# An Approach Proposed for Detecting Users activities from Recorded Log

Deepti Sahu
Institute of Technology and Management

Rishi Soni
Institute of Technology and Management

## ABSTRACT

The development of the web has created a big challenge for directing the client to the website pages in their area of interest. Accordingly, just option is to capture the intuition of the client and provide them a list of recommendation. Most specifically, online navigation activities develop with day by day; consequently extract information with the capability of intelligence, from these activities is a tedious job. Webmaster of an organization ought to utilize methods of web mining to fetch intuition, Web usage mining (WUM) is one among them.WUM is designed to operate on web server logs; logs contain client's navigation history which is very useful for the web recommendation. Recommendation is an application of web usage mining. Consequently, recommendation system can be utilized to forecast the navigation pattern of client and recommend those to client in a form of recommendation list. This paper, suggest a recommendation principal that recommends a list of pages on the basis of client's past navigation history (recorded within the web log). This approach brings the advance within the precision of displayed pages for the client or users.

## Keywords
Recommendation, IP protocol, list of recommendation

## 1. INTRODUCTION
It becomes more difficult to fetch related information from the vast collection of information; similarly to fetch related information through the WWW due to the unstable expansion of information existing on the WWW is a difficult task. Data mining is a technique that is really very helpful for fetching relevant and hidden information from the ocean of information with an efficient way. It has turned out to be considerably harder to get to relevant data on the WWW due to the unstable development of data. Web personalization is a probable solution to solve this problem [1]. This methodology concentrates on building up a smart recommender framework offer a web service which is personalized for getting interrelated web pages with more effectiveness. One among the main subject under data mining be a mining of Sequential pattern is comprehensively useful in multiple areas for data extracting. Mining of the Sequential pattern is the procedure of determining frequent group of items from the complete set of record. Frequent set of items means a set of items appear frequently in a particular order [2]. This paper proposed a Recommender system that used sequential pattern mining technique for the calculation of frequent item sets. A recommender system is a system that intelligently presents a set of recommendation or a system with the talent of guiding users towards their desired choices and services based on the user's navigation pattern history lying on the web. Input of the recommender system is a web log file taken from the Website of Texec Company. Weblogs contain IP address of client, time of request, requested URL, id of user, status code, referrer etc. Web log of different servers have different formats [3].

It is not possible to mined directly web log data. For the mining of web log first of all applied data preprocessing technique then frequent patterns of web access are extracted from the sequential pattern mining technique. Results of sequential mining technique are than applied for the closed sequential pattern mining. Afterward, a pattern tree is created by means of the closed sequential patterns of the a web access. For the web page access pattern of current user, web recommendations is provided to the user's through the matching of pattern tree sequence with the current access sequence. To deliver better recommendation to the user the web personalization system's developers must identify what the customers' behavior when they surf on the web.

Rest of this paper is arranged in following section. Current research that has been carried out in this area is described in Section 2. Section III presents the system architecture of Recommender system. Performance evaluation and results discusses in section IV. And last, section concludes the paper.

## 2. RELATED WORK
ThabetSlimani, and Amor Lazzez Present a study about various techniques of sequential pattern mining algorithms; they mainly work on five ample categories of these algorithms. First category is study based on Apriori algorithm, second is the study of Breadth first searching algorithm, Depth first search method is the third category of they studied, fourth is the study of improved sequential pattern mining at last they studied incremental pattern mining algorithm. Toward the end comparative evaluation chart is prepared on the premise of major key points. This gave an amendment in the understanding of the strategies of sequential pattern mining [4].

Zhou. B et al. [5] presented a smart recommender system of web with the capability of intelligence is recognized as Sequential Web access-based recommender system (SWARS) designed in taking account sequential pattern mining. They proposed CS-mine, a proficient sequential pattern mining method made to identify patterns fetched sequentially and those are frequent. Afterwards Frequently fetched patterns were put in a compact tree structure for coordinating as well as producing Web links in the form of recommendation. They was evaluate performance of the system through precision, Time taken and satisfaction They also introduced effective sequentially accessed sample mining algorithm, acknowledged as Conditional Sequence Base mining (CSB-mine). Presented approach of CSB-mine work with the conditional access sequence of each frequent event and eliminate the need of WAP tree. This upgraded approach of mining is effective in correlation to calculations based on WAP tree mining. Cui Wei et al. [7] has presented a hybrid system of web personalization which used the principal of grouping and adjoining (contiguous) of sequential patterns. For this they grouped log file to analyzed the essential structural engineering of web sites, and for every group, they utilized adjoining sequential pattern mining to boost the

topologies of websites further. They have exhibited two assessing parameters to measure the execution behavior of their system. Zhenglu Yang et al. [8] have proposed an proficient sequential pattern mining algorithm (LAPIN_WEB:LAst Position Induction for WEB log), which is an extension of prior LAPIN algorithm to take out access patterns of user's from the traversal path in Web log files. Mining of Web log comprises of preprocessing of data, mining of sequential patterns and visualization.

## 3. SYSTEM ARCHITECTURE

In proposed system of recommender system as shown in fig.1 first step is to collect web server logs from the web servers.web server logs are the file in ASCII format or in form of text file, store the web access activities of all user's. These log files maintain the information about the IP address of clients who access the website, time of request, time stamp, ID of user, requested URL, status code which indicates the successful completion or failure status during the web search. There are number of pattern mining algorithms are available for finding frequently accessed web access patterns from recorded web server log files. In Proposed system improved sequential pattern mining algorithm is used for finding frequently accessed web access patterns. Both the consecutive steps improved sequential mining and construction of pattern tree are offline processes. Pattern tree can be rebuilt as often as possible to fuse new access data [8].
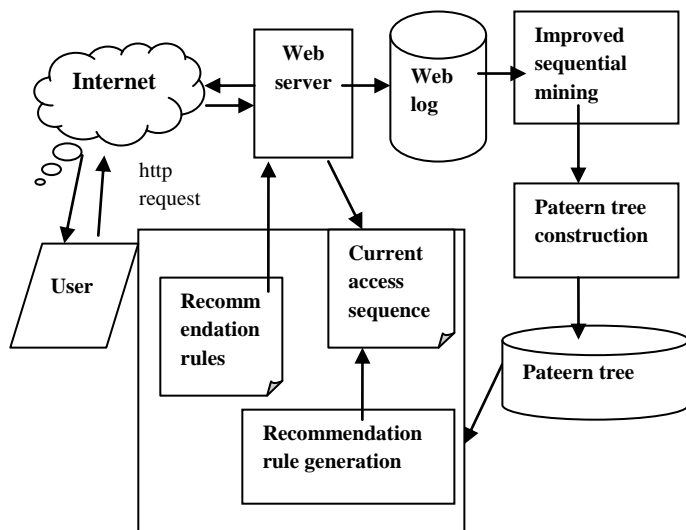


**Fig.1 System Architecture**

At that point when a user access the site, HTTP request of the user in present browsing session are recorded all together and the present access arrangement is developed. Every user get access the website can be recognized his/her IP address. In this system Users current access sequence is matched from the access sequence recorded in the pattern tree for the generation of recommendation rule. If the access sequence is matched then the recommendation rule component generates the rules for recommendation. From the recommendation rule principle, recommended or related links will be will be progressively implanted into the current requested page. Description of proposed system Architecture is given below.

## 3.1 Improved sequential pattern mining

Mining of sequential patterns is an imperative area of data mining that is widely applied into the number of areas like for searching frequently access pattern, for market analysis etc. [9]. Most often, sequential pattern mining is outlined as

selecting the entire group of regular sub patterns in a set of all patterns [10, 11]. Sequential patterns is a chain of object that frequently come out in a specific order, such that, objects in the same object set are relied upon to have the identical transaction time value within a specific time interval. Improved sequential pattern be a special sequential pattern which has no super-sequence with the matching occurrence frequency i.e. minimum support [12]. There exist two strategies for improved frequent pattern mining:

(1) Find out the last or final improved pattern set

(2) Perform post-pruning on the chosen improved pattern set

It is desirable to make use of the first strategy due to the reason that the next strategy desired storage area to store the discovered patterns and carry out post-pruning; although, when the patterns come to be more difficult, it's tough to ensure that each and every generated pattern is improved without analyzing the formerly learned patterns [13]. Thus second strategy is used. Mining of improved sequential pattern have entranced the greater part of the analysts for its Principal of web recommender system are drawing attention to a larger extent as an suitable process to counter info overload and assist the users of the Web who search the information to seek out what they want in a very quicker way [17]. These recommender principals are advantageous in instructing the users to the goal or objective pages. On the other side, utilizing proposal frameworks to accessible Web destinations needs impressive work and information [19]. There exist numerous strategies to assemble such frameworks. Among these approaches, building of web recommendation system with the help of web access log is an accepted approach [20], [21]. On this paper, now we've designed a web recommendation system taking into account improved sequential pattern mining. The proposed framework will be utilized for generating personalized Web information for the user. Input for the proposed system will be the data of web server log.

Web personalization approach focuses on developing an intelligent recommender system, that offer personalized web service for accessing interrelated web pages more efficiently and effectively. Web personalization is any activity that changes the information or services offerings by a Web site according to the necessities of a particular user or a set of users getting advantage of the knowledge acquired from the user's navigational conduct and individual interests, along with the content and the structure of the Web site [15]. The intent of a Web personalization system is to supply users with the information they require, without anticipating from them to ask for it clearly [16]. Web Recommender framework is a sort of personalized web application which supplies extensive user esteem by personalizing a number of sites on the Web [17].

In improved sequential pattern mining inputs are cleaned database (which obtained after Preprocessing phase, Preprocessing phase comprise of cleaning of data, identification of Session and User, content retrieval) , Input sequence i.e. a sequence for which recommendation rules are generated and user-specific minimum support, minimum support means quantity of data items pattern contained. And output is the frequent web access patterns means patterns repeated frequently in web log or patterns mostly searched.

Let E be a group of distinctive access events, represents the web assets accessed by users that is to say web pages, uniform resource locator (URL's) categories or topics. A web access

sequence S = e1e2…en (ei ∈ E) for 1≤i≤n is an ordered set of access events, and n is called the length of the web access sequence. Improved sequential pattern mining mines only frequent improved subsequences specifically the sequences contain no super sequence with the similar support, instead of mining whole set of frequent subsequences. Subsequences are denoted as S'. Result of improved sequential pattern mining is a set of frequent item set L = {a1, a2 … … an}.In our proposed system, the improved sequential pattern algorithm has been explained like this:-

**Algorithm 1:  Call_improvedseq()**

**L is the Array initialize as NULL.**

**Input: Cleaned database D; Input sequence S; Minimum support min_supp;**

**Output: Frequent web access patterns store in L ;**

**Algorithm:**

1. **Sub-sequences s' are mined and assigned unique ids.**

2. **If(super_patt or sub_patt exists)**

3. **Then peform pruning & Modify in L**

**Else**

**Add in L**

4. **Scan D and find the frequent item set ai so as to append to form sequential pattern.**

5. **If ($a != valid$)**

   **Return;**

   **Else**

   **Call_improvedseq()**

6. **Return**

## 3.2  Pattern tree creation

Pattern-tree creation phase is anticipated for organizing or putting sequential web access patterns succinctly in order that is acclimated for coordinating with the current access sequence additionally, creating recommendations efficiently inside of the Recommendation Rules Generation phase. In practice, gathering of sequential web access patterns are often treated as a group of strings over a set of predetermined alphabet. Each node of the Pattern-tree is marked as a access event with a corresponding supporting value. To build a Pattern-tree, there is a need to perform scanning of entire successive web access patterns. pattern tree is built from the mined frequently web access sequence. created pattern tree is employed for recommend, recommendation for a user's current web access pattern. tree structure is employed to store a collection of frequent items however regular tree store individual character in every node.

**Example** Let CSP be a closely accessed sequential web pattern with their corresponding support value $CSP = \{3\ 5:\ 2, 3\ 5\ 5:\ 2, 3\ 6\ 5:5, 55:\ 3, 3\ 3\ 5\ 5:2\ , 7\ 7:3\}$

Firstly, a root node is created initialize as null. Afterwards, all the events or items are inserted into the next level as shown in fig. pattern tree for the given example is shown in Fig 3.
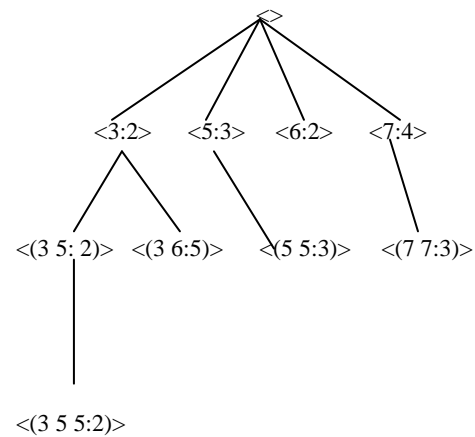


**Fig 2:  Pattern-tree**

Algorithm of  the pattern tree in proposed system is as follows:

**curr_node: current node.**

**R: empty root node**

**Pattern tree:PT**

**Input: Frequent item sets from improved sequential mining are L.**

**Output: PT**

**Algorithm:-**

**1. Initially empty root node is taken for the construction of pattern tree PT.**

**2. Consider every sequence S, S∈L, symbolized as S = e1e2…en do**

   i. **Fix curr_node on R.**
   ii. **For  c = 1 to n do**

   **If (curr_node have child marked as ec)**

**Utilized max_supp between S and ec and set curr_ node as ec.**

   **Else**

**Construct a new child_node with support S and set curr_node point to $e_C$**

**3. Return pattern tree T.**

## 3.3  Recommendation Rule Generation

Recommendation Rule Generation Recommendation Rule Generation phase generally look for advantageous matching access path within the Pattern tree with the assistance of current access sequence of user. Generally, if the current access sequence of user is massive in size then the opportunity to found matching path is lesser is lesser from the tree. For enhancing the relevancy of advice in this phase, suffix sequence of this access sequence are going to be taken into consideration once the matching path of the whole access sequence can't be found. In such case, going to ask for the matching path supported an equivalent access sequence by removing the primary item repeatedly until an identical path is found or once know a lot of item may be far from the access sequence. Additionally, the length of the longest path within

the Pattern-tree is that the depth of the Pattern-tree. The matching path won't exist once the length of this present access sequence is longer than the depth of the Pattern-tree. Thusly, some antecedent items can be removed to accomplish this access sequence shorter than the depth of the Pattern-tree before the sequence matching process start [22].

Algorithm 3: Call_Rec_rule_gen()

**MinLen: minimum length to be accessed.**

**MaxLen: Maximum length to be accessed;**

**curr_node: current node.**

**Input: - PT: pattern-tree based on min threshold; frequent itemset L;**

**Output: R: generation of recommendation rules**

**Algorithm:**

1. **Initialize R= null;**

2. **If (L ›MaxLen)**
   Remove the first$|L| - MaxLen + 1$ *from L*;
3. **If(L ‹ MinLen)**
   **Return R**

   **Else**

   **Set root node as current node of PT.**

4. **For(a$_i$: 1 to length(L))**
   If($curr\_node \rightarrow child\_node$)

   $Child\_node = ei$
   **else**
   drop item from L.
   **Repeat from step 3.**
5. **If(curr_node ->child_node)**
   **Insert child_node into R ordered by minimum supports.**
6. **Return R.**

The process of recommendation rule generation works n a particular algorithm that calls various algorithms within itself. The basic criteria behind the calling function are the whole process that is being taking place for the process. The main process can be explained with the help of pseudo-code that represents the sequence of the processes taking place and the order in which the algorithms are being called.

The pseudo-code for the recommendation system proposed can be explained as below:-

**Input:- Items from the database D;**

**Output:- Patterns generated from periodical mining Algorithm**

1. **Data Cleaning**
2. **Loading Cleaned dataset.**
3. **Identification of unique URL's and assigning unique ids and abbreviations to the urls and the country names respectively.**
4. **Applying algorithm of closed sequential pattern mining so that frequent access patterns are generated**

   **Call_closedseq();**

5. **Pattern tree construction for making the searching faster**
   **Call_PTC();**

6. **Entering prefix from the user and searching for the recommendation rule generation**
   **Call_Rec_rule_gen();**

7. **Enter the suffix sequence from the user for the recommendation rule generation to find precision and satisfaction.**
   **Formulas:-**

$$Precision = \frac{|Rc|}{R}$$

$$Satisfaction\ (m) = \frac{|Rs(m)|}{R}$$

**Where, R are total rules generated.**
***Rc* is subset of R which are correct rules.**
**R$_s$ is the subset of R, consists of satisfactory recommendation rules.**

8. **A new recommendation system is created from recommender system for efficient searching.**

## 4. PERFORMANCE EVALUATION
## 4.1 Evaluation Measures

Let A = b1b2…bkbk+1…bn be present a web page fetching sequence. Consider prefix sequence Aprefix = b1b2…bk(k ≥ minimum Length) and for this, generate a recommendation rule ROR = {r1, r2, …,rm} through the Pattern-tree, where all events are arrange in order of their support. If bk + 1 ∈ ROR, mark the recommendation rule as acceptable otherwise unacceptable. If there exists bi∈ROR (k+1≤ i ≤ k+1+m, m>0), than the recommendation rule is describe as m-step satisfactory. Else, recommendation rule is descrie as m-step unsatisfactory. Let R = {ROR1, ROR2, …,RORl} be a set of recommendation rules, where RORi(1≤ i ≤l) is a recommendation rule. |R| = l is the sum number of recommendation rules in R.

Following evaluation measures are considered for the Performance Evaluation [23].

**Definition 1**: Let Rc is the subset of total rules generated (R), consists of all right(correct) recommendation rules.

Precision of recommendation is defined as

$$Precision = \frac{|Rc|}{R}$$

Rc is the correct recommendation rules

Precision is the measures of probability that how a user will access one of the recommended page.

**Definition 2:** Let Rs (m) is the subset of R, consists of all satisfactory recommendation rules., m-step satisfaction is define as

$$Satisfaction(m) = \frac{|Rs(m)|}{R}$$

Where, R are total rules generated.

$R_c$ is subset of R which are correct rules.

3. Consumed time

# 5. EXPERIMENTAL RESULTS

This section presented the results of the experiment in proposed recommendation system. This Paper present a test of base and proposed system against different sequences, suffixes and satisfaction parameters and plot time consuming graph and Precision measure graph as shown in fig 3 and fig. 4.
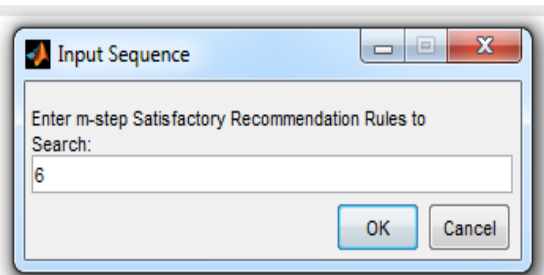
## 5.1 Procedure

Step 1: User is asked to input sequence to search
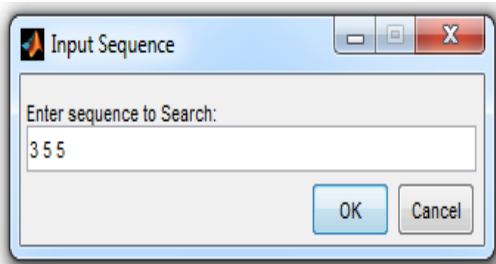


step 2:- Then asked for the suffix that he wants to search
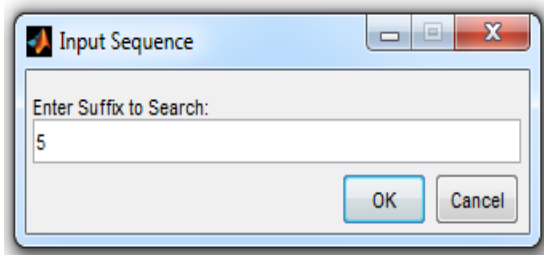


Step 3:- And then user is asked for m-step satisfactory recommendation rule to be searched



These above steps are common for both the base and proposed papers.

### For Base Paper:

The output rules generated with the performance evaluation measures i.e. precision, satisfaction and CPU time consumption i.e. t_base are shown below in the first screenshot attached.

### For Proposed Paper:

The output rules generated with the performance evaluation measures i.e. precision, satisfaction and CPU time consumption i.e. t_propose are shown below in the second screenshot attached.

```
output_rule =

     5     22
     5     21
     5     21
     5     21
     5     20
     5     19
     5     17
     5     13

precision = 0.375000
satisfaction = 0.000

t_base =

   150.8336
```

**Base Screenshot**

```
output_rule =

     5     22
     5     21
     5     21
     5     21
     5     20
     5     19
     5     17
     5     13

precision = 0.375000
satisfaction = 0.000

t_propose =

   125.9452
```

**Proposed Screen shot**

Base / Proposed Time consumption representation in graphical form as shown in below fig.
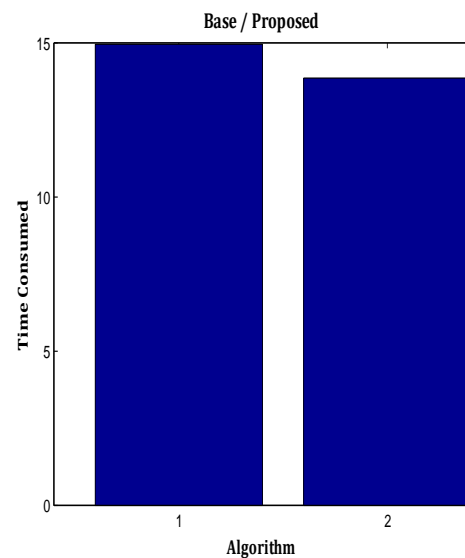


**Fig.3 Consumed time graph**

## 5.2 Graphical Representation of Precision In Base And Proposed Paper.

Precision of Base and Proposed paper measure on the Following data

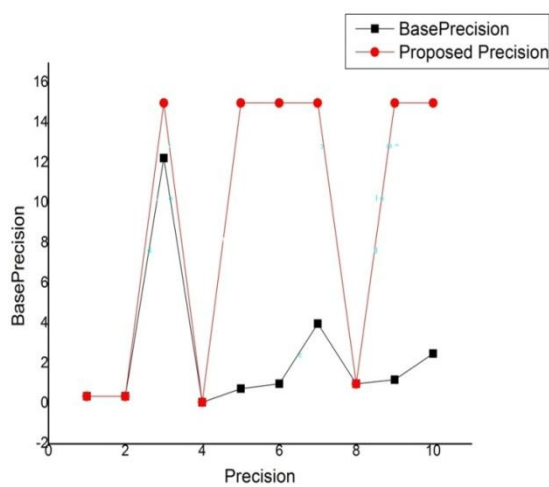| s.no. | seque nce | Suffi x | mstep | basepresisi on | Proposed presision |
|-------|-----------|---------|-------|----------------|--------------------|
| 1 | 2 5 5 | 5 | 6 | 0.375 | 0.375 |
| 2 | 3 5 5 | 5 | 2 | 0.375 | 0.375 |
| 3 | 5 3 5 | 5 | 4 | 12.2492 | 15 |
| 4 | 9 9 9 | 9 | 5 | 0.078947 | 0.079847 |



**Fig.4 Precision measure graph**

## 6. CONCLUSION

This paper presented a recommendation system work with the principal of improved sequential mining of patterns. For this, we have worked on Improved-Pattern mining method for the efficient mining of sequentially stored web access Patterns obtained from the preprocessing, applied on the web server log files and then with the help of improved sequential pattern mining, improved sequential web access patterns are got .And in the last step A Pattern tree is created using the patterns created, where during the improved sequential pattern mining process. Recommended or targeted recommendation in the form of web links given to the current user, who they seek the information on the WWW by means of matching the current access sequence content with the sequence marked on pattern tree. Proposed web recommendation system is validate with the web log of Texac company and the results of experiment show that the results of the proposed system are better with fine precision and satisfaction in comparison of Base system

## 7. REFERENCES

[1] M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley & Sons Inc., New York, 2002.

[2] R. Agrawal and R. Srikant, "Mining Sequential Patterns", In Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, Taipei, Taiwan, 1995.

[3] M. Lie and L. Fan, "A Web Personalization System Based on Users'Interested Domains,"Proc. 7th IEEE Int. Conf.on Cognitive Informatics (ICCI'08)-IEEE 2008, pp.153-159.

[4] ThabetSlimani, and Amor Lazzez, "Sequential Mining: Patterns And Algorithms Analysis", Computer Science, Taif University & LARODEC Lab, Saudia Arabia, 2 Computer Science,Taif University , Saudia Arabia.

[5] Zhou. B,Hui. S.C and Chang.K, "An intelligent recommender system using sequential Web access patterns", IEEE Conference on Cybernetics and Intelligent Systems, vol. 1, pp. 393 - 398, December 2004.

[6] Cui Wei, Wu Sen, Zhang Yuan and Chen Lian-Chang, "Algorithm of mining sequential patterns for web personalization services", ACM SIGMIS Database, vol. 40 , no. 2, pp. 57-66,May 2009.

[7] Zhenglu Yang, Yitong Wang and Masaru Kitsuregawa, "An Effective System for Mining Web Log", LNCS, vol. 3841, pp.40-52, 2006.

[8] Mobasher, H. Dai, T. Luo and M. Nakagawa, "Effective personalization based on association rule discovery from web usage.

[9] SizuHou, Xianfei Zhang, "Alarms Association Rules Based on Sequential Pattern Mining Algorithm," In proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 2, pp.556-560, Shandong, 2008.

[10] R. Agrawal and R. Srikant, "Mining Sequential Patterns", In Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, Taipei, Taiwan, 1995.

[11] Zhenglu Yang, Yitong Wang and MasaruKitsuregawa, "An Effective System for Mining Web Log", LNCS, vol. 3841, pp.40-52, 2006.

[12] Shengnan Cong, Jiawei Han and David Padua, "Parallel Mining Of Closed Sequential Patterns", in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 562 – 567, Chicago, Illinois, USA, 2005.

[13] Yan. X, Han. J and Afshar. R, "CloSpan: mining closed sequential patterns in large datasets", In Proceedings of the 3rd SIAM International Conference on data mining, pp. 166-177, San Francisco, CA, May 2003.

[14] Ming-Yen Lin, Sue-Chen Hsueh and Chia-Wen Chang, "Mining Closed Sequential Patterns with Time Constraints", Journal of Information Science and Engineering, vol. 24, pp. 33-46, 2008.

[15] MagdaliniEirinaki and MichalisVazirgiannis, "Web Mining for Web Personalization", ACM Transactions on Internet Technology (TOIT), vol. 3, no.1, pp. 1 – 27 , 2003.

[16] J. Ben Schafer, Joseph A. Konstan and John T. Riedl, "Recommender Systems for the Web", In Visualizing the Semantic Web, Springer, pp.102-123, 2006.

[17] OlfaNasraoui, Zhiyong Zhang, and EsinSaka, "Web Recommender System Implementations in Multiple Flavors: Fast and (Care) Free for All", in Proceedings of SIGIR Open Source Information Retrieval Worskhop,

Seattle, WA, July 2006.

[18] J. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-Commerce", in Proceedings of ACM Conference on Electronic Commerce (EC-99), pp.158-166, 1999.

[19] Hiroshi Ishikawa, Manabu Ohta, Shohei Yokoyama, Junya Nakayama, and Kaoru Katayama, "On the Effectiveness of Web Usage Mining for Page Recommendation and Restructuring", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol. 2593, pp. 253-267, 2009.

[20] Osmar R. Zaiane, Jia Li, and Robert Hayward, "Mission-Based Navigational Behaviour Modeling for Web Recommender Systems", in proceedings of the 6th International workshop on knowledge discovery, Seattle,

WA, USA, August 22-25, 2004.

[21] Mehdi Adda, PetkoValtchev, RokiaMissaoui and ChabaneDjeraba, "Toward Recommendation Based on Ontology Powered Web-Usage Mining", Internet Computing, vol. 11, no.4, pp. 45-52, 2007.

[22] Sarabjot Singh Anand and BamshadMobasher, "Intelligent Techniques for Web Personalization", Lecture Notes in Computer Science, Vol. 3169, Springer, 2005.

[23] Utpala Niranjan1, Dr.R.B.V. Subramanyam2, Dr.V.Khanaa, "An Efficient System Based On Closed Sequential Patterns for Web Recommendations", IJCSI International Journal of Computer S cience Issues, Vol. 7, Issue 3, No 4, May 2010.