Association Rule Mining From Server Log File

[1]Kalpana Wani , [2]J.W.Bakal

PIIT New Panvel, Navi Mumbai, SSJCOE Dombivali, Mumbai
Email: [1]kjwani82@gmail.com, [2]bakaljw@gmail.com

Abstract— As we know World Wide Web plays vital role in serving the needs of the user's on web. The web log files are generated as a result of an interaction between the client and the service provider on web. Web log file contains the massive hidden valuable information pertaining to the visitors, if mined can be used for predicting the navigation behavior of the users. However the task of discovering frequent sequence patterns from the web log is challenging. This system focuses on adopting an intelligent technique that can provide personalized web service for accessing related web pages more efficiently and effectively, so that it can be determined which web pages are more likely to be accessed by the user in future. This system uses two intelligent algorithms for predicting the user behavior's namely FP Growth and Eclat.These algorithms save the time and space problem of existing system. Further from the frequent pages pattern Direct and Indirect Association Rules are generated and based on that Ranking is provided to pages which will help recommendation system to recommend similar search pages.

Index Terms— Association Rule, Indirect Association Rule, Parsing, Preprocessing.

I. INTRODUCTION

The World Wide Web serves as a vast, widely distributed, global information service center for advertisement, consumer information, e-commerce, education, financial management, government, news and many other information services. So, it has become much more difficult to access relevant information from the web with the explosive growth of information available on the internet. Therefore, further research work needs to be carried out for extracting the appropriate content as per the user's needs. This can be performed using Web mining as it helps in extracting the common sequences of the user's accessed web pages. In the web environment, association rules are typically applied to HTTP server log data that contain historical user sessions. Web sessions are gathered without any user involvement and additionally, they reliably reflect user behavior while navigating throughout a web site. For that reason, web sessions can be regarded as an important source of information about users. Association rules that reveal similarities between web pages derived from user behavior can be simply utilized in recommender systems. The main goal of such a recommendation is to suggest to the current user some web pages that appear to be useful.

II. LITERATURE SURVEY

In literature survey we discuss regarding various methods of Frequent Pattern generation and Mining direct and indirect association rule from weblog data.

A. Frequent Pattern Generation:

Bina Kotiyalt, Ankit Kumar, Bhaskar Pant, R.H. Goudar,Shiv ali Chauhan and Sonam Junee(2013)[1] proposed a web page recommendation system by finding frequent pattern using Apriori and Eclat algorithm and compare both the algorithm for generating frequent pattern.

Dr. S. Vijayarani1, Ms. P. Sathya(2013)[2] Compare the Éclat and Rapid Association Rule mining algorithm for finding the frequent item sets in data streams. Éclat and Rapid Association Rule mining algorithm are used for finding the frequent item sets in data streams.

N. Venkatesan, Ramaraj. (2012) [4] proposed the novel bit search technique which is implemented in the existing association rule mining algorithms. Frequent item sets are generated with the help of apriori & Eclat based bit search technique.

Karthikeswaran D, Dinakar S. (2010)[8] Proposed recursive algorithm uses this doubly Linked tree to efficiently find all access patterns that satisfy user specified criteria.

B. Association Rule Generation:

Ms Shweta1, Dr. Kanwal Garg (2013)[3] compares the Association rule algorithms to find out the best combination of different attributes in any data. In this paper author uses Apriori to find association rule. Here author consider three association rule algorithms: Apriori

Association Rule, Predictive Apriori Association Rule and Tertius Association Rule. Author compares the result of these three algorithms and presents the result. According to the result obtained using data mining tool author find that Apriori Association algorithm performs better than the Predictive Apriori Association Rule and Tertius Association Rule algorithms.

Rachna Somkunwar (2012) [6] proposed a new technique known as Hash-set technique that will help to increase the efficiency of algorithm by avoiding multiple scanning the database. Our Hash Techniques based algorithm is used to reduce the frequently scanning of item set also it helps to mine interesting association rules on the basis of lift and conviction technique.

PRZEMYSŁAW KAZIENKO (2009)[13] proposed algorithm to extracts complete indirect association rules with their important measure confidence, using pre-calculated direct rules. Both direct and indirect rules are joined into one set of complex association rules, which may be used for the recommendation of web pages. Performed experiments revealed the usefulness of indirect rules for the extension of a typical recommendation list. They also deliver new knowledge not available to direct ones. The relation between ranking lists created on the basis of direct association rules as well as hyperlinks existing on web pages is also examined.

III. PROPOSED SYSTEM

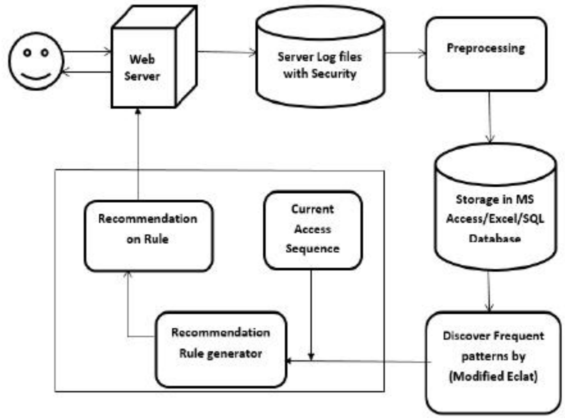

Web Server
Server Log files with Security
Preprocessing
Storage in MS Access/Excel/SQL Database
Recommendation on Rule
Current Access Sequence
Recommendation Rule generator
Discover Frequent patterns by (Modified Eclat)

Fig.1. Architecture Diagram of Proposed System

Following steps are used for implementation:

Step 1: Data Cleaning & Preprocessing

Preprocessing is necessary, because Log file contain noisy &ambiguous data which may affect result of mining process. Some of web log file data are unnecessary for analysis process and could affect detection of web attack. Data preprocessing is an important steps to filter and organize only appropriate information before applying any web mining algorithm. Preprocessing reduce log file size also increase quality of available data. The purpose of data preprocessing is to improve data quality and increase mining accuracy. Preprocessing consists of field extraction, data cleansing, user identification, session identification.

Step 2: Finding Frequent Pattern

First count the occurrences of pages in database and then reorder the pages with higher count page first and so on. Draw the FP Growth tree and start mining FP tree by considering the suffix as page with lowest count first. Find the conditional pattern base for each page and if length of conditional pattern base is less than some threshold value then find the frequent pattern using Éclat algorithm and if length is greater than or equal to the threshold value then use FP Growth algorithm for finding the frequent pattern of visited pages.

Step 3: Direct Association Rule Generation between the Frequent Access Pages

After getting the frequently visited page sequence filter the pattern by applying some minimum confidence and support value and form the one page association rule between pages of frequent pattern and calculate its support and confidence values and filter it with respect to minimum confidence and support value.

Step 4: Indirect Association Rule Generation between the Frequent Access Pages

In direct association rule of pages there are many pages which are connected to each other indirectly via some transitive pages. Find such pages which are indirectly connected and filter it with respect to minimum confidence and support value.

Step 5: Complex Association Rule Generation

A complex association rule exists if at least one of two component rules exists, i.e., either direct or complete indirect or both of them. The main quality features of both direct and in indirect rules—confidences—are combined within complex association rules. A complex association rule is characterized by the complex confidence, $con*(d_i \to * d_j)$, as follows

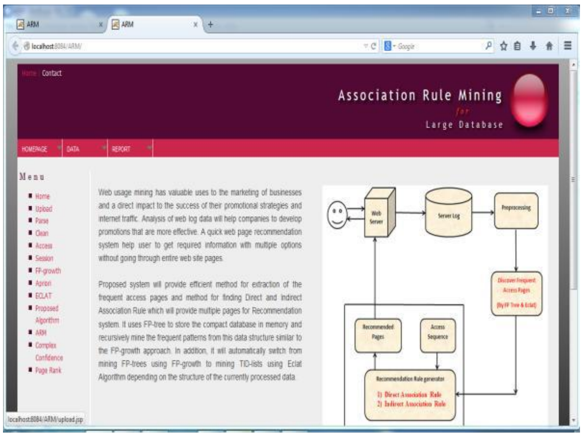
$$Con*(di \to *dj) = \alpha\ .Con(di \to dj)+(1 - \alpha)\ .\ Con\#(di \to \#dj)$$

Where α is the direct confidence reinforcing factor $\alpha \in [0, 1]$

Calculate the complex confidence value for different values of α. Setting α we can emphasize or damp the direct confidence at the expense of the complete indirect one. The greater the value of α, the closer the complex confidence to the direct one. Depending on highest value of α rank is

IV. SYSTEM IMPLEMENTATION

In our system instead of using any readymade tool for data preprocessing we have developed our own processing method which process web log data present in W3c format i.e. in Extended Log Format. Home page of proposed system is shown below. It consist of the display of the Architecture Diagram of the system and a brief explanation of the working of the system.



Web usage mining has valuable uses to the marketing of businesses and a direct impact to the success of their promotional strategies and internet traffic. Analysis of web log data will help companies to develop promotions that are more effective. A quick web page recommendation system help user to get required information with multiple options without going through entire web site pages.

Proposed system will provide efficient method for extraction of the frequent access pages and method for finding Direct and Indirect Association Rule which will provide multiple pages for Recommendation system. It uses FP-tree to store the compact database in memory and recursively mine the frequent patterns from this data structure similar to the FP-growth approach. In addition, it will automatically switch from mining FP-trees using FP-growth to mining TID-lists using Eclat Algorithm depending on the structure of the currently processed data.

Fig.2 Home Page of Proposed System

On left hand side of the home page Menu options are given which are listed below

1) Home
2) Upload
3) Parse
4) Clean
5) Access
6) Session
7) FP-Growth
8) Apriori
9) Éclat
10) Proposed Algorithm(FEM)
11) ARM
12) Complex Confidence
13) Page rank

Upload

Second option is Upload, which facilitates to upload the weblog file in w3c format. We can browse the file from its location and click on upload button as shown below.

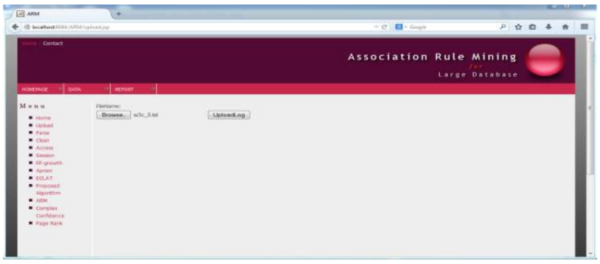
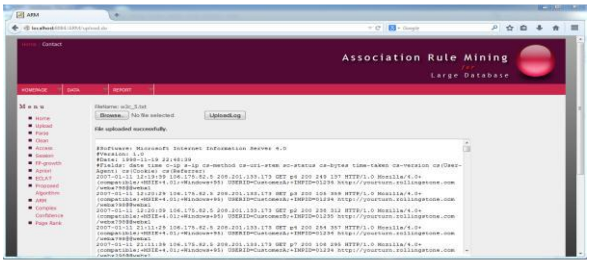


Fig.3 Web Log File Upload

After clicking on the upload button if file is uploaded successfully then message will be displayed about successful uploading of file and file contents are displayed as shown below. If file type is not matching then it display the error.



File uploaded successfully.

```
#Software: Microsoft Internet Information Server 4.0
#Version: 1.0
#Date: 1998-11-19 22:48:39
#Fields: date time c-ip s-ip cs-method cs-uri-stem sc-status cs-bytes time-taken cs-version cs(User-
Agent) cs(Cookie) cs(Referrer)
2007-01-11 12:19:39 106.175.82.5 208.201.133.173 GET p4 200 249 137 HTTP/1.0 Mozilla/4.0+
(compatible;+MSIE+4.01;+Windows+95) USERID=CustomerA;+IMPID=01234 http://yourturn.rollingstone.com
/webx7988@webx1
2007-01-11 12:20:29 106.175.82.5 208.201.133.173 GET p3 200 105 359 HTTP/1.0 Mozilla/4.0+
(compatible;+MSIE+4.01;+Windows+95) USERID=CustomerA;+IMPID=01234 http://yourturn.rollingstone.com
/webx7988@webx1
2007-01-11 12:20:59 106.175.82.5 208.201.133.173 GET p2 200 238 312 HTTP/1.0 Mozilla/4.0+
(compatible;+MSIE+4.01;+Windows+95) USERID=CustomerB;+IMPID=01235 http://yourturn.rollingstone.com
/webx7988@webx1
2007-01-11 21:11:29 106.175.82.5 208.201.133.173 GET p4 200 254 357 HTTP/1.0 Mozilla/4.0+
(compatible;+MSIE+4.01;+Windows+95) USERID=CustomerA;+IMPID=01234 http://yourturn.rollingstone.com
/webx7988@webx1
2007-01-11 21:11:39 106.175.82.5 208.201.133.173 GET p7 200 108 295 HTTP/1.0 Mozilla/4.0+
(compatible;+MSIE+4.01;+Windows+95) USERID=CustomerA;+IMPID=01234 http://yourturn.rollingstone.com
/webx7988@webx1
```

Fig.4 Web Log File ContentsParse

Once weblog file is uploaded successfully then we can start preprocessing of the file by selecting the Parse option from the Menu. We have to use Parse option for uploaded file only, if we try to use it without uploading a file it will give error. Following snapshot show that when we select Parse option for uploaded file it automatically display the file name.

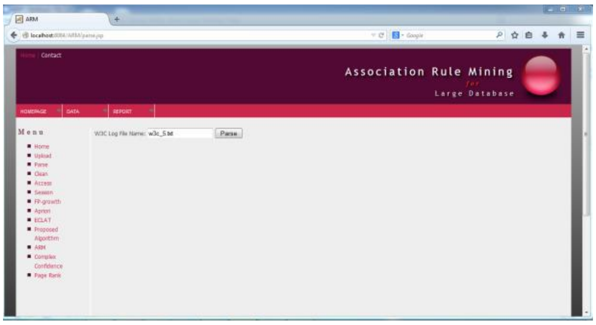
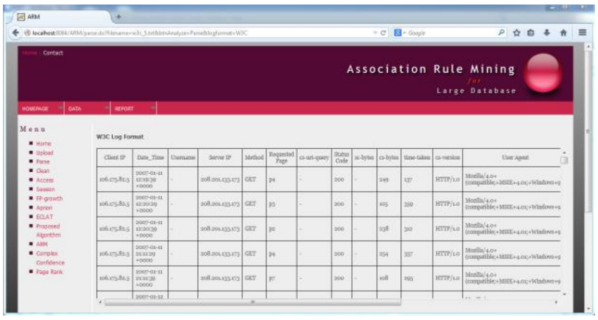


Fig.5 Web Log File Selection for Parsing of Data

After getting required file name click on parse button which will convert W3C format weblog file information in a tabular format as shown below.



W3C Log Format

Client IP	Date_Time	Username	Server IP	Method	Requested Page	cs-uri-query	Status Code	sc-bytes	cs-bytes	time-taken	cs-version	User Agent
106.175.82.5	2007-01-11 12:19:39 +0000	-	208.201.133.173	GET	p4	-	200	-	249	137	HTTP/1.0	Mozilla/4.0+ (compatible;+MSIE+4.01;+Windows+9
106.175.82.5	2007-01-11 12:20:29 +0000	-	208.201.133.173	GET	p3	-	200	-	105	359	HTTP/1.0	Mozilla/4.0+ (compatible;+MSIE+4.01;+Windows+9
106.175.82.5	2007-01-11 12:20:59 +0000	-	208.201.133.173	GET	p2	-	200	-	238	312	HTTP/1.0	Mozilla/4.0+ (compatible;+MSIE+4.01;+Windows+9
106.175.82.5	2007-01-11 21:11:29 +0000	-	208.201.133.173	GET	p4	-	200	-	254	357	HTTP/1.0	Mozilla/4.0+ (compatible;+MSIE+4.01;+Windows+9
106.175.82.5	2007-01-11 21:11:39 +0000	-	208.201.133.173	GET	p7	-	200	-	108	295	HTTP/1.0	Mozilla/4.0+ (compatible;+MSIE+4.01;+Windows+9
	2007-01-12											

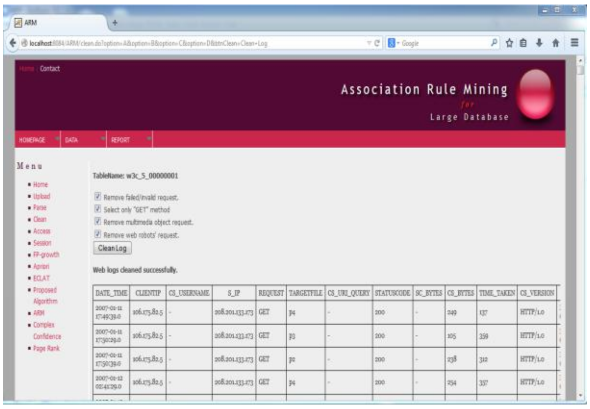
Fig.6 Tabular Representation of Web Log file data

Clean

After converting web log data into a tabular format we need to clean the data to know accurate information. In web log data many entries are not completed or sometimes we want information about some specific data format etc. Depending upon our requirement we can clean the data and retrieve the information. Different cleaning option provided in this system are given below.

a) Remove failed/invalid request.
b) Select only "GET" method
c) Remove multimedia object request.
d) Remove web 'Robots' request.

We can select either one or multiple option depending upon our data requirements as shown below.



TableName: w3c_5_00000001

☑ Remove failed/invalid request.
☑ Select only "GET" method
☑ Remove multimedia object request.
☑ Remove web robots' request.

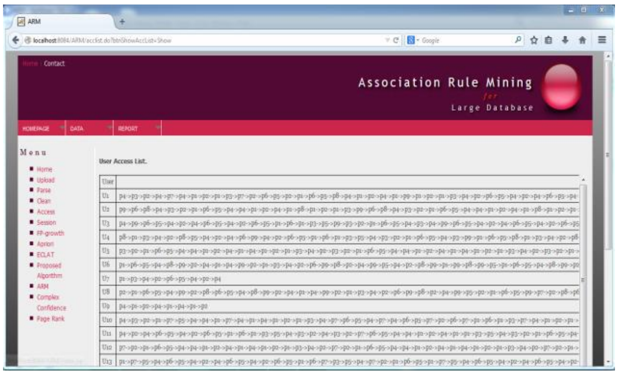
Clean Log

Web logs cleaned successfully.

DATE_TIME	CLIENTIP	CS_USERNAME	S_IP	REQUEST	TARGETFILE	CS_URI_QUERY	STATUSCODE	SC_BYTES	CS_BYTES	TIME_TAKEN	CS_VERSION
2007-01-11 17:49:39.0	106.175.82.5	-	208.201.133.173	GET	p4	-	200	-	249	137	HTTP/1.0
2007-01-11 17:50:29.0	106.175.82.5	-	208.201.133.173	GET	p3	-	200	-	105	359	HTTP/1.0
2007-01-11 17:50:39.0	106.175.82.5	-	208.201.133.173	GET	p2	-	200	-	238	312	HTTP/1.0
2007-01-12 02:41:29.0	106.175.82.5	-	208.201.133.173	GET	p4	-	200	-	254	357	HTTP/1.0

Fig.7 Web Log Data after cleaning Access

To know the user access pages list we can use this option. It will display list of users as well as their access pages as shown below.

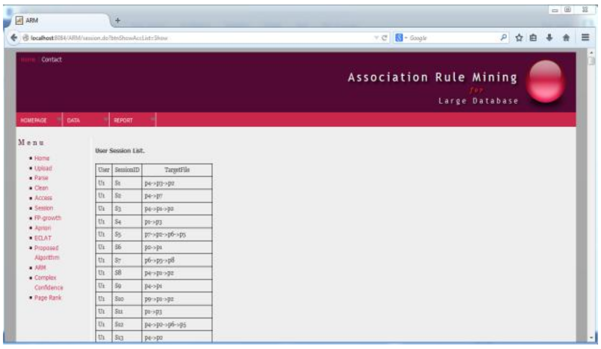


User Access List.

User	
U1	p4->p3->p2->p4->p7->p4->p1->p2->p1->p3->p7->p2->p6->p5->p2->p1->p6->p5->p8->p4->p1->p2->p4->p1->p9->p1->p2->p1->p3->p4->p2->p6->p5->p4->p2->p4->p6->p5->p4
U2	p9->p6->p8->p4->p3->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p8->p1->p2->p1->p3->p9->p6->p8->p4->p3->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p8->p1->p2->p1
U3	p4->p9->p6->p5->p4->p2->p4->p6->p5->p4->p2->p6->p5->p1->p6->p1->p3->p5->p9->p3->p2->p1->p6->p5->p4->p3->p4->p9->p6->p5->p4->p2->p4->p6->p5->p4->p2->p6->p5
U4	p8->p1->p3->p4->p2->p8->p5->p4->p2->p4->p6->p9->p4->p2->p6->p5->p1->p6->p1->p3->p5->p4->p3->p2->p1->p6->p5->p4->p3->p9->p1->p6->p5->p8->p1->p3->p4->p2->p8
U5	p3->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p4->p1->p2->p1->p3->p4->p2->p3->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p4->p1->p2->p1->p3->p4->p2->p3->p2->p1
U6	p1->p6->p5->p4->p8->p9->p2->p4->p1->p4->p9->p2->p1->p3->p4->p2->p6->p9->p8->p2->p4->p9->p5->p4->p2->p8->p9->p1->p2->p8->p9->p5->p1->p6->p5->p4->p8->p9->p2
U7	p1->p3->p4->p2->p6->p5->p4->p2->p4
U8	p2->p1->p6->p5->p4->p9->p2->p8->p6->p5->p4->p8->p9->p2->p4->p1->p4->p9->p2->p1->p3->p4->p2->p6->p9->p8->p2->p4->p9->p5->p2->p1->p6->p5->p9->p7->p2->p8->p6
U9	p4->p1->p2->p4->p1->p4->p1->p2
U10	p4->p3->p2->p1->p7->p5->p4->p4->p1->p7->p4->p1->p4->p1->p2->p1->p3->p4->p7->p6->p5->p4->p7->p4->p6->p5->p7->p2->p6->p7->p1->p6->p1->p3->p7->p4->p1->p2->p1
U11	p4->p2->p4->p6->p5->p4->p2->p6->p5->p1->p6->p1->p3->p5->p4->p3->p2->p4->p3->p2->p7->p6->p5->p4->p4->p1->p2->p4->p1->p1->p3->p5->p4->p3->p2->p1->p6->p5->p4
U12	p7->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p4->p1->p2->p1->p3->p4->p2->p7->p2->p1->p6->p5->p4->p4->p1->p2->p4->p1->p4->p1->p2->p1->p3->p4->p2->p7->p2->p1
U13	p1->p7->p5->p4->p6->p5->p4->p2->p4->p6->p5->p4->p2->p6->p5->p1->p6->p7->p3->p5->p4->p7->p2->p1->p6->p5->p1->p7->p5->p4->p6->p5->p4->p2->p4->p6->p5->p4->p2

Fig 8. User Access List Session

This option gives the details of visited users their sessions and the visited pages in respective sessions as shown in figure.



User Session List.

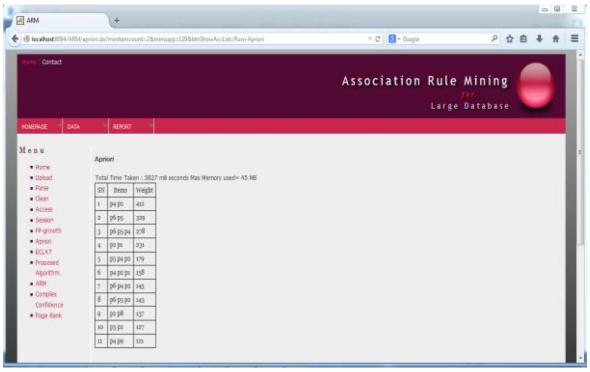
User	SessionID	TargetFile
U1	S1	p4->p3->p2
U1	S2	p4->p7
U1	S3	p4->p1->p2
U1	S4	p1->p3
U1	S5	p7->p2->p6->p5
U1	S6	p2->p1
U1	S7	p6->p5->p8
U1	S8	p4->p1->p2
U1	S9	p4->p1
U1	S10	p9->p1->p2
U1	S11	p1->p3
U1	S12	p4->p2->p6->p5
U1	S13	p4->p2

Fig 9 User Session

Next we can execute any of the given algorithms to find frequent patterns from clean weblog data. Apriori, Éclat and FP-Growth algorithms implementation is not a part of proposed system. Only implementation of proposed algorithm is a part of this project and only for proposed FEM algorithm Direct and Indirect Association rules are generated. Apriori, Éclat are used in Existing system to generate frequent patterns. Apriori, Éclat and FP-Growth algorithms are implemented to compare performance of proposed system with respect to existing system.

Following snapshot shows the execution of all above algorithm for support value 20 except Apriori as it gives memory problem for low support value.

Apriori Algorithm



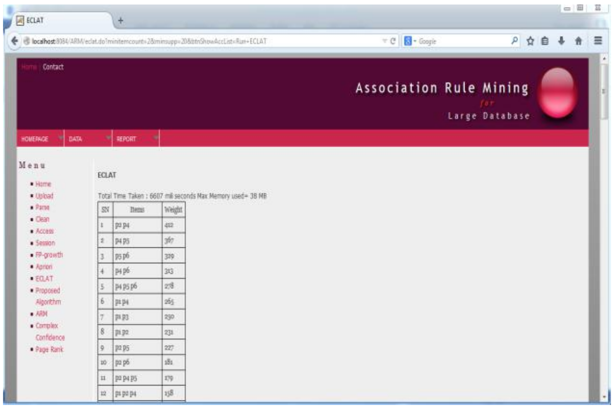
Apriori

Total Time Taken : 5827 mili seconds Max Memory used= 45 MB

SN	Items	Weight
1	p4 p2	412
2	p6 p5	329
3	p6 p5 p4	278
4	p2 p1	231
5	p5 p4 p2	179
6	p4 p2 p1	158
7	p6 p4 p2	145
8	p6 p5 p2	143
9	p2 p8	137
10	p3 p2	127
11	p4 p9	121

Fig .10 Apriori Algorithm

Éclat Algorithm



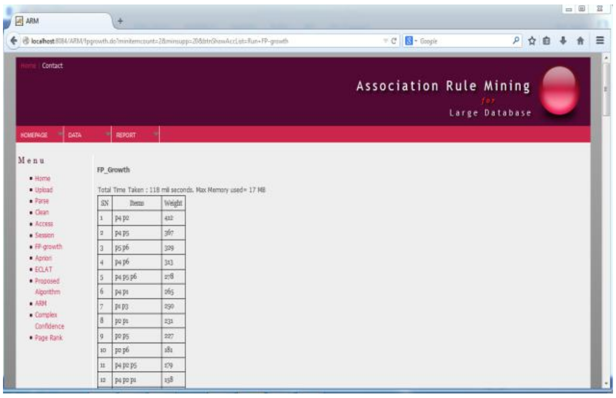
ECLAT

Total Time Taken : 6607 mili seconds Max Memory used= 38 MB

SN	Items	Weight
1	p2 p4	412
2	p4 p5	367
3	p5 p6	329
4	p4 p6	313
5	p4 p5 p6	278
6	p1 p4	265
7	p1 p3	250
8	p1 p2	231
9	p2 p5	227
10	p2 p6	181
11	p2 p4 p5	179
12	p1 p2 p4	158

Fig. 11 Éclat Algorithm

FP-Growth Algorithm



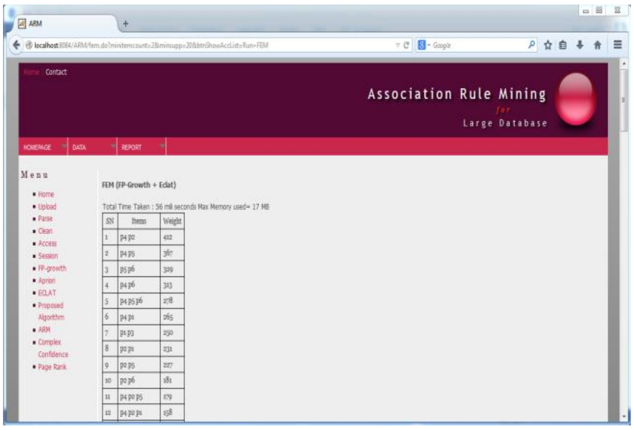
FP_Growth

Total Time Taken : 118 mili seconds. Max Memory used= 17 MB

SN	Items	Weight
1	p4 p2	412
2	p4 p5	367
3	p5 p6	329
4	p4 p6	313
5	p4 p5 p6	278
6	p4 p1	265
7	p1 p3	250
8	p2 p1	231
9	p2 p5	227
10	p2 p6	181
11	p4 p2 p5	179
12	p4 p2 p1	158

Fig .12 FP-Growth Algorithm

Proposed FEM Algorithm



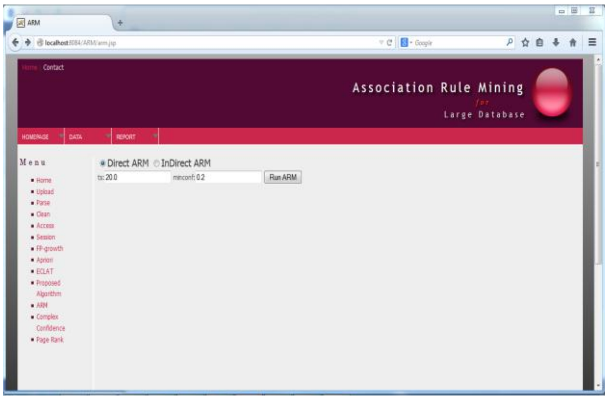
FEM (FP-Growth + Eclat)

Total Time Taken : 56 mili seconds Max Memory used= 17 MB

SN	Items	Weight
1	p4 p2	412
2	p4 p5	367
3	p5 p6	329
4	p4 p6	313
5	p4 p5 p6	278
6	p4 p1	265
7	p1 p3	250
8	p2 p1	231
9	p2 p5	227
10	p2 p6	181
11	p4 p2 p5	179
12	p4 p2 p1	158

Fig 13. Proposed FEM Algorithm

After executing FEM algorithm we can find direct and indirect Association rule snapshot for that given below. Depending on your selection it generate direct and indirect association rule as shown below.

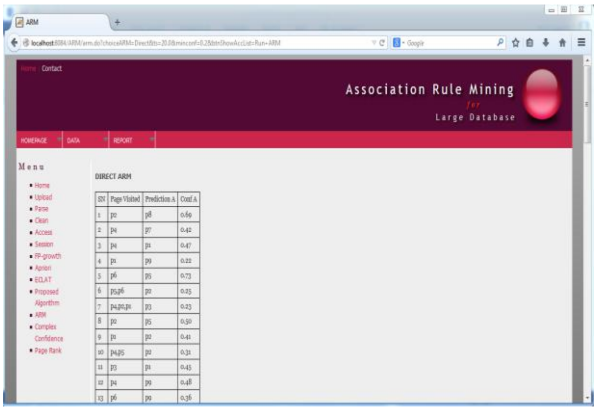


● Direct ARM ○ InDirect ARM

ts: 20.0 minconf: 0.2

Fig..14 Generation of Direct and Indirect Association Rule

Direct Association Rule

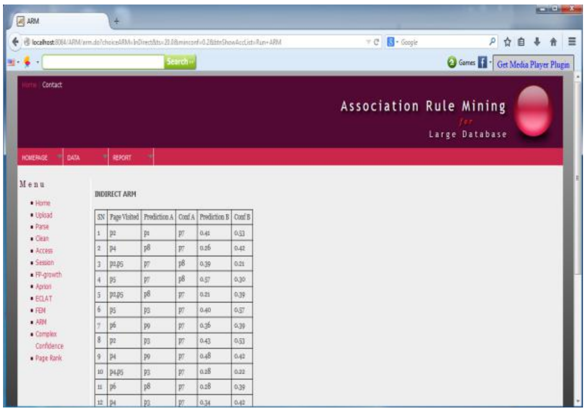


DIRECT ARM

SN	Page Visited	Prediction A	Conf A
1	p2	p8	0.69
2	p4	p7	0.42
3	p4	p1	0.47
4	p1	p9	0.22
5	p6	p5	0.73
6	p5,p6	p2	0.25
7	p4,p2,p1	p3	0.23
8	p2	p5	0.50
9	p1	p2	0.41
10	p4,p5	p2	0.31
11	p3	p1	0.45
12	p4	p9	0.48
13	p6	p9	0.36

Fig..15 Generation of Direct Association Rule

Indirect Association Rule

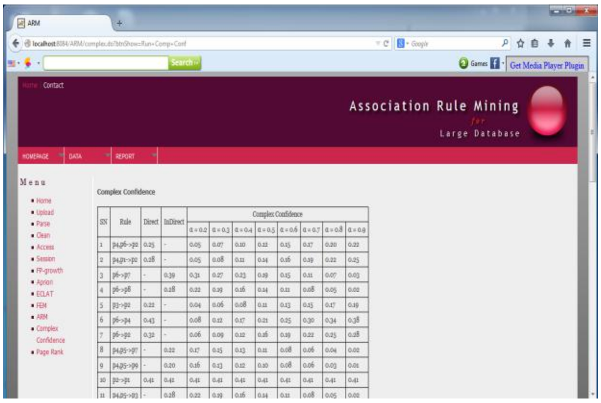


INDIRECT ARM

SN	Page Visited	Prediction A	Conf A	Prediction B	Conf B
1	p2	p1	p7	0.41	0.53
2	p4	p8	p7	0.26	0.42
3	p2,p5	p7	p8	0.39	0.21
4	p5	p7	p8	0.57	0.30
5	p2,p5	p8	p7	0.21	0.39
6	p5	p3	p7	0.40	0.57
7	p6	p9	p7	0.36	0.39
8	p2	p3	p7	0.43	0.53
9	p4	p9	p7	0.48	0.42
10	p4,p5	p3	p7	0.28	0.22
11	p6	p8	p7	0.28	0.39
12	p4	p3	p7	0.34	0.42

Fig..16 Generation of Indirect Association Rule

Complex Confidence

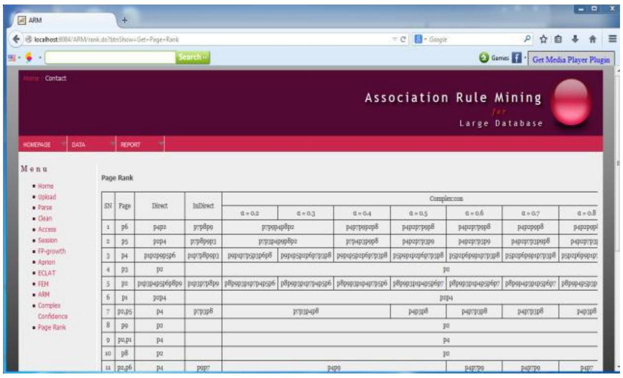


Complex Confidence

SN	Rule	Direct	InDirect	Complex Confidence							
				α = 0.2	α = 0.3	α = 0.4	α = 0.5	α = 0.6	α = 0.7	α = 0.8	α = 0.9
1	p4,p6->p2	0.25	-	0.05	0.07	0.10	0.12	0.15	0.17	0.20	0.22
2	p4,p1->p2	0.28	-	0.05	0.08	0.11	0.14	0.16	0.19	0.22	0.25
3	p6->p7	-	0.39	0.31	0.27	0.23	0.19	0.15	0.11	0.07	0.03
4	p6->p8	-	0.28	0.22	0.19	0.16	0.14	0.11	0.08	0.05	0.02
5	p3->p2	0.22	-	0.04	0.06	0.08	0.11	0.13	0.15	0.17	0.19
6	p6->p4	0.43	-	0.08	0.12	0.17	0.21	0.25	0.30	0.34	0.38
7	p6->p2	0.32	-	0.06	0.09	0.12	0.16	0.19	0.22	0.25	0.28
8	p4,p5->p7	-	0.22	0.17	0.15	0.13	0.11	0.08	0.06	0.04	0.02
9	p4,p5->p9	-	0.20	0.16	0.13	0.12	0.10	0.08	0.06	0.03	0.01
10	p2->p1	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
11	p4,p5->p3	-	0.28	0.22	0.19	0.16	0.14	0.11	0.08	0.05	0.02

Fig..17 Generation of Complex Confidence

Page Rank



Page Rank

SN	Page	Direct	InDirect	Complex:con						
				α = 0.2	α = 0.3	α = 0.4	α = 0.5	α = 0.6	α = 0.7	α = 0.8
1	p6	p4p2	p7p8p9	p7p9p4p8p2		p4p7p9p2p8	p4p2p7p9p8	p4p2p7p9p8	p4p2p9p8	p4p2p9p
2	p5	p2p4	p7p8p9p3	p7p3p4p9p8p2		p7p4p3p9p8	p4p2p7p3p9	p4p2p7p3p9	p4p2p7p3p9p8	p4p2p7p3
3	p4	p1p2p9p5p6	p1p7p8p9p3	p9p1p7p5p3p6p8	p9p1p5p2p6p7p3p8	p9p1p5p2p6p7p3p8	p5p9p1p2p6p7p3p8	p5p2p6p9p1p7p3p8	p5p2p6p9p1p7p3p8	p5p2p6p9p1
4	p3	p2		p2						
5	p2	p1p3p4p5p6p8p9	p1p3p7p8p9	p8p9p3p1p7p4p5p6	p8p9p3p1p7p4p5p6	p8p9p3p1p4p7p5p6	p8p9p3p1p4p5p6p7	p8p9p3p1p4p5p6p7	p8p9p4p3p1p5p6p7	p8p9p4p5p3p
6	p1	p2p4		p2p4						
7	p2,p5	p4	p7p3p8	p7p3p4p8			p4p3p8	p4p7p3p8	p4p7p3p8	p4p3p8
8	p9	p2		p2						
9	p2,p1	p4		p4						
10	p8	p2		p2						
11	p2,p6	p4	p9p7	p4p9				p4p7p9	p4p7p9	p4p7

Fig.18 Page ranking based on Complex Confidence

IV. CONCLUSION

In today's world there is boom of online shopping and because of that if companies want to increase their business it's required to analyze the user behavior. This is possible only with help of web mining and its further processing to find frequent pattern and the association rule between them. This system will help to recommend appropriate page to user.

In future we are planning to complete remaining work by providing security to log file

V. ACKNOWLEDGMENT

I would like to thank my guide Dr.J.W. Bakal to provide a valuable guidance and inspiring me to publishing paper in International Journals and Conferences.

REFERENCES

[1] Bina Kotiyalt, Ankit Kumar, Bhaskar Pant, R.H. Goudar, Shiv ali Chauhan and Sonam Junee" User Behavior Analysis in Web Log through Comparative Study of Eclat and Apriori", Proceedings of7'h International Conference on Intelligent Systems and Control (ISCO 2013)

[2] Dr. S. Vijayarani, Ms. P. Sathya "An Efficient Algorithm for Mining Frequent Items in Data Streams",in International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 3, May 2013

[3] Ms Shweta, Dr. Kanwal Garg ,"Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms"in Volume 3, Issue 6, June 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

[4] N. VENKATESAN, RAMARAJ," FREQUENT ITEMSET MINING WITH BIT SEARCH", Journal of Theoretical and Applied Information Technology, 15 July 2012. Vol. 41 No.1

[5] Ravi Bhushan and Rajender Nath," Recommendation of Optimized Web Pages to Users Using Web Log Mining Techniques", 978-1-4673-4529-3/12/$31.00c 2012 IEEE

[6] Rachna Somkunwar" Hash-Set Technique of Association Rules"in Volume 2, Issue 11, November 2012 ISSN: 2277 128X ,International Journal of Advanced Research in Computer Science and Software Engineering.

[7] Chowdary Farha ahmed, Byeong-Soo Jeong, "Efficient mining of high utility patterns over data streams with a sliding window model".Springerlink.com, 2011

[8] Yo unghee Kim, Won Young Kim and Ungmo Kim "Mining frequent item sets with normalized weight in continuous data streams". Journal of information processing systems. 2010.

[9] L. Zhou, Z. Zhong, J. Chang, J. Li, J.Z. Huang, S. Feng,"Balanced parallel FP-Growth with MapReduce", in Conference on Information Computing and Telecommunications, IEEE, pp. 243 – 246, 2010.

[10] Ding, Yau, S. S.: TCOM, "An innovative data structure for mining association rules among infrequent items", pp. 290-301. Comput. Math. Appl. 57, 2 (2009).

[11] Ding, 1., Yau, S. S.: TCOM, an innovative data structure for mining association rules among infrequent items, pp. 290-301. Comput. Math. Appl. 57, 2 (2009).

[12] Sang Lin ,Hu-yan Cui ,Ren Ying ,Zhou-lin Lin ," Algorithm Research for Mining Maximal Frequent Itemsets Based on Item Constraints", 2009 Second International Symposium on Information Science and Engineering, ISBN: 978-0-7695-3991-

[13] PRZEMYSŁAW KAZIENKO "MINING INDIRECT ASSOCIATION RULES FOR WEB RECOMMENDATION", Int. J. Appl. Math. Comput. Sci., 2009, Vol. 19, No. 1, 165–186

◈◈◈