Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 8 (2017) pp. 2289-2308

http://www.ripublication.com

An Efficient Web Personalization Approach based on Periodic Accessibility and Web Usage Mining

Y.Raju[1], Dr. D. Suresh Babu[2] and Dr. K. Anuradha[3]

[1]Department of Information Technology, Geethanjali College of Engineering and Technology, Hyderabad, India.
[2]Department of Computer Science, Head, Kakatiya Government College, Kakatiya University, India.
[3]GRIET, CSE Department , Hyderabad, India.

Abstract

The growth of the web and its usage is remarkable as in today needs. Users find it very complicated to extract useful and relevant information from huge amounts of information. Regular web personalization will aim to analyze the user's periodic browsing log patterns in web usage logs and to suggested the generally significant resources to users for a specific period of time. In this paper, an efficient web personalization approach based on user browsing time interval and utilizing web usage log is proposed. It initially creates a user activity model knows as individual web usage pattern utilizing the web log and fuzzy concept analysis. Depends on the individual web usage pattern it can efficiently suggest resources that users will be most interested in over a period of time. The practical evaluation of results suggests that periodic based web personalization is minimized the run-time processing load over the server and improvise the user satisfactory level.

Keywords: Web Personalization, Periodic Accessibility, Web usage mining, Fuzzy concept analysis.

1. INTRODUCTION

Personalization is based on the procedure of collecting and processing information regarding the site visitors, analyzing the information, and providing appropriate information to an individual visitor at the accurate interval supported. Many

personalization technologies [1], [7], [10], [11], allow it to aim site-targeted ads, product promotions, news feed customizations, document recommendations, appropriate advice, and email targeting. As the explosive of information over the web has made it complicated to retrieve appropriate information from the Web. Because of the web vastness and complex structure users often fails to find the target page when surfing the web. Web personalization is a ways to resolve this difficulty by leveraging the knowledge gained from analyzing the user's browsing activity to tailor the content and formation of the website to the needs [8].

A number of techniques have been explored in the web usage logs for web personalization to discover various web accessibility patterns. Even though, the periodic nature of browsing patterns is repeatedly ignored. Actually, various valuable browsing patterns happen repeatedly during a specific time period, this can depict a user surfing habits, but not during other time periods. These browsing patterns are called "periodic browsing patterns". A user's periodic browsing patterns build it simple to suggest resources that are most likely to be of interest to users over a period of time. Personalized resources can then be provisioned and expressed to users for a specific period of time. It can call as a problem because regular Web personalization utilizing regular web usage mining activities [14], [23]. It also not depends on the user's present browsing information, which can be ambiguous as largely web personalization approaches and other, periodic web personalization approaches can be just an intermediate web page prior to the current browsing to reach the goal page. Many different techniques have been considered to discover patterns from web browsing logs for web personalization. It consist of "rule-based filtering approaches", "content-filtering database to approach", "collaborative filtering approach", and "hybrid approaches" [9].

The personalization of a web page dynamically changes the content of a web page according to the preferences of the client or client, so that each client gets the information that suits their needs. Periodic Web personalization is the process by which a Web server and its associated applications vigorously modify content for a specific user depend on information about their activity on a website above a specific period of time. However, creating a regular Web personalization at runtime increases processing overhead as concurrent users grow. To overcome this problem, we propose "periodic web usage mining for efficient web personalization" to minimize processing overload at runtime. This paper proposes a new periodic web usage mining to construct user personalization page and user activity model in support of fuzzy concept theory [26] to support effective periodical web personalization.

This papers structured in four sections. Section-2 present Related Works, Section-3 presents Proposed Web Personalization Approach, section-4 present Performance Evaluation and section-5 presents the Conclusion of the work.

2. RELATED WORKS

Several research activities manage web-enabled mining and web personalization is performed in the past. Most attempts targets on retrieving constructive patterns and regulations by means of data mining practices to recognize the navigation activity of users. It allowing people to make decisions about site reorganization or modification. In several cases, the recommended engine will help users navigate the site. Some sophisticated systems offer much additional utility by establishing the concept of adaptive websites and by suggested resources to actively change the structure of the site. All research attempts merge to one or more of the previously described Web personalization such as "user profiling", "Web-enabled mining technology", "content management" and "publishing mechanisms".

2.1 Web Personalization

In a nutshell, Web personalization can be characterized as a task of customizing the information or provisions that a Web site provides to an each or group of users, supported by the facts gained by browsing activity evidenced in the Web site's logs. This information is repeatedly collective with the content and arrangement of the website in addition to the user's interests/preferences. Personalization has moved up during a number of stages. Initially, personalization was applied to give site visitors the opportunity to explore the site, advertise and promote their products. In the next step, the visitor wanted to increase the amount spent per visits by providing additional costly products or associated products. Nowadays, personalization is gradually utilized as resources of quickly communicating information to visitors, making them more useful and attractive.

The foremost module of the Web personalization construction is log mining and processing [15]. "Log analysis" and "Web Usage Mining" is a practice that enables it to apply useful statistics and data mining practices such as "clustering", "association rule discovery", "classification", and "sequential pattern" retrieval to process information accumulated in the web server logs to disclose valuable patterns and it able to analyze later [16]. These patterns depend on the method used and the input facts and can be correlations between "users page clusters, usage patterns, user groups and web pages". It can later store these patterns in a storage to perform query systems or OLAP work with visualization techniques.

In [5], the author proposed a framework for personalizing web mining based on web content along with web usage data and a site structure for more precisely predicting future requests from users. It suggested an algorithm known as "modified IncSpan" for efficient mining of chronological patterns in growing databases. This algorithm can detect database-based sequential patterns in sequential patterns, occur in the insert and append databases, and clogged sequential patterns come from result sequential

patterns.

In [18], M. Jalali et. al., has provided a solution supported by the LCS algorithm to analyze and process user search patterns for subsequent web page predictions. Their architecture improves classification accuracy and provides efficient online forecasting. Some evaluation techniques applied to evaluate the quality of found predictions. In [19], G. Fang et. al., to progress the effectiveness of the proposed algorithm and reduce database search time, we proposed a dual algorithm for Web-enabled mining based on sequential numbers that are appropriate for mining interval patterns. They used an algorithm that changed the user's interval pattern to binary. We then used the inconsistent search strategy to generate frequent itemset candidates twice. They also calculate and maintain for the sequence number aspect to scan the user's interval pattern once, unlike the existing dual search mining algorithm. Their experiments indicate that the competence of the algorithm is faster and additional proficient than suggesting similar algorithms.

In [21], the authors proposed an efficient and new architecture for Web search personalization by Web-enabled mining without user's explicit feedback. The author performs web personalization utilizing the "sequential browsing pattern mining algorithm" and "the Apriori algorithm". The Apriori algorithm provides frequent patterns and explosive candidate sequences but requires a lot of space for processing. We will use the FP-growth tree and the Markov model to reduce the limitations of this algorithm. It utilized prNG Graph to find user browsing patterns. Use FP-growth trees instead of prNG graphs to improve performance.

2.2 Web Usage Mining

The intention of web-enabled mining is to disclose unknown details in the web server's log file. It can be done by pertaining statistics and data mining schemes over web log data, it can also identify the remarkable patterns of user navigation activity, such as users and page clusters, in addition to likely relationships among web pages and user groups. The web usage mining procedure [17], [20] are viewed in three-step procedure consisting of "data preparation", "pattern discovery", and "pattern analysis". Highly developed data mining techniques and algorithms suitably tuned for utilized in a various web domain which consists of "association rules", "sequential pattern discovery", "clustering", and "classification".

In [4] a combines data mining practices with web accessibility log analysis to provide insights into the collaborative effects of Web-enabled mining and the potential for achieving its use in Web personalization. This work describes the information "pre-processing", "pattern discovery", and "pattern analysis" phases as three basic steps of the WUM process that web designers must follow in information extraction. In data mining algorithm is most commonly to declare that it always based on the problem in

hand as a result of the probable outcome of the WUM method. In spite of the dissimilarity of the framework, the information concealed in the click stream data found in the WUM method utilized for personalizing the Web and further calibrate and make the resulting improvements in previous application designs [6].

Fang Liu et. al. [12] proposed a technique to unite user profile and generic profile. This technique involves configuring user profiles based on the configuration of a generic profile supported on search history and the ODP category hierarchy. A user's search intent is represented by combining the two profiles and mapping the user's query to a collection of categories. The technique of combining the two profiles shows that it produces more accurate results than utilizing only a single profile.

Web Usage Mining [3], [8] is a potential technology for periodic web personalization, mining web usage logs for user activity modeling and personalization. In research works [2], [13], [9] was performed to find periodic patterns in a time series database. Most of these tasks, however, focus on regular pattern mining instead of pertaining these patterns to realistic applications. They only converse a few prospective applications for periodic accessibility patterns. In addition, Li and Deogun [22] presents that cyclic patterns will be applied for data classification as well as for upcoming periodic calculation. Nevertheless, as far as we identify, no approach has been anticipated for the periodic accessibility based web personalization.

3. PROPOSED WEB PERSONALIZATION APPROACH

Fig.1 illustrates the architecture of the proposed Web personalization structure based on periodic Web usage log facilitated mining. The entire web accessibility requests are initially written to the web server and accumulated in the web usage log storage.

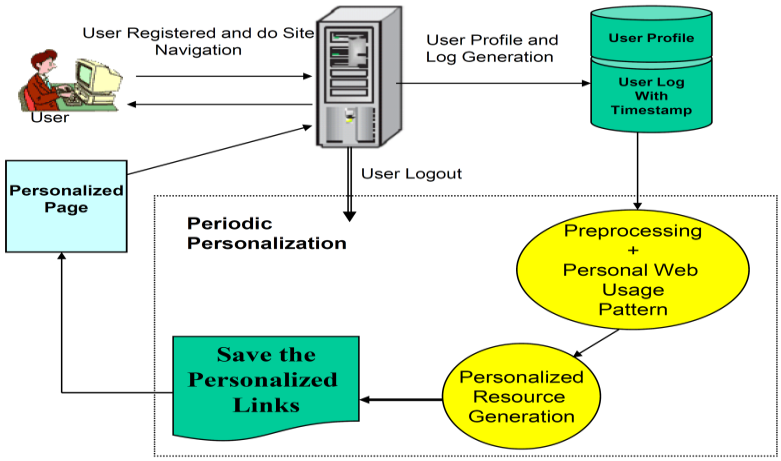

Fig.1: Proposed Web Personalization Approach Architecture

The configure of "User Activity Model" module is configured for the individual web usage pattern model. When a period condition is particular to an individual user or a web server, the module for periodic log-based web personalization generation from a personalize resource and an individual web usage pattern. The web server afterward provisions the configuration and presents the resource to the user for a specified period of time. The proposed system consist of three modules as, Data preprocessing, Constructing web Usage patterns, and Constructing Periodic Personalized Resources.

A. Data Pre-processing

The URL recorded in the web usage log contains very modest semantic information regarding the web content that the user accessed [24], [26], [27]. This study assumes that every requested URL in the web usage log has its semantic information interpreted with one or more predefined subjects or type, such as "news, sports, and entertainment". These tasks can be done effortlessly by the website creator or manager, either manually or semi-automatically. An instance of a semantically rich web usage log is shown in Table-1.

Table 1: An illustration of web usage log

UserIP	Time Stamp	Method	URL
192.168.200.135	01/Jan/2014:06:58:33 +0530	GET	/news/Police2.htm HTTP/1.1
192.168.201.204	01/Jan/2014:07:56:50 +0530	GET	/news/seagate_march.htm HTTP/1.1
192.168.202.212	01/Jan/2014:22:40:00 +0530	GET	/news/ITT.htm HTTP/1.1
192.168.202.203	01/Jan/2014:23:11:22 +0530	GET	/travel/restaurents.htm HTTP/1.1

If a user periodically accesses a specific resource, the user can be considered to have a web accessibility, which can be understood as "a particular user is interested in a particular resource for an assured period of time". Therefore, it can use the periodic property "M_p" and the resource property "M_r" to symbolize the web accessibility activity.

In this activity, we described eight actual sequential conceptions as, "early morning", "morning", "noon", "early afternoon", "late afternoon", "evening", "night", and "late night" with regular properties. Otherwise, one can employ days of the week and other actual sequential concepts of periodic properties for more common intentions. All pre-defined subjects for a site can be utilized as resources to illustrate the properties of Web browsing activities that are considered. The choice of suitable domain-specific subjects based on the application. For instance, the item name in the list can be utilized for e-commerce sites, and also for news categories for news sites.

In preprocessing, we need to process the innovative Web usage log sequentially to recognize all individually browsed intervals for the individual user. It utilizes the preprocessing works for the traditional web server logs as considered in [1] namely for "data cleaning", "user identification", and "interval identification".

- **Data Cleaning:** Depend on the particular resource properties M_r, it will remove unwanted records concerning with the Semantic Annotation field. Entries are considered useful in web usage logs when in any case one of their resource properties is mentioned in their requested URLs.
- **User identification:** It instruct to analyze individual browsing activity, the unique user should be recognized with the "UserIP" value. To personalize a website, the methods should have to be capable in differentiating among dissimilar individual or groups users.
- **Interval identification:** It defines the gap involved in the timestamp of two repeating requests from the similar user in an interval, it also assumed that the latest surfed interval has started. It applies a constant 30 minutes limit for the timeout.

An browsing interval "$S = \{(URL_1, t_1), (URL_2, t_2), \ldots, (URL_n, t_n)\}$" is a repetition of requested URL_i having a timestamp "t_i $(1 \leq i \leq n)$". The interval d_i of URL_i calculated basically as "$d_i = (t_i{+}1 - t_i)$". For URL_n that does not have "$t_n{+}1$", it's time interval will be computed by standard recent interval, i.e., "$d_n = (d_1 + d_2 + \ldots + d_{n-1}) / (n - 1) = (t_n - t_1) / (n - 1)$". To find out the d_n, it only needs to maintain the correlated interval having $n > 1$, which contains one or more requested URLs. In addition, it explores the session time from beginning to end as S with "t_1" and "$(t_n + d_n)$)" correspondingly. For continuous exploring, it overlooks the information within the session times of the interval and converts it to a value from 0 to 24. For instance, the interval start time "21/02/2014 09:25:01" can depict as 9.25.

The period of the accessibility interval $p(S)$ is described by successive time intervals of the start time as "$t_s(S) \in [0, 24]$" and the end time "$t_e(S) \in [0, 24]$". It is indicated as,

$$p(S) = \begin{cases} [ts(S), te(S)], & \text{if } ts(S) \leq te(S) \\ [0, te(S)] \cup [ts(S), 24], & \text{otherwise} \end{cases}$$

Assume that each URL_i in the browsing interval S is correlated with a resource feature collection as "$M_{ri} \subseteq M_r$", that represents the meaning of the contents of the URL_i. In this consideration, each browsing interval can be considered as a sequence of a group resource features as "M_{ri}", and denoted as "$S = \{(M_{r1}, t_1, d_1), (M_{r2}, t_2, d_2), \ldots, (M_{rn}, t_n, d_n)\}$". The entire interval for every "$m_k \in M_r$" is utilized to estimate the user's point of interest in the resource for the period of the browsing interval S, which can be calculated as follows,

$$d(S, m_k) = \sum_{i=1}^{n} \alpha_{ki} d_i, \text{ where } \alpha_{ki} = \begin{cases} 1, & \text{if } m_k \in M_{ri} \\ 0, & \text{otherwise} \end{cases}$$

B. Constructing Web Usage Patterns

In individual web usage pattern configuration, we recognize the user's web browsing activity and configure the individual web usage pattern in the user's individual browsing interval. In fact, others may have different interpretations of the temporal concept of real life. For the requested resource, there is a similar issue because it is complicated to accurately understand the user's intent through the browsing interval. Fuzzy theory [13] is a suitable technique for describing the uncertainty of information and is integrated into a formal concept analysis [3] to represent features both periodic and resources.

A fuzzy periodic web usage perspective is presented as, "$K = (G,Mp,Mr, I)$", where "G" is a collection of browsing intervals for a user, "Mp" is a collection of periodic features, "Mr" is a collection of resource features and "$I = R(G \times (Mp \cup Mr))$" is a fuzzy collection in the domain of "$G \times (Mp \cup Mr)$" correspond to fuzzy associations among browsing intervals "$g \in G''$ and features "$m \in Mp \cup Mr\ ''$. Each fuzzy association "$R(g,m) \in I$ " is symbolized by elements of "$\mu(g,m) \in [0, 1]$", where,

$$\mu(g,m) = \begin{cases} \mu_p(g,m), & \text{if } m \in M_p \\ \mu_r(g,m), & \text{if } m \in M_r \end{cases}$$

Based on the above explanation, each browsing interval "$g \in G''$ in addition it indicated as a fuzzy collection in the domain of "$Mp \cup Mr$ ", i.e., "$g = \{m, \mu(g,m) \mid m \in Mp \cup Mr\ \}''$. The elements of "$\mu p(g,mp)$" for a periodic feature "$mp \in Mp$" in an browsing interval "$g \in G$ " can be computed utilizing the period of g, i.e., "$p(g)$".

In this work, the associated utility is characterized as "$\mu_p(g,m_p) = \max_{t \in p(g)} \{\mu_p(t, m_p)\}''$, where "$\mu_p(t,m_p)$" is described in Fig.2, which is personalized from [26].

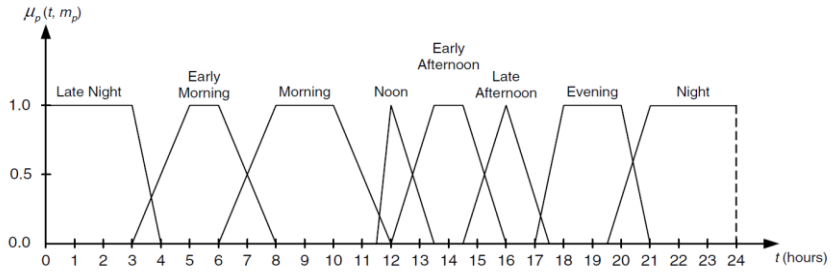

Figure 2: Associated utility $\mu_p(t,m_p)$

In the browsing interval, the associated value "$\mu r(g,mr)$" for the resource feature "$mr \in Mr''$ can be calculated utilizing the entire interval of "mr'', i.e, "$d(g,mr)$". The relevant functions in this study are described as follows: "$Z(mr)$" is the ratio of the entire interval of browsing to the resource "mr" in all browsing intervals of the user,

which represents the user's overall interest in the resource "*mr*". The "*z*(*g*,*mr*)" is described as the ratio of how long to browsing resource "*mr*" in browsing interval "*g"*, which represents the user's confined interest in resource "*mr*".

The fuzzy intervallic web usage perspective can be symbolized as a cross table with rows and columns, where rows present browsing intervals and columns period and resource features. The elements of "$\mu(g,m) \in [0, 1]$" of row "*g"* and column "*m"* have the fuzzy relationship among browsing interval "*g"* and feature " *m"*. It can organize the individual web usage patterns supported by the fuzzy periodic web usage context. The web usage pattern "U_P " based on the user's web usage perspective "$K = (G,Mp,Mr, I)$" is "$U_P = (W_P,<_{W_P})$", where "W_P" is a collection of all web browsing activities, and "$<_{W_P}$ " is a partial activity on "W_P" to correspond to the order of the web browsing activities.

C. Construction of Periodic Personalization Resources

Personalization not only saves users unnecessary browsing time but also gives them the opportunity to get specific information relevant to their interests, additionally improve overall website user experience, improving structure, functionality and browsing experience. To organize the page periodically, it needs to search the individual web pattern. A cyclic circumstance is described as a constant time period having a start and end time interval. The process of the periodic personalization flow is shown in Fig.3.

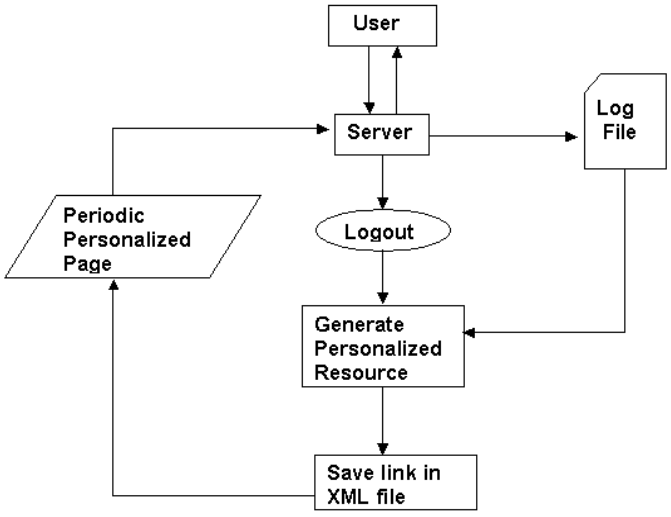

Figure 3: Process Flow for periodic Personalization Resources

This will perform pre-processing in support of the web browsing logs, organize web page patterns, and personalized resources to measure the "applicability" and "satisfaction" of the users.

Algorithm-1 gives the algorithm *PR_Generation* to generate an ordered personalized resource collection for a given periodic "p_c" supported by the Individual Usage Pattern of the user "U_P". In the algorithm *Search_Periodic_Activity* given in Algorithm-2. It retrieves all duration support activities based on the "U_P". All retrieved duration support activities are arranged in high to low of a priority. Because all personalized resources are from the same period supporting activities, and they are arranged high to low utilizing fuzzy values.

Algorithim-1: ***PR_Generation*** **(p_c, U_P)**

Input:
p_c - a period condition;
$U_P = (W_P, <W_P)$ - a Personal Web Usage Pattern;
Output:
$PRo(p_c, U_P)$ - a set of ordered personalized resources;
Process:
Initialize $PRo(pc, U_P) \leftarrow \{\emptyset\}$;
for all $v(L) \in W_P$ **do**
$v(L).\ status \leftarrow unvisited$;
end for
Acquire period-supported activities $PSA_p(p_c, U_P) \leftarrow$ ***Search_Periodic_Activity*** $(p_c, U_P, v(\emptyset))$;
for all $v(B) \in PSA_p(p_c, U_P)$ in priority top to bottom order **do**
for all $m_r \in B \cap M_r$ in priority top to bottom order **do**
if $m_r \notin PRo(p_c, U_P)$ **then**
Append m_r into $PRo(p_c, U_P)$;
end if
end for
end for
return $PRo(p_c, U_P)$;

Algorithm-2 provides a ***Search_Periodic_Activity*** algorithm for periodic support activity searches for a specified periodic circumstance "p_c", depends on the individual web usage pattern of a specific user "U_P". To increase the effectiveness of the periodic support activities search process, the algorithm begins exploring at the high node of the pattern and recursively searches only the through the sub-activities that satisfy the periodic circumstance "p_c".

Algorithim-2: ***Search_Periodic_Activity*(p_c, U_P, $v(L)$)**

Input:
p_c - a period condition;
$U_P = (W_P, <W_P)$ - a Personal Web Usage Pattern;
$v(L)$ - the current web access activity;
Output:
$PSA_p(p_c, U_P)$ - a set of period-supported activities;
Process:
Initialize $PSA_p(p_c, U_P) \leftarrow \{\emptyset\}$;
for all direct sub-activities of $v(L)$, i.e., $v(B_i) \in W_P$ with $v(L_i) \in W_P$ $v(L)$ **do**
if $v(L_i).status = unvisited$ **and** $v(L_i) \cap P_f(p) \neq \emptyset$ **then**
if $v_p(L_i) \subseteq P_f(p_c)$ **or** $v_p(L_i) \supseteq P_f(p_c)$ **then**
$PSA_p(p_c, U_P) \leftarrow \{v(L_i)\} \cup$ ***Search_Periodic_Activity*** $(p_c, U_P, v(L_i))$;
else
$PSA_p(p_c, U_P) \leftarrow$ ***Search_Periodic_Activity*** $(p_c, U_P, v(L_i))$;
end if
$v(L_i).\ status \leftarrow visited;$
end if
end for
return $PSA_p(p_c, U_P)$;

All the web browsing requests are first written to the web server and accumulated in the web usage log. The "User Activity Model Configuration" module then configures the individual Web usage pattern model. In a periodic circumstance specified by a user or a web server for the periodic web personalization module generates a personalized resource from the individual web usage pattern. The web server then being configured to presents the resource to the user for a specified period of time.

4. PERFORMANCE EVALUATION

The evaluation made by the user logs collected from the "www.reachouthyderabad.com" website it provides all variety of information related to the twin cities. An average of 9 Lakhs hits for a month from around the world it records. Three-month user log records are collected from January to March 2014. It contents 4847 records, a sample of its is shown in Fig.4. Because it makes problems to evaluate online, we build an offline system for the evaluation. We have defined eight actual hours as, "*Early Morning*", "*Morning*", "*Noon, Early Afternoon*", "*Late Afternoon, Evening*", "*Night* and *Late Night*" to indicate the temporal nature of the web activity.

```
192.168.200.135 - - [01/Jan/2014:06:56:07 +0530] "GET / HTTP/1.1" 200 1494
192.168.200.135 - - [01/Jan/2014:06:56:08 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:56:40 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:56:50 +0530] "GET /news/seagate_march.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:57:09 +0530] "GET /news/Police2.htm HTTP/1.1" 404 294
192.168.200.135 - - [01/Jan/2014:06:57:10 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:57:22 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:57:43 +0530] "GET /news/seagate_march.htm HTTP/1.1" 404 308
192.168.200.135 - - [01/Jan/2014:06:58:33 +0530] "GET /news/Police2.htm HTTP/1.1" 404 294
192.168.200.135 - - [01/Jan/2014:06:58:42 +0530] "GET /morenews.htm HTTP/1.1" 404 326
192.168.200.135 - - [01/Jan/2014:06:58:58 +0530] "GET /news/ca.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:56:07 +0530] "GET / HTTP/1.1" 200 1494
192.168.201.204 - - [01/Jan/2014:07:56:08 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:56:40 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:56:50 +0530] "GET /news/seagate_march.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:57:09 +0530] "GET /news/Police2.htm HTTP/1.1" 404 294
192.168.201.204 - - [01/Jan/2014:07:57:10 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:57:22 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:57:43 +0530] "GET /news/seagate_march.htm HTTP/1.1" 404 308
192.168.201.204 - - [01/Jan/2014:07:58:33 +0530] "GET /news/Police2.htm HTTP/1.1" 404 294
192.168.201.204 - - [01/Jan/2014:07:58:42 +0530] "GET /morenews.htm HTTP/1.1" 404 326
192.168.201.204 - - [01/Jan/2014:07:58:58 +0530] "GET /news/ca.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:40:00 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:40:08 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:41:04 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:41:13 +0530] "GET /news/ITT.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:41:48 +0530] "GET /astro/index.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:42:40 +0530] "GET /astro/index.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:43:22 +0530] "GET /cheers/index.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:44:42 +0530] "GET /cheers/index.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:45:23 +0530] "GET /news/vizai.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:46:11 +0530] "GET /travel/beyondhyd.htm HTTP/1.1" 404 326
192.168.202.212 - - [01/Jan/2014:22:47:41 +0530] "GET /travel/district_services.htm HTTP/1.1" 404 326
```

Fig 4: Web Usage Log File

A weblog is a copy of a transaction recorded by the server for a collection of users. Fig.4 illustrates a few lines of such distinctive log transaction. For past reasons, several weblogs utilize the identical structure. Several fields are quoted and some are split by a space, usually a particular space.

```
<?xml version='1.0' encoding='iso-8859-1'?>
<userDetails>
<uip>192.168.200.135</uip>
<EARLY-MORNING>
    <head-id>4</head-id>
    <shead-id>10</shead-id>
    <link>/students_den.htm</link>
    <link-name>List of City Colleges  |Courses and Training|
    Profile of academic institutes|
    IAS Coaching|Interviews|Writing a resume|MBA|MCA </link-name>
    <access-date>25/Feb/2014</access-date>
</EARLY-MORNING>
<EARLY-MORNING>
    <head-id>2</head-id>
    <shead-id>5</shead-id>
    <link>/astro/index.htm</link>
    <link-name>Your Horoscope/Forecast for Today/Month</link-name>
    <access-date>27/Feb/2014</access-date>
</EARLY-MORNING>
<EARLY-MORNING>
    <head-id>2</head-id>
    <shead-id>5</shead-id>
    <link>/astro/index.htm</link>
    <link-name>Know Your Birthday Tree/General Star Sign</link-name>
    <access-date>27/Feb/2014</access-date>
</EARLY-MORNING>
<EARLY-MORNING>
    <head-id>3</head-id>
    <shead-id>8</shead-id>
    <link>/ithyderabad/index.htm</link>
    <link-name>IT Companies |IT Schools|IT Newsmakers</link-name>
```

Fig. 5: Periodically organized Personalized Resource

To achieve the preprocessing four frequent users with most browsing intervals are identified and a new log file utilizing Java I/O API is created. To speed the filter process we build a sitemap lookup table known as a lookup. Utilizing lookup table we try to find how persistently a user visited a particular link in a specified period. In support of the similar web Browsing Activity, Periodic Supported Activities and organize the resources it generates the personalized resources in a specified period of the interval we consider it as user interest towards it. We record the links into a temporary buffer utilizing Hashtable to building the patterns. Running pre-process on the log file it constructs an organized Personalized resource tree as shown in fig.5. This XML file parsed at runtime utilizing SAX Parser to generate and delivered to the user personalized page.

A. Evaluation Measures

To simplify web browsing for end users, many techniques being proposed for personalizing the web. However, real-time web personalization is a costly approach and delays in page serving slow user experience. Periodic personalization satisfies users more effectively. Periodic browsing patterns allow it to easily infer resources that the users might be most interested in over a period of time. Providing time-based real-time services required massive computing with increased server overhead, which could result in long delays and dissatisfaction with users if the server makes more requests in an assured period of time.

- **User Applicability**: Let "$PR_{all} = \{PR_1, PR_2, \cdots, PR_n\}$" be a gathering of collections of personalized resources for in general web personalization, and "PR_a" is the subset of "PR_{all}" consist of all appropriate collections of personalized resources. The applicability of in general web personalization is described as,

$$applicability = \frac{|PR_a|}{|PR_{all}|} \qquad (1)$$

- **Personalization Process Load:** It measures the processing load "L_S", on the server for generating each user Period Web Personalization at run time in a number of personalized resource *PR*, utilizing the following equation.

$$PLoad(PL_S) = \frac{((R_Q + PR) \times \beta)}{S_P} \times 100 \qquad (2)$$

where,

$$PR = \frac{R_Q}{3}$$

and,

$R_Q \rightarrow$ No. of Request at a particular time

$PR \rightarrow$ Number of Personalization Resource

$\beta \rightarrow$ Load Constant for Personalization generation

$S_P \rightarrow$ Server Process Handing Capacity

- **Satisfaction:** Let S_T be the collection of all browsing intervals for testing. "$SS_p(PR_i)$" is the subset of S_T comprising all "period-supported intervals" of the personalized resources "PR_i", and "$SS_r(PR_i)$" is the subset of "$SS_p(PR_i)$" comprising all "resource-supported intervals". The satisfaction of the personalized resources "PR_i" is described as,

$$satisfaction(PR_i) = \begin{cases} 0, & \text{if } SS_p(PR_i) = \emptyset \\ \frac{|SS_r(PR_i)|}{|SS_p(PR_i)|}, & \text{otherwise} \end{cases}$$

The "*satisfaction*" for in general web personalization is described as,

$$satisfaction = \frac{\sum_{PR_i \in PR_a} satisfaction(PR_i)}{|PR_a|}$$

where PR_a is the personalized resources of a user.

B. Result Analysis

- **User Applicability**: Applicability of the user in general web usage log is evaluated supported on the usages made on the different durations of periods. Fig.6 shows the applicability distribution of the users and at the different period interval. The distribution results illustrate that the anticipated periodic web personalization approach will achieve satisfactory coverage for predefined time period circumstances over a reasonable period of time.

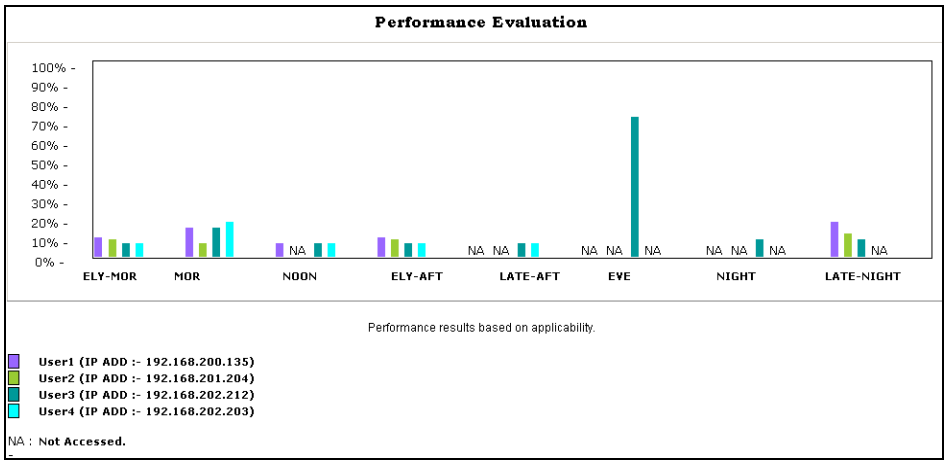

Figure 6: Users applicability distribution

- **Load Evaluation:** To evaluate we have considered two scenarios. In the first scenario, we compute the load performance in run-time (i.e., on request), in the second scenario we measure periodic personalization generated in advance. The comparison results are shown in Fig.7 and 8. It shows a clear indication that with periodic and in advance personalized resource *PR* generated, minimize the processing load on the server. We configure Server maximum process handling capacity, $S_P = 1000$. In the run-time scenario, we consider $\beta = 3$ and in the second scenario $\beta = 1$ because at run-time server need to do 2 times more processing then the advance *PR* generated.

Table 1: Personalization Load Processing Comparison

No. of Requests (R_Q)	Number of Personalization Resource (*PR*)	Load (PL_s) at Run-time (%)	Load (PL_s) at periodic and adv. generation (%)
100	33	40	13
150	50	60	20
200	67	80	27
250	83	100	33

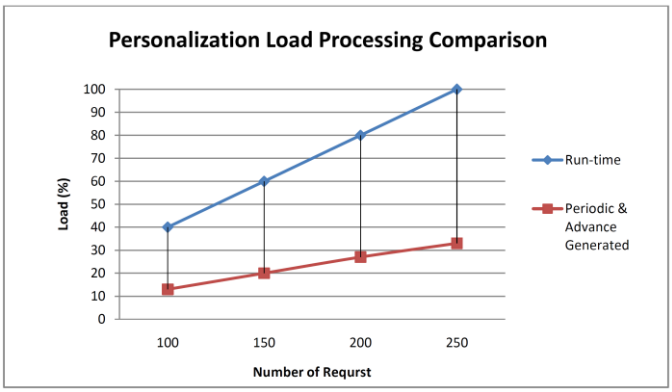

Figure 7: Personalization Load Processing Comparison

In the evaluation result in Fig.7, illustrates that as the number of requests increases, the load on the server increases at the same time, causing the server to overload. Server overloading delays request processing and providing the personalized resources. The delay in the service leads to the dissatisfactions to the users. To overcome this, it must efficiently minimize the personalization load on the server at runtime. To do so, it needs to prepare resources in advance to make the service easier and efficient.

- **Satisfaction:** The satisfaction measures how possible a user is concerned in one of the personalized resources in the period maintained intervals. We calculate the satisfaction for the users varying the number of PR and the time period from 1 to 5 hours. The result obtains are presented in Fig. 8, 9 and 10. The result depicted that with increase *PR* and time period satisfaction level increases.

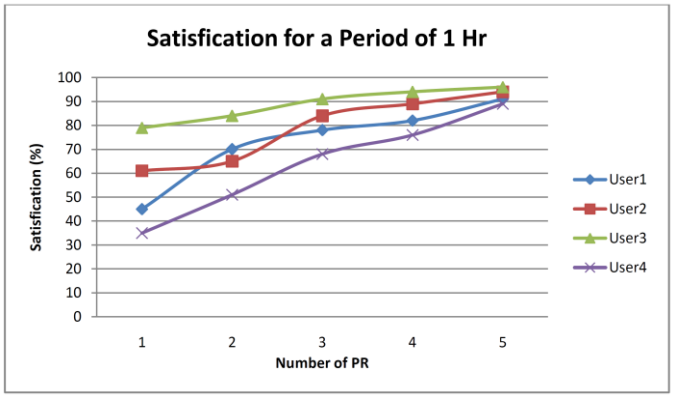

Fig.8: User Satisfaction for a Period of 1 hr

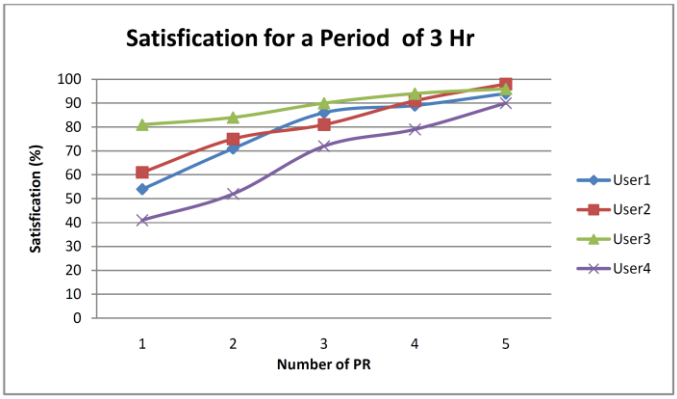

Fig.8: User Satisfaction for a Period of 3 hr

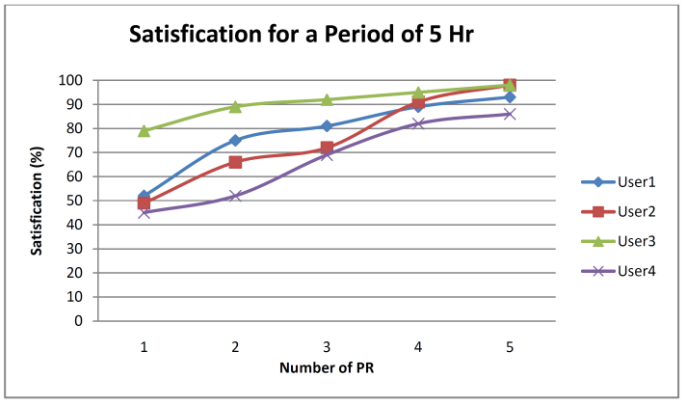

Fig.9: User Satisfaction for a Period of 5 hr

Fig.7, 8 and 9 show the overall satisfaction of different web personalization over a period of one to five hours, depend on a dissimilar number of personalized resources also beginning 1 to 5. Satisfaction also increases when an increase in the number of personalized resources all figures shown. However, if the number of personalized resources is 4 or more, the increase is not significant. It can improve user satisfaction with personalized resources, but various personalized resources will influence the resource pre-work process that will be constructive to the user.

5. CONCLUSION

This paper presents an efficient web personalization approach depending on the periodic accessibility and web usage log mining. It will support periodic web personalization utilizing individual users' periodic browsing patterns. Unlike the periodic approach, the proposed approach allows users to proficiently established which resources they are mainly concerned in for a given interval of time exclusive of utilizing the user's existing accessibility information. As a result, it can perform personalized resource provisioning that is more costly in the unrealistic scenarios. The testing evaluation results suggested that the proposed approach quite effective in terms user applicability and satisfactory measure in various time interval utilizing web usage log. In future, it can be evaluated in the longer interval in multiple periods to measure the effectiveness.

REFERENCE

[1]. F. Cai,Shuaiqiang Wang,Maarten de Rijke, "Behavior-Based Personalization in Web Search", Journal Of The Association For Information Science And Technology, 68(4):855–868, 2017.

[2]. X. Liu, , & H. Turtle, "Real-time user interest modeling for real-time ranking", Journal of American Society for Information Science and Technology, 64(8), 1557–1576, 2013.

[3]. C. Liu, R.W. White & S.Dumais, "Understanding web browsing behaviors through log analysis of dwell time", In Proceedings of the 33rd ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 379–386, 2010.

[4]. K. Srinivasa Rao, A. Renuka Kalyani, M. Krishnamurthy, "A New Approach of Inferring User Interest for User Search Goals with Web Log Sessions", Journal of Web Development and Web Designing, Volume-1 Issue-2, Page 1-15, 2016.

[5]. S. Khonsha and M. Hadi Sadreddini, "New hybrid web personalization framework", In Proc. of IEEE 3rd International Conference on Communication Software and Networks, pp.86-92, 2011.

[6]. T. Thanh Sang Nguyen, H. Yan Lu, and J. Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge" IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 10, October 2014.

[7]. M. Dhandi, R. K. Chakrawarti, "A Novel Approach for Web Personalization", International Journal of Computer Applications (0975 – 8887) Volume 151 – No.8, October 2016.

[8]. V. Dongre, J. Raikwal, "An Improved User Browsing Behavior Prediction Using Web Log Analysis", International Journal of Advanced Research in Computer Engineering & Technology, Volume 4 Issue 5, May 2015.

[9]. S. Jagan, S.P. Rajagopalan, "A Survey on Web Personalization of Web Usage Mining" International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 01, March-2015.

[10]. Beatric, R. Fernandes, Leo. J. Peo, N. Kamat, S. Miranda, "New Approaches to Web Personalization Using Web Mining Techniques", International Journal of Computer Science and Information Technologies, Vol.5 (2),2195-2201, 2014.

[11]. M. Eirinaki and M. Vazirgiannis, "Web mining for web personalization", ACM Transactions on Internet Technology, 03(01):1-27, February 2003.

[12]. Fang Liu, Clement Yu , Weiyi Meng, "Personalized Web Search For Improving Retrieval Effectiveness" , CIKM02, pp. 1-35, 2002.

[13]. K. Sudheer Reddy, G. Partha Saradhi Varma, M. Kantha Reddy, "An Effective Preprocessing Method for Web Usage Mining", International Journal of Computer Theory and Engineering, Vol. 6, No. 5, pp. 412-415, October 2014.

[14]. M. Jafari, F. Soleymani Sabzchi, A. Jalili Irani, "Applying Web Usage Mining Techniques to Design Effective Web Recommendation Systems: A Case Study", ACSIJ International Journal, Vol. 3, Issue 2, No.8, pp. 78-90, March 2014.

[15]. Robert F. Dell, Pablo E.Roman, and Juan D.Velasquez, "Web User Session Reconstruction Using Integer Programming", IEEE International Conference on Web Intelligence and Intelligent Agent, 2008.

[16]. C. I. Ezeife and Y. Lu, "Mining Web log sequential patterns with position coded pre-order linked WAP-tree", IEEE Data Minining Knowledge Discovery, Vol. 10, no. 1, pp. 5–38, 2005.

[17]. S. Bosnjak, M. Maric, Z. Bosnjak, "The Role Of Web Usage Mining In Web Application Evaluation", Management Information Systems, Vol. 05(01):31-36, 2010.

[18]. M. Jalali, Norwati Mustapha, Md. Nasir B Sulaiman, Ali Mamat, A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems, 12th International Conference Information Visualisation, 2008.

[19]. G. Fang, Jia-le Wang, Hong Ying, Jiang Xiong, "A double algorithm of Web usage mining based on sequence number", IEEE, 2009.

[20]. B. Liu, B. Mobasher, and O. Nasraoui, "Web usage mining in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data", Springer-Verlag,Berlin, Germany, pp. 527–603, 2011.

[21]. K. P. Adhiya And S. R. Kolhe, "An Efficient And Novel Approach For Web Search Personalization Using Web Usage Mining", In Journal Of Theoretical And Applied Information Technology, Vol. 73, No. 2, March 2015.

[22]. D. Li and J. S. Deogun, "Discovering Partial Periodic Sequential Association Rules with Time Lag in Multiple Sequences for Prediction", In Proc. of the 15th Intl. Symposium on Foundations of Intelligent Systems, pages 332–341, 2005.

[23]. B. Mobasher, R.Cooley and J. Srivastava, "Automatic personalization based on web usage mining", ACM Communication,43(08):142-151, August 2000.

[24]. S. Salin and P. Senkul, "Using semantic information for web usage mining based recommendation", In Proc. 24th ISCIS, Guzelyurt, Turkey, pp. 236–241, 2009..

[25]. D. Olaru and B. Smith, "Modelling Daily Activity Schedules with Fuzzy Logic", In Proc. of the 10th Intl. Conf. on Travel Behaviour Research, 2003.

[26]. G. Stumme, A. Hotho, and B. Berendt, "Usage mining for and on the Semantic Web", In Data Mining: Next Generation Challenges and Future Directions. Menlo Park, CA, USA: AAAI/MIT Press, pp. 461–480, 2004.

[27]. H. Dai and B. Mobasher, "Integrating semantic knowledge with web usage mining for personalization", In Web Mining: Applications and Techniques, IGI Global, pp. 205–232, 2005.