

# WEB USAGE MINING USING FP GROWTH ALGORITHM: E-LEARNING PORTAL

S.Iswarya

*J.J College of Arts and Science, Pudukkottai.*

*Department of Computer Science*

*simbuaishwarya15@gmail.com Science*

S.Sudha

*Assistant Professor and Head*

*Department of Computer Science*

*J.J College of Arts and Science(Autonomous)*

*Pudukkottai 622 422*

## Abstract

Web usage mining is the part of web mining. The information is collected as result in horrendously huge data in web. The information is gathered by the quantity of guests in the internet searcher. Utilization information catches the personality or root of Web clients alongside their perusing conduct at a Website. Web use mining itself can be arranged further contingent upon the sort of utilization information considered. They are web worker information, application worker information and application level information. Web worker information compare to the client logs that are gathered at Web worker. A portion of the run of the mill information gathered at a Web worker incorporate IP addresses, page references, and access season of the clients and is the principle contribution to the current explores. E- Learning is one of the Web based application where it will look with huge measure of information. To create the E-Learning entrance utilization examples and client practices, this paper executes the elevated level cycle of Web Usage mining utilizing essential Association Rules calculation known as FP development Algorithm.

**Keywords--:** *Web Usage Mining, Web server logs, Data Preprocessing, Pattern discovery, FP growth.*

## INTRODUCTION

Web use mining is the utilization of information mining methods to find intriguing use designs from web information, to comprehend and better serve the necessities of Web-based applications. Utilization information embodies the character or root of web clients. Web use mining itself can be ordered further contingent upon the sort of utilization information considered. They are web worker information, application worker information and application level information. Web worker information compare to the client logs that are gathered at Web worker. The

urgent data extricated is found with the utilization of affiliation rules about clients' practices. The data gathered includes IP addresses, page references, and access season of the clients. This work focuses on web use mining and specifically centers around finding the web use examples of sites from the worker log documents. Web utilization mining is the use of information mining that apply information mining procedures to find the personal conduct standard utilizing web information. Web use mining measure is by and large separated into three undertakings: preprocessing, design investigation and example revelation. Preprocessing incorporates the combination and synchronization distinguishing proof, client ID, meeting ID (or sessionization), scene ID, and the reconciliation of snap stream information with request information sources, for example, content or semantic data. In the example examination stage fascinating information is separated from regular examples and these outcomes are utilized for site changes. In pattern discovery phase, frequent pattern discovery

algorithm are applied on crude information. For discovering the data that is covered up in web logs, a few information mining procedures are applied on web worker logs. A similar report between the two affiliation rule calculations to be specific, FP Growth algorithm and web log is shown in this work.

It very well may be said that, the World Wide Web is today the most significant media for gathering, sharing and appropriating data. Advanced education (AE) is one of the Departments where online innovation has been rapidly and effectively embraced. The incredible proposition of online courses that, these days, is offered by colleges is one proof of that. Considerably more, totally virtual colleges are showing up. Overseeing and following understudies, planning courses, making assessments, and so on requires explicit frameworks which are called Learning Management Systems (LMS).

CMS encourage web conveyance and the executives for educator drove themes and incorporate conferencing frameworks, surveying and test modules, virtual workspaces and different devices for estimating results and announcing progress for individual or gathering of understudies. They will in general be extremely printed and format arranged to give convenience, however restricting adaptability. They ordinarily add solid coordinated creating devices and parts to associate with information base frameworks. Numerous CMS have been created and are being used the world over. Anyway they don't uphold apparatuses which permit teachers to altogether follow and survey all the exercises performed by all students, nor to assess the structure of the course substance and its adequacy on the learning cycle. Truth be told, these conditions furnish the instructor with access rundown data, for example, more visited pages, most loved specialized strategy, and different insights. By and by this data isn't enough to analyze the behavior of each student and his evolution.

The web is no more a term that needs to be looked into the dictionary for its meaning. It is such a variant, vast and dynamic data repository that comprises of mostly raw data which is a source to the enormous supply of information. The

information excavated from this repository is watched out by, users, web service providers, business analysts, thus making it even complex to be dealt with. The web users hence, want to have the effective search tools to find relevant information easily and precisely. Data mining is the process of excavation for finding out knowledge from data. Web mining is the process of excavating information and patterns from web. It is used to understand customer behavior, evaluate the effectiveness of a particular web site, and help quantify the success of a marketing campaign. It also allows looking for patterns in data through content mining, structure mining, and usage mining. web usage mining is applied to many real world problems to discover interesting user navigation patterns for improvement of web site design by making additional topic or recommendations observing user or customer behavior. There are three methods which are applicable for web mining

- 1) Web content mining
- 2) Web structure mining
- 3) Web usage mining
- 4) Data mining (DM) is a step from Knowledge Discovery in Database (KDD) process, which is defined as a "nontrivial process of identifying valid, novel, potentially useful and ultimately understandable pattern in data". The term pattern here refers some abstract representation of a subset data of the data, that is, an expression in some language describing a data subset or a data subset or a model applicable to that subset.
- 5) Data mining efforts associated with the Web, called Web mining, can be broadly categorized into three areas of interest based on which part of the Web to mine; Web Content mining, Web Structure mining, and Web Usage Mining. In Web mining, data can be collected at the server-side, client-side, proxy servers or a consolidated Web/business database. The information provided by the data sources described above can be used to construct several data abstractions, namely users, page-views, click-streams and server

sessions

Web Usage Mining is defined as the process of applying data mining techniques to the discovery of usage patterns from Web logs data which to identify Web user's behavior. Web Usage Mining is the type of Web mining activity that involves an automatic discovery of user access patterns from one or more Web servers.

Web usage mining process is generally divided into three tasks:

- Preprocessing,
- Pattern analysis
- Pattern discovery.

Preprocessing includes the fusion and synchronization identification, user identification, session identification (or sessionization), episode identification, and the integration of click stream data with order data sources such as content or semantic information. The uploaded data will be fragment and encrypt.

In the pattern analysis phase interesting knowledge is extracted from frequent patterns and these results are used for website modifications.

In pattern discovery phase, frequent pattern discovery algorithms are applied on raw data. For finding out the information that is hidden in web logs, several data mining techniques are applied on web server logs. This research work concentrates on web usage mining and in particular focuses on discovering the web usage patterns of websites from the server log files using FP growth algorithm.

## II. EXISTING SYSTEM

The research work was initiated through a system study and analysis phase, where significant study was conducted to understand the existing system. Using Apriori algorithm for weblog mining is a novel technique. The explosive growth of the World Wide Web (WWW) in recent years has turned the web into the largest source of available online data.

- Situations like a few disconnected themes in a solitary page may prompt disarray

and make harder to arrive at the data that the guests are searching for.

- The plan of the entire site (interface, content, structure, convenience, and so forth) is perhaps the main angles for any foundation that needs to get by in the internet.
- Understand the manner in which client peruses the site and discover which is the most regular utilized connection and example of utilizing the highlights accessible in the site.

All these information is available online but are hidden for the users. Existing system doesn't focus on the hidden information. This research work uses web usage mining FP growth based approach for analyzing the visitor browsing behavior.

## Limitations

Apriori algorithm, in spite of being simple, has some limitations. They are,

- It is costly to handle a huge number of candidate sets. For example, if there are 104 frequent 1- item sets, the Apriori algorithm will need to generate more than 107 length-2 candidates and accumulate and test their occurrence frequencies. Moreover, to discover a frequent pattern of size 100, such as  $\{a_1, \dots, a_{100}\}$ , it must generate  $2^{100} - 2 \sim 10^{30}$  candidates in total. This is the inherent cost of candidate generation, no matter what implementation technique is applied.

It is tedious to repeatedly scan the database and check a large set of candidates by pattern matching, which is especially true for mining long patterns.

## III. PROPOSED SYSTEM

The main goal of the proposed system is to identify usage pattern from web log files of a website. FP growth algorithm is used for this purpose. The main goal of the proposed system is to identify usage pattern from web log files of a website. This algorithm is influential for mining frequent item sets.

## Advantages

The major advantages of FP- Growth algorithm is:

- Uses compact data structure
  - Eliminates repeated database scan
- FP-growth is an order of magnitude faster

than other association mining algorithms and is also faster than tree researching. This algorithm reduces the total number of candidate item sets by producing a compressed version of the database in terms of an FP growth.

The algorithm consists of two steps:

- Compress a large database into a compact, Frequent Pattern tree (FP-tree) structure
  - Highly condensed, but complete for frequent pattern mining and avoid costly database scans
- Develop an efficient, FP -tree-based frequent pattern mining method (FP-growth)
  - A divide-and-conquer methodology: decompose mining tasks into smaller ones and avoid candidate generation: sub-database test only

#### IV. FREQUENT PATTERN

##### GROWTH ALGORITHM

FP growth algorithm generates frequent item sets from FP Tree by traversing in bottom up fashion. It allows frequent item set discovery without candidate item set generation. It is a two step approach.

Step 1: Build a compact data structure called the FP-tree. It is built using 2 passes over the data-set. Step 2: Extracts frequent item sets directly from FP-tree Traversal through FP-Tree

##### Algorithm:

**Input:** A database DB, represented by FP-tree constructed and a Minimum support threshold.

**Output:** The complete set of frequent patterns.

**Method:** call FP-growth (FP-tree, null).

Procedure FP-growth (Tree, a) {

- 1) if Tree contains a single prefix path then // Mining single prefix-path FP-tree {
- 2) let P be the single prefix-path part of Tree;
- 3) let Q be the multipath part with the top branching node replaced by a null root;
- 4) for each combination (denoted as  $\beta$ ) of the nodes in the path P do
- 5) generate pattern  $\beta \cup a$  with support = minimum support of nodes in  $\beta$ ;
- 6) let freq pattern set(P) be the set of patterns so

generated; }

7) else let Q be Tree;

8) for each item  $a_i$  in Q do { // Mining multipath FP-tree

9) generate pattern  $\beta = a_i \cup a$  with support =  $a_i$ .support;

10) construct  $\beta$ 's conditional pattern-base and then  $\beta$ 's conditional FP-tree Tree  $\beta$ ;

11) if Tree  $\beta \neq \emptyset$  then

12) call FP-growth(Tree  $\beta$ ,  $\beta$ );

13) let freq pattern set(Q) be the set of patterns so generated; }

14) return(freq pattern set(P)  $\cup$  freq pattern set(Q)  $\cup$  (freq pattern set(P)  $\times$  freq pattern set(Q)))

##### Advantages:

- 1) It uses Compact data structure.
- 2) It eliminates repeated database scan.
- 3) It is faster than Apriori algorithm.
- 4) It reduces the total number of candidate item sets by producing a compressed version of the database in terms of an FP tree. The 2nd module is performed in two steps.

➤ FP Tree generation

➤ Applying FP Growth to generate association rules

FP tree is a compact data structure that stores important, crucial and quantitative information about frequent patterns. The fundamental segments of FP tree are:

- It comprises of one root marked as "root", a bunch of thing prefix sub-trees as the offspring of the root, and a regular thing header table.
- Each hub in the thing prefix sub-tree comprises of three fields: thing - name, tally, and hub interface, where thing - name registers which thing this hub speaks to, tally enrolls the quantity of exchanges spoke to by the bit of the way arriving at this hub, and hub - connect connections to the following hub in the FP tree conveying a similar thing - name, or invalid if there is none. Each entry in the frequent-item header table consists of two fields, (1) item - name and (2) head of node link, which points to the first node in the FP-tree carrying the item - name.

Second, an FP -tree-based pattern -fragment growth mining method is developed, which starts from a frequent length-1 pattern (as an initial suffix pattern), examines only its conditional - pattern base (a



“sub-database” which consists of the set of frequent items co-occurring with the suffix pattern), constructs its ( conditional ) FP - tree, and performs mining recursively with such a tree. The pattern growth is achieved via concatenation of the suffix pattern with the new ones generated from a conditional FP -tree. Since the frequent item set in any transaction is always encoded in the corresponding path of the frequent- pattern trees, pattern growth ensures the completeness of the result.

### ORIGIN OF WEB USAGEMINING

It very well may be said that, the World Wide Web is today the most significant media for gathering, sharing and appropriating data. Advanced education (HE) is one of the Departments where electronic innovation has been rapidly and effectively received. The incredible proposition of online courses that, these days, is offered by colleges is one proof of that. Significantly more, totally virtual colleges are showing up. Overseeing and following understudies, planning courses, making assessments, and so on requires explicit frameworks which are called Learning Management Systems (LMS).

Course Management Systems (CMS), Enterprise Learning Management (ELMS) and Learning Content Management Systems (LCMS). CMS encourage web conveyance and the executives for teacher drove points and incorporate conferencing frameworks, surveying and test modules, virtual workspaces and different devices for estimating results and revealing advancement for individual or gathering of understudies. They will in general be exceptionally literary and format arranged to give usability, however restricting adaptability. They commonly add solid incorporated writing devices and parts to interface with information base frameworks. Numerous CMS have been created and are being used far and wide. Anyway they don't uphold instruments which permit teachers to completely follow and survey all the exercises performed by all students, nor to assess the structure of the course substance and its adequacy on the learning cycle. Indeed, these conditions furnish the teacher with access synopsis data, for example, more visited pages, most loved

specialized technique, and different insights. By and by this data isn't sufficient to examine the conduct of every understudy and his advancement. The web is not any more a term that should be investigated the word reference for its significance. It is a particularly variation, immense and dynamic information store that contains generally crude information which is a source to the colossal inventory of data. The data exhumed from this vault is kept an eye out by, clients, web specialist co-ops, business examiners, accordingly making it even complex to be managed. The web clients thus, need to have the compelling pursuit devices to discover pertinent data effectively and absolutely. Information mining is the cycle of unearthing for discovering information from information. Web mining is the way toward unearthing data and examples from web. It is utilized to comprehend client conduct, assess the adequacy of a specific site, and help measure the achievement of an advertising effort. It likewise permits searching for designs in information through substance mining, structure mining, and use mining. web utilization mining is applied to numerous certifiable issues to find fascinating client route designs for development of web architecture by making extra point or proposals noticing client or client conduct. There are three strategies which are appropriate for web mining.

- 1) Web content mining
- 2) Web structure mining
- 3) Web usage mining

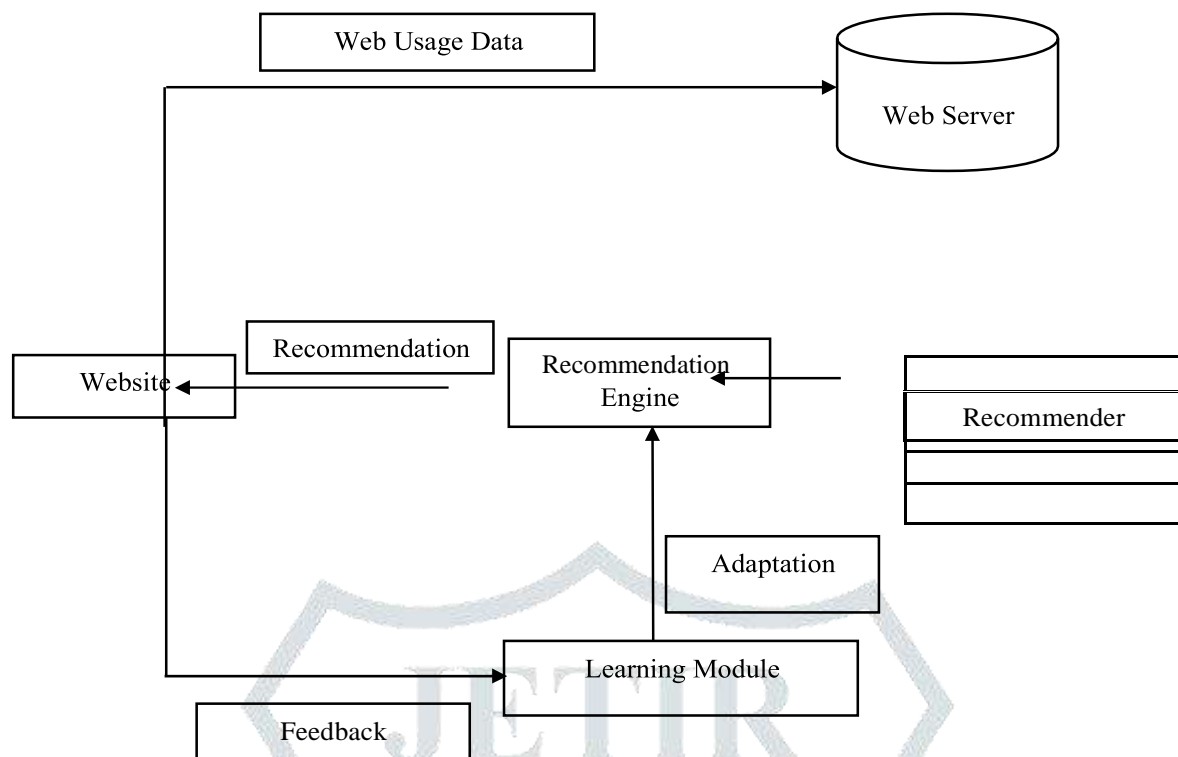


Figure 1: Process of Web Usage Mining

## DATA CONTENT MINING, WEB STRUCTURE MINING AND WEB USAGE MINING

Data mining (DM) is a stage from Knowledge Discovery in Database (KDD) measure, which is characterized as a "nontrivial cycle of distinguishing legitimate, novel, possibly valuable and eventually justifiable example in information". The term design here alludes some theoretical portrayal of a subset information of the information, that is, an articulation in some language depicting an information subset or an information subset or a model material to that subset.

Data mining efforts associated with the Web, called Web mining, can be broadly categorized into three areas of

interest based on which part of the Web to mine; Web Content mining, Web Structure mining, and Web Usage Mining. In Web mining, data can be collected at the server-side, client-side, proxy servers or a consolidated Web/business database. The information provided by the data sources described above can be used to construct several data abstractions, namely users, page-views, click-streams and server sessions.

## PROCESS OF WEB USAGE MINING

Web Usage Mining is characterized as the way toward applying information mining strategies to the disclosure of use designs from Web logs information which to recognize Web client's conduct. Web Usage Mining is the sort of Web mining action that includes a programmed disclosure of client access designs from at least one Web workers. Web usage mining process is generally divided into three tasks:

- Preprocessing,
- Pattern analysis
- Pattern discovery.

Preprocessing includes the fusion and synchronization identification, user identification, session identification (or sessionization), episode identification, and the integration of click stream data with order data sources such as content or semantic information. The uploaded data will be fragment and encrypt.

In the pattern analysis phase interesting knowledge is extracted from frequent patterns and these results are used for website modifications.

In pattern discovery phase, frequent pattern discovery algorithms are applied on raw data. For finding out the information that is hidden in web logs, several data mining techniques are applied on web server logs. This research work concentrates on web usage mining and in particular focuses on discovering the web usage patterns of websites from the server log files using FP growth algorithm.

## METHODOLOGY OF THE RESEARCH

The Web Usage Mining process proposed a major guide line upon project implementation. Fig.3 shows the general flow of the project methodology.

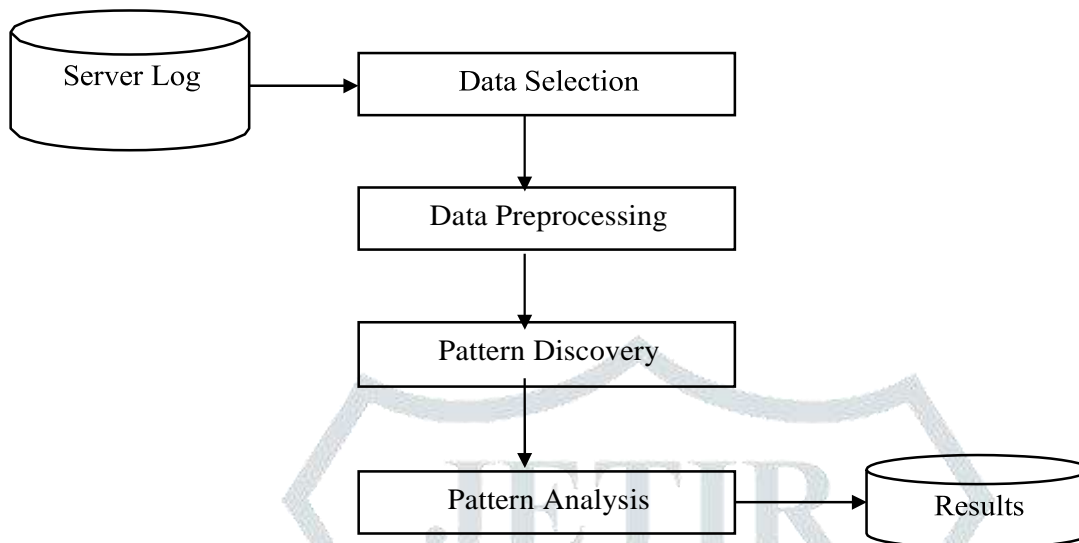


Figure 2: Research Methodology

### 4.1 WEB SERVER LOG FILES

The principle point of this work is to change over client situated contribution to a PC based configuration for example to get familiar with the client's advantages by changing over the log records into gainful results. A log document is a book document in which each page demand made to the web worker is recorded. Log records are documents that rundown the activities that have been happened. These log documents live in the web workers, web intermediary workers and customer programs. The web log document has the expansion .log and contains ASCII characters.

Log documents contain the accompanying data:

This paper explores the use of Web Usage Mining techniques to analyze Web log records collected from E-Learning portal. Using commercial data Web mining tools it have identified several Web access pattern by applying well known data mining techniques (FP growth Algorithm) to the access logs of this educational

portal. This includes descriptive statistic and Association Rules for the portal including support and confidence to represent the Web usage and user behavior for E learning.

The outcomes and discoveries for this test examination can be use by the Web organization and might be upper level in the E-learning network to design the overhauling and upgrade to the entryway introduction. As of late, the coming of information digging procedures for finding use design from Web data(Web Usage Mining) demonstrates that these strategies can be a practical option in contrast to customary dynamic Web Usage Mining is the way toward applying information mining methods to the revelation of utilization designs from Web information and is focused towards applications. Web Usage Mining mines the auxiliary information (Web worker access logs, program logs, client profiles, enrollment information, client meetings or exchanges, treats, client questions, mouse clicks and some other information as the aftereffect of cooperation with the Web) got from the associations of the clients during certain time .



V.

## ARCHITECTURE DIAGRAM

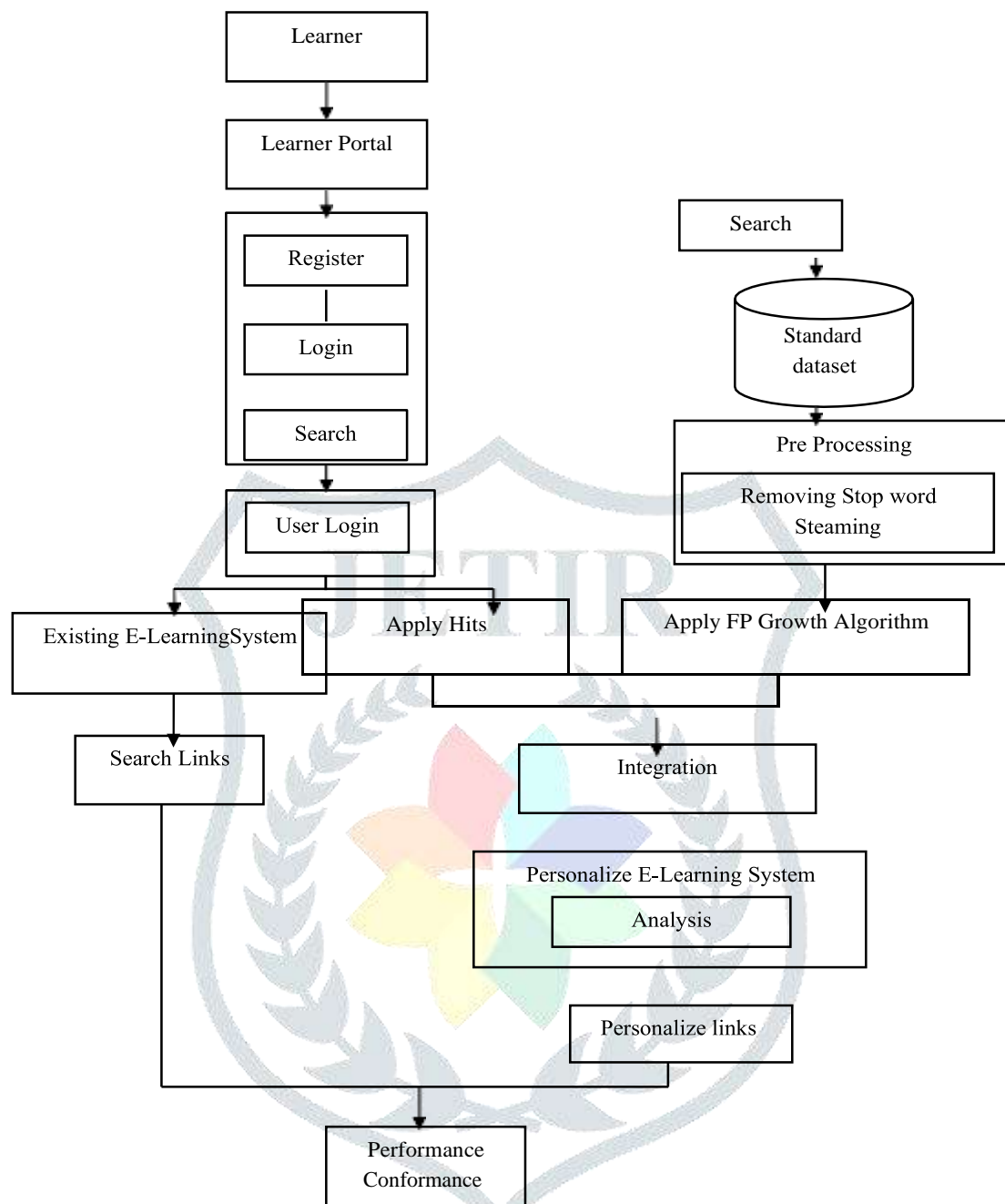


Fig3: Architecture Diagram

## VI. WEB SERVER LOG FILES

The main aim of this work is to convert user oriented input into a computer-based format i.e. to learn the user's interests by converting the log records into beneficial outcomes. A log file is a text file in which every page request made to the web server is recorded. Log files are files that list the actions that have been occurred. These log files reside in the web servers, web proxy servers and client browsers. The web log file has the extension .log and contains ASCII characters. Log files contain the following information:

This paper explores the use of Web Usage Mining techniques to analyze Web log records collected from E-Learning portal. Using commercial data Web mining tools it have identified several Web access pattern by applying well known data mining techniques (FP growth Algorithm) to the access logs of this educational portal. This includes descriptive statistic and Association Rules for the portal including support and confidence to represent the Web usage and user behavior for E learning. The results and findings for this experimental analysis can be use by the Web administration and may be upper level in the E-learning community in order to plan the upgrading and enhancement to the portal presentation. Recently, the advent of data mining techniques for discovering usage pattern from Web data (Web Usage Mining) indicates that these techniques can be a viable alternative to traditional decision making Web Usage Mining is the process of applying data mining techniques to the discovery of usage patterns from Web data and is targeted towards applications. Web Usage Mining mines the secondary data (Web server access logs, browser logs, user profiles, registration data, user sessions or transactions, cookies, user queries, mouse clicks and any other data as the result of interaction with the Web) derived from the interactions of the users during certain period of Web sessions.

### □ *Registration (Data owner/User)*

This module is responsible for user registration. Users can register by giving their name, address, contact number and email. Data owner can register by giving organization name, address, contact and email. The username and password will sent to the Data User's registered

email id.

### □ *Data Preprocessing*

Data Pre processing phase is one of the most challenging phase in this study. The major task in this phase are includes handling missing values, identifying outliers, smooth out noisy data and correct inconsistent data. Data Preprocessing consists of all the actions taken before the actual Pattern Analysis phase process starts.

The Data Preprocessing phase is being done by using available software in the market. On early stage of this phase, Macro tool in Microsoft Access have been selected to assist the preprocessing tasks and for the following data preprocessing task, filter tool in Microsoft Excel becomes the selected tool.

### □ *Pattern Analysis*

During the Pattern Analysis phase, the descriptive method is being used analyze the data such as general summary of the Web usage and customer behaviors. This general summary includes the most active users using the portal. The analysis also tries to find out the top visitors for each facility or option that being provided by the E- learning portal. Beside the option analysis, the sever log files also trace the information of documents that was downloaded.

### □ *Pattern Discovery*

Given a server log files that represent E-learning portal activities, the main purpose of Association Rules is to generate all Association Rules that have support and confidence greater than the user specified minimum support (called min\_sup) and minimum confidence (called min\_conf) respectively. An algorithm for finding all Association Rules, henceforth, referred to as the FP algorithm the selected of FP algorithm is because of the performance where it able to run the mining process in short period. Currently, FP algorithm is commonly used for generating the Association Rules for Web Usage Mining and this experimental study focus on exploratory of Web Usage Mining in E-Learning portal (E-learning portal).

□ **User name:**

This identifies who had visited the web site. The identification of the user mostly would be the IP address that is assigned by the Internet Service provider (ISP). This may be a temporary address that has been assigned. Therefore here the unique identification of the user is lagging. In some web sites the user identification is made by getting the user profile and allows them to access the web site by using a user name and password. In this kind of access the user is being identified uniquely so that the revisit of the user Canalso be identified.

□ **Time stamp:** The time spent by the user in

each web page while surfing through the web site. This is identified as the session.

□ **Page last visited:** The page that was visited by the user before he or she leaves the web site.□ **Success rate:** The success rate of the web site can be determined by the number of downloads made and the number copying activity under gone by the user. If any purchase of things or software made, this would also add up the success rate.□ **Results:**

As stated above, this study will focus on Web Usage Mining of E-learning portal. The results of this study are divided into two sections where the first section will discuss about the general descriptions of the access pattern and users behaviors of E-learning portal (descriptive statistic). Another section will display the supports and confidences of the different level in E-learning portal. All the results will display using certain chart for such as pie and bar chart to make it easier understand.

**VII.****CONCLUSION**

This research has attempted for the purpose of web usage mining. The proposed methods were successfully tested on the web log files. In this research, the problem is solved easily in server log files. The simulation result shows that the FP-Growth algorithm is used for finding the most frequently access pattern generated from the web log data, by using the concept of web usage mining and the problem can easily find out that the user's interest. So that website can be improve and more easily accessible for the users. The main goal of the proposed system is to identify usage pattern

from web log files.

**VIII.****FUTURE WORK:**

In order to enhance and continue this project, the suggested methodology can be implemented for system development purposes. The system may perform and implement the Web usage mining phase including data selection, data preprocessing, pattern discovery and analysis. FP growth algorithm may be a part of the pattern discovery sub function.

Lastly, for future work, the another method for analyzing sparse data can be used in the study of E-Learning Web log access, use of different similarity Association Rules and conclude about the most suitable alternatives for knowledge extraction from Web log data.

**IX.RESULTS AND DISCUSSION**

As stated above, this study will focus on Web Usage Mining of E-learning portal. The results of this study are divided into two sections where the first section will discuss about the general descriptions of the access pattern and users behaviors of this research (descriptive statistic). Another section will display the supports and confidences of the different level in this web usage mining. All the results will display using certain chart for such as column chart to make it easier understand.

**Scalability chart****REFERENCES**

- [1.] R. Cooley, B. Mobasher, and J. Srivastava "Web mining information and pattern discovery on the World Wide Web", 8 Nov 1997.
- [2.] Jiawei Han and Micheline Kamber, "Data mining Concepts and Techniques", Elsevier publication, Edition 2006.
- [3.] RajanChattamvelli, "Data Mining Methods", Narosa publications, Edition 2009.

- [4.] Santhosh Kumar and Rukumani, "web usage mining", ijana publication, vol.1, pages 400-404, Edition 2010.
- [5.] Rahul Mishra and AbhaChoubey, "FP from web log data using FP Growth for web usage mining", ijarcsse publications, vol.2, Edition 2012.
- [6.] Divya and Vinod Kumar, "AIS, Apriori and FPTree Algorithm", ijcsmr publication, vol.2, paper 30.
- [7.] G.Sudamathy and C.Jothivenkateshwaran, "An efficient hierarchical frequent pattern analysis approach for web usage", ijca publication, vol.43, Edition 2012.
- [8.] JianhanZhu, Jun Hong and John G. Hughes, "Page clustering: Mining conceptual link hierarchical from web log files for adaptive websites navigation", ACM publication, vol.4, Edition 2004.
- [9.] Harish Kumar and Anil Kumar, "Clustering Algorithm Employ in Web Usage Mining: An Overview", INDIA Com publication, Edition 2011.
- [10.] Idams, "Divisive analysis (DAINA) Algorithm", Unesco publication, chapter 7.1.5, Edition 2005.
- [11.] Hussain T, "A hierarchical cluster based preprocessing methodology for web usage mining", IEEE publication, Edition 2010.
- [12.] Ashok Kumar D, Loraine Charlet Annie M.C, "Web log mining using K-Apriori algorithm", ijca publication, vol.41 Edition march 2012.
- [13.] ShyamSundarMeena, "Efficient discovery of frequent pattern using KFP-Tree from web logs", ijca publication, vol.49, Edition July 2012.
- [14.] Ravindra Gupta, Prateek Gupta "Fast preprocessing of web usage mining with customized web log pre-processing and modified frequent pattern tree" International Journal of Computer Science- 2012 vol 1.

