IJCAT - International Journal of Computing and Technology
Volume 1, Issue 1, February 2014
www.IJCAT.org

Improving Website Usability by Mining Trending and Popular Web Resources Based on User Access Patterns from Weblogs

[1]Deepika Jain, [2]Bhawna Nigam

[1] Department of Information Technology, SGSITS,
Indore (M.P.), India

[2] Department of Information Technology, IET-DAVV,
Indore (M.P.), India

Abstract - Websites are now mode of education, business and entertainment for almost every entity on Internet. To retain interests of users in the website, it is necessary to provide advance usability features. For example, in online shopping, if information about frequently visited brands of a specific item is provided to the users, then it will help users to identify the purchasing trends of that item, and help them to buy a product. Also, it will help companies to identify popularity of products. But, as interests of users frequently changes with time, this paper proposes a web usage mining technique that retrieves popular and trending visited web resources of website on daily basis, to improve site usability. Experimental result shows trending and popular web resources for different access patterns of users.

Keywords **- Web Usage Mining, Prediction, Recommendation, Web logs, Trending, popular.**

1. Introduction

In today's world, there are numerous websites available on Internet for a single application. To sustain in this competitive environment, it becomes very important to retain interest of users in a website. Therefore, better usability features are needed to be incorporated in the website. Web Usage Mining is used to identify common interest of users, by determining the traversals performed by them and utilize this information for future users. One usability feature, which is already present in many of the websites, is to identify popularly visited resources with "maximum hits" or "next access of web resources", and provide their links to the users. This is done by analyzing previous history of access logs.

A. Anitha uses a log file and integrates Markov Model, Sequential Pattern Mining and Clustering for predicting next page access from similar access patterns to engage users from the beginning of the site traversal [1]. R. Mabroukeh and Christie I. Ezeife also uses lower order markov model with semantic information on medium and large size data for next page access prediction the results of which can be useful for prefetching and caching [6]. Faten Khalil, Jiuyong Li and Hua Wang integrate Markov Model and association rules for predicting web page accesses [13] on logs of year 1999. [5] suggests algorithms for mining minimized set of valid association rules by considering the structural knowledge of website. In their later work, Faten Khalil, Jiuyong Li and Hua Wang have also used the concept of clustering to provide better prediction accuracy [17]. [18] also used Markov Model and association rules for prediction and used support and confidence values for the same. [11] used the concept of dynamic nested markov model to predict next accessed web page whose analysis is done on different schemes of prefetching and caching [15].

Log files of an University are analyzed by Sartaj Ahmad and M.Z. Khan on weekly basis [2] to evaluate pages (that consists of topics of particular subject) on the basis of their "hits per week per topic" with "maximum session time spent on that topic per week" and conducted test on topics with hits and session time spent less than the average value, to improve online teaching techniques for the topic in which students' performance is not up to mark. Clustering and Association rules are used for E-Learning Personalization based on recent access interest of users [7] (for 100 learning sessions), online automatic recommendations for active users, without their explicit feedback [9] (taking logs of six months) and improving student's performance [8] (for log of 11 months).

For speeding up Web Access, Abhinav Srivastava, Abhijit Bhosale and Shamik Sural suggests placing frequently accessed web pages in Server Cache [3] using Weighted

Association Rules by capturing both user's habit and interests. They used log of eight hours. Navin Kumar Tyagi, A. K. Solanki and Manoj Wadhwa emphasizes on Website improvement that can be done by observing the maximum occurring status error when next page can't be visited from referrer page, i.e. link to next page fails and gives error upon access [4].The logs of eight days are analyzed for this task.

Bamshad Mobasher Robert Cooley, Jaideep Srivastava used association rules and transaction clustering for automatic web personalization [10] recommending next accesses for a session using hypergraphs. The logs taken are older logs for "n" number of days. C. Umapathi, M. Aramuthan, and K. Raja proposed techniques for prefetching [12] to enhance web services by predictive caching. [14] proposes a predictive model to offer better prediction latency. [16] proposes prediction model that considers order of web pages as well as time spent by users on that page and clusters are created for user sessions whose results are represented in form of click stream trees. The logs used were of year 1999. [19] performs survey of various methods for future request prediction. [20] incorporates knowledge of semantic web along with prediction for recommending list of products to the users. [22] proposes architecture for online recommendation system that first analyzes navigational patterns in offline mode and perform classification and future prediction for current user activity. [21] uses information retrieval, natural language processing and data mining to evaluate trending results of products for prediction purpose.

The Weblogs used in approaches mentioned above either uses older logs or logs of "n" number of days at one instance. But behavior and interest of users may vary as time passes. So older logs may not be suitable to be used for improving usability feature. Also, many of these approaches focuses on finding "popular" resources i.e. frequently used resources for the whole site or "popular resource" for user access patterns, but these "popular web resources" may be different from current trend of user's traversal. Therefore an incremental approach is proposed to improve website usability, that directly uses the trail information obtained from pre-processing of web logs, which predicts two categories of results -

(a) "Trending web resources" for current trend of user access patterns, which are evaluated irrespective of previous history from the logs. Access logs are processed only for real time web logs on daily basis.

(b) "Popular web resources" for current as well as past user access patterns, from real time logs processed on daily basis and appending this data with past usage data.

For each Web resource, weighted support is calculated regarding each user access pattern, and recommendation of top "n" resources will be done on the basis of decreasing order of weighted support. The result of trending and popular web resources generated in current day will be recommended on next day to the users.

The organization of paper is according to the following sequence – Section 2 contains terminologies used throughout the paper, Section 3 shows the Architecture of system, Section 4 contains working of the proposed algorithm, Section 5 shows the details of Data Set, Section 6 represents Experimental results and Section 7 is related to conclusion and Future work.

2. Terminologies –

(I)Web Resource - Web resource are fundamental elements of a site that can be requested by a user, such as web pages such as html pages, aspx pages, pdf files or multimedia (audio/video) files etc.

(II) Trail - Sequence of web resources visited by a user in session.

(III)Support - (Number of times users visited a particular trail)/ (total number of trails)

(IV)User Access Pattern - Web resource or combination of web resources (except the last visited web resource) that can be followed by a user to traverse the website. For e.g. if a->b->c trail is visited by user, then user access patterns can be a, a->b and b.

(V)Next Visited Web Resource - Web resource that can be visited 'next' by the user for an already traversed access pattern. For e.g. if a->b->c trail is visited by user, then next visited web resources for user access patterns i.e. a, a->b and b are {a=b}, {a->b=c} and {b=c}.

(VI)Weighted Support/Incremental Weighted Support - For a particular User Access Pattern -

(a) If a new entry for next visited web resource appears, then Weighted Support of next visited web resource = support value of new entry

(b) If same visited web resource (as previously accessed) reoccurs, then Incremental Weighted Support of next visited web resource = support value of previous entry + support value of new entry

(VII) Trending web resource - These are the next visited web resources for a specific User Access Pattern (recommended in decreasing order of weighted support) whose weighted support is calculated in incremental manner for real time web logs that are processed on daily basis. These web resources are independent of previous history of access logs. The result of current day trending

web resources will be recommended on next day to the users.

(VIII) Popular web resource - Popular web resource for a specific User Access Pattern are the next visited web resources (recommended in decreasing order of weighted support) whose weighted support is calculated in incremental manner by mapping results of trending web resources of current day with past usage data. These results are also recommended on the next day.

3. Architecture – Fig 1 shows the architecture of the system that is used for evaluating as well as

recommending "trending" and "popular" web resources. The Data Set used in this work belongs to website of an Engineering Institute (IET), which is explained in Section 5. As log file is processed on daily basis consecutively, so log files of each day are taken as input in day wise sequence.

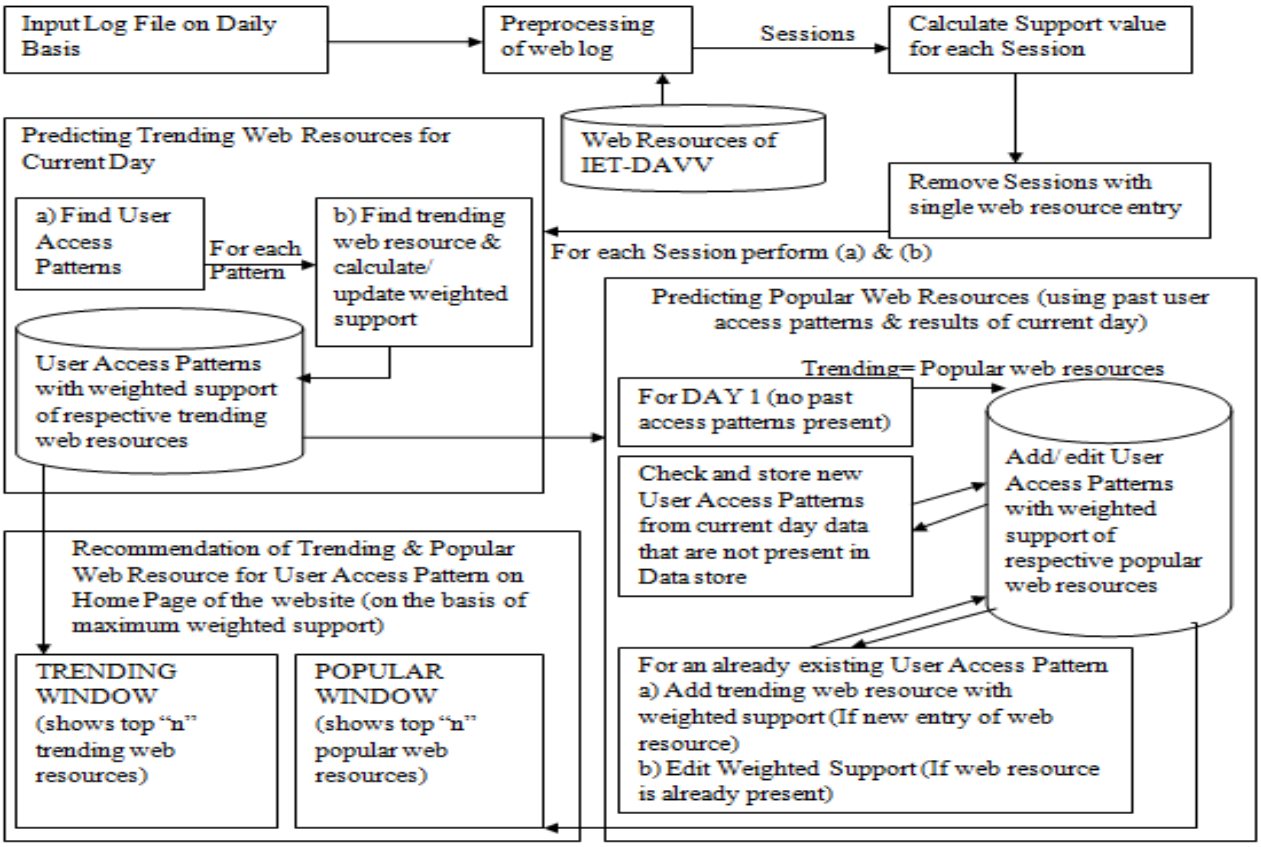


Fig 1: Architecture of System

4. Proposed Algorithm – The implementation of architecture is shown step by step-

Step 1: Select the log file (of current day).

Step 2: For each entry of web log, perform pre-processing task (i) to (iv) as follows –

(i) Data Cleaning - The log entry will be discarded if -

(a) The status code is 4XX or 5XX (client side or server side error)

(b) Number of bytes received by user is zero

(c) Requested web resource is root i.e. "/", and referrer field doesn't contain keyword "IET" (it means that some other website's home page is requested by user)

(d) The requested web resource is not valid (according to the entries of IET) If Data is valid, then go to (ii), else continue with (i).

(ii) Validate Time Entry - If this the first valid entry of log, then go to (iii) else using the time detail of current entry, check whether the time of existing sessions exceeds 20 minutes or not (by subtracting both the time values). If any Session/sessions are expired, then store their trail information along with number of times that particular trail information is visited, in a separate data structure. Move to (iii).

(iii) User Identification - On the basis of IP address, Identifier field, Username, Version of protocol used by user and Browser used by user i.e. User Agent field, User information is stored.

Also, the session information i.e. web resource which is requested and accessed by user is stored and time information is also stored in a suitable data structure. Move to (iv).

(iv)Session Identification (Referrer Based Method) - If it is the first entry, then user is a new user. Else, check the Referrer field. If Referrer is "-", or any website other than www.iet.dauniv.ac.in, then it is again a new user. If this is not the case, then check the information of existing users. If user information is matched, then check whether the last visited web resource of that user is same as the referrer web resource of current entry or not. If yes, then it is existing session, else it is a new session.

Continue, (i) to (iv) till the end of log file reaches. If end of file is reached, then remove the sessions one by one and store the trail information with the number of times they are visited, in a data structure.

Step 3: Calculate the value of support for each trail.

Step 4: Filter the obtained trails - If trail contains single entry of web resource (for e.g. abc.html), then discard that entry, because it can't be inferred from such trails, that from which user access pattern it will be referred or which web resource will be visited next from this web resource.

Step 5: Find trending web resources for User's Access Patterns- Find user's access patterns and evaluate trending web resources for them. For each trail information, perform the following -

(i) Find all possible unique User Access Patterns and store them.

(ii) Find all possible unique next visited web resources that can be accessed.

(iii) For each particular User Access Pattern, check the next web resource that can be accessed, known as trending web resource (from the current trail information) and store this information with the value of weighted support. If that trending web resource already exists, then calculate the incremental weighted support and update the existing information.

Step 6: Find popular web resources for User's Access Patterns - If it is the first time the popular web resources are evaluated, then trending web resources = popular web resources for user access patterns. Else, for each next day entry, perform the following -

(i) Match each new User Access Patterns obtained for current day log entries with past usage patterns - If a new access pattern is not present in previous entries, then add this information along with trending web resource' information (which will be known as popular web resource in this case) and their weighted support.

(ii) If User Access Patterns matches with past usage patterns - For each User Access Pattern, check the next visited web resources' entry for finding popular web resources.

(a) If next visited web resource entry is new, then add this information (known as popular web resource in this case) along with its weighted support for the User Access Pattern entry.

(b) If next web resource entry i.e. popular web resource is already present, then calculate the incremental weighted support and update the existing information.

The links of top n "trending web resources" as well as "popular web resources" for different user access patterns will be recommended in decreasing order of weighted support. For the next day, perform the same steps (1 to 6) for the current log and so on. The algorithm is demonstrated in the following example.

Example - Suppose after data pre-processing and calculation of support values, trail details are given for 2 days. Following tables represents the trending web resource and popular web resources that are predicted for user access patterns.

DAY 1: Trails with decreasing order of Support Values -

1. a->b->c = 0.6 2. a->b->d->c = 0.4

Table 1: Evaluating Trending Web Resources along with their weighted support for different User Access Patterns for DAY 1

No.	Trail	User Access Patterns	User Access Pattern with trending web resources & their weighted support
1	a->b->c = 0.6	a	a={b=0.6}
		a->b	a->b={c=0.6}
		b	b={c=0.6}
2	a->b->d->c = 0.4	a	a={b=0.6+0.4=1.0}
		a->b	a->b= {c=0.6,d=0.4}
		a->b->d	a->b->d={c=0.4}
		b	b={c=0.6,d=0.4}
		b->d	b->d={c=0.4}
		d	d={c=0.4}

As shown in Table 1, firstly user access patterns are generated from trail information, and then mapping of each user access pattern along with trending web resources is shown. When second trail information is explored, then the result obtained from first trail information is also appended.

In first trail, trending web resource for user access pattern 'a' is 'b' with weighted support = 0.6. Now, when the second trail information is analyzed, then for same user access pattern 'a', trending web resource is again 'b'. Therefore weightage of b should be increased after 'a' is traversed. Hence weights of 'b' are added which gives resultant weight = 0.6 + 0.4 = 1.0, which means that after 'a' is traversed, weight of 'b' is maximum i.e. it is likely that the next visited web resource after 'a' will be 'b' and hence should be recommended as one of the trending web resource for 'a'. Trending as well as popular web resources will be provided to the users in decreasing order of weighted support. In the same Table 1, for user access pattern 'b', trending web resource c will be recommended first because weighted support of 'c' is maximum, whenever web resource

'b' is accessed. For DAY 1, Popular Web Resources will be same as Trending Web Resources, as shown in Table 2, because there is no previous history data which is to be analyzed.

Table 2: RESULT - Trending and Popular Web Resources for DAY 1

Trending Web Resources along with weighted support for different User Access Patterns	Popular Web Resources along with weighted support for different User Access Patterns
a={b=1.0}	a={b=1.0}
a->b= {c=0.6,d=0.4}	a->b= {c=0.6,d=0.4}
a->b->d={c=0.4}	a->b->d={c=0.4}
b={c=0.6,d=0.4}	b={c=0.6,d=0.4}
b->d={c=0.4}	b->d={c=0.4}
d={c=0.4}	d={c=0.4}

The results of DAY 1 will be recommended on DAY 2 to the users. Then for the next day i.e. DAY 2, Trending Web Resources will be evaluated, irrespective of what users have accessed previously. While popular web resource will be evaluated by combining results of Weighted Support of Trending Resources of DAY 2 and popular web resources of DAY 1.

DAY 2: Trails with decreasing order of Support Values-

1. b->c->a = 0.7 2. a->b->d = 0.3

Table 3: Evaluating Trending Web Resources along with their weighted support for different User Access Patterns for DAY 2.

No.	Trail	User Access Pattern	User Access Pattern With trending web resources & their weighted support
1	b->c->a = 0.7	b	b={c=0.7}
		b->c	b->c={a=0.7}
		c	c={a=0.7}
2	a->b->d = 0.3	a	a={b=0.3}
		a->b	a->b={d=0.3}
		b	b={c=0.7,d=0.3}
			b->c={a=0.7}
			c={a=0.7}

Working of evaluating trending web resources shown in Table 3 is same as that explained for Table 1. Following table shows how popular web resources are derived for different access patterns using data of DAY 1 and DAY 2.

Access patterns obtained in DAY 1 for predicting popular web resource are mapped with patterns obtained in DAY 2 for predicting trending web resources. If User access pattern is present in DAY 1 as well as DAY 2, then next web resources of both days' entries are mapped and weighted/incremental weighted support is calculated to find popular web resource. If any User access pattern is not available in either of two days, then also it will be stored for finding popular web resource. For e.g.'d' is one of the user access pattern visited by users in DAY 1 but not in DAY 2. So this information is not important for mining User Access Patterns and trending web resources for DAY 2, but as it is the part of previous usage data, it will be

Table 4: Evaluating Popular Web Resources for DAY 2 based on popular web resources for DAY 1 & trending web resources for DAY 2

Popular Web Resources for User Access Patterns (DAY 1)	Trending Web Resources for User Access Patterns (DAY 2)	Popular Web Resources for User Access Patterns (DAY 2)
a={b=1.0}	a={b=0.3}	a={b=1.0+0.3=1.3}
a->b = {c=0.6,d=0.4}	a->b={d=0.3}	a->b ={d=0.4+0.3=0.7,c=0.6}
a->b->d= {c=0.4}	b= {c=0.7,d=0.3}	a->b->d={c=0.4}
b= {c=0.6,d=0.4}	b->c={a=0.7}	b={c=0.7+0.6=1.3, d=0.4+0.3=0.7}
b->d={c=0.4}	c={a=0.7}	b->d={c=0.4}
d={c=0.4}		d={c=0.4}
		b->c={a=0.7}
		c={a=0.7}

consider in DAY 2 for evaluating popular web resources for different user access pattern. Table 5 shows the final patterns generated in Table 4.

Table 5: RESULT -Trending and Popular Web Resources for DAY 2.

Trending Web Resources for User Access Patterns (Day 2)	Popular Web Resources for User Access Patterns(Day2)
a={b=0.3}	a={b=1.3}
a->b={d=0.3}	a->b= {d=0.7,c=0.6}
b={c=0.7,d=0.3}	a->b->d={c=0.4}
b->c={a=0.7}	b={ c=1.3, d=0.7}
c={a=0.7}	b->d={c=0.4}
	d={c=0.4}
	b->c={a=0.7}
	c={a=0.7}

5. Data Set - Log files are taken from Devi Ahilya Vishwavidyalaya (DAVV) University, Indore. The Access Logs are stored in Apache Combined Log Format. Sample Log Entry is shown below –

117.229.45.197 - - [11/Mar/2012:07:36:01 +0530] "GET /BE%20Syllabus.htm HTTP/1.1" 200 8764 "http://www.iet.dauniv.ac.in/" "Mozilla/5.0 (SymbianOS/9.4; Series60/5.0 NokiaC6-00/41.0.010; Profile/MIDP-2.1 Configuration/CLDC-1.1) AppleWebKit/533.4 (KHTML, like Gecko) NokiaBrowser/7.3.1.31 3gpp-gba")

Explanation of above entry is shown in Table 6 –

Table 6: Parameters in Apache Combined Log Format

Parameter Name (value mentioned in example)	Description
Host (117.229.45.197)	IP Address of client (remote host) which made the request to the server.
Rfc1413 (-)	Identifier used to identify the client making HTTP Response
User (-)	User Name of the user requesting for a web resource
[date:time] ([11/Mar/2012:07:36:01 +0530])	Date and Time of request
"request" ("GET /BE%20Syllabus.htm HTTP/1.1")	User's Requested Web Resource
Status (200)	Status Code (in terms of HTTP)
Bytes (8764)	Size of object returned to the client
"Referer"("http://www.iet.dauniv.ac.in/")	Referrer field (the web resource from which another web resource is requested)
"User-agent" ("Mozilla/5.0 (SymbianOS/9.4; Series60/5.0 NokiaC6-00/41.0.010; Profile/MIDP-2.1 Configuration/CLDC-1.1) AppleWebKit/533.4 (KHTML, like Gecko) NokiaBrowser/7.3.1.31 3gpp-gba")	User Agent or Browser Information of Client

The log files of 26 days, i.e. from 11 March 2012 to 5 April 2012 are processed on daily basis to determine trending and popular web resources for different user access patterns. These logs contain entries for sites of various Schools in DAVV University. But proposed work is done only for one Institution, i.e. Institute of Engineering & Technology (IET-DAVV) Indore, so as to observe the effect of proposed work on a single website.

As Web log contains entries for various sites, some marker needs to differentiate entries of IET-DAVV from other entries. Therefore, in order to get complete and accurate results of trail information, names of all web resources belonging to IET (html, pdf or txt files) are already stored in a text file. During this period, information about web resources is fixed and there were four hundred twenty web resources present in the site.

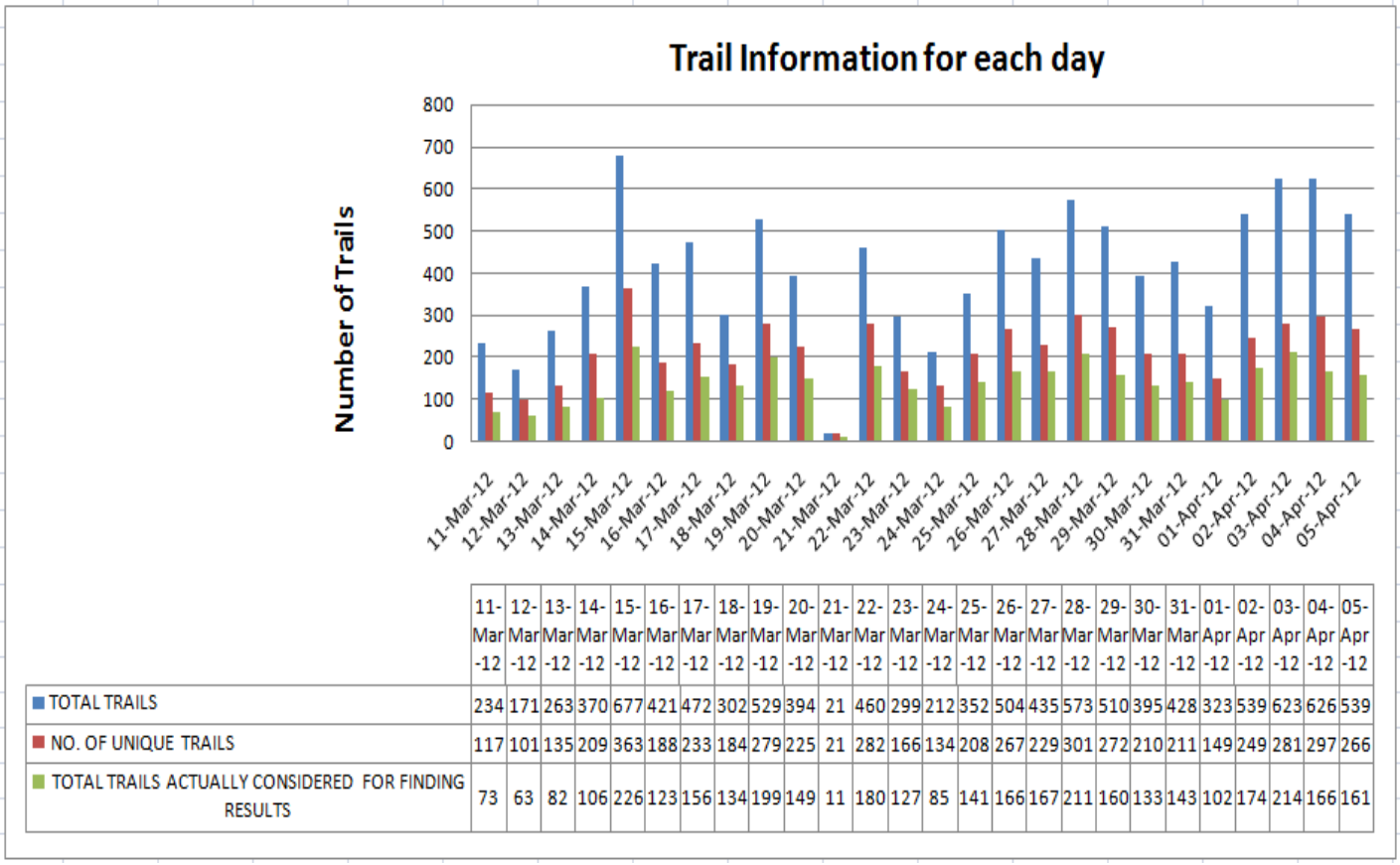
6. Experimental Results –

Fig 2 and Fig 3 shows statistics of log entries for 26 days. Fig 2 shows the count of trails generated, number of unique trails obtained and number of filtered trails. Fig 3 shows the number of different user access patterns that were generated for evaluating "trending" web resources and number of different user access patterns that were maintained when evaluating "popular" web resources.

Number of Recommended web resources will depend upon the popularity of website, number of resources in website, and choice of developer. For e.g. table 7 shows the analysis of a particular User Access Pattern i.e. Syllabus page for B.E. (/be%20syllabus.htm) is given for "top ten" links (i.e. internal links of web resources) in decreasing order of weighted support, for intermediate dates i.e. 12-March, 23-March and 5-April for evaluating both trending as well as popular web resources. These top links can be shown at the Home Page of the site, with "appropriate names" of links to the users, inside two windows namely – Trending Window and Popular Window.

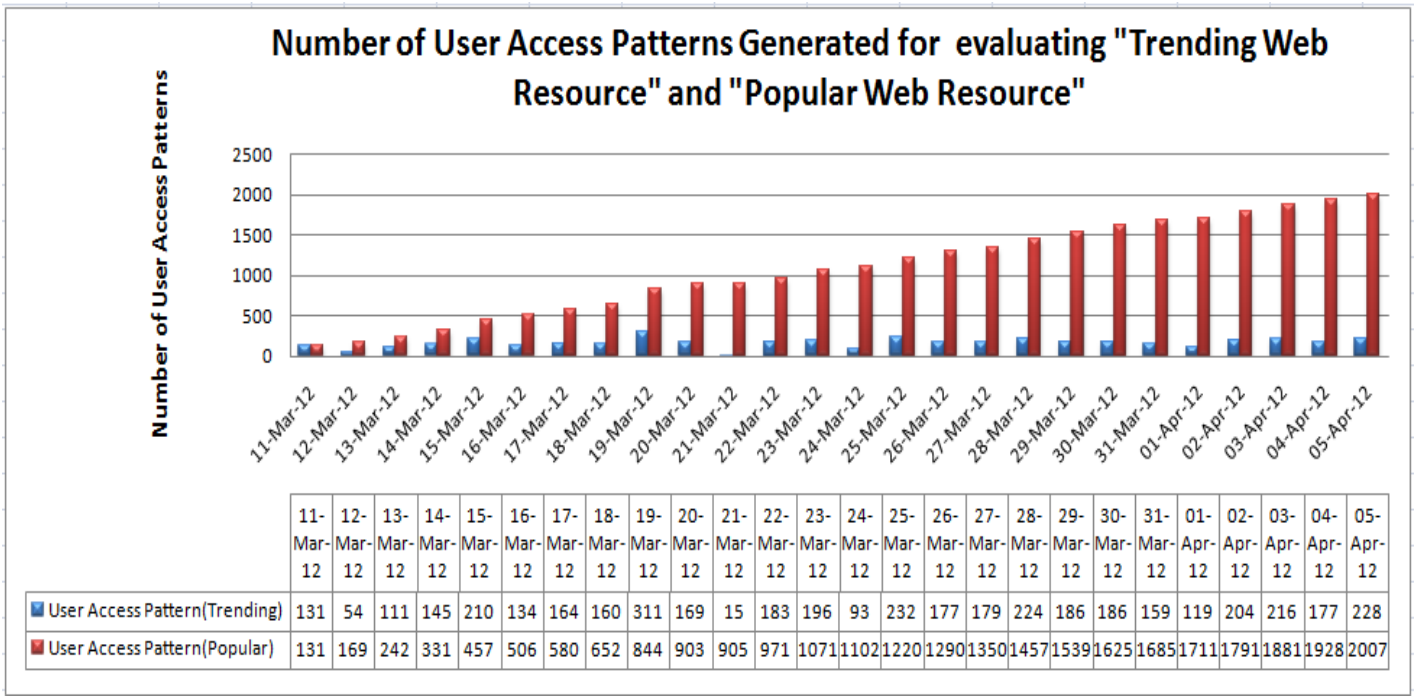
Fig 4, fig 5 and fig 6 represent trending web resources visited after Syllabus page for B.E. on respective days. In above table, some upcoming web resources have same weighted support, so such links will be shown in the order in which they are visited in the website first.

As shown in fig 4, the first recommended trending web resource after Syllabus page of B.E. will be syllabus of Ist year (all branches), because on 12th march, the syllabus page was accessed by most of the students while looking into the syllabus and schemes section. For the same user access pattern i.e. Syllabus page of B.E., results to top ten trending web resources are given for 23rd March and 5th April.



	11-Mar-12	12-Mar-12	13-Mar-12	14-Mar-12	15-Mar-12	16-Mar-12	17-Mar-12	18-Mar-12	19-Mar-12	20-Mar-12	21-Mar-12	22-Mar-12	23-Mar-12	24-Mar-12	25-Mar-12	26-Mar-12	27-Mar-12	28-Mar-12	29-Mar-12	30-Mar-12	31-Mar-12	01-Apr-12	02-Apr-12	03-Apr-12	04-Apr-12	05-Apr-12
TOTAL TRAILS	234	171	263	370	677	421	472	302	529	394	21	460	299	212	352	504	435	573	510	395	428	323	539	623	626	539
NO. OF UNIQUE TRAILS	117	101	135	209	363	188	233	184	279	225	21	282	166	134	208	267	229	301	272	210	211	149	249	281	297	266
TOTAL TRAILS ACTUALLY CONSIDERED FOR FINDING RESULTS	73	63	82	106	226	123	156	134	199	149	11	180	127	85	141	166	167	211	160	133	143	102	174	214	166	161

Fig 2: Trail Information for Web Logs [11 March to 5 April]



	11-Mar-12	12-Mar-12	13-Mar-12	14-Mar-12	15-Mar-12	16-Mar-12	17-Mar-12	18-Mar-12	19-Mar-12	20-Mar-12	21-Mar-12	22-Mar-12	23-Mar-12	24-Mar-12	25-Mar-12	26-Mar-12	27-Mar-12	28-Mar-12	29-Mar-12	30-Mar-12	31-Mar-12	01-Apr-12	02-Apr-12	03-Apr-12	04-Apr-12	05-Apr-12
User Access Pattern(Trending)	131	54	111	145	210	134	164	160	311	169	15	183	196	93	232	177	179	224	186	186	159	119	204	216	177	228
User Access Pattern(Popular)	131	169	242	331	457	506	580	652	844	903	905	971	1071	1102	1220	1290	1350	1457	1539	1625	1685	1711	1791	1881	1928	2007

Fig 3: Number of User Access Patterns for evaluating Trending and Popular Web Resources

Table 7: Trending Window showing top ten links for Syllabus page for B.E.

12-March-2012	23-March-2012	5-April-2012
/syllabus/iyr_all.pdf = 0.03508	/syllabus/iyr_all.pdf = 0.01956	/syllabus/iyr_all.pdf = 0.04267
/syllabus/ivyr_comp.pdf = 0.01754	/syllabus/iiyr_it.pdf = 0.00869	/syllabus/iiyr_comp.pdf = 0.01484
/syllabus/iiyr_comp.pdf = 0.01169	/syllabus/iiiyr_comp.pdf = 0.00869	/syllabus/iiyr_it.pdf = 0.01113
/syllabus/iyr_ptdc.pdf = 0.01169	/syllabus/ivyr_etc.pdf = 0.00434	/syllabus/iiiyr_it.pdf = 0.00927
/result_dec2011.htm = 0.01169	/syllabus/iiyr_etc.pdf = 0.00434	/syllabus/ivyr_it.pdf = 0.00927
/syllabus/iiiyr_civil.pdf = 0.00584	/syllabus/iiiyr_etc.pdf = 0.00434	/syllabus/iiiyr_comp.pdf = 0.00927
/syllabus/iiyr_etc.pdf = 0.00584	/syllabus/iiyr_comp.pdf = 0.00217	/syllabus/ivyr_comp.pdf = 0.00742
/be_iyr_rolllist.htm = 0.00584	/aboutdavv.htm = 0.00217	/syllabus/iiiyr_mech.pdf = 0.00556
	/syllabus/iiyr_ei.pdf = 0.00217	/syllabus/iyr_ptdc.pdf = 0.00371
	/syllabus/iiyr_mech.pdf = 0.00217	/syllabus/iiiyr_etc.pdf = 0.00371

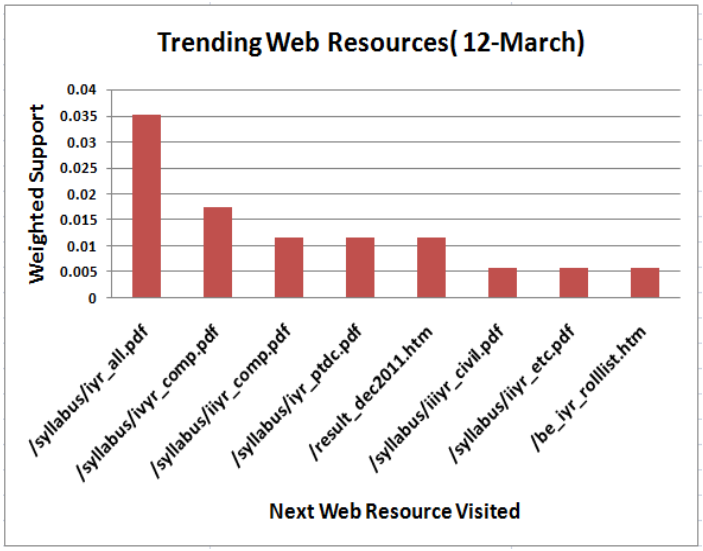


Fig 4: Graphs for Top 10 Trending Web Resource recommended after Syllabus Page of B.E. (12-Mar)

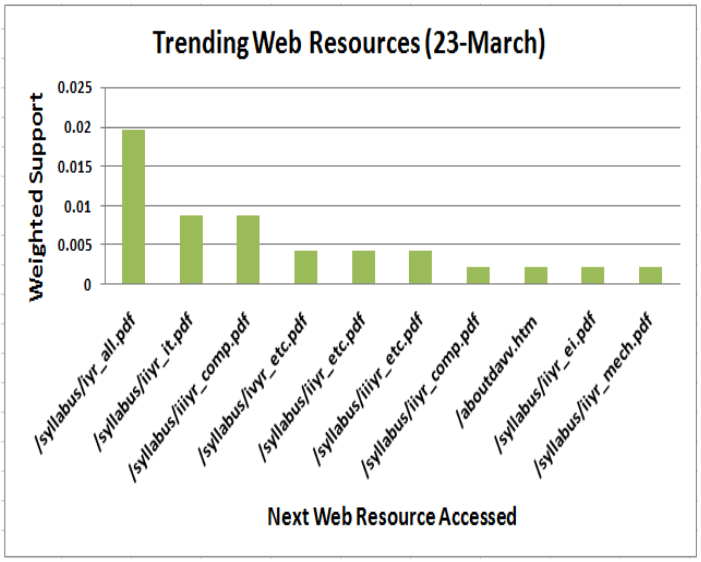


Fig 5: Graphs for Top 10 Trending Web Resource recommended after Syllabus Page of B.E. (23-Mar)

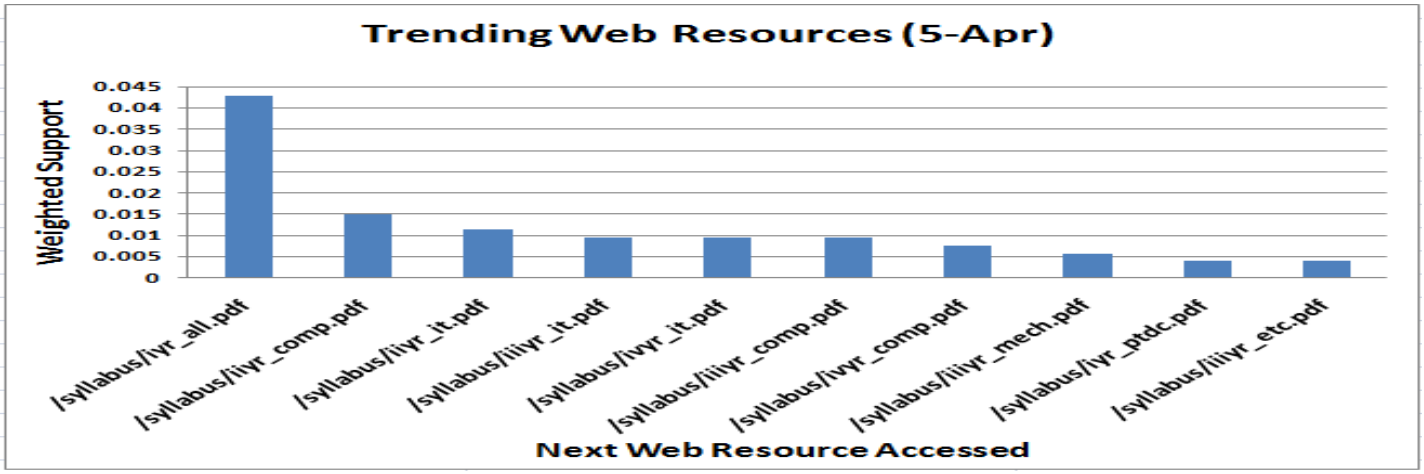


Fig 6: Graphs for Top 10 Trending Web Resource recommended after Syllabus Page of B.E. (5-Apr)

The web resource that is referred after Syllabus page of B.E. is Syllabus Page of first year (all branches) (Fig 5 and 6), because it was the time when academic session of all engineering branches was over except first year. So, maximum hits after B.E. Syllabus page was that of syllabus of first year B.E. Table 8 shows the popular web resources for same user access pattern i.e. Syllabus Page for B.E. and for same intermediate dates i.e. 12-March, 23-March and 5-April.

Table 8: Popular Window showing top ten links for Syllabus Page of B.E.

12-March-2012	23-March-2012	5-April-2012
/syllabus/iyr_all.pdf = 0.06927	/syllabus/iyr_all.pdf = 0.17410	/syllabus/iyr_all.pdf = 0.41973
/syllabus/ivyr_comp.pdf = 0.02181	/syllabus/iiyr_it.pdf = 0.09719	/syllabus/ivyr_comp.pdf = 0.17810
/syllabus/iyr_ptdc.pdf = 0.01596	/syllabus/iiiyr_etc.pdf = 0.07102	/syllabus/iiiyr_mech.pdf = 0.15147
/result_dec2011.htm = 0.01169	/syllabus/iiyr_etc.pdf = 0.06594	/syllabus/iiyr_it.pdf = 0.15142
/syllabus/iiyr_comp.pdf = 0.01169	/syllabus/ivyr_comp.pdf = 0.05963	/syllabus/iiiyr_comp.pdf = 0.13497
/syllabus/iiiyr_civil.pdf = 0.01012	/syllabus/iiiyr_mech.pdf = 0.05893	/syllabus/iiyr_comp.pdf = 0.11875
/calbe.htm = 0.00854	/syllabus/iiyr_mech.pdf = 0.05096	/syllabus/iiiyr_it.pdf = 0.10278
/syllabus/iiyr_it.pdf = 0.00854	/syllabus/iiiyr_comp.pdf = 0.04478	/syllabus/iiiyr_etc.pdf = 0.09767
/syllabus/iiyr_etc.pdf = 0.00584	/syllabus/iiyr_civil.pdf = 0.03691	/syllabus/iiiyr_ei.pdf = 0.08100
/be_iyr_rolllist.htm = 0.00584	/syllabus/ivyr_etc.pdf = 0.03600	/syllabus/iiyr_mech.pdf = 0.07493

Result of popular web resources will depend upon the trending links of current day as well as data obtained from previous day (which is calculated for evaluating popular links.). For example popular web resources for 12th march are derived from result of popular web resources for 11th march and trending web resources for 12th march. Graphs for popular web resources are shown in Fig 7, 8 and 9.

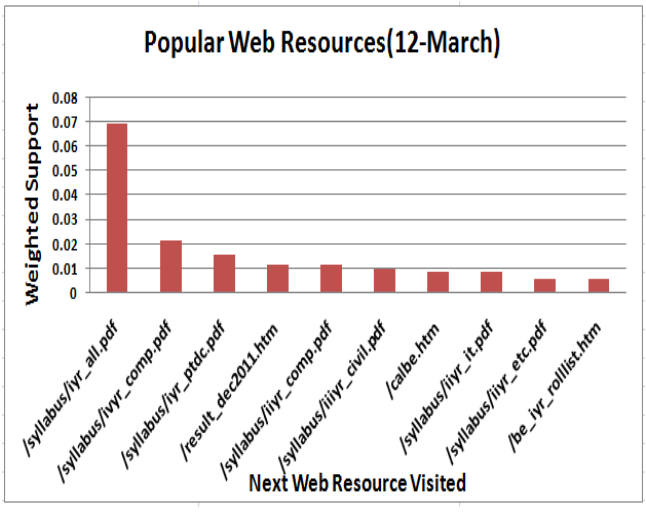


Fig 7: Graphs for Top 10 Popular Web Resource recommended after Syllabus Page of B.E. (12-Mar)

Fig 7 shows that the first popular web resource for Syllabus page in 12th March is same as that in Trending web resource. But there is a difference of value of weighted support. In case of trending resource, the weighted support of first year syllabus page is 0.03508, whereas in case of popular web resource, the weighted support is 0.06927. So it will always depend upon the log entries and past usage data, that what will be the popular web resources. Similar analysis is given for 23-march and 5-april for top ten links of popular web resources.

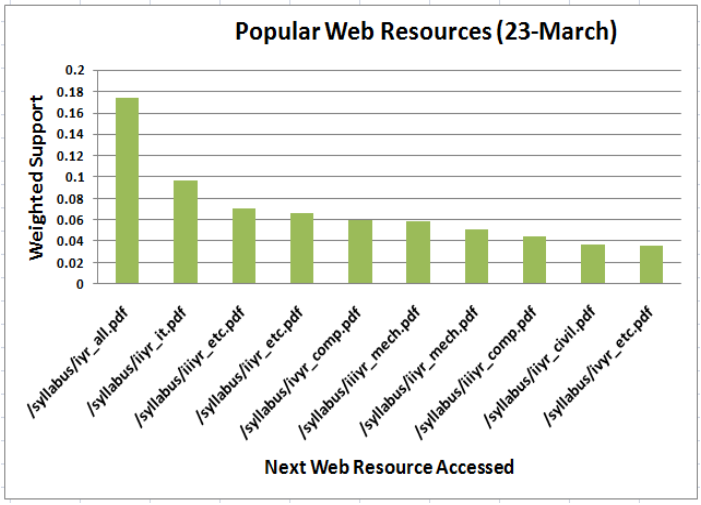


Fig 8: Graphs for Top 10 Popular Web Resource recommended after Syllabus Page of B.E. (23-Mar)

There are numerous user access patterns generated, so it is very difficult to show results of all user access patterns. Hence for five User Access Patterns, top five trending & popular web resources are shown in Table 9, 10 and 11.These results are shown for same three intermediate

dates of Experimental Data i.e. 12-Mar, 23-Mar & 5- April.

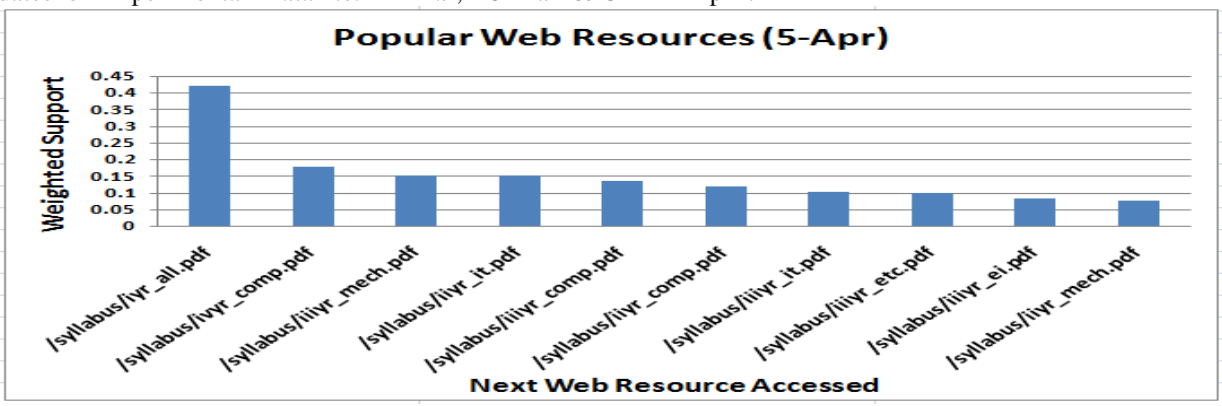


Fig 9: Graphs for Top 10 Popular Web Resource recommended after Syllabus Page of B.E. (5-Apr)

Table 9: Recommended Trending and Popular Web Resources for User Access Patterns (12-March)

TOP FIVE TRENDING AND POPULAR WEB RESOURCES ON 12TH MARCH FOR FOLLOWING USER ACCESS PATTERNS				
User Access Pattern	Trending web Resource	Weighted Support	Popular web Resource	Weighted Support
/ (Home Page)	Seminar- Next Generation Wireless N/W	0.07017544	Seminar- Next Generation Wireless N/W	0.10008997
	B.E. Class Time Table	0.029239766	B.E. Class Time Table	0.050607286
	Academic Programme (M.E.)	0.011695907	B.E. Schemes and Syllabus	0.050157446
	Exam Results-Dec 2011	0.011695907	Campus Recruitment Record - LG	0.042735044
	B.E. Schemes and Syllabus	0.011695907	Exam Results-Dec 2011	0.041610435
B.E. Class Time Table	B.E. Schemes and Syllabus	0.011695907	Academic Calendar (B.E. II,III,IV,PTDC)	0.012820513
	Time Table-B.E. IV Yr I.T.	0.005847953	B.E. Schemes and Syllabus	0.011695907
	Time Table-B.E. IV Yr C.S (Section A)	0.005847953	Time Table-B.E. IV Yr I.T.	0.010121457
	Fee Structure (B.E.)	0.005847953	Time Table-B.E. II Yr E&TC(Section A)	0.008547009
	Academic Calendar (B.E. I Yr)	0.005847953	Time Table-B.E. IV Yr C.S (Section A)	0.005847953
B.E. Schemes and Syllabus	B.E. I YR Schemes and Syllabus	0.03508772	B.E. I YR Schemes and Syllabus	0.06927575
	Syllabus- B.E. IV Yr C.S.	0.01754386	Syllabus- B.E. IV Yr C.S.	0.021817364
	Syllabus- B.E. II Yr C.S.	0.011695907	Syllabus- B.E. I Yr PTDC	0.01596941
	Syllabus- B.E. I Yr PTDC	0.011695907	Exam Results-Dec 2011	0.011695907
	Exam Results-Dec 2011	0.011695907	Syllabus- B.E. II Yr C.S.	0.011695907
/->B.E. Class Time Table	Time Table-B.E. IV Yr I.T.	0.005847953	Time Table-B.E. IV Yr I.T.	0.010121457
	Time Table-B.E. IV Yr C.S (Section A)	0.005847953	Time Table-B.E. IV Yr C.S (Section A)	0.005847953
	Academic Calendar (B.E. I Yr)	0.005847953	Academic Calendar (B.E. I Yr)	0.005847953
	B.E. Schemes and Syllabus	0.005847953	B.E. Schemes and Syllabus	0.005847953
			Time Table-B.E. II Yr IT (Section B)	0.004273505
/->B.E. Schemes and Syllabus	B.E. I YR Schemes and Syllabus	0.0116959065	B.E. I YR Schemes and Syllabus	0.028789924
			Syllabus – B.E. II Yr I.T.	0.008547009
			Syllabus- B.E. IV Yr C.S.	0.004273505

		Syllabus – B.E. III Yr Civil Engg	0.004273505
		Syllabus – B.E. II Yr Mech Engg	0.004273505

The User Access patterns that are evaluated in these tables are – **(i)** Home Page of IET-DAVV, **(ii)** Page for class time table of various branches, years and their sections of B.E., **(iii)** Schemes and Syllabus page for various years and branches of B.E., **(iv)** Home Page of IET-DAVV followed by class time table of various branches, years and their sections of B.E. (i.e. (i) followed by (ii)) and **(v)** Home Page of IET-DAVV followed by Schemes and Syllabus page for various years and branches of B.E.

According to Table 9, there are four Trending Web Resources and not five, after /->B.E. Class Time Table, and there is single Trending Web Resource after /->B.E. Schemes and Syllabus. It means that maximum links of resources that will be displayed is five and minimum links will depend upon count of traversed resources for that day. Results of trending and popular web resources 23rd March and 5th April are shown in Table 10 and 11 for same user access patterns.

Table 10: Recommended Trending and Popular Web Resources for User Access Patterns (23-March)

TOP FIVE TRENDING AND POPULAR WEB RESOURCES ON 23RD MARCH FOR FOLLOWING USER ACCESS PATTERNS				
User Access Pattern	Trending web resource	Weighted Support	Popular web resource	Weighted Support
/ (Home Page)	Seminar- Next Generation Wireless N/W	0.060200658	Seminar- Next Generation Wireless N/W	0.36759457
	B.E. Schemes and Syllabus	0.02341137	B.E. Class Time Table	0.31813484
	Academic Programme (M.E.)	0.020066889	B.E. Schemes and Syllabus	0.21249327
	B.E. Class Time Table	0.020066889	Academic Programme (M.E.)	0.21118206
	Exam Results-Dec 2011	0.020066889	Exam Results-Dec 2011	0.19393064
B.E. Class Time Table	Students Club	0.006688963	B.E. Schemes and Syllabus	0.04378016
	Time Table- B.E. II Yr C.S. (Section B)	0.006688963	Time Table- B.E. II Yr E&I	0.040380474
	Time Table- B.E. III Yr Mech Engg	0.003344482	Time Table- B.E. II Yr C.S. (Section B)	0.038453635
	Time Table- B.E. IV Yr E & TC	0.003344482	Time Table- IY Yr Mech Engg	0.026320385
	Time Table- B.E. IY Yr Mech Engg	0.003344482	Timetable- B.E. III Yr E&Tc (Section B)	0.02611121
B.E. Schemes and Syllabus	B.E. I YR Schemes and Syllabus	0.040133778	B.E. I YR Schemes and Syllabus	0.20576844
	Syllabus- B.E. IV Yr E & I	0.010033445	Syllabus – B.E. II Yr I.T.	0.0887162
	Syllabus- B.E II Yr E & I	0.010033445	Syllabus- B.E. III Yr Mech Engg	0.06350135
	B.E. Class Time Table	0.006688963	Syllabus- B.E. IV Yr C.S.	0.06297642
	Syllabus- B.E. III Yr Mech Engg	0.006688963	Syllabus- B.E. III Yr E & TC	0.062551826
/->B.E. Class Time Table	Recent Placement Updates	0.003344482	B.E. Schemes and Syllabus	0.01881082
	Time Table-B.E. II Yr PTDC (Mech Engg)	0.003344482	Time Table-B.E. IV Yr C.S. (Section B)	0.018069578
	Time Table-B.E. III Yr E& I	0.003344482	Timetable- B.E. II Yr E & I	0.01804087
	Time Table-B.E. III Yr C.S. (Section A)	0.003344482	Time Table-B.E. IV Yr I.T.	0.017116718
	Time Table-B.E. IV Yr C.S. (Section B)	0.003344482	Time Table- B.E. II Yr I.T.(Section A)	0.014715896
/->B.E.	B.E. I YR Schemes and Syllabus	0.006688963	B.E. I YR Schemes and Syllabus	0.06317177

Schemes and Syllabus	Syllabus – B.E. III Yr Civil Engg	0.003344482	Syllabus – B.E. II Yr I.T.	0.02086992
	Syllabus- B.E. III Yr Mech Engg	0.003344482	Syllabus-B.E. III Yr C.S.	0.016690122
	Syllabus- B.E. II Yr C.S.	0.003344482	Syllabus- B.E. II Yr C.S.	0.01563338
	Syllabus- B.E. II Yr E & I	0.003344482	Syllabus – B.E. III Yr Civil Engg	0.013047889

Table 11: Recommended Trending and Popular Web Resources for User Access Patterns (5-April)

TOP FIVE TRENDING AND POPULAR WEB RESOURCES ON 5TH APRIL FOR FOLLOWING USER ACCESS PATTERNS				
User Access Pattern	Trending web resource	Weighted Support	Popular web resource	Weighted Support
/ (Home Page)	B.E. Class Time Table	0.040816326	Seminar- Next Generation Wireless N/W	0.6893222
	Seminar- Next Generation Wireless N/W	0.025974026	B.E. Class Time Table	0.6633657
	Academic Programme (M.E.)	0.016697587	B.E. Schemes and Syllabus	0.4819691
	B.E. Schemes and Syllabus	0.012987013	Academic Programme (M.E.)	0.38796613
	Faculty Profiles Mech Engg Dept	0.009276438	Exam Results-Dec 2011	0.38527352
B.E. Class Time Table	Time Table – B.E. I Yr	0.005565863	Time Table – B.E. II Yr E & I	0.06917134
	Time Table - B.E. III Yr IT (Section B)	0.003710575	B.E. Schemes and Syllabus	0.068153895
	B.E. Schemes and Syllabus	0.003710575	Academic Calendar (B.E. II,III,IV,PTDC)	0.05228133
	Time Table-B.E. IY Yr Mech Engg	0.001855288	Time Table- B.E. II Yr I.T.(Section A)	0.051238332
	Time Table – B.E. I Yr PTDC Mech Engg	0.001855288	Time Table- B.E. II Yr C.S. (Section B)	0.049481142
B.E. Schemes and Syllabus	B.E. I YR Schemes and Syllabus	0.042671613	B.E. I YR Schemes and Syllabus	0.4197351
	Syllabus- B.E. II Yr C.S.	0.014842301	Syllabus- B.E. IV Yr C.S.	0.17810978
	Syllabus – B.E. II Yr I.T.	0.011131725	Syllabus- B.E. III Yr Mech Engg	0.15147012
	Syllabus – B.E. III Yr I.T.	0.009276438	Syllabus – B.E. II Yr I.T.	0.15142924
	Syllabus – B.E. IV Yr I.T.	0.009276438	Syllabus-B.E. III Yr C.S.	0.13497698
/-> B.E. Class Time Table	Time Table – B.E. I Yr	0.003710575	Time Table-B.E. II Yr E & I	0.03412733
	Time Table - B.E. III Yr IT (Section B)	0.001855288	Time Table-B.E. IV Yr C.S. (Section B)	0.026245655
	Academic Calendar (B.E. II,III,IV,PTDC)	0.001855288	B.E. Schemes and Syllabus	0.024562402
	Time Table-B.E. III Yr E& I	0.001855288	Academic Calendar (B.E. II,III,IV,PTDC)	0.023496225
	Time Table-B.E. III Yr C.S. (Section A)	0.001855288	Time Table- B.E. II Yr I.T.(Section A)	0.022273786
/->B.E. Schemes and Syllabus	Syllabus-B.E. III Yr E & TC	0.003710575	B.E. I YR Schemes and Syllabus	0.11235324
	Syllabus-B.E. II Yr I.T.	0.001855288	Syllabus-B.E. III Yr C.S.	0.04734108
	Syllabus-B.E. IV Yr C.S.	0.001855288	Syllabus -B.E. II Yr I.T.	0.045505866
	Syllabus-B.E. III Yr Mech Engg	0.001855288	Syllabus- B.E. IV Yr C.S.	0.040890794
	Syllabus-B.E. III Yr C.S.	0.001855288	Syllabus- B.E. II Yr C.S.	0.030928817

7. Conclusion and Future Work – The usability feature of trending and popular web resources is designed in this work, keeping in mind the generalized set of users. The key concept behind this feature is that as interest of users gets frequently changed with time in a website, therefore results of popular web resources derived from older logs may not fit according to current demand of

users. So, for different user access patterns, web logs are proposed on daily basis to provide "Trending web resources" according to current trend of access and "Popular web resources" i.e. overall popular resources, considering history as well as current data. Future work is to determine accuracy of prediction results for finding both trending as well as popular web resources. Proposed prediction algorithm can be also used for personalized users either on the basis of IP Address or User Id field. Moreover, concept of content mining i.e. semantic web can also be added into this work to obtain more accurate results so that only relevant links of next web resources(according to the content of web resource) will be provided to generalized as well as personalized users.

References

[1] A. Anitha: *A New Web Usage Mining Approach for Next Page Access Prediction.* International Journal of Computer Applications (0975 – 8887) Volume 8– No.11, (October 2010)

[2] Sartaj Ahmad,M.Z. Khan: *Improving Effectiveness of Online Teaching(An Application of Web Usage Mining).* International Journal of Computer Applications (0975 – 8887) Volume 8– No.13, (October 2010)

[3] Abhinav Srivastava, Abhijit Bhosale, and Shamik Sural: *Speeding Up Web Access Using Weighted Association Rules.* S.K. Pal et al. (Eds.): PReMI 2005, LNCS 3776, pp. 660–665, 2005. Springer-Verlag Berlin Heidelberg (2005)

[4] Navin Kumar Tyagi, A. K. Solanki and Manoj Wadhwa: *Analysis of Server Log by Web Usage Mining for Website Improvement.* IJCSI International Journal of Computer Science Issues, Vol 7, Issue 4, No 8, (July 2010).

[5] Bhawna Nigam, Suresh Jain and Sanjiv Tokekar: *Mining Association Rules from Web Logs by Incorporating Structural Knowledge of Website.* International Journal of Computer Applications (0975 – 8887) Volume 42– No.11 (March 2012)

[6] Nizar R. Mabroukeh and Christie I. Ezeife: *Using Domain Ontology for Semantic Web Usage Mining and Next Page Prediction*. CIKM'09, Hong Kong China, Copyright 2009 ACM. (November 2–6, 2009).

[7] A.Anitha and Dr.N.Krishnan: *A Dynamic Web Mining Framework for E-Learning Recommendations using Rough Sets and Association Rule Mining*. International Journal of Computer Applications (0975 – 8887) Volume 12– No.11, (January 2011)

[8] Amelia Zafra and Sebasti´an Ventura: *Web Usage Mining for Improving Students Performance in Learning Management Systems*. N. Garc´ıa-Pedrajas et al. (Eds.): IEA/AIE 2010, Part III, LNAI 6098, pp. 439–449, 2010.c_Springer-Verlag Berlin Heidelberg (2010)

[9] Mohamed Koutheaïr Khribi, Mohamed Jemni and Olfa Nasraoui: *Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval*. Educational Technology & Society, 12 (4), 30–42. (2009).

[10] Bamshad Mobasher, Robert Cooley and Jaideep Srivastava: *Automatic Personalization Based on Web Usage Mining.* Commun. ACM 43(8): 142-151 (2000)

[11] B. Nigam, S. Jain, "*Generating a New Model for Predicting the Next Accessed Web Page in Web Usage Mining*", in Proceedings of the 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET '10), Washington, DC, USA, pp.485-490, 2010.

[12] C. Umapathi, M. Aramuthan, and K. Raja: *Enhancing Web Services Using Predictive Caching.* International Journal of Research and Reviews in Information Sciences (IJRRIS) Vol. 1, No. 3 Science Academy Publisher, United Kingdom, www.sciacademypublisher.com, ISSN: 2046-6439(Sep 2011)

[13] Faten Khalil, Jiuyong Li and Hua Wang: A Framework of Combining Markov Model with Association Rules for Predicting Web Page Accesses. Proc. Fifth Australasian Data Mining Conference (AusDM2006)

[14] Debajyoti Mukhopadhyay, Priyanka Mishra, Dwaipayan Saha and Young-Chon Kim: A Dynamic Web Page Prediction Model Based on Access Patterns to *Offer Better User Latency.* The 6th International Workshop on Multimedia Signal Processing & Transmission; Jeonju, Korea (2006).

[15] Bhawna Nigam and Dr. Suresh Jain: *Analysis Of Markov Model On Different Web Prefetching And Caching Schemes.* International Conference on Computational Intelligence and Computing Research, Coimbatore (28-29 December, 2010).

[16] Sule Gunduz and M. Tamer Ozsu: *A Web Page Prediction Model Based on ClickStream Tree Representation of User Behaviour.* ACM 1581137370/03/0008(2008).

[17] Faten Khalil, Jiuyong Li and Hua Wang: *Integrating Recommendation Models for Improved Web Page Prediction Accuracy.* Thirty-First Australasian Computer Science Conference (ACSC 2008), Wollongong, Australia.

[18] Siriporn Chimphlee, Naomie Salim, Mohd Salihin, Bin Ngadiman and Witcha Chimphlee: *Using Association Rules and Markov Model for Predict Next Access on Web Usage Mining.* T. Sobh and K. Elleithy (eds.), Advances in Systems, Computing Sciences and Software Engineering, 371–376. Springer (2006)

[19] Ujwala Patil and Sachin Pardeshi: *A Survey on User Future Request Prediction: Web Usage Mining.* International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 2, Issue 3, March 2012)

[20] Sneha Y. S, G Mahadevan and Madhura Prakash: *A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web.* International Journal of Computer Theory and Engineering Vol. 4, No. 2 (April 2012)

[21] Natalie S. Glance, Matthew Hurst and Takashi Tomokiyo : *BlogPulse: Automated Trend Discovery for Weblogs.* ACM, New York, NY USA.(May 17–22, 2004)

[22] T.SubMasthan Rao, Y.Ravindra,U.Satish Kumar and S.Sandeep,K.Srikanth : *An Effective Framework For*

Identifying Personalized Web Recommender System By Applying Web Usage Mining. International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 2, Issue 3, pp. 307-312(May-June2012)