A
Major Project Report II
On

An Effective Periodic Web Content Recommendation Based on Web Usage Mining

Submitted in Partial Fulfillment of the Requirement

For the Award of the Degree of

Master of Technology

In

Software Engineering

By

Ravi Khatri
University Roll No. 2K13/SWE/15

Under the Esteemed Guidance of

Dr. Daya Gupta
Computer Science & Engineering Department, DTU

2013-2015

COMPUTER SCIENCE & ENGINEERING DEPARTMENT

DELHI TECHNOLOGICAL UNIVERSITY

DELHI – 110042, INDIA

DECLARATION

I hereby declare that the major project – II work entitled **"An Effective Periodic Web Content Recommendation Based on Web Usage Mining"**, which is being submitted to Delhi Technological university, in partial fulfillment of requirements for the award of degree of Master Of Technology (Software Engineering) is a genuine report carried out by me. The material contained in the report has not been submitted to any university or institution for the award of any degree**.**

Ravi Khatri

2K13/SWE/15

CERTIFICATE

This is to certify that the project report entitled **"An Effective Periodic Web Content Recommendation based on Web Usage Mining"** is a bona fide record of work carried out by Ravi Khatri (2K13/SWE/15) under my guidance and supervision, during the academic session 2013-2015 in partial fulfillment of the requirement for the degree of Master of Technology in Software Engineering from Delhi Technological University, Delhi.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any Degree or Diploma.

DATE…….

Dr. Daya Gupta

Professor

Department of Computer Engineering

Delhi Technological University

ACKNOWLEDGEMENT

I would like to express my deep sense of respect and gratitude to my project supervisor Dr. Daya Gupta for providing me the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to him for the support, advice and encouragement he provided me without which the project could not have been a success.

Last but not the least I would like to express sincere gratitude to my parents and friends for constantly encouraging me during the completion of work.

Ravi Khatri

M.Tech (SWE)

2K13/SWE/15

Dept. of Computer Science & Engineering

DTU, Delhi

TABLE OF CONTENTS

List of Figures

List of Tables

Abstract

Now a day's use of internet has been increased tremendously, so providing information relevant to a user at particular time is very important task. Periodic web personalization is a process of recommending the most relevant information to the users at accurate time. In this paper we are proposing an improved personalize web recommender model, which not only considers user specific activities but also considers some other factors related to websites like total number of visitors, number of unique visitors, numbers of users download data, amount of data downloaded, amount of data uploaded and number of advertisements for a particular URL to provide a better result. This model consider user's web access activities to extract its usage behavior to build knowledge base and then knowledge base along with prior specified factors are used to predict the user specific content. Thus this advance computation of resources will help user to access required information more efficiently and effectively.

Keywords—Web Usage Mining, Web usage logs, web recommendation, Knowledge Base, Periodic personalization.

CHAPTER 1: INTRODUCTION

1.1 Overview

As we know today web is one of the largest sources of information and it is expanding with unprecedented rate, web is not only used for gathering information but also used for making purchases (E-Commerce) and also a new way of expanding business. Along with development in internet technology there is also possibility that users may lose in order to find relevant resources. Traditionally users gather information via search engines or by uniform resource locator (URL). But in order to provide information efficiently and effectively for better use of resources in timely manner we have to understand the users web access behavior and his/her navigational pattern for which web mining activity is carried out in order to recommend the resources. User's web access activity varies according with time i.e. a user searches for one piece of information in morning, second in afternoon and some different information in evening. So providing global best available resources by understanding the user's time varied web access behavior is main task in this project.

The concept of recommendation system is important because traditional search process sometime is not very effective because it depends on user search query, if user's query is not very effective than it results in poor results and also some websites are paid to search engines in order to display their link at top because of this recommendation system come into existence and play a very important role in today's internet technology.

1.2 General Concepts

i) *Data Mining***:** -Data mining [1] is a process of extracting hidden knowledge and patterns from large amount of data for decision making process. And this knowledge and patterns is used to predict the future, which allows us to make decisions regarding an organisation to which data is associated. Data mining technique has profound impact because results obtain by this technique is much better than results obtain by writing queries on the relationship define by user. Basically this technique allows us to make smarter decisions, finding new opportunities, solving and dealing with problems more effectively.

ii) *Techniques of Data mining : -*There are different techniques [2] which are applied over large set of data in order to extracts knowledge. They are as follow

a) *Outlier Detection:* -Outlier detection means finding data items that is not matching the pattern formed by the rest of data items i.e. outliers are those data items which show totally different behaviour from remaining data set and such data items require additional analysis.

b) *Association Rule Learning:* -Association rule learning helps us to find interesting relations among different data sets variables or we can say it is used to find hidden patterns in the data set variables and then these patterns are used to find variables which occur with high frequency.

c) *Classification analysis:* -Classification analysis is an approach to obtain information about data based on certain attributes in order to separate them in different groups. This information helps to set up different categories where different types of data follows different category.

d) *Cluster analysis:* -Cluster analysis is the process of identifying different classes or groups for objects of unknown classes and they are separated in such a manner that objects with similar characteristics are in the same class which are separated from other objects. On the basis of this division different cluster are defined by their cluster description in which objects common features are summarized.

e) *Regression analysis:* -Regression analysis is an analysis in which relationship between different variables are to find out where one variable is dependent upon on other but reverse is not true. Relationship is from independent variables to dependent variables.

iii) *Web Mining:* -Web mining [3, 4] is a part of data mining, which deals with extraction of knowledge and useful patterns from internet or WWW. Web mining helps us to retrieve information from web very effectively and efficiently. It is further categories in three section web content mining, web structure mining, and web usage mining, [3, 4] performs survey on web mining techniques and deal with how to extract knowledge from web. Categories of web mining are defined as follow

a) *Web Content Mining*: - Web Content Mining [5] is extraction of information from web page contents basically it is the extension of search engines. Along with the text content of pages it also includes multimedia data like images, audio, video, etc. Information Retrieval provides a wide variety of statistical methods for web content mining.

b) *Web Structure Mining* : -As its name suggest that it deals with the structure of web or we can say that web structure mining [6] deals with modelling and discovering the hyperlink structure of web. As we known web pages are connected through hyperlinks and in order to find related information we have to understand this hyperlink structure of web pages.

c) *Web Usage Mining*: - Web Usage Mining [7] deals with mining of web server logs in order to understand the user behaviour with the web to serve users in better manner and also to increase economy benefit. It is also defined as understanding the user interaction behaviour with the websites. Log files contain information like IP address, Web sites, timestamp, and important keywords associated with that particular websites. So analyzing such log files will helps us to find resource requirement of user in near future, help to improve web site structure, also help to improve advertising content.

iv) *Recommendation System:* -Recommendation System is a system which explores the resource utilization of users in different ways and extracts their usage behavior and based on this behavior recommends or predicts the resources which will be required in the future

or simply we can say it is an application of data mining process. It [8] has a very profound impact on today's internet world because information is increasing at a very rapid rate

over the web and finding useful information in effective and timely manner is very challenging issue, so in order to provide relevant information to the users we have to analyze navigational behavior of users.

v) *Periodic Web Content Recommendation:* -Periodic web content recommendation [9] is another step in this direction, this is because an user's web access behavior follows certain pattern i.e. which is repeated daily over certain fixed interval of time and understanding this behavior is important for personalization and to accomplish this task temporal periodic attributes has been considered for a particular user in order to provide better result e.g. every morning if you are searching for devotional music than this model will recommend you the URL's which are related to devotional content in the morning.

1.3 Motivation

Recommendation system [10] is an application of the data mining process and as we know web is hug and it is continuously increasing with rapid rate and lots of work has been done in this direction, so in order to satisfy the increasing demand of users various recommendation system has been developed based on different user requirements and targeted users. An important work has been done in this direction by considering the periodic attributes in recommendation system which helps recommendation system to predict the resources corresponding to a user based on current timestamp. In this recommendation system it considers users access behaviour and recommend the resource URL's which are local best i.e. only URL's which are previously access by that user means if we have better

URL available for a particular resource then it in not recommended to user which is the drawback of the system which we have eliminated by considering certain factors which effectively find resource URL's which are globally best rather locally best.

1.4 Related Work

There is significant amount of work has been done in the direction of recommendation system, L. Zhang, X. Liu, X. Liu [11] describes about personalized instructing recommendation system, which is based on collaborating filtering for recommendation. X. Wang, Y. Ouyang, X. Hu [12] finds frequent access patterns based on the user web access behavior with the revised FP-tree algorithm called FAP-mining. S. Puntheeranurak, H. Tsuji [13] explains framework which build user profile, which has two bifurcations one contain factual information about user and second describe the behaviour of the user. W. Xiao-gang, L. Yue [14] describes algorithm which uses sequential access pattern mining and these patterns are used for recommendation.

But problems with these approaches are that either they have not considered the periodic activity of the user i.e. repeated web access behavior over time or they recommending the local resource URL's i.e. only URL's which are used by the user earlier or both.

In order to consider the repeated web access behavior A.C.M. Fong, B. Zhou, S.C. Hui, G.Y. Hong, T.A. Do [17], propose an algorithm which recommends resources on the basis of user periodic web access behavior by considering periodic attributes but the drawback associated with this algorithm is that it recommends the local best resource URL's not the global best URL's.

1.5 Problem Statement

Recommendation System is one of the important applications of data mining process because predicting resources which are required by the users in advance is very significant task now a day's.

Personalised web content recommendation is an another version of recommendation system which analyze the web access activity of an individual user by using its web server log files and trying to find out future requirement of resources of user.

The problem is that when a user access a particular resource like news during a particular time periods of the day like morning then the system should recommends the resource URL's at that particular period of time.

Periodic web content recommendation [17] is very significant piece of work done in the direction of recommendation of resources based on the particular time period, this is because user's web access behaviour changes with time means a user searches for news in the morning, some educational topic in the afternoon and searching music or movies in the evening. So predicting resources to the user as per its periodic web access activity is very important.

But problem with this approach [17] is that it only predicts resource URL's which are only previously used by that user, if an URL exist which is not used by that user and providing better content then it is not recommended to the user, which is a drawback of this system. This thesis proposes an enhanced version of periodic web content recommendation by considering six parameters which helps us to find that global acceptance of URL's corresponding to a particular resources and then system recommends the URL's which gives best possible resource.

Hence the problem of this thesis is

"**Proposing an enhance version of periodic web content recommendation using web usage logs files by considering global factors related to the URL's of a particular resource**".

1.6 Scope of work

The algorithm proposed by A.C.M. Fong, B. Zhou, S.C. Hui, G.Y. Hong, T.A. Do in [17] has been enhanced by incorporating the set of six parameters. These parameters are evaluated over the resource URL's, which are generated on the basis of the periodic condition over the web usage lattice also called as knowledge base.

Knowledge base generation and extended periodic web personalization are the two main components of this recommender system architecture. Knowledge base generation is responsible for constructing web usage lattice, which is representation of knowledge base. Based on the periodic condition, personalized resource generator will generate resources and then we are finding URL's corresponding to these resources and theses URL's are evaluated based on the six parameters and recommends the URL having highest fitness value.

The scope of work can be summarized as

- Proposing an enhanced version of algorithm that returns URL's by considering the access session of other users for that particular topic.
- Elicitation of set of parameters that play a key role in identification of the global best solution means recommending URL's, which are even not used by that user previously by considering access session of other users.
- Implementing proposed algorithm in C language.

- Validating results of proposed system.

1.7 Thesis Organisation

Further thesis is organised as follow.

Chapter 2: This chapter provides description about the various recommendation system, their architecture and principle.

Chapter 3: This chapter provides the research background of this research work.

Chapter 4: In this chapter we described the proposed approach.

Chapter 5: This chapter deal with implementation part of this research work.

Chapter 6: This chapter contain the results derives in this research work and its comparison with methodology described in chapter 3 which describes research background of this work.

Chapter 7: This chapter deals with conclusion part of this thesis and improvements in future.

Chapter 8: This chapter deals with publication from this research work.

CHAPTER 2: LITERATURE REVIEW

In this section we have performed literature survey over different recommendation algorithms which are proposed by researcher over few years.

2.1 Personalized instructing recommendation system (PIRS).

It is a novel approach [11] for web based learning, which uses collaborating filtering along with Apriori Algorithm for recommendation. This model first understand the user leaning habits, this is done by asking the different users to rate different items and there are some items which are not rated by users so we use collaborating filtering algorithm to predict the rating of the items whose rating is not predicted by the user.

As we know users web access activity is recorded in the server logs in the form of URL. This URL's are pre processed in order to remove noise and convert it into Maximal forward reference path (MFR). After this Apriori algorithm are applied to find frequently occurring sequences. These frequently occurring sequences along with learning habits build user model data, which helps to find patterns recommendation for a particular user.

The architecture of this model is shown in Figure. 1 which have composed of some important components they are like Testing learning style, Mining Frequent sequences, user data model, Pattern database, personalized recommendation system.

i) *Testing Learning Style*: - This component is responsible for finding the learning style of different users because every user may have different learning style and making our recommendation system adaptable to these learning gives better results. This model used Kolb's style and Solomon's style of learning. But as we known usually users not

rate all the factors of this learning style so in order to enhance the results and proper understanding of learning style collaborating filtering is used to rate all un-rated factors.

ii) *Mining Frequent Sequence:* - In this component first of all web access sequences are pre processed to maximum forward reference (MFR) path and then Apriori algorithm is used to mine frequents patterns.

iii) *User Model Data:* - After finding the learning style of user and frequent sequences, these two are combined to form user model data.

iv) *Pattern Database:* - It is database which store the frequent patterns obtain from the above components.

v) *Personalized Recommendation System:* - This component is responsible for extracting the pattern from database based on the user who is making request resource and recommend it to user so that the user will get information more accurately and in timely manner.

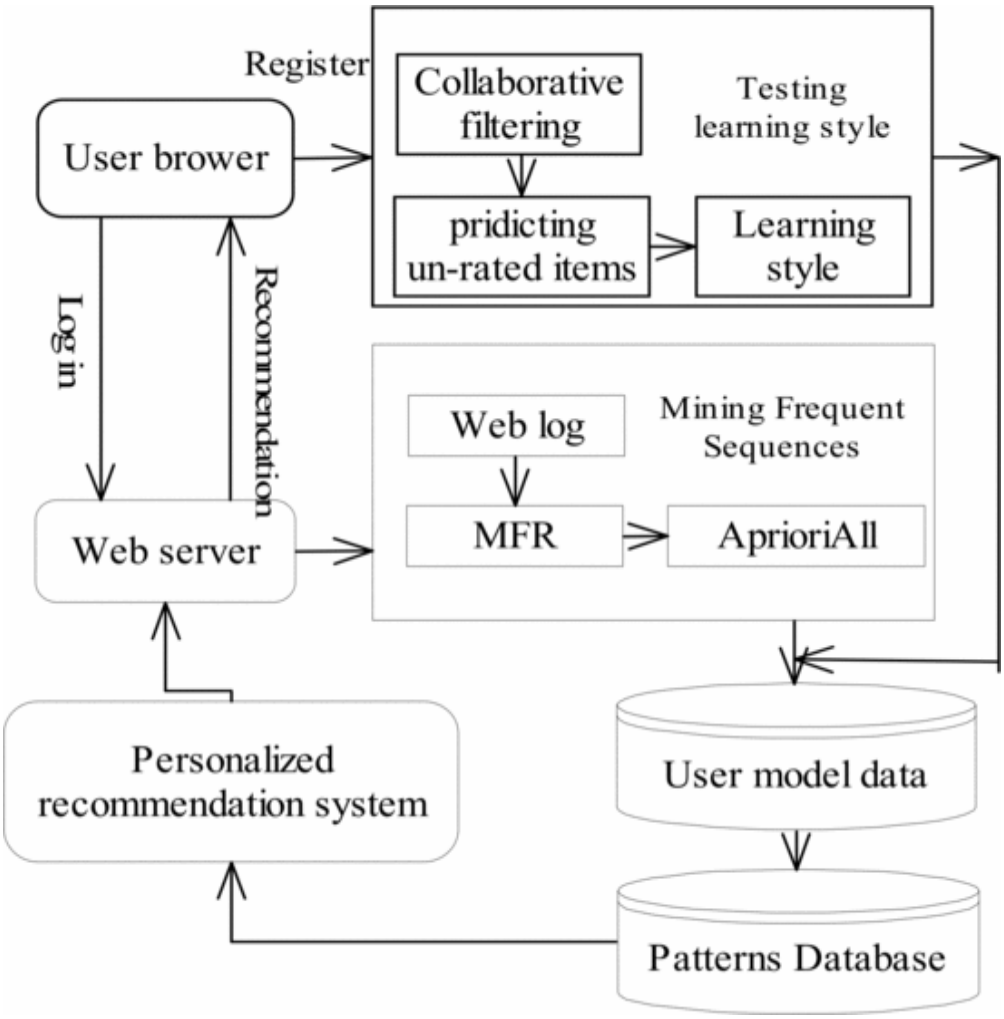


Figure 1 Architecture of PIRS [11]

Analysis of this model

i) This model has not considered the periodic attributes corresponding to the user due to this it is not able recommend the resources in timely manner.

ii) This model is also recommending resources from its previous used history means if a better URL is available corresponding to the resources then it is not recommended by the system.

\

2.2 Discovery of user frequent access patterns on web usage mining.

This algorithm [12] is used to find frequent access patterns based on the user web access behavior with the revised FP-tree algorithm called FAP-mining.

Here access pattern is based on the user access path. As we known web log files contain web access behaviour of users corresponding to its session because a user can use the system in different sessions in terms of sequences of websites.

As shown in Figure 2 access path of a user is A-B-C-D-B-G-E-H-G-C-A-I-K-I-D, here nodes represent a particular website and link represent hyperlink through which users can traverse from one page to other. There is difference between user access path and user access pattern, user access pattern is denoted as forward reference of user access path.

In Figure 2 user access path for H is A-B-C-D-B-G-E-H whereas user access pattern is A-B-G-E-H, so it can be concluded that user access pattern always represent the web access behaviour of the user in simpler way.

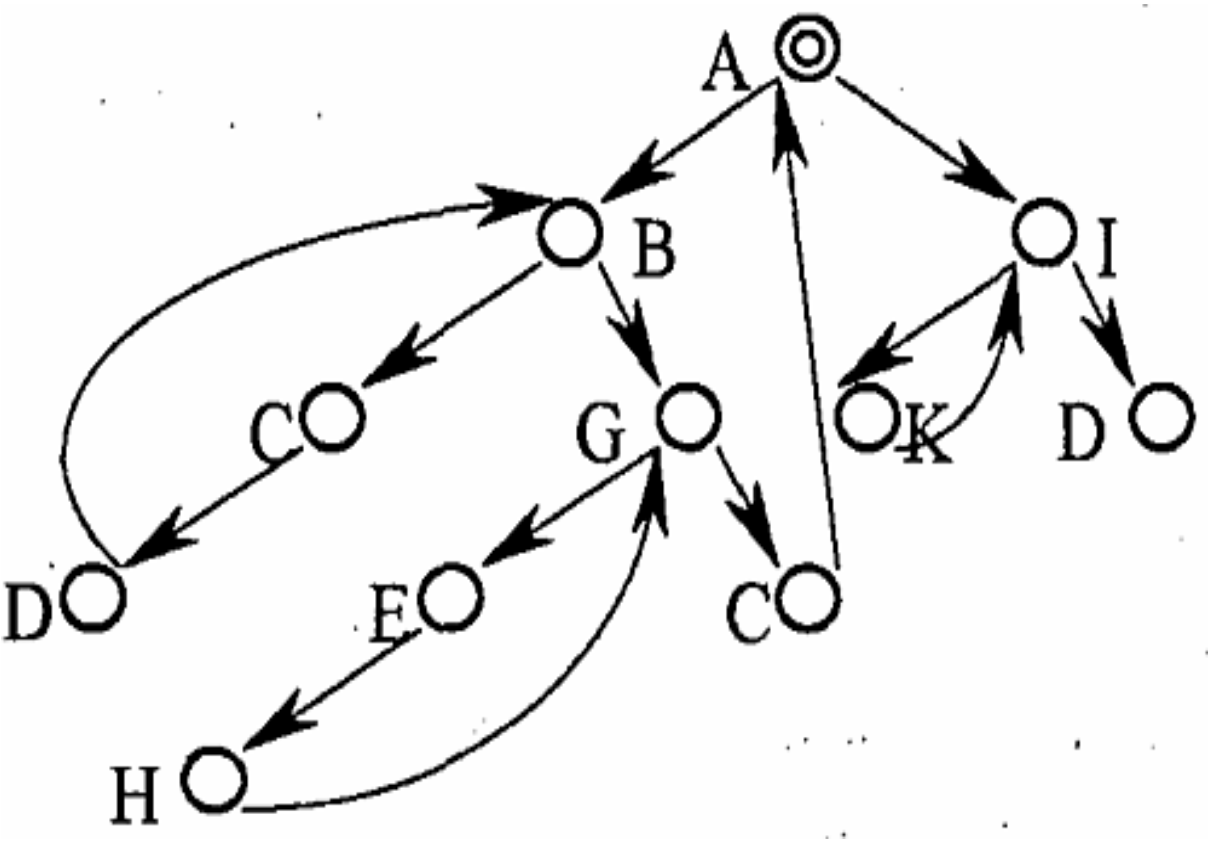


Figure 2 Access Path or Browsing path of user [12]

In association rule mining there is not any order in between elements of item whereas access pattern mining having certain order.

FAP-Mining is divided into two parts

1. In first part we build Frequent Access Pattern Tree based on the access path obtain from web usage logs and also maintain counts for access of each page.
2. In second part FAP tree is used to mine short and long access pattern.

2.3 Mining web logs for a personalised recommender system

This is a framework [13] based on web usage mining for personalized recommendation system. In this framework we build user profile, which has two bifurcations one contain factual information about user and second describe the behaviour of the user.

Traditional way of finding information is searching over search engines by user queries, search engines gives results on the basis of query. In this case it is very tedious for the user to find relevant information because sometime information is not given by search engines properly because of the improper query structure, some websites are paid to search engines so they appear before irrespective of that whether it content more relevant or not.

A Personalized recommendation system framework is shown in Figure 3, which composed of different components like Monitoring agent, Classification agent, Learning agent, and Recommendation agent.

The personalize recommendation system is categories into two processing unit i.e. offline processing and online processing.

i) *Monitoring Agent*: - It is a part of online processing unit. Registration and web log filtering are two main task of this agent. Its collects information about user through registration mechanism and this information is used to create user profile. This agent monitors the behaviour of user by looking at the web log files, which record web access activity of users.

ii) *Classification Agent*: -Basically this agent is responsible to classifying the different webpage based on their content and also certain important keywords are identified based on the document keyword selection technique, which gives importance of keywords in a

particular document. This agent is also responsible for applying rule discovery methods to user's web log files by using association rule mining. This agent gives certain rules with pre-condition and post-condition. User profile information and rules obtain together called consumer's behaviour.

iii) *Recommendation Agent*: - This agent based on two approaches i.e. collaborating filtering algorithm and content based filtering. It uses the similarity between different user's profiles. First it uses content based filtering for finding the recommended pages and then user provide rating to these pages and then collaborating filtering along with clustering is used to find more web pages for recommendation.

iv) Learning Agent: - This module is responsible for understanding the user's changing interest. Users provide their feedback whether it is positive or negative it helps learning agent to change user's Meta profile according to which recommendation changes.

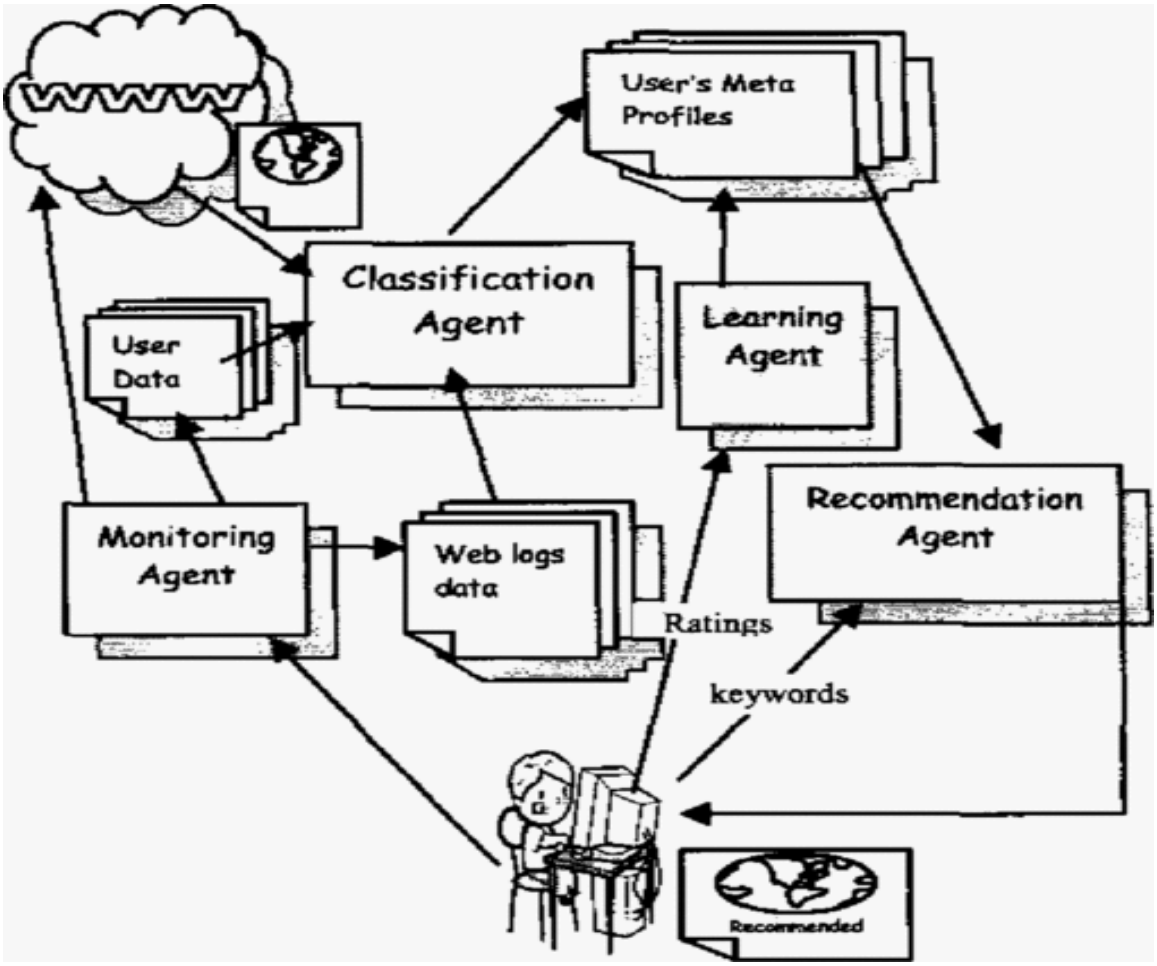


Figure 3: Framework of Recommender System [13]

2.4 Web mining based on user access patterns for web personalization.

Most of the recommendation algorithm uses clustering and association rule mining whereas this model [14] uses sequential access pattern mining and then patterns are stores in tree like structure called pattern tree and this generated tree is used for recommendations.

As we see system architecture is represented in Figure 4.

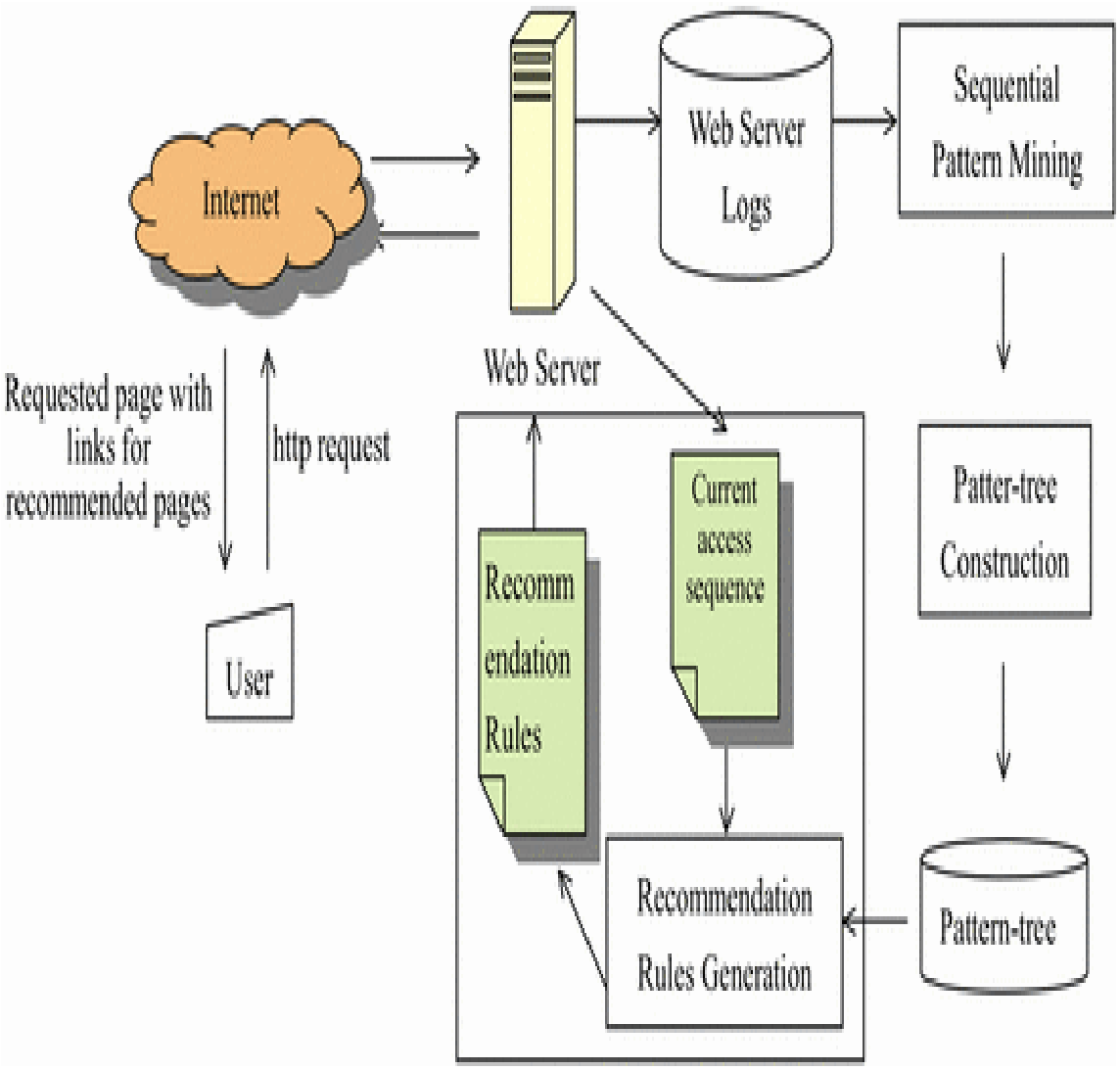


Figure 4: Architecture of Recommendation System [14]

i) Sequential Pattern Mining: - As we known during a particular session user perform certain browsing activity, which is define as sequence of web pages. Sequential pattern mining is applied over web server logs in order to obtain sequential pattern which satisfy certain minimum support threshold value.

ii) Pattern tree construction: - Basically this model is used to store sequential web access pattern obtained in previous step and then used to generate recommendation rules by considering current access pattern

iii) Recommendation rule generation: - This component is responsible for finding the recommended access path of the user which is matched more closely with any of sequence access path in the sequence tree.

CHAPTER 3: RESEARCH BACKGROUND

This chapter basically provides background for the research work presented in this thesis.

All models which we have seen in previous chapters do recommend the resources to the user as per web access activity with some variations but none of them considered the time attributes i.e. recommending the resources as per timestamp at which user is using the system.

This model considers the web access activity along with time attributes in order to recommend the resources in efficient and timely manner.

In actual practice a user's web access behaviour varies with time that means a user searches for devotional music or news in the morning while movie or some rock music in the evening so web access behaviour changes with person to person and it also changes from time to time for a particular user. One more important point about this approach is that it not depends upon the current web access activity because there may be chances that the current web access activity is just an intermediate step in order to reach to final web page.

This is a novel approach which based on two technique i.e. fuzzy theory [15] and formal concept analysis FCA [16].

Figure 5 shows architecture of the model which is proposed in the paper [17]. It has mainly two components "user beahvior model construction" and "periodic web personalization".

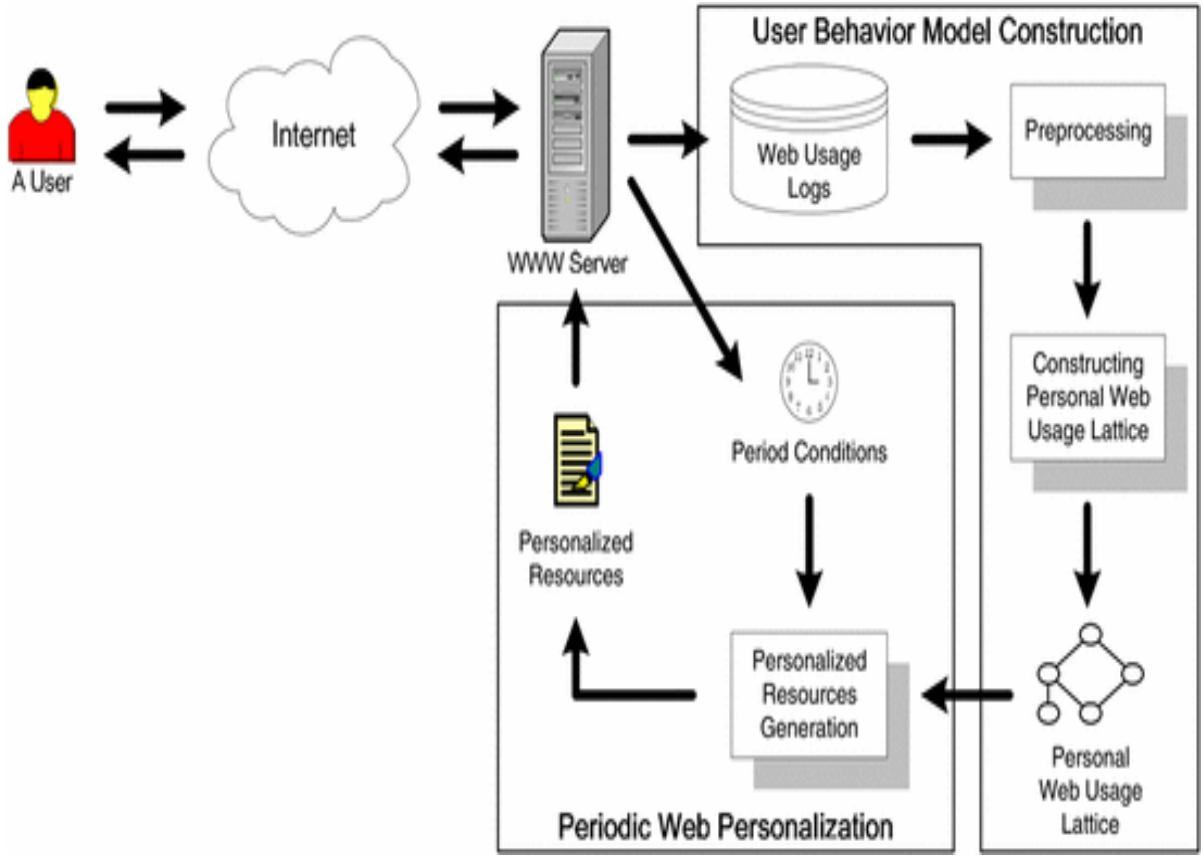


Figure 5: Architecture of Recommendation System [17]

In this approach user behaviour model is constructed from usage log files.

3.1User Behavior Model Construction

There are number of steps involved in this component like pre processing, constructing web usage lattice. Basically web usage logs contain information about usage behaviour of the user in terms of URL's access by them. In this model it is also assumed that each of the URL is annotated by website administrator with some semantic information like topic associated with the URL's or category of the URL's like music, news, movies etc.

Table 1 Web usage logs

User Id	Date	Timestamp	URL	Topic

1	21/8/2014	08:20:01	U1	R1,R2,R3
1	21/8/2014	08:22:05	U2	R1,R2,R4
1	21/8/2014	08:27:30	U4	R1,R4,R5
2	21/8/2014	09:01:30	U3	R1,R2,R6
2	21/8/2014	09:02:40	U7	R1,R3
2	21/8/2014	09:03:50	U8	R1,R4
1	21/8/2014	21:30:00	U5	R7,R8
1	21/8/2014	21:32:10	U6	R7
1	22/8/2014	08:15:10	U1	R1,R2,R3
1	22/8/2014	08:17:10	U2	R1,R2,R4
2	22/8/2014	21:35:05	U9	R8,R9
2	22/8/2014	21:40:05	U10	R8

Table 1 represents an example of web usage logs which is enriched semantically.

Here we are interested in periodic access pattern i.e. user is interested in particular resource in some specific time. So in order to understand periodic behaviour author represents periodic attributes (P_i) which corresponds to eight real life temporal concepts they are late night (LN), early morning (EM), morning (M) ,noon (N), early afternoon (EA), late afternoon (LA), evening (E), night (N) respectively.

In table 1 column with heading 'Topic' represent resources (M_r) in which user is interested during web access activity i.e. R1, R2....R9.

i) *Pre-processing*: - In pre processing basically data is processed before its use and it involves steps likes removing redundant data i.e. data cleaning, identifying different user's corresponding their user id, identification of session corresponding to users.

(1) *Data Cleaning*: - Data cleaning is a process of removing unnecessary, incomplete, duplicate and redundant entries in order to make our data more clean and workable.

(2) *User identification*: -web server logs contain web access behavior of numbers of users so in order to differentiate between users we have an attribute in web server logs named userid, which is used to identify different users. This is required because in order to provide web personalization we have to understand the web access behavior of users individually.

(3) *Session identification*: - In this paper author divide the web access activity of users in different sessions, which helps to categories these sessions in different temporal concepts, here session is defined on the basis of the web access request timestamp if the timestamp varies within predefined threshold value then that particular web access activity is included in a particular session.

A session is defined as $S = \{(u_1, t_1), (u_2, t_2), \ldots., (u_n, t_n)\}$ where u_i is an URL which is requested at timestamp t_i.

Along with session we also keep record of duration of access of a particular URL i.e. $d_i = (t_{i+1} - t_i)$ and duration for last url is calculated by taking average of all other durations i.e.

$$d_n = {(d_1 + d_2 + \ldots\ldots d_{n-1})}/{n-1} = {t_n - t_1}/{n-1}$$

Start and end time of session is defined as t_1 and t_n+d_n respectively.

ii) *Knowledge Base Generation*: - In knowledge base generation basically we identified user's web access activities and on the basis of these activities we construct web usage lattice. But in real life different people have different rendition corresponding to requested resources in access sessions. So in order to handle this wispiness in both temporal attributes and resource attributes we use fuzzy theory [15] technique and integrated with formal concept analysis [16]. Knowledge base generation process is further comprises of following steps.

(1) *Web Usage Context*: - Context is defined as K= (G, M, I) where G is a set of objects and M is a set of attributes and I is called incidence which is defined as binary relation over G and M.

In case of web access activity we define fuzzy periodic web usage context where G is a set of access sessions for a user and M is defined as union of M_p and M_r ($M = M_p \cup M_r$). And binary relation over G and $M_p \cup M_r$ is defined as

$$I = R\big(G \times (M_p \cup M_r)\big)$$

And I is a fuzzy set defined on the domain of G X (M_p U M_r) and fuzzy relation over particular session i.e. $g \epsilon G$ and attributes $m \epsilon (M_p \cup M_r)$ is defined as R(g,m) belongs to I and it is represented by membership function $\mu(g, m) \epsilon [0,1]$

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

Membership value for periodic attributes $\mu_p(g,m)$ is defined as

$$\mu_p(g,m_p) = max_{t \in p(g)}\{\mu_p(t,m_p)\}$$

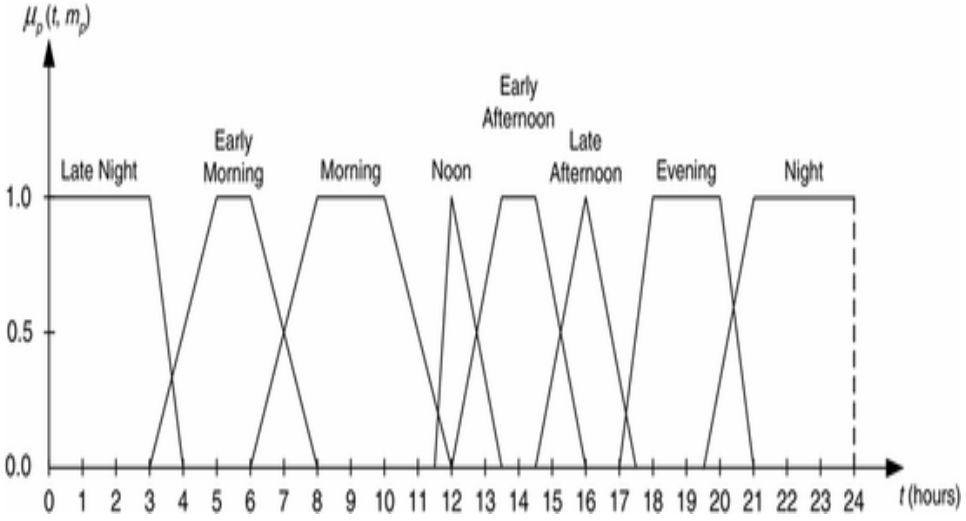


Figure 6: Membership Function of Periodic attributes

Where $\mu_p(t,m_p)$ is defined in figure 6 and $p(g)$ is period of g, which is defined as difference between start and end time of a session.

$$p(g) = \begin{cases} [t_s(g), t_e(g)] & if\ t_s \leq t_e \\ [0, t_e(g)] \cup [t_e(g), 24] & otherwise \end{cases}$$

Where $t_s(g)$ is start time of a session and $t_e(g)$ is end time of a session.

Membership value for resource attributes $\mu_r(t,m_r)$ depend upon the time interval for which that resources are used i.e. $d(g,m_r)$ which is defined as follow

$$d(g,m_k) = \sum_{i=1}^{n} \alpha_{ki} d_{i,} \quad where\ \alpha_{ki} = \begin{cases} 1, if\ m_k \epsilon M_{ri} \\ 0, otherwise \end{cases}$$

And membership value is defined as

$$\mu_r(g, m_r) = \begin{cases} 0, & if\ z(g, m_r) < \frac{1}{2} Z(m_r) \\ \frac{2\, z(g, m_r)}{Z(m_r)} - 1, & if\ \frac{1}{2} Z(m_r) \leq z(g, m_r) \leq Z(m_r) \\ 1, & if\ z(g, m_r) > Z(m_r) \end{cases}$$

Where

$$z(g, m_r) = \frac{d(g, m_k)}{t_e\,(g) - t_s(g)} \qquad \text{And}$$

$$Z(m_r) = \frac{\sum_{g_k \in G} d(g_k, m_k)}{\sum_{g_k \in G} (t_e(g_k) - t_s(g_k))}$$

$Z(m_r)$ is defined as ratio of the sum of total time duration of the resources used in different sessions to the sum of periods of different sessions. It helps to draw the global interest level of a user corresponding to the resource.

$z(g, m_r)$ is defined as ratio of the total time duration of the resource used in a particular session to the period of that session. It helps to draw the local interest level of a user corresponding to the user.

(2) Web Usage Lattice: - Lattice is a graphical way of representing binary relation. As we define web usage context that is binary relation on set of access session and attributes. So the lattice defined over this web usage context is called web usage lattice.

Now web usage lattice can be derived from web usage context by using formal concept analysis [ ].

As we know web usage context i.e. $K = (G, M_p, M_r, I)$ and if we have a pair (P, Q) such that $P \subseteq G$, $Q \subseteq (M_p \cup M_r)$ and $P^* = Q \; and \; Q^* = P$ then fuzzy set on Q is called web access activity which defined as

$$v(Q) = \{m, \mu(Q, m) | m \epsilon Q\}$$

Where

$$\mu(Q, m) = \; max_{g \epsilon Q^*}\{\mu(g, m)\}$$

Now we define support and confidence of the web access activity.

$$\sup\big(v(Q)\big) = \sup(Q)$$

$$Conf\big(v(Q)\big) = prob(\big(Q \cap M_r \big| Q \cap M_p\big))$$

Support defined corresponding to a web access activity is equal to the support value defined over the attributes.

$$\sup(Q) = \frac{\sum_{g \in Q^*}\big(\mu_p(g) \times \mu_r(g)\big)}{|G|}$$

In order to generate web usage lattice $\big(L_p\big)$ there are some FCA methods, one of them is TITANIC algorithm [18]. In this paper author used TITANIC algorithm to generate web usage lattice from web usage context.

3.2Periodic web personalization

This component is responsible for generating the personalized resources corresponding to a particular user based on the web usage lattice $\big(L_p\big)$ for a periodic condition (p_c). Now based

on this periodic condition we search for the web access activities in the web usage lattice (L_p) and find the resources which belongs to the periodic condition.

Now there are number of web access activities in the web usage lattice, which get matched with the periodic condition, so in order to find the best web access activity which matched with periodic condition we define similarity index in between them, maximum the similarity index more accurate will be the personalized resources corresponding to a user.

$$Sim_p(v(Q), p_c) = \frac{|v_p(Q) \cap P_f(p_c)|}{|v_p(Q) \cup P_f(p_c)|}$$

Where $P_f(p_c)$ is fuzzy set over fuzzy periodic condition.

$$P_f(p_c) = \left\{ m_p, \mu_p(p_c, m_p) \middle| m_p \epsilon M_p \text{ and } \mu_p(p_c, m_p) = \max_{t \in p_c} \{\mu_p(t, m_p)\} \right\}$$

Where $\mu_p(t, m_p)$ is defined in fig. fuzzy Figure 6.

Similarity index calculated is used to find the priorities of the different web usage activities in the web usage lattice, more similarity index means more priority of the web access activity. So the resource attributes corresponding to more prior web access activity are needs to be recommended first for better results and customer satisfaction.

If we have two web access activities corresponding to given periodic condition i.e. $v(Q_i)$ *and* $v(Q_j)$ then $v(Q_i)$ is said to be of higher priority than $v(Q_j)$, if

- $Sim_p(v(Q_i), p_c) > Sim_p(v(Q_j), p_c)$ *or*

- $Sim_p(v(Q_i), p_c) > Sim_p(v(Q_j), p_c)$, $but\ Conf(v(Q_i)) > Conf\left(v(Q_j)\right)$ or
- $Conf(v(Q_i)) = Conf\left(v(Q_j)\right), but\ Sup(v(Q_i)) > Sup\left(v(Q_j)\right)$.

Personalized resources which are identified from web usage lattice based on the periodic condition are defined as

$$PR(p_c, L_p) = \{m_r | \exists v(Q_i) \in SA_p(p_c, L_p), such\ that\ m_r \in Q_i \cap M_r\}$$

Where $SA_p(p_c, L_p)$ is defined as web access activity corresponding to time period p_c.

3.3 Personalized Resource Generator Algorithm

The algorithm for generation of the personalized resources is as follow

Input:

- A periodic condition (p_c)
- A personal web usage lattice (L_p)

Output:

- Personalized resources ($PR_o(p_c, L_p)$)

Process:

1. Initialization of $PR_o(p_c, L_p)$ to $\{\emptyset\}$
2. For all web access activities nodes in web usage lattice mark as unvisited.
3. Now we find all web access activity corresponding to the periodic condition such activities are called supported activities

$$SA_p(p_c, L_p) \leftarrow PSA\_Search(p_c, L_p, v(\emptyset))$$

4. Repeat steps 5 to 9 for all web access activity belongs to the supported activities in decreasing order of priority.
5. Repeat steps 6 to 8 for all the resource attributes belonging to the supported activities ($m_r \in B \cap M_r$) in decreasing order of priority.
6. if $m_r \notin PR_o(p_c, L_p)$ then

   Append m_r into $PR_o(p_c, L_p)$
7. end if
8. end of step 5.
9. End of step 4.
10. return $PR_o(p_c, L_p)$

Algorithm PSA_Search ($p_c, L_p, \upsilon(Q)$)

Input:

- A periodic condition (p_c).
- A web usage lattice (L_p).
- Current web access activity (υ(Q)).

Output:

- We obtain a set which contain all web access activity corresponding to periodic condition ($SA_p(p_c, L_p)$).

Process

1. Initialization of $SA_p(p_c, L_p) \leftarrow \{\emptyset\}$.
2. Repeat step 3 to 11 for all the sub activities of υ(Q).
3. ***if*** $\upsilon(Q_i).mark = unvisited$ ***and*** $\upsilon(Q_i) \cap P_f(p_c) \neq \emptyset$ ***then***
4. **if** any of the periodic attribute of $v_p(Q_i)$ matches with periodic condition $P_f(p_c)$ then

   $SA_p(p_c, L_p) \leftarrow \{\upsilon(Q_i)\} \cup$ PSA_Search ($p_c, L_p, \upsilon(Q)$)

5. **else** $SA_p(p_c, L_p) \leftarrow$ PSA_Search ($p_c, L_p, \upsilon(Q)$)
6. **end if**.
7. $\upsilon(Q_i).mark \leftarrow visited$
8. **end if**.
9. **End of step 2**.
10. **return** $SA_p(p_c, L_p)$.

CHAPTER 4: PROPOSED APPROACH

This chapter is based on the proposed approach for periodic web personalization. This approach is extended version of the approach described in chapter 3. As discussed in [17], authors describe an approach for periodic web personalization in which they consider temporal attributes in order to recommend the resources efficiently and effectively in timely manner. But problem with this approach is that it only recommends the local resources, in order to provide global best recommendations here we are introducing certain parameters.

4.1 Parameter Associated with URL's

There are many parameters which are associated with a particular website and these parameters have different degree of importance for different resource providers. Here we have discussed six parameters, which have impact over web recommendation.

a) *Total Numbers of User's (P1):* - One of the important parameter is number of user preferred to use that particular website, more number of user means websites is more popular as compare to other websites having less number of user. So this parameter helps us to find website which provide better content as compare to others. Here we use human access behavior because a user always preferred websites which provide better content. We can also associates a coefficient C1 with this parameter, whose value lies in between 0 and 1 which decides the importance of parameter i.e. how much this parameter is effective for that particular website.

b) *Number of Unique Visitors (P2):* - Difference between total number of user and number of unique visitor is that number of unique visitor is always less than or equal to

total number of user for a particular website. This parameter helps us to not trap in local maxima or local minima i.e. suppose a particular website is used by certain number of

users repeatedly while an another website which provide same resource having more number of unique visitors or distinct users than later one is more preferable as compare to prior one. C2 is a coefficient which is associated with this parameter. It is better to give more preference to P2 as compare to the P1.

c) *Number of User's download Data:* - This parameter is useful in case when user is looking for the resources which are required to download from respective websites. Suppose user wants to download the resource it can be anything like video, audio, image, text document etc. Now if number of user download data from a particular website is higher than either that website provide resource which are more relevant to the user demand or may be it provide data freely or any other reason due to which more users prefer that website.

d) *Amount of Data Downloaded:* - Sometimes amount of data downloaded is more important than the number of user downloads the data. Sometimes there is less known websites which having less numbers of user but these users uses this websites very frequently because of availability of quality data which results in increase of amount of data downloaded that signify the importance of resource provider.

e) *Amount of Data uploaded:* - Importance of this parameter is similar to that of explained in previous paragraph but only difference is that in case of amount of data downloaded importance is from user perceptive whereas in case of amount of data

uploaded the importance is from resource provider perceptive i.e. more data uploaded means more availability of data.

f) *Number of Advertisement:* - Popularity of any websites can also be thought as more number of advertisements indicates more popular is the websites, this is because generally advertisement are shown on the websites where users are more likely to visit or vice versa.

Different parameters have different importance to the users and resource providers. This importance is represented by coefficient values i.e. C1, C2, C3, C4, C5, C6, these coefficients associated with different parameters P1, P2, P3, P4, P5, P6 respectively. Larger value of coefficient means more importance is given to it. E.g. if a user searches for educational content then parameter P2 gets more importance by assigning higher value to C2 and if a user searches for movie than giving importance to parameter P1 or P2 is of no value because we have to recommend that website which having higher number of users, who have downloaded content from it, so in this case parameter P3 gets higher value. But here we assigning certain fixed value to these parameters to avoid complexity but it still provide improved results.

4.2 System Overview

System overview is represented in Figure 7, where knowledge base generation and extended periodic web personalization are two main components of the system.

Knowledge base generation is responsible for generating knowledge base in the form of web usage lattice, which is described in detail in chapter 3.

The extended periodic web personalization component uses knowledge base generated by the knowledge base generation component and it is responsible for recommending URL's by considering the web access session of other users in order to recommend URL, which provide more relevant content i.e. best solution by considering global domain of URL's.

The extended periodic web personalization component will first generate resource v/s URL matrix and based on this matrix, this component will find URL's corresponding to the required resources and these URL's are evaluated based on the parameters explained in section 4.1 of this chapter. In evaluation part we are finding the fitness value of URL's based on the parameters and system will recommends the URL having highest fitness value to avail maximum relevant content to the user. For example if a user is searching for news in every morning then this proposed system not only recommends URL's which are previously accessed by that user but first recommend URL's which provide more relevant content to the user.

Web usage logs stores user's web access records, which gives idea about the user's browsing pattern, in order to generate knowledge base, steps are shown in Figure 7 like pre-processing, web usage context, web access pattern and finally construction of knowledge base, these steps are explained in detail in chapter 3 Research Background.

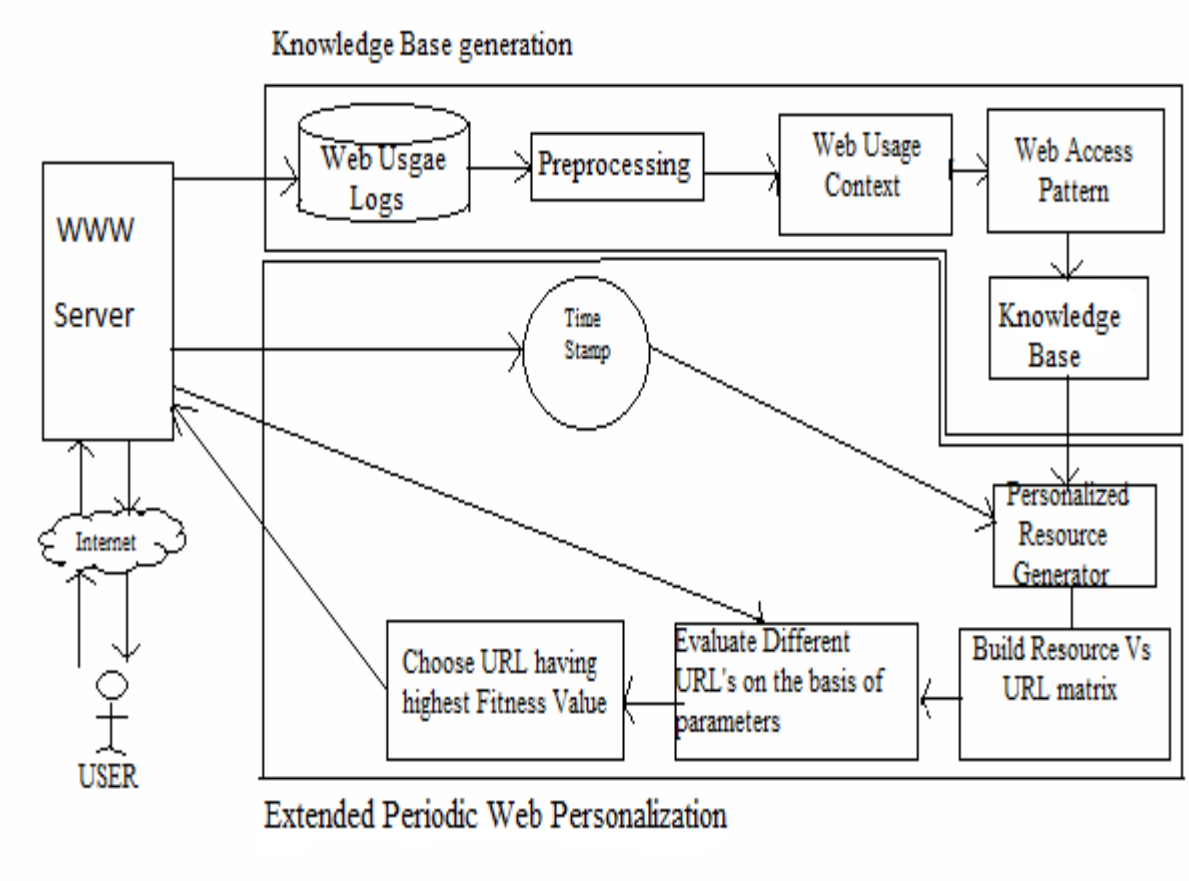


Figure 7: System Overview of Personalized Recommendation System

4.2.1 Knowledge Base Generation

Knowledge Base Generation component is primary component of the system and it is similar to explain in [17], which initially deals with web server. There are sub components of the knowledge base generation they are web usage logs repository, Pre-processing, web usage context, and construct knowledge base. Overview of Knowledge base generation is shown in Figure 8.

i)Web Usage Logs: - Web Usage Logs collect the user's access behavior in terms of its userid, date, time stamp, URL, Topics, Table 2 represents Web usage logs. Shown in Table 2.

UserID: - It is a unique identifier corresponding to a user which is used to identify a user uniquely, every user has its own userid which is recorded in web server at the time when user is making request to the server for example IP address is an example of userid.

Date: - As its name suggest that its keep record of the date of the web access activity.

Timestamp: - Basically it records the time of the web access activity of the users.

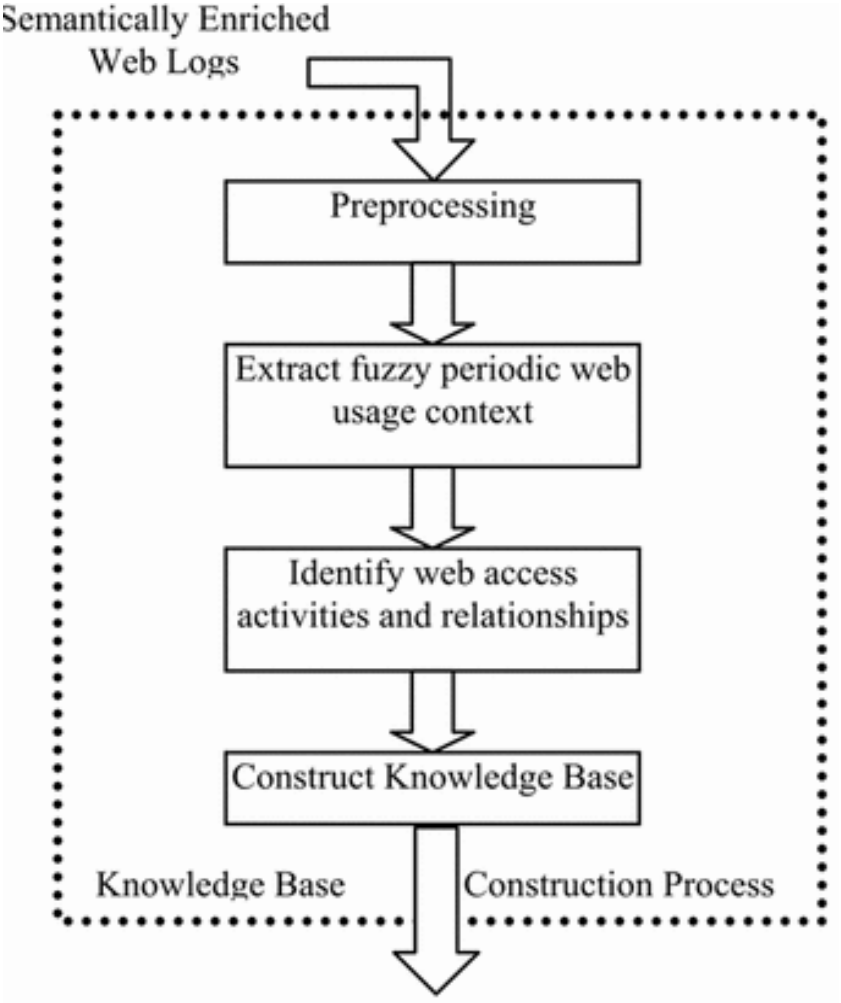
Uniform Resource Locator (URL): - It represents the websites which is accessed by the user in order to use the resources.

Topics/Resources: - Topics are the contents or the resources which is available at a particular website.

ii) *Pre-processing:* - As its name suggest that processing of data before its use and it is applied in order to remove redundant and incomplete data from the web logs, so that there is not any discrepancy in the data. There are further sub activities involve in pre-processing they are Data cleaning, User Identification, Session Identification.

iii) *Extract Fuzzy Periodic Web Usage context:* - Web Usage Context is defined as binary relation over set of access session and attributes. Here we use term "Fuzzy Periodic" because value of web usage context is defined by using fuzzification this is because we divide 24 hours of a day in 8 temporal concepts and value of periodic attributes is determined by the fuzzy function which is described in chapter in 3.

Similarly value of resource attributes corresponding to session is based on the membership function, which is also defined in chapter 3.



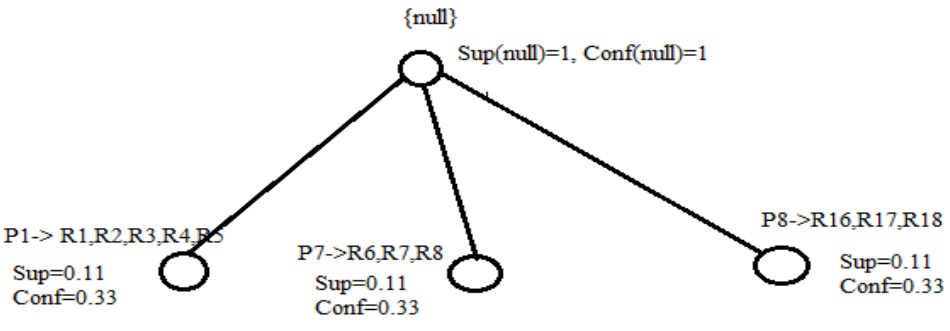


Figure 8.Overview of Knowledge Base Generation

Table 2 Web Usage logs

User Id	Date	Timestamp	URL	Topic
1	21/8/2014	08:20:01	U1	R1,R2,R3
1	21/8/2014	08:22:05	U2	R1,R2,R4
1	21/8/2014	08:27:30	U4	R1,R4,R5
2	21/8/2014	09:01:30	U3	R1,R2,R6
2	21/8/2014	09:02:40	U7	R1,R3
2	21/8/2014	09:03:50	U8	R1,R4
1	21/8/2014	21:30:00	U5	R7,R8
1	21/8/2014	21:32:10	U6	R7
1	22/8/2014	08:15:10	U1	R1,R2,R3
1	22/8/2014	08:17:10	U2	R1,R2,R4
2	22/8/2014	21:35:05	U9	R8,R9
2	22/8/2014	21:40:05	U10	R8

iv) Identify Web Access Activities: - Web Access Activity is also called web access pattern, these are the activities which are performed by the users in order to access resources from World Wide Web. By using Formal Concept Analysis [16] we identify web access activities based on web usage context.

v) Construct Knowledge Base: -Knowledge Base construction means representing identified web access activities in such a way that it provides a way to identify the web access activities which resembles to the current periodic condition in

order to recommend the resources this representation of web access activities is called web usage lattice. Once we identified the web access activity corresponding to a periodic condition then we find out the resources which are accessed in that activity and

then these are resources which are most likely to be used in the current periodic condition.

vi) Personalized Resource Generation: -Personalized resource generator is the component which is responsible for predicting the resources which are required by the users at particular periodic condition. This is done by searching for resources over the web usage lattice.

4.2.2 Extended Periodic Web Personalization

Extended Periodic Web Personalization is the proposed part of the algorithm in which we are finding the URL's, which are providing the demanded resources in more efficiently and

effectively by incorporating the parameters, which are described in Section 4.1 of this chapter. Proposed algorithm is shown in Figure 9 in the form of flow chart.

i)Build Resource Vs URL Matrix: - In this section we build a matrix, which represents resources and their corresponding URL's from where they are accessible. This matrix helps us to identify all the URL's to a particular resource from where it is accessible. This matrix helps us to find global best recommendation because here we consider all those URL's which provide access to the desired resource. In previous approach it either recommends resources or URL's from local usage logs i.e. it only considers URL's,

which are previously accessed by the user not considers other URL's, which provides more promising content or resource to the user.

Table 3 Resource v/s URL matrix

	U1	U2	U3	U4	U5	U6	U7
R1	1	1	0	1	0	1	0

R2	1	0	0	0	0	1	1
R3	0	1	1	0	1	0	0
R4	0	1	0	1	1	0	1
R5	1	0	1	1	0	0	0

Table 3 is an example of resource vs URL matrix, where rows represents resources and columns represent URL's. An entry corresponds to i^{th} row and j^{th} column is either 1 or 0. If entry is 1 means i^{th} resource is accessible from jth URL and if entry is 0 means i^{th} resource is not accessible from j^{th} URL.

ii) Evaluate URL's on the Basis of Parameters: - Now after step I, we get all those URL's which provide us the required resource. After this we evaluate all URL's on the basis of parameters defined in section III (A). For this evaluation we define fitness function $FF(U_{ij})$.

$$FF(U_{ij}) = \sum_{k=1}^{6} C_k * P_{kj} * f(R_{ij})$$

$FF(U_{ij})$ represents fitness value of resource R_i corresponding to j^{th} URL (U_j). As we known personalize resource generator will generates resources corresponding to a periodic condition of a user, so the URL's belong to these resources are evaluated based on this fitness function formula.

$$f(R_{ij}) = 1 \; if \; R_i \; resource \; is \; accessible \; from \; URL \; U_j$$

$$0 \quad otherwise$$

k^{th} parameter value corresponding to j^{th} URL is represented by P_{kj} and coefficient of these parameters are represented by C_k.

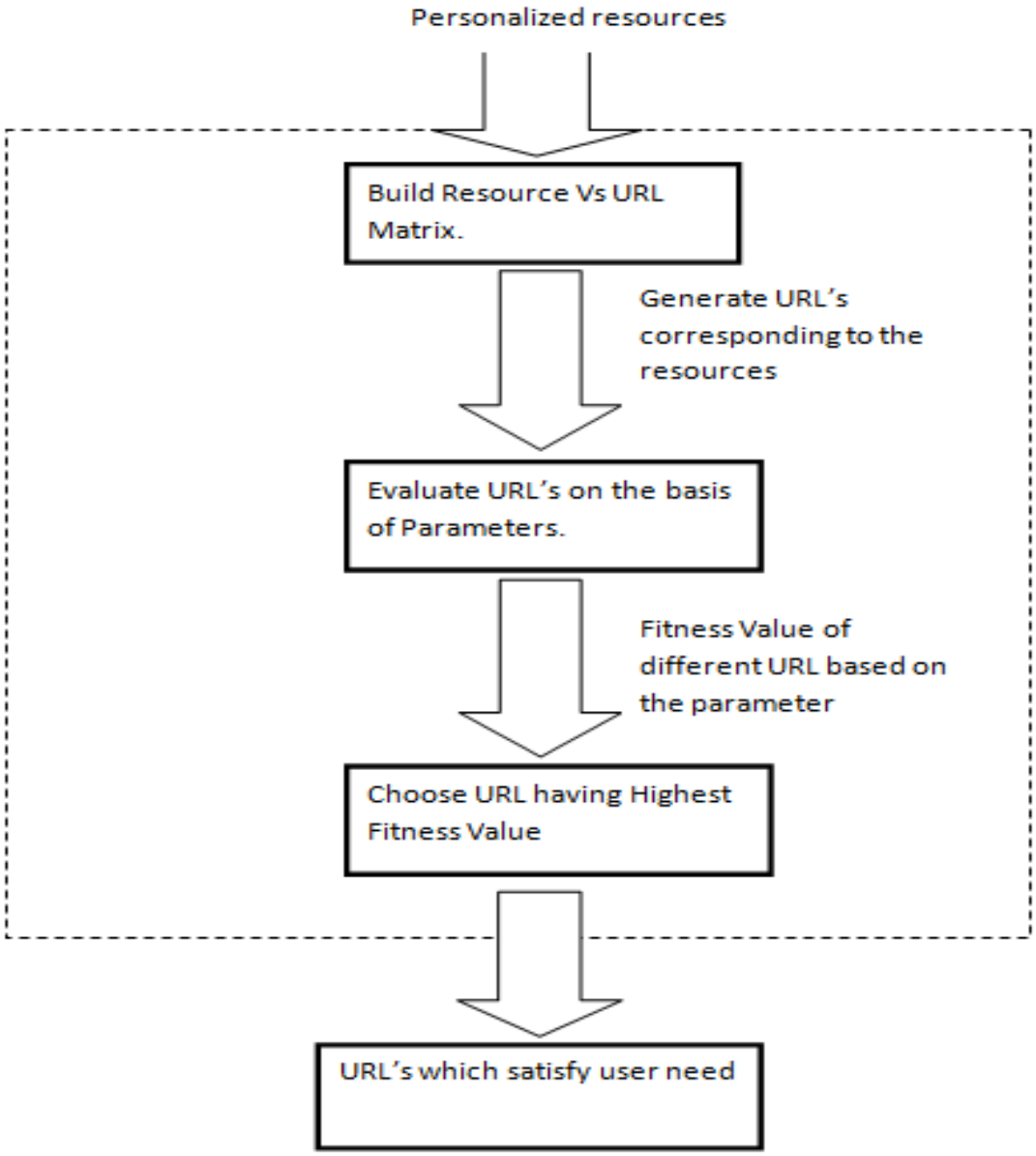


Figure 9 Extended Periodic web Personalization Architecture

iii) Make Selection for URL: - After following previous two steps we got the resource URL's along with their fitness value, now we choose the URL having highest fitness value in order to provide user information in efficient and effective manner. So

after following all these steps we recommends the URL's which provides resources more relevant to the user.

CHAPTER 5: IMPLEMENTATION

This chapter deals with implementation details of the algorithm along with intermediate data and table which are generated and finally identify the URL's which are most appropriate in order to access resources efficiently. Following are the steps involved in generation of knowledge base along with the evaluation of URL based on six parameters.

5.1 Steps in Algorithm

a) Identify user based on their user id.

b) Identify session and periods corresponding to different users.

c) Build web usage context.

d) Identify web access activities.

e) Build web usage lattice based on format concept analysis.

f) Based on the periodic condition identify resources which are required by the user at particular time instant.

g) Build resources v/s URL matrix from using web usage logs.

h) Now consider each resource identified in step (f) and find corresponding URL's by using matrix drawn in previous step.

i) Evaluate each URL identified in previous step corresponding to six factors and sort them in decreasing order of their fitness value.

j) Recommends URL's with high fitness value first.

5.2 Implementation details along with intermediate data.

Initially we have web logs files which contain the records of user web access behavior. This log file is in text format, where each line represents web access record of a user and each record is in the form of **"userid-date-month-year-hour-minute-second-URL-resource1-resource2-resource3"**.In this research we have taken log file, which having 300 numbers of records. Snapshot of web log data file is shown in Figure 10.

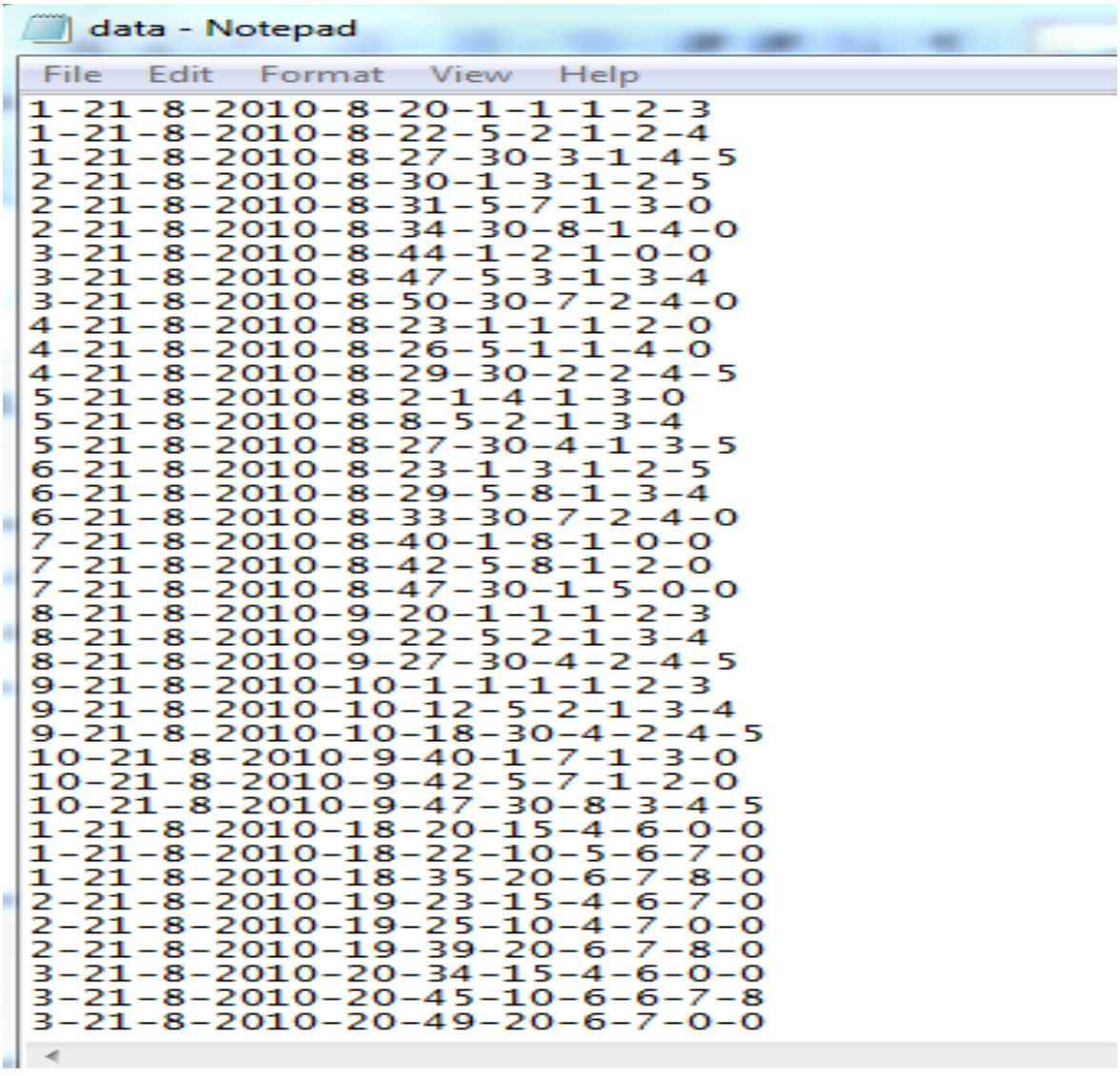


Figure 10 Snapshot of web log data.

Steps:

a) First of all we identify different user, this can be done on the basis of their userid.

b) Identify different session corresponding to different user. Here we also define some threshold value based on which we differentiate among multiple sessions. Threshold

value is 30 minutes, it means that two different web access record are considered in same session until is not differ by more than 30 minutes.

Entry of a session is in form of (u, t, d) where 'u' represent accessed URL, 't' represents start timestamp and 'd' represents duration of accessing URL U_i.

Sessions identified for user 1 are

S_{11}=	{	(1, 8.336, 0.044),	(2, 8.380, 0.153),	(3, 8.533, 0.099)	}
S_{21}=	{	(4, 18.37, 0.019),	(5, 18.38, 0.024),	(6, 18.64, 0.131)	}
S_{31}=	{	(16, 21.63, 0.036),	(16, 21.667, 0.250),	(17, 21.96, 0.143)	}
S_{41}=	{	(1, 8.336, 0.044),	(2, 8.380, 0.153),	(3, 8.533, 0.099)	}
S_{51}=	{	(4, 18.37, 0.019),	(5, 18.38, 0.024),	(6, 18.64, 0.131)	}
S_{61}=	{	(16, 21.63, 0.036),	(16, 21.667, 0.250),	(17, 21.96, 0.143)	}
S_{71}=	{	(1, 8.336, 0.044),	(2, 8.380, 0.153),	(3, 8.533, 0.099)	}
S_{81}=	{	(4, 18.37, 0.019),	(5, 18.38, 0.024),	(6, 18.64, 0.131)	}
S_{91}=	{	(16, 21.63, 0.036),	(16, 21.667, 0.250),	(17, 21.96, 0.143)	}

Finding period of a session, it is defined as time difference between end timestamp of last activity in a session and start timestamp of first activity in a session.

Period of sessions denoted by $Per_i(S_j)$, which represents period of a session corresponding to the i[th] user for j[th] session.

Period of sessions corresponding to user having userid 1.

$Per_1(S_1)$=8.533-8.336+0.99=0.296

$Per_1(S_2)$=0.395

$Per_1(S_3)$=0.429

$Per_1(S_4)$=0.296

$Per_1(S_5)$=0.395

$Per_1(S_6)$=0.429

$Per_1(S_7)$=0.296

$Per_1(S_8)$=0.395

$Per_1(S_9)$=0.429

c) Building web Usage context for periodic attributes and resource attributes. Table 4 represents web usage context of Periodic attributes for user 1 and Table 5 represents web usage context for resource attributes for user 1.

Table 4 Web usage context of Periodic attributes for user 1

	P1	P2	P3	P4	P5	P6	P7	P8
S1	0	0	1	0	0	0	0	0
S2	0	0	0	0	0	0	1	0
S3	0	0	0	0	0	0	0	1
S4	0	0	1	0	0	0	0	0
S5	0	0	0	0	0	0	1	0
S6	0	0	0	0	0	0	0	1
S7	0	0	1	0	0	0	0	0
S8	0	0	0	0	0	0	1	0
S9	0	0	0	0	0	0	0	1

Table 5 Web usage context of Resource attributes for user 1

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
S1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

S2	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
S3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
S4	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S5	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
S6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
S7	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S8	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
S9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0

d) Identify web access activities and also find support and confidence value of those activities.

Web access activities for user with userid 1

$v(Q_{11})$={ (P3,1), (R1,1), (R2,1), (R3,1), (R4,1), (R5,1) }

$v(Q_{21})$={ (P7,1), (R6, 1), (R7,1), (R8,1) }

$v(Q_{31})$={ (P8,1), (R16,1), (R17,1), (R18,1) }

e) Build web usage lattice by using formal concept analysis and TITANIC algorithm.

Web usage lattice for user 1 shown in figure 11.

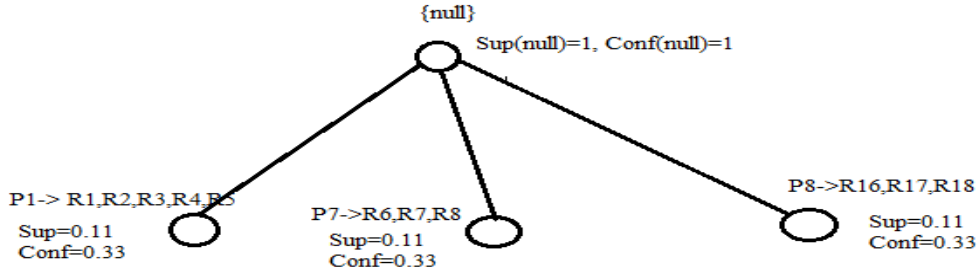


Figure 11 Web Usage Lattice

f) Personalized resource generator will generate the resources based on periodic condition by searching over the web usage lattice. E.g. if periodic condition is equivalent to the P3 in case of user 1 then it will recommend resources R1,R2,R3,R4,R5.

g) Build resource v/s URL matrix, Table 6 represents resource v/s URL matrix. U_i represents URL having name 'i'.

Table 6 Resource v/s URL matrix

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
R1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
R2	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
R3	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
R4	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
R5	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
R6	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R7	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R8	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
R9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
R10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
R18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
R19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

h) Now evaluate URL's corresponding to the recommended resources which are generated by the personalized resource generator on the basis of parameter explained in previous chapter and then recommends those URL, which having highest fitness value.

$$FF(U_{ij}) = \sum_{k=1}^{6} C_k * P_{kj} * f(R_{ij})$$

$$FF(U_{11}) = (C_1 * P_{11} + C_2 * P_{21} + C_3 * P_{31} + C_4 * P_{41} + C_5 * P_{51} + C_6 * P_{61}) * f(R_{11})$$

Here FF(U_{11}) is defined as fitness value for resource R_1 corresponding to the URL U_1, where C_1, C_2, C_3, C_4, C_5, and C_6 are coefficient which define the importance of parameter. In this research we work consider normalize value.

CHAPTER 6: RESULTS AND COMPARISON

This chapter deals with results obtained from previous and proposed approach and also comparison in between them. Both approaches are implemented in C language. The experiments are carried very carefully so that we can determine the behavior of proposed approach.

6.1 Performance Measures

Fitness function is one of the main performance measures which is based on upon six different parameters, they are

a) Total number of user.

b) Number of unique visitor.

c) Number of user downloads data.

d) Amount of data downloaded.

e) Amount of data uploaded

f) Number of advertisement.

Value of fitness function to a resource R_i corresponding to the URL U_j decides its effectiveness, more value of fitness function means more relevant is the URL to the user needs.

6.2 Experimental Results

Figure 12 represents Access Pattern of user having userid 1 for session 1, basically it shows that during different sessions which resources user access and from which URL's they are accessed, similarly Figure 13 also represents access pattern for user having userid 2 for session 2.

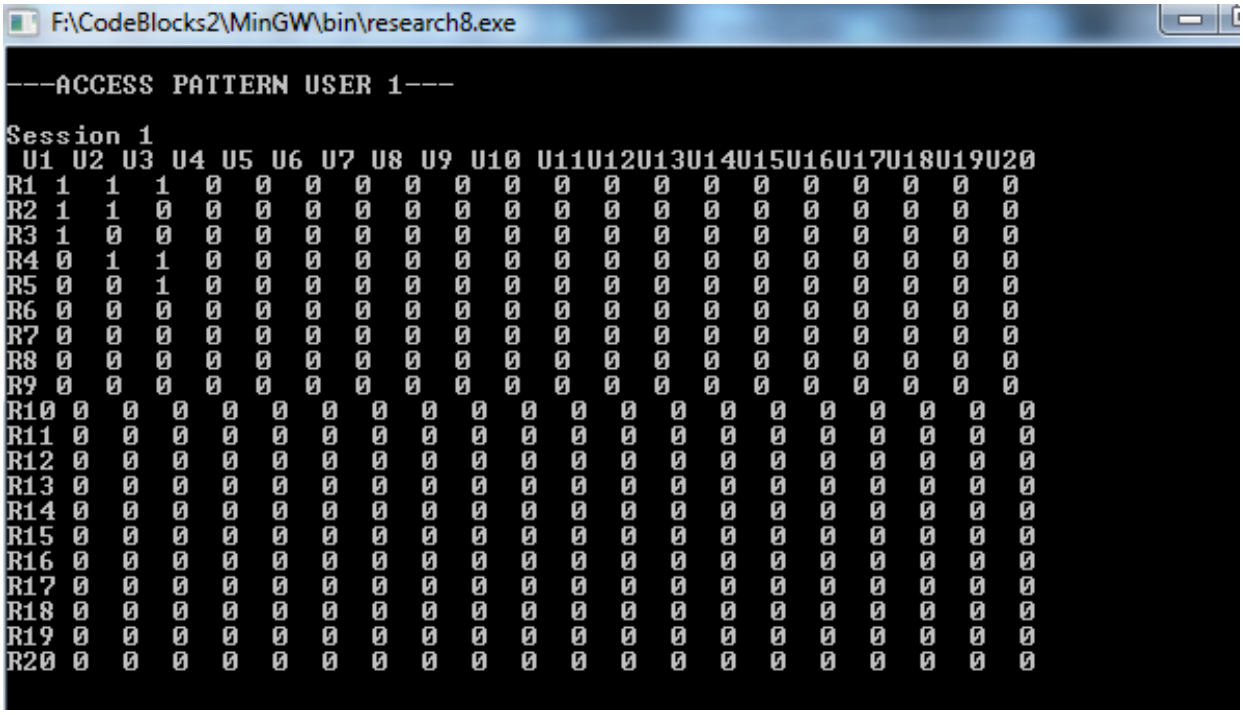


Figure 12 Access Pattern of user 1 for session 1

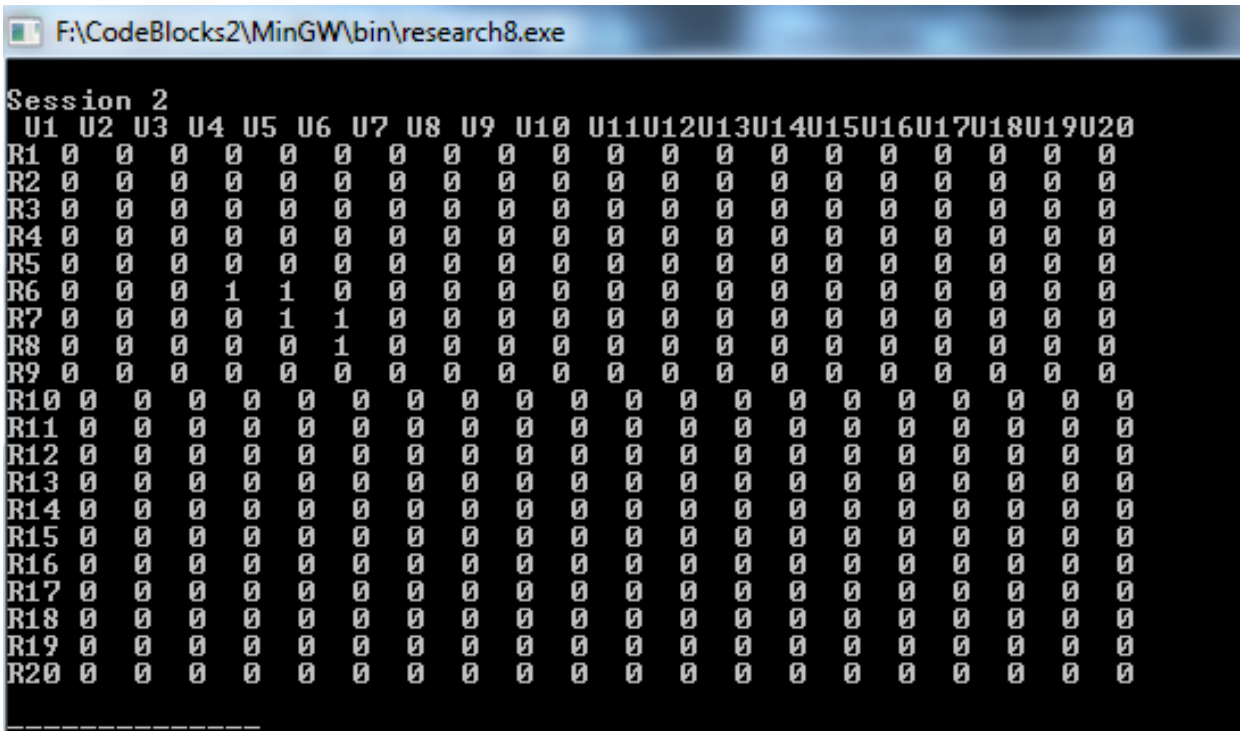


Figure 13 Access Pattern of user 1 for session 2.

Figure 14 Results of previous and proposed approach

Figure 14 shows the result obtained from previous and proposed approach. Table 9 and Table 10 shows recommended URL's along with highest fitness value among those URL's in both previous approach explained in [17] and proposed approach.

Table 7 Previous Approach Results

Resource(R_i)	Recommended URL's	Fitness value(Highest)
R1	U1,U2,U3	75.75
R2	U1,U2	75.75
R3	U1	46.85
R4	U2,U3	75.75
R5	U3	67.30

While in case of proposed approach recommended URL's are shown in table in T.

Table 8 Proposed Approach Results

Resource(R_i)	Recommended URL's	Fitness value(Highest)
R1	U7,U2,U8,U3,U4,U1	86.15
R2	U7,U2,U8,U3,U4,U1	86.15
R3	U7,U2,U8,U3,U4,U1	86.15
R4	U7,U2,U8,U3,U4,U1	86.15
R5	U2,U8,U3,U4,U1	75.75

6.3 Comparative Study

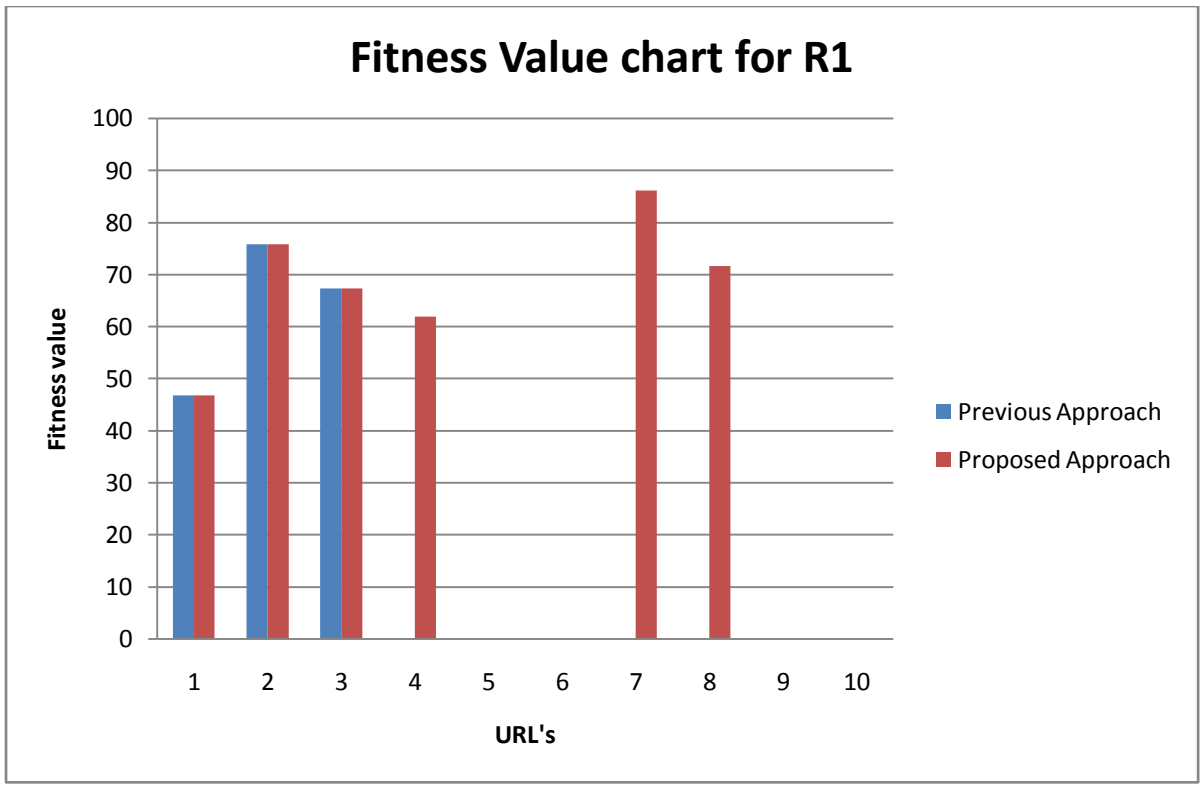


Figure15. Fitness Value Chart for Resource (R1)

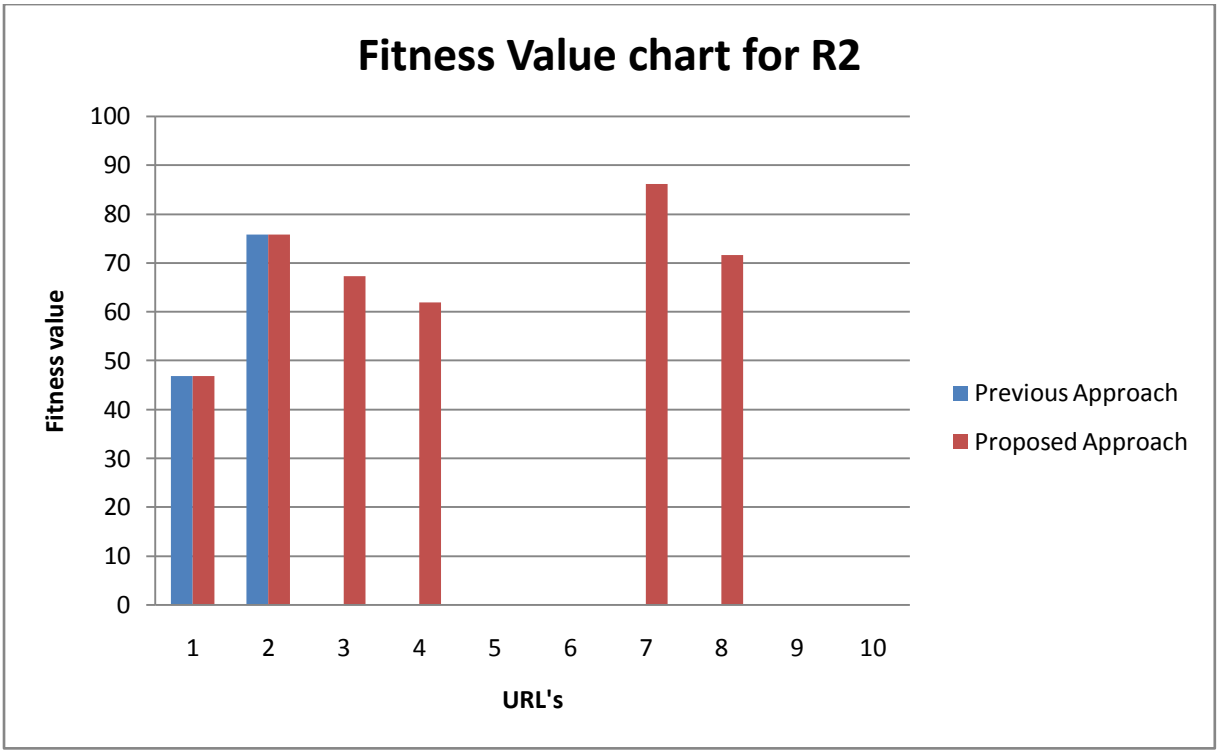


Figure 16 Fitness Value Chart for Resource (R2)

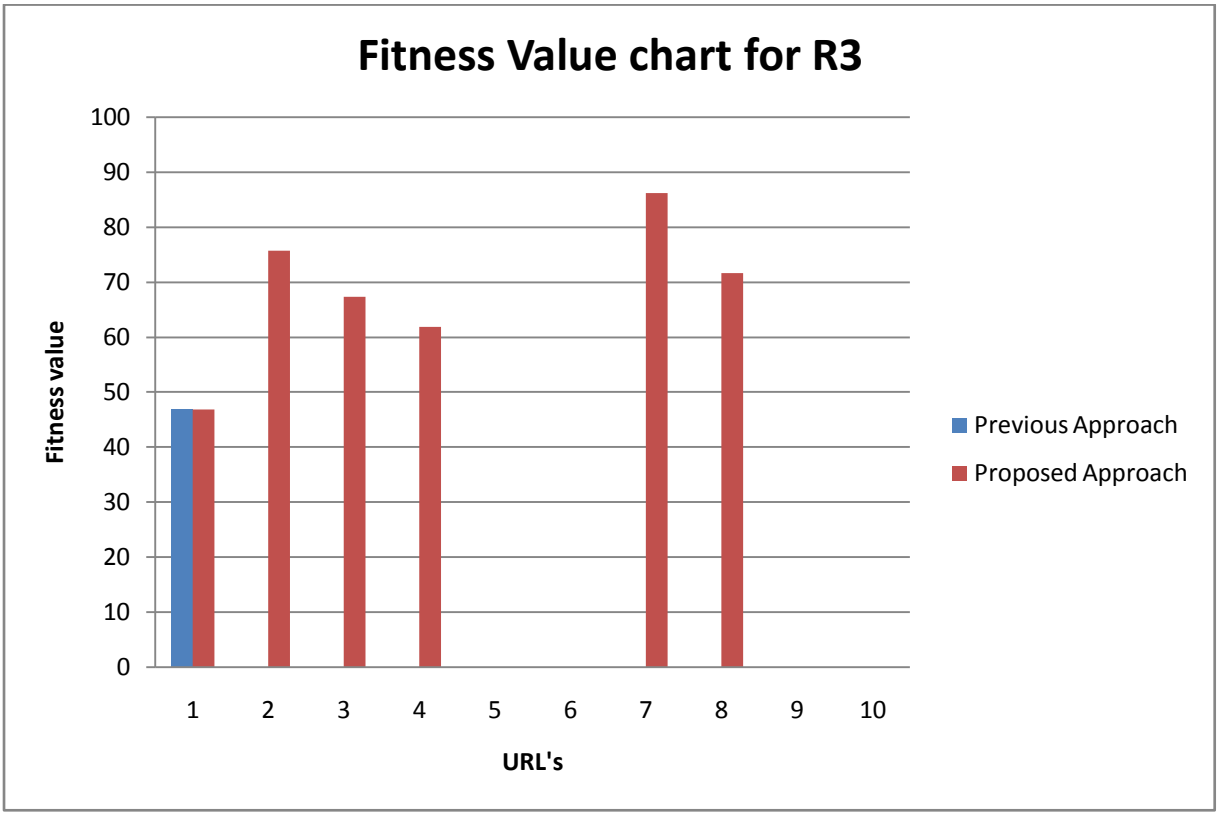


Figure 17 Fitness Value Chart for Resource (R3

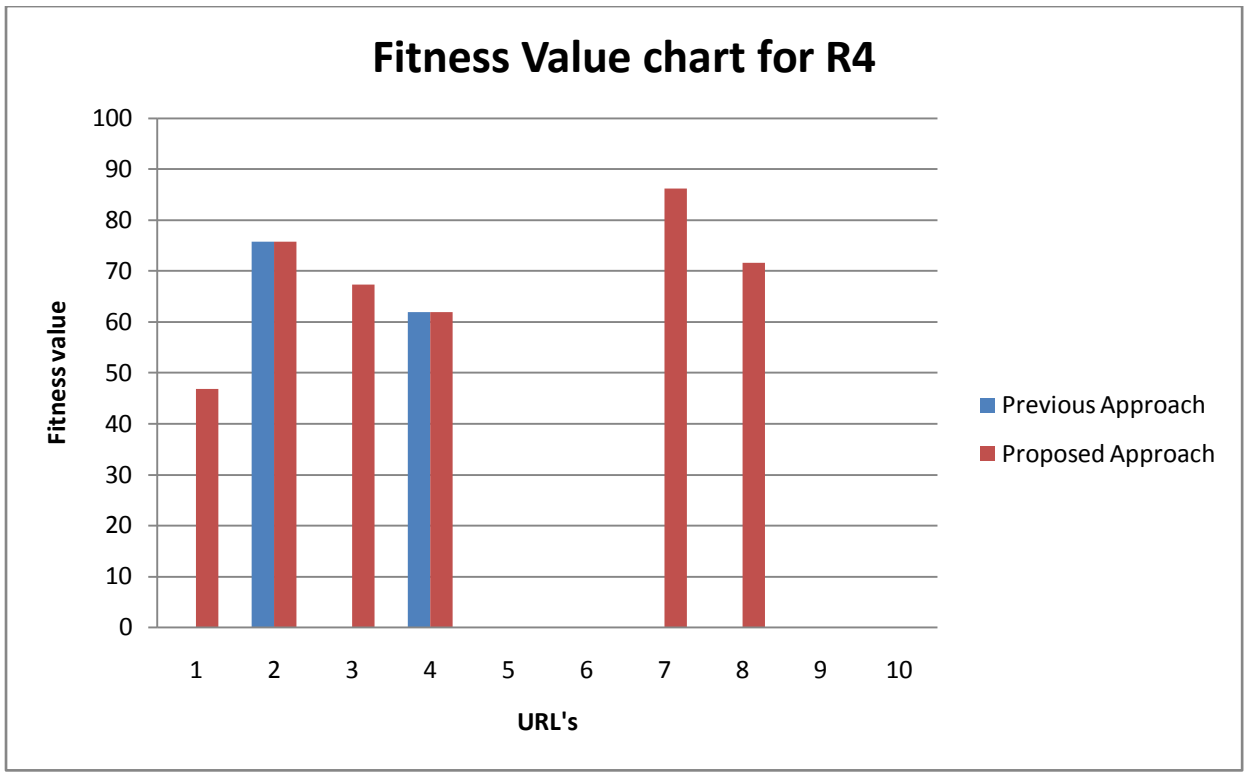


Figure 18 Fitness Value Chart for Resource (R4)

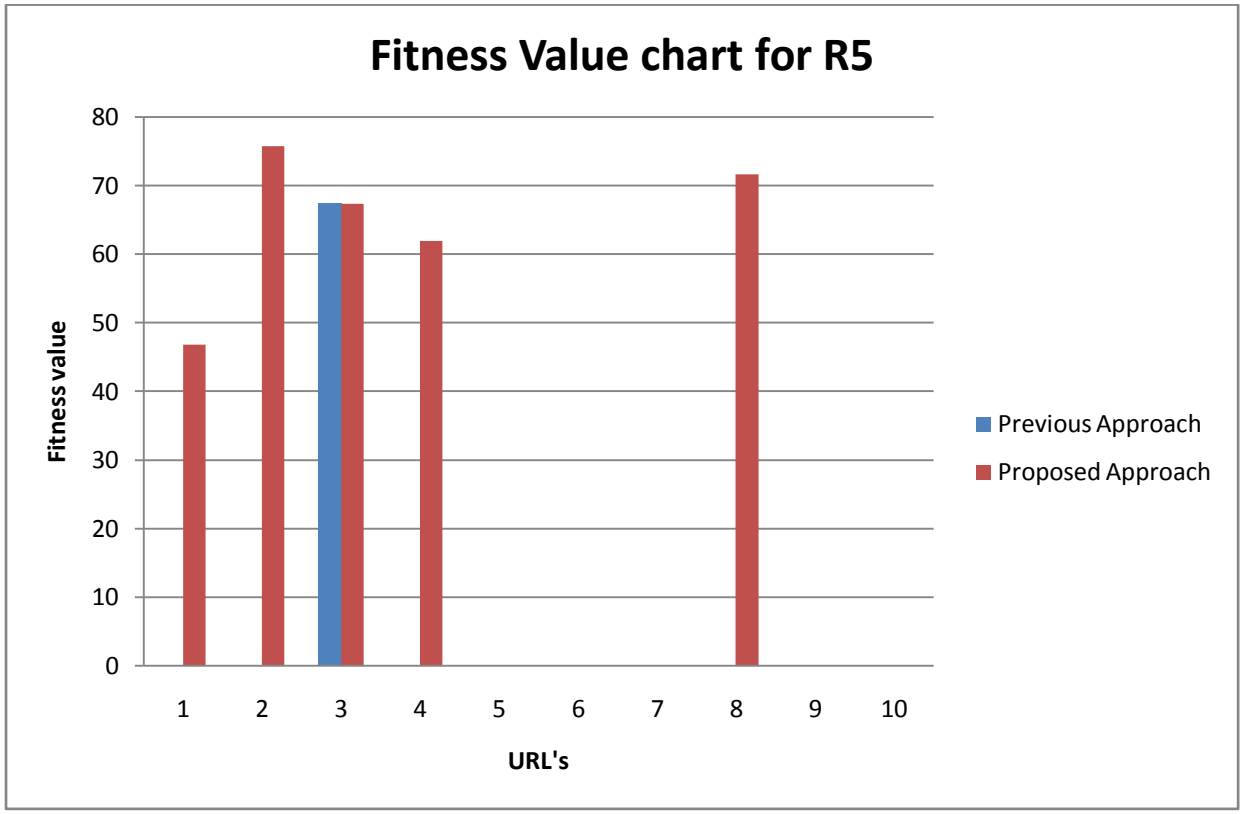


Figure 19 Fitness Value Chart for Resource (R5)

Comparison between two approaches is shown in bar chart graph form. There are five graphs which correspond to the five different resources. This bar chart graph is drawn in between URL's and fitness value calculated for a particular resource. Blue bar is representation of previous approach results whereas red bar is representation of proposed approach.

Figure 15, we can see that there are only three blue bars and blue bar with URL 2 having the highest fitness value so according to previous approach URL 2 is most efficient recommendation. While there are six red bars and red bar with URL 7 having the highest fitness value. So there is enhanced recommendation in two ways firstly proposed approach recommends more URL's as compare to previous approach and secondly proposed approach recommends URL with highest fitness value, which is well suited for the user needs. In this way similar comparison has been shown in further figures.

CHAPTER 7: CONCLUSION AND FUTURE WORK

This chapter deals with conclusion infered and also the future expansion of this research work.

7.1 Conclusion

As we known web is increasing with very tremendous rate so getting informatin relevant to the user in efficient and effective manner is very important task. So in order to provide information to the user efficiently recommendation system come into existance as we see diifferent type of recommendation, periodic personalized web recommendation system is a type of recommender system whcih considers periodic attributes for making recommendation. But problem with such system is that either they not consider peridic attibutes or not provide global best recommendation. In this research work, we have proposed an extended algorithm for periodic web personalization, which not only considers users periodic access patterns but also considers different parameters related to different websites. These parameters help to evaluate different websites in order to provide recommendation, which is globally best among all the available URL's.

7.2 Future Work

In future we can also provide better recommendation by considering others factors, which are related to websites and users profile. In this research we assume that web usage logs are semnatically enhanced and in proper formate but this is not the ideal case, in order to improve results we can perform pre-processing to get better results.

CHAPTER 8: PUBLICATION FROM THE RESEARCH

This chapter briefly states the publication that has been done during this research work.

Khatri R., Gupta D., "**An Effective Periodic Web Content Recommendation Based on Web Usage Minin**g" has been accepted in IEEE Conference "**Recent Trends in Information System (ReTIS-15)**", **Kolkata**.

REFERENCES

[1] S.P. Bora, "Data Mining and Data Ware Housing", 3rd International Conference on Electronics Computer Technology (ICECT- 2011).

[2] O.A. Nassar, N.A. Al Saiyd, "The Integrating Between Web Usage Mining and Data Mining Techniques", 5th International Conference on Computer Science and Information Technology (CSIT-2013).

[3] B. Singh, H. Kumar, "Web Data Mining Research: A Survey", Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference.

[4] K. Sharma, G. Shrivastava, V. Kumar, "Web Mining: Today and Tomorrow", Electronics Computer Technology (ICECT), 2011 IEEE International Conference.

[5] K. Pol, N. Patil, S. Patankar, C. Das, "A Survey on Web Content Mining and Extraction of Structured and Semi structured Data".

[6] S. Boddu, V.P. Krishna, R. Rao, D.K. Mishra, "Knowledge Discovery and Retrieval on World Wide Web using Web Structure Mining", 4th International Conference on Mathematical /Analytical Modelling and Computer Simulation 2010.

[7] R. Omar, A. Osman, Z. Abdullah, "Web Usage Mining: A Review of Recent Works", 5th International Conference Information and Communication Technology (ICT4M2015).

[8] J. A. Konstan, B. N. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. Grouplens: Applying Collaborative Filtering to Usenet News. Comm. of ACM, 40(3):77-87, 1997.

[9] Q. Yang, S. Zhang, B. Feng, "Research on Personalized Recommendation System of Scientific and technological Periodical Based on Automatic Summarization", 1st IEEE International Symposium on Information Technologies and Application in Education (ISITAE 2007).

[10] T.S. Nguyen, H. Lu, J. Lu, "Web Page Recommendation Based on Web Usgae and Domain Knowledge", IEEE Transaction on Knowledge and Data Engineering.

[11] L. Zhang, X. Liu, X. Liu, "Personalized Instructing Recommendation System based on Web Mining", 9th International conference for Young Computer Scientist 2008.

[12] X. Wang, Y. Ouyang, X. Hu, "Discovery of User Frequent Access Patterns on Web Usgae Mining", 8th International Conference on Computer Supported Cooperative Work in Design.

[13] S. Puntheeranurak, H. Tsuji, "Mining Web logs for a personalized Recommender System", 3rd International conference on Information Technology: Research and Education (ITRE 2005).

[14] W. Xiao-gang, L. Yue, "Web Mining Based on User Access Patterns for Web Personalization", International Colloquium on Computing, Communication, Control and Management 2009.

[15] L.A. Zadeh, Fuzzy Sets, Journal Information and Control.

[16] B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundation", ACM TOIT, 3(1):1-27,2003.

[17] A.C.M. Fong, B. Zhou, S.C. Hui, G.Y. Hong, T.A. Do, "Web Content Recommender System Based on Consumer behaviour Modelling", IEEE Transaction on Consumer Electronic, 2011.

[18] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal. "Computing Iceberg Concept Lattices with TITANIC", Data Knowledge Engineering, 42(2):189-222.2002.