University
10	Usr_1	log1_9	http://www.annauniv.edu/centres.php	Home - Anna University
11	Usr_1	log1_10	http://www.annauniv.edu/index.php	Home - Anna University
12	Usr_1	log1_11	http://www.annauniv.edu/index.php	Home - Anna University
13	Usr_1	log1_12	http://www.annauniv.edu/department/index.php	Home - Anna University
14	Usr_1	log1_13	http://www.annauniv.edu/department/index.php	Anna University - Departments
15	Usr_1	log1_14	http://www.annauniv.edu/department/index.php#	Anna University - Departments
16	Usr_1	log1_15	http://www.annauniv.edu/department/index.php#	Anna University - Departments
17	Usr_1	log1_16	http://www.annauniv.edu/department/index.php#	Anna University - Departments
18	Usr_1	log1_17	http://www.annauniv.edu/department/index.php#	Anna University - Departments
19	Usr_1	log1_18	http://www.annauniv.edu/department/index.php#	Anna University - Departments
20	Usr_1	log1_19	http://www.annauniv.edu/department/index.php#	Anna University - Departments
21	Usr_1	log1_20	http://www.annauniv.edu/department/index.php#	Anna University - Departments
22	Usr_1	log1_21	http://www.annauniv.edu/department/index.php#	Anna University - Departments

Figure 2: Relevant Data

(2) Improved PSO Clustering

For user clustering, In web usage mining process, grouping of users is a significant task. For user clustering, PSO methodology is connected. Particle Swarm Optimization is a Swarm Intelligence based optimization system which was created by Kennedy and Eberhart in the mid-1990s [15]. PSO approach replicates the movement and running of flying creatures. In this technique [14], every user is referred as particle and it is initialized with an arrangement of 50 users randomly. Here the users search with same Interest Change Rate (ICR) is considered for clustering. The average of ICR is assessed by utilizing the cosine similarity measure. This similarity is ascertained by applying the the accompanying mathematical statement:

$$\text{Similarity\_Value(Query,Cluster}_i) = \frac{\Sigma i\, c_{Q,j} c_{i,j}}{\sqrt{\Sigma c_{Q.j}^2}\sqrt{\Sigma c_{i,j}^2}} \quad (1)$$

This value is computed between the active user query category and category of the keywords in the travelled pages of the existing users. These categories were found by utilizing the ODP taxonomy. Similarity value is in the range from 0 to 1. These similarity values of all users are assigned as the initial value of the ICR.

Each user is evaluated by discovering their fitness value. The accompanying equation is applied for fitness calculation:

$$Fit\_Value = \alpha * \gamma\left(P_l(i)\right) + \beta * \frac{|V| - |U|}{|V|} \tag{2}$$

where $P_l(i)$ is the arrangement of pages went by l^{th} user at i^{th} cycle, $\gamma(P_l(i))$ is the clustering quality of the pages viewed, |U| is the quantity of user sessions. |V| is the aggregate number of pages viewed. From this fit_value calculation, the best value of the l^{th} user is denoted by p_l, which is called as the personal best. Best value from the values of all users is taken as the Global best which is illustrated by G. At every iteration, ICR and the new category can be found by utilizing the accompanying mathematical statements:

$$ICR_1^{(i)} = \omega * ICR_1^{(i-1)} + e_1 * r_1\left(p_1 - C_1^{(i-1)}\right)$$

$$+ e_2 * r_2\left(G - C_1^{(i-1)}\right) \tag{3}$$

$$C_1^{(i)} = C_1^{(i-1)} + ICR_1^{(i)} \tag{4}$$

where ω is the inertia value that controls the effect of past and current ICR of each user. r1 and r2 are uniformly distributed random variables and e1 and e2 are acceleration coefficients which controls the maximum step size between successive cycles. The estimations of ICR and category of the users can be changed in successive iterations. In this way, the user may change to another cluster by the estimation of ICR. The iterative process was stopped when the convergence was met. Fig. 3 presents the procedure for user clustering.

Input : User access data of 50 users
Output : Clusters *Ci* of users
Parameters : Inertia value ω, random variables r1, r2, acceleration coefficients e1, e2
Procedure :
Step 1 : Assign the initial values for ω, r1, r2, e1, e2
Step 2 : **for** each user U **do**
Find similarity score using Eqn. (1)
Initialize the similarity score as Interest Change Rate (ICR) of users
Find fitness value using Eqn. (2)
Initialize fitness value as category for all users
Find overall best position.
Step 3 : **end for**
Step 4 : **while**(Convergence Met) **do**
Step 5 : **for** each iteration **do**
Step 6 : **for** each user U **do**
Renovate Interest Change Rate(ICR) using Eqn. (3)
Renovate Category of users using Eqn. (4)
If New Category is better than Current Category **then**
Current category = New Category
end if
Step 7 : **end for**
Step 8 : **end for**
Step 9 : **end while**

Figure 3: Procedure for User Clustering

This improved PSO algorithm was applied for grouping the users into a set of clusters with their similarity in web travels [13]. The clustered users with their viewed pages are picturised in Fig. 4.

Cluster	Userid and URLs
COMPUTER	User-2 : Computers/Software/Databases/Data_Mining User-4 : Computers/Artificial_Intelligence/Fuzzy User-9 : Computers/Software/Databases/Oracle User-12 : Computers/Multimedia/Music_and_Audio/Audio_Formats/MP3/News_and_Media/MP3.com User-13 : Computers/Software/Information_Retrieval/Web_Clustering User-15 : Computers/Artificial_Intelligence/Agents/Publications/Software_Engineering User-23 : Adult/Computers/Internet/Searching/Directories User-24 : Computers/Software/Globalization/Operating_Systems/Windows User-29 : Computers/Data_Formats/Document User-30 : Computers/Programming/Languages/C User-34 : Computers/Data_Formats/Document/Text/Word
EDUCATION	User-0 : Reference/Education/Colleges_and_Universities/Asia/India/Tamil_Nadu/Anna_University User-31 : Reference/Education/Colleges_and_Universities/North_America/United_States/Texas/Texas_A&M_University/Admissions/Conferences
SPORTS	User-1 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-6 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-10 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-11 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-18 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-33 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket User-36 : Business/Consumer_Goods_and_Services/Sporting_Goods/Cricket

Figure 4: Clustered Users

(B) Active User Categorization

Active user categorization is the process of discovering the cluster name or category based on their search concept. The motivation behind this categorization is to recover the similar user viewed pages for improving the web search performance. The active user is classified in view of the existing users searching travels. With the end goal of active user categorization, the ODP taxonomy is used. The ODP dataset incorporates the top level categories of Adult, Arts, Business, Computers, Games, Health, Home and so on. In this ODP dataset, the path categories of every last one of keywords are indicated [16]. The related cluster name is extricated and the respective pages are recovered from the corresponding user cluster.

(1) Keyword Extrication

Keywords are regularly used for web searchers and record databases to discover data and make sense of whether two bits of test are connected with each other. Keyword extrication from source data is an ordinary tool which is used for web engines and indexes alike to quickly sort and find specific data [17]. For the purpose of active user categorization, the keywords are extricated from the active user search concept. The procedure for the keyword extrication is shown in Fig. 5.

Input : Active user search query
Output : Set of Keywords
Parameters : Active user query q, Keyword w
Procedure :
Step 1: **GET** query q
Step 2: **for** each keyword w **do**
Step 3: **if** w is Stop word **then**
DISCARD w
else
EXECUTE Stemming process
Step 4: **if** w is occurred in multiple times **then**
RETRIEVE keyword w with one time
else
STORE w in keyword set
end if
end if
end for

Figure 5: Procedure for Keyword Extrication

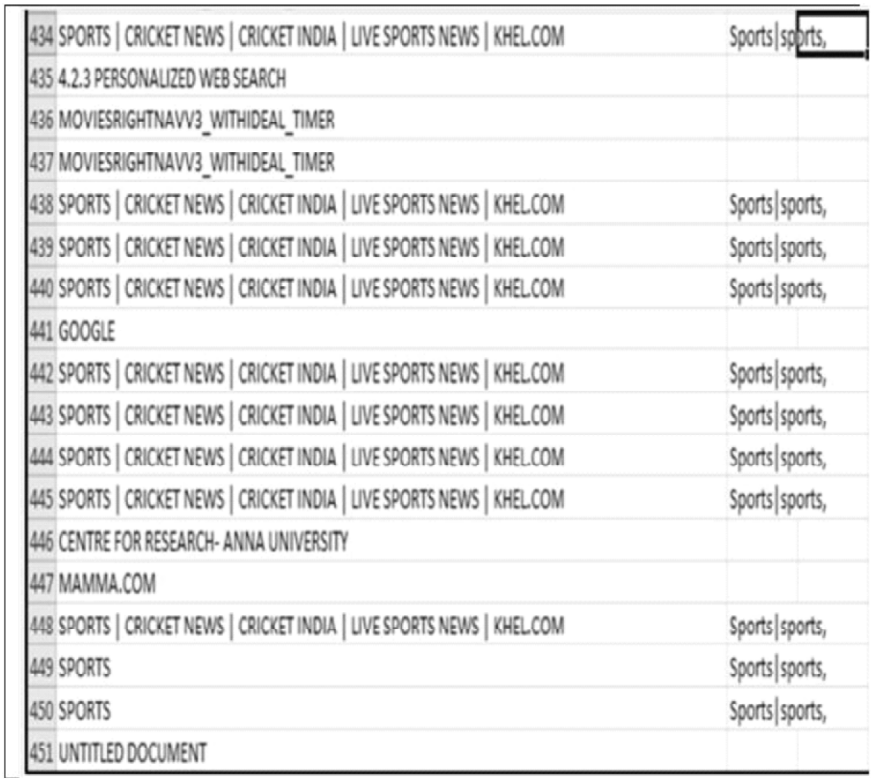
(2) Keyword Matching and Categorization

The extricated keywords of the active user search query are matched with the ODP taxonomy and the related cluster name is extricated. At that point this cluster name is matched with existing user clusters for recovering suitable viewed pages for improving the browsing performance. Based on this recovered cluster name the pages are extricated for the active user. The procedure for keyword matching and categorization of active user is presented in Fig. 6.

Input : Set of Keywords in a Query
Output : Cluster of related pages
Parameters : ODP hierarchy H_i, Keyword w, vertex v_i, Cluster name C_i
Procedure :
Step 1: **for** each keyword w **do**
Step 2: **for** each Taxonomy hierarchy H_i **do**
if $w = v_i$ **then**
$C_i = H_i$
end if
Step 3: **for** cluster C_i **do**
GET and SAVE searched pages from C_i
end for
end for
end for

Figure 6: Procedure for Keyword Matching and Categorization

Figure 7: Shows the searched pages or suitable pages for active user reference.



434	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
435	4.2.3 PERSONALIZED WEB SEARCH	
436	MOVIESRIGHTNAVV3_WITHIDEAL_TIMER	
437	MOVIESRIGHTNAVV3_WITHIDEAL_TIMER	
438	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
439	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
440	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
441	GOOGLE	
442	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
443	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
444	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
445	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
446	CENTRE FOR RESEARCH- ANNA UNIVERSITY	
447	MAMMA.COM	
448	SPORTS | CRICKET NEWS | CRICKET INDIA | LIVE SPORTS NEWS | KHEL.COM	Sports|sports,
449	SPORTS	Sports|sports,
450	SPORTS	Sports|sports,
451	UNTITLED DOCUMENT	

Figure 7: Searched Pages for Active user

IV. EXPERIMENTAL RESULTS

To evaluate the accuracy of extracted suitable pages result in our categorization system, the measuring criteria of Page Correctness Rate (α) is applied to examine the predicting sequence.

The page correctness rate is used to calculate the coverage of predicting sequence in the realistic further browsing patterns and is computed as

$$\text{Page Correctness Rate } (\alpha) = \frac{Count\ A}{l} \tag{5}$$

where *count A* is the number of the same pages between testing pages and categorized pages, and *l* is the length of the testing pages.

For example, T = {A, B, C, D, E, F, G, H, I, J} is one user actual browsing sequence, and N = {B, D, C, A, G, F, E, I, J, S} is the identified pages retrieved by our categorization system. The Page Correctness Rate (α) is 9/10 = 0.90. Table I shows the comparative analysis of various models of identification of user's behavior.

Table I
Performance analysis of Active user categorization

1.	Pre-fetching of web pages into cache	Prediction Enabled Web Server/WMO	0.65
2.	AR, FS and FGS	FS produce optimal outcomes	0.69
3.	Clustering Sequential Association Rule	Behavior Model	0.72
4.	Trend Based Prediction	Prediction System	0.75
5.	Online/Offline phases of architecture Clustering	Online prediction of user's movement	0.80
6.	LCS alogorithm, Clustering	WebPUM	0.85
7.	Sequential Pattern mining with Kth order Markov model clustering	Prediction system to determine next page access	0.86
8.	Modified Markov model with Association Rule Mining	Two-tier framework with EC	0.89
9.	PSO clustering, Keyword based approach	Active user Categorization System	0.90

From the above table the performance of the proposed model is useful to extract the more suitable pages for the active user.

V. CONCLUSION

A keyword based categorization system was proposed to identify the active user interests. We design a PSO based categorization to cluster the existing user based on their similar searching travels. The search query of active users can be obtained to compare the similarity with existing user clusters formed in cluster generation phase. Then, the pages of matched cluster could be extricated to improve the browsing performance. The results show the performance of our proposed model is useful to extract the 90% suitable pages for the active user.

REFERENCES

[1] Q. Yang., H. H. N. Zhang., T. Y. Li, "Mining web logs for prediction models in WWW caching and prefetching." Int'l conf. on knowledge discovery and data mining, 2001, pp. 473-478.

[2] Y. H. Wu., A. L. P. Chen, "Prediction of web page accesses by proxy server log." WWW, vol. 5, no. 1, 2002, pp. 67-88.

[3] J. H. Zhu., J. Hong., J. G. Hughes, "Using markov models for web site link prediction." in proc. of the 13th ACM conf. on hypertext and hypermedia, 2002, pp. 169-170.

[4] V. S. Tseng., K. W. Lin, "Efficient mining and prediction of user behavior patterns in mobile web systems." Information and s/w tech. vol. 48, no. 6, 2006, pp. 357-369.

[5] Alexandros Nanopoulous., Dimitris Katsaros., Yannis Manolopoulous, "Effective prediction of web-user accesses : A data mining approach." in Proc. Workshop WEBKDD, 2001.

[6] Piatetsky-Shapiro, Gregory. "Discovery, analysis, and presentation of strong rules." Knowledge discovery in databases, 1991, pp.229-238.

[7] Mathias Gery., Hatem Haddad. "Evaluation of web usage mining approaches for user's next request prediction." in WIDM ACM conf. New Orleans, Louisiana, USA, Nov. 7-8, 2003, pp.74-81.

[8] Enrique Frias-Martnez., Vijay Karamcheti, "A customizable behaviour model for temporal prediction of web user sequences." in WEBKDD 2002,LNAI 2703, 2003, pp 66-85.

[9] Jan., Nien-Yi., Nancy P. Lin. "Web user behaviors prediction system using trend similarity." Intl' conf. on simulation, modelling and optimization, World scientific and engineering academy and society (WSEAS), 2007, pp. 69-74.

[10] Mehrdad Jalali., Norwati Mustapha., Ali Mamat., Md.Nasir B. Sulaiman, "A new classification model for online predicting user's future movement." Intl' Symp. on Info. Tech. Aug. 26-28,2008, pp. 1-7.

[11] Mehrdad Jalali., Norwati Mustapha., Ali Mamat, Md. Nasir B. Sulaiman, "WebPUM : A new web-based recommendation system to predict user's future movements." in Int. Jour. Expert sys with Applns, 2010, pp. 6201-6212.

[12] A.Anitha, "A new web usage mining approach for next page access prediction." Intl. Jour. of comp. applications, vol.8, no.11, Oct. 2010, pp.7-10 .

[13] Mamoun A.Awad., Issa Khalil, "Prediciton of user's web browsing behaviour :application of markov model." IEEE transactions on systems, man and cybernetics-part B: cybernetics, vol. 42, no. 4, 2012, pp. 1131-1142.

[14] Sumathi Ganesan., Sendhilkumar Selvaraju. "Application of particle swarm optimization and user clustering in web search." Computational intelligence in data mining-vol. 2. Springer India, 2015. pp. 427-437.

[15] Kennedy. J., Eberhart, R. "Particle swarm optimization." Intl. conf. neural networks, 1995, pp. 1942–1948.

[16] Zhang, D., Lee, W.S., "Web taxonomy integration using support vector machines." Intl. conf. world wide web, ACM, 2004, pp. 472–481.

[17] Durairaj. M., Kumar. K.M. "News recommendation systems using web mining : a survey." Int. conf. of engg. trends and technology,12(6), 2014, pp. 293-299.