

A SURVEY ON WEB USAGE MINING: PROCESS, APPLICATION AND TOOLS

Neeraj Kandpal¹, Ripu Ranjan Sinha², M. S. Shekhawat³

¹Research Scholar, Suresh Gyan Vihar University, Jaipur, Rajasthan.

²Prof.(Research), Suresh Gyan Vihar University, Jaipur, Rajasthan.

³Department of Physics, Govt. Engineering College, Bikaner, Rajasthan.

Abstract: An enormous amount of data is available in form of web documents over the World Wide Web and is increasing day by day. Web mining is used to extract useful information from web documents. Web mining is categorized into three types, namely, web content mining, web structure mining and web usage mining. Web usage mining is the data mining technique to mine the web log data from World Wide Web to extract useful information. Web usage mining useful for the applications like e-commerce to do personalized marketing, fight against terrorism, fraud detection, to identify criminal activities, web design etc. This paper is going to explain in detail about the process involved in Web Usage Mining, Web Usage Mining applications and tools.

Keywords-Web Usage Mining, Web log files, Web usage mining process, Web usage mining applications and Tools.

INTRODUCTION

Web Mining is the use of data mining techniques to automatically discover and extract information from Web documents and services. Web mining is the application of data mining techniques to extract knowledge from Web data including Web documents, hyperlinks between documents, usage logs of web sites etc. A common taxonomy of web mining defines three main research lines: Web Content Mining, Web Structure Mining and Web Usage Mining. Thus Web Mining can be categorized in to three broad areas of mining[1].

a. Web Content Mining: Web Content Mining (WCM) is responsible for exploring the proper and relevant information from the contents of web. It focuses mainly inner document level.

b. Web Structure Mining: Web Structure Mining (WSM) is the process by which we discover the model of link structure of the web pages.

c. Web Usage Mining: Web usage mining is a research field that focuses on the development of techniques and tools to study users web navigation behavior[1]. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web based applications. It also called as Web log mining.

SOURCE OF DATA FOR WEB USAGE MINING

Server log file is a text file which automatically created, when a user requests a page from the web site. It is a file to which the Web server writes information each time a user requests a resource from that particular site. When a user sends queries to the server, requested databases will be retrieved. At the same time ,the user session including the URL, Client's IP address, accessing date and time , query

stem will be recorded in the server logs. These server logs can be preprocessed and mined in order to get some insight into the usage of a server site as well as the user's behavior[2].

Analyzing this log data may find out many interesting patterns. A Web log file records activity information when a Web user submits a request to a Web Server. A log file resides at three different places in system[3]:

(i) Web Servers (ii) Web proxy Servers and (iii) Client browsers.

(i) Web Server Log: These log files resides in web server and notes activity of the user browsing website. There are four types of web server logs i.e., access logs, agent logs, error logs and referrer logs.

(ii) Web Proxy Server Log: These log files contains information about the proxy server from which user request came to the web server.

(iii) Client browser Log Files: These log files resides in client's browser and to store them special software are used.

TYPES OF WEB SERVER LOG FILES

There are four types of web server log files[4]:

a. Access log file: Access log is used to capture the information about the user and it has many numbers of attributes like Date, Time, ,Client IP Address, User Authentication, ,Server name, Server IP address, Server Port, Server Method (HTTP Request), URI Stem, Server URI Query, Status, Bytes Sent, Bytes Received, Time stamp, Protocol version, Host, User Agent, Cookies, ,Referer etc. It is possible to analyze the variables in the access log using web usage mining technique. The data from Access Logs provides a broad view of a Web server's and users. Such analysis enables server administrators and decision makers to characterize their server's audience and usage patterns[2].

b. Error log file: When user click on a particular link and the browser does not display the particular page or website then the user receives error 404 not found. The analysis of Error log data can provide important server information such as missing files, erroneous links, and aborted downloads. This information can enable server administrators to modify and correct server content, thus decreasing the number of errors users encounter while navigating a site[2].

c. Agent log file: Agent log is used to record the details about online user behaviour, user's browser, browser's version and operating system. It is a standard log file while comparing the access log

d. Referrer log file: Referrer log is used to store the information of the URLs of web pages on other sites that link to web pages.

Types of Log File Format

There are mainly three types of log file formats in general that are used by majority of the servers[8,6,4].

a. Common Log File Format (NCSA common log file format): It is the standardized text file format that is used by most of the web servers to generate the log files. The configuration of common log file format is given below:

```

"%h %l %u %t \"%r\" %>s %b" common CustomLog
logs/access_log common
e.g.: 127.0.0.1 RFC 1413 frank [10/Oct/2000:13:55:36 -
0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
    
```

b. Combined Log Format (NCSA combined log file format): It is same as the common log file format but with three additional fields i.e., referral field, the user_agent field, and the cookie field. The configuration of combined log format is given below:

```

LogFormat "%h %l %u %t \"%r\" %>s %b
\"%{Referer}i\" \"%{Useragent}i\"" combined CustomLog
logs/access_log combined
e.g. : 127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700]
"GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en]
(Win98; I ;Nav)"
    
```

c. Multiple Access Logs (W3C extended log file format/IIS log file format): It is the combination of common log format and combined log file format but in this format multiple directories can be created for access logs. Configuration of multiple access logs is given below:

```

LogFormat "%h %l %u %t \"%r\" %>s %b" common
CustomLog logs/access_log common CustomLog
logs/referer_log "%{Referer}i -> %U" CustomLog
logs/agent_log "%{User-agent}i"
    
```

ARCHITECTURE OF WEB USAGE MINING

The whole procedure of using Web usage mining for Web recommendation consists of three steps, i.e. data collection and pre-processing, pattern mining (or knowledge discovery) as well as knowledge application. Figure (1) depicts the architecture of the web usage mining[7].

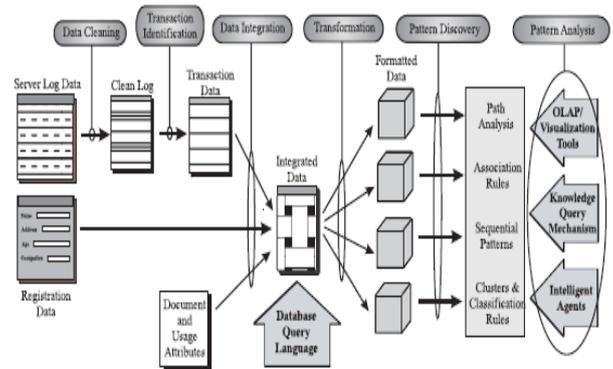


Figure (1).Architecture of Web Usage Mining

Steps Involved in Web Usage Mining (Process of Web Usage Mining)

The steps involved in Web Usage Mining are as follows:

(A) Data collection : It is the very first step of Web usage mining. It involves extraction of log data from server log files. Data can be basically collected from three sources[1,25,26]:

a. The server side: These logs usually contain basic information e.g.: name and IP of the remote host, date and time of the request, the request line exactly as it came from the client, etc. This information is usually represented in standard format.

b. The Proxy Side: Many Internet Service Providers (pydISPs) give to their customer proxy server services to improve navigation speed through caching. The main difference with the server side is that proxy servers collect data of groups of users accessing groups of web servers.

c. The Client Side: Access data can be tracked also on the client side by using JavaScript or applets, or even modified browsers. These techniques avoid the problems of session identification.

(B) Data Integration: Integrate multiple log files into a single file is defined as data integration.

(C) Data preprocessing : Real world data may be noisy or inconsistent so we have to preprocess them to make them consistent and reliable. So preprocessing phase is very important step of web usage mining [9]. The main steps of preprocessing are:

a. Data Cleaning/ Data Reduction: The purpose of data reduction process is to remove unwanted data that may affect the overall mining process[4]. A algorithm is devised by Helmy et al. [10] in which they remove any extensions like gif, jpg, css in target URL. Use of this algorithm, these types of useless data is removed and the mining process gets be evaluated results comparatively fast. The HTTP status code is also a concern for data reduction. In the web personalization area a researcher take the entire data log that contains success code 200 series. As in web intrusion detection, all the status code of server errors is most important because in successful status code an almost no margin to find a suspect. Suneetha et al. [11] give details of HTTP status code. In anomaly user behavior investigation the failure error i.e 400 series code and server error 500 series code is important. So in the web log entries that contains 400 and 500 status code is not eliminated in data reduction phase.

b. User Identification: User identification refers to identify unique users. Users with different ip address are considered unique users. According to Chaofeng [12], each IP address represents one user. An IP address represents a different user, if a page is requested by a referrer link; there is another user with the same IP address. Cooley et al. [13] proposed a heuristic that if a web page is directly accessed without any hyperlink by same IP, assumed as a different user.

c. Session Identification: Session identification refers to differentiate the web log entries into different user sessions by a session timeout. Once a user was identified then click stream is divided in to clusters. This method of division is called Session Reconstruction or Sessionization. Some researchers [4] have coined that there is a new session if time limit is exceed more that 30 minutes.

d. Path Completion: This step is used to check the missing pages after constructing transactions. The missing page problem is due to proxy servers and caching problem of clients.

(D) Pattern discovery : In Pattern Discovery phase, data mining techniques like association rule mining and clustering applied on web log files after preprocessing to discover the useful pattern. Firstly the web logs are converted into relational data base and then three main operations Association, Clustering and Sequential Analysis, are performed on data for pattern discovery[24].

a. Association rule mining: Association rule mining is one of the data mining technique which is used to discover useful pattern. It works on generating frequent pattern and rules. In web log file number of URL visit by number of users so we can identify frequently accessed web pages by users which can help to understand user needs. Two basic parameters of association rule are support and confidence. The association rule is mainly focuses on

discovery of relations between pages visited by users on web site. Association rule can be used to relate the web page is most often used by the single server session. Several algorithms like Apriori, Eclat, Frequent Pattern tree etc. to perform association rule mining.

b. Clustering Analysis: Clustering Analysis is used to group the data or items which have similar attributes or characteristics. Clustering is unsupervised learning technique. Clustering analysis defined as similar characteristics users are group together without knowledge of group definition. Clustering will help us to find group of common behavior users. Clustering of web pages are very important for internet service provider to analyze the behavior of users. Clustering can also be used for anomaly detection. Once the data has been segmented into clusters, you might find that some cases do not fit well into any clusters. These cases are anomalies.

c. Sequential pattern analysis: Sequential pattern analysis is used to find that a suspected user visit a particular link A followed by link B in a time ordered set of sessions [19]. By using this approach we can predict the suspected user psychology which is useful in crime detection. There are several algorithms like Apriori All, GSP, SPADE, Prefix Span and Spam are used for sequential pattern analysis.

d. Classification: In this method web server data is classified according to some common attributes like hour of the day in which data accessed. Classification is a mapping method of data that could be one or several predefined data.

(E) Pattern Analysis: The main purpose of pattern analysis is to analyze the pattern which is identified during pattern discovery phase. Its main purpose is to find out a valuable model or standard pattern for specific web usage mining application. Some important techniques used for pattern analysis are visualization technique, OLAP techniques, data and knowledge querying and usability analysis[4,14].

a. OLAP (Online Analytical Processing Technique) is a powerful paradigm for strategic analysis of relational database which is very useful in business systems [4]. OLAP is part of the broader category of business intelligence, which also encompasses relational reporting and data mining. Typical applications of OLAP include business reporting for sales, marketing, management reporting, business process management, budgeting and forecasting, financial reporting and similar areas, with new applications coming up, such as agriculture.

b. Data and Knowledge Querying: Query mechanism such as SQL is the most common method of pattern analysis. This is an important part of web usage mining in which we analyze the different reasons of anomaly behaviors of users. By the use of SQL we find some specific results from database like suspected session in database created by the users like failure status code of http protocol in very short interval of time. The http status code in web server logs help in identifying the suspected users that

triggers a lot of errors during browsing the site. When a user makes a lot of errors during login on any e-commerce site, it may be a malicious user that wants to guess the password.

c. Usability analysis is a modeling technique to accessing the behavior of user on the web site. Barse et al. [15] proposed some Fraud indication is analyzed by the analysis of web log files, when the ratio between transmitted and received data is suspiciously high, a great deal of data is transmitted (some period of time) after data has been received, a great number of downloads are done. Anomaly behavior of user is also traced by using this information. When a user requests a page and if the returned bytes are different from other request for the same page, it is an indication of anomaly behavior of user. An intruder may also tamper the database with the help of SQL injection, XPath injection, Cross site scripting (XSS). These are the some specific web attacks that are commonly encountered[4]. Salama et al. [16] proposed a framework for SQL injection detection. SQL Injection attack (SQLIA) is a type of code-injection attack in which an attacker can leak, modify, and delete information stored in the database. The main reason of SQL injection attack is an inefficient input validation in the database.

d. Visualization Technique: Visualization Technique is a method that used to understanding the behavior of web users by graphical method.

APPLICATIONS OF WEB USAGE MINING

The main applications of Web Usage Mining are :

(i) Personalization of web content: The objective of a Web personalization system is to “provide users with the information they want or need, without expecting from them to ask for it explicitly” [17]. Web Personalization [18] is the task of making Web based information systems adaptive to the needs and interests of individual users, or groups of users. Typically, a personalized Web site recognizes its users, collects information about their preferences and adapts its services, in order to match the user's' needs. According to D. Antoniou et al. [19] some types of Personalization are:

a. Web usage data mining personalization: The customer preference and the product association are automatically learned from click stream.

b. Computational Intelligent combinations: Provide the different information system which have been designed to provide Web users with the information they search, without expecting them to ask for it explicitly.

c. Novel online recommender system: It builds profiling models and offers suggestions without the user taking the lead.

d. Helping Online Customers Decide through Web Personalization: The goal of a personalized website is to take advantage of the knowledge obtained from the analysis of the user's navigational behavior in combination with other information collected, such as the user's location, previous navigation patterns, and items purchased.

(ii) Web design: This helps in the designing the web structure based on the user's query and can help to retrieve only relevant documents.

(iii) E-commerce: Web Usage Mining retrieves the user information from the various web logs. This information may be related to their personal information like age, qualification, their interests, their economy, their usage time, frequently accessing sites, their buying behavior etc. This information useful in ecommerce for various reasons like, to improve their marketing, to improve their production, to improve customer relationships etc.

(iv). Web Advertising/Marketing: Web advertising also referred to as an online advertisement. The use of popular websites can be an effective way of introducing new a product to the customer.

(v) Pre-fetching and caching: Efficiently delivering web content, i.e., caching and prefetching. Web Usage Mining can be used to develop proper pre-fetching and caching strategies so as to reduce the server response time. Caching refers to the practice of saving content in memory in the hope that another user will request the same content in near future, while involves guessing at which content will be of interest to the user, and loading it into memory.

(vi) Transaction Analysis: New environment brings new changes in the current economic model as it changes the relationship between operators and customers from the traditional physical store to electronic transactions on the internet. Analysis of e-commerce uses clickstream data to determine the marketing effectiveness of the site by quantifying user behavior while actually visiting the site visitor browsing the site recording the translation in a sales transaction [20].

(vii) Modification of web site : For successful website, modification according to user need is essential. Required modification is successfully determined by the web usage mining of the server log data.

(viii) Fraud detection: Unauthorized users can be traced using search results of web log data. A user unsuccessfully trying the access to any web site may be an intruder tries to break the password of restricted area of website.

(ix) Customer Relationship Management: Customer Relationship Management is becoming standard terminology. It focuses creating value for the customer and

company over the long term and the relationships are built with the customers, which provide value for services [9].

(x) **Product/Site recommendation:** Web site and various products can be recommended to users according to the user interest using web usage mining.

(xi) **Identify Web Robots:** Web Robots are software programs behaves like human for target website. These programs are very harmful for websites because they may crack a password or may breakdown the site by continuous fake requests.

(xii) **To improve web server program's performance:** Web usage mining is very useful for improvement of performance of the of web server. E.g. access time for the particular site can be improved by mining server log data of the user access log.

TOOLS USED IN WEB USAGE MINING

Some tools used to explore Web Usage Mining are:

(i) **Web Utilization Miner (WUM):** WUM uses mining language MINT[1]. "MINT" is the mining language serving as interface between the user and the miner. MINT supports the specification of criteria of statistical, structural and textual nature[1]. To discover the navigation patterns satisfying the expert's criteria, WUM exploits an innovative aggregated storage representation for the information in the web server log. Knowledge about the navigation patterns occurring in or dominating the usage of a web site can greatly help the site's owner or administrator in improving its quality.

Web Utilization Miner WUM employs an innovative technique for the discovery of navigation patterns over an aggregated materialized view of the web log. This technique offers a mining language as interface to the expert, so that the generic characteristics can be given, which make a pattern interesting to the specific person. There are two major modules: the Aggregation Service prepares the web log data for mining and the MINT-Processor does the mining. The Aggregation Service extracts information on the activities of the users visiting the website and groups consecutive activities of the same user into a transaction. It then transforms transactions into sequences. The MINT-Processor mines the aggregated data according to the directives of the human expert

(ii) **KOINOTITES:** KOINOTITES, is a software system that exploits Web Usage Mining and user modeling techniques for the customization of information to the needs of individual users. More specifically, KOINOTITES processes the Web server log files, and organizes the information of a Web site into groups, which reflect common navigational behavior of the Web site visitors. KOINOTITES is a software tool, which exploits Web Usage

Mining techniques in order to create user communities from Web data. KOINOTITES is based on a modular architecture, and comprises the following two main components: i) A mining component, that consists of the modules that perform the main functions of the system, i.e., data preprocessing, session identification, pattern recognition and knowledge presentation. ii) A Graphical User Interface (GUI) component, supplemented by wizards and on-line help that is used for user interaction with the system. Both components have been implemented using the Java programming language[1].

(iii) **Web miner [6]:** A general and flexible framework for web usage mining to extract relationship from data collected in large web data repositories.

(iv) **Web Site Information Filter System (Web SIFT):**

Web SIFT system uses content and structure information from the web site in order to identify potentially interesting results from the mining usage data. This is a framework for web usage mining to perform preprocessing and knowledge discovery and automatically defines a belief set. The information filter uses this belief set to find interesting patterns. Web SIFT system is based on the Web Miner prototype divides the web usage mining into three parts: Preprocessing, Pattern Discovery and Pattern Analysis. Web SIFT system implemented using a Relational database, Procedural SQL and Java programming language[1].

(v) **WebViz:** Pitkow et al. [21] developed a tool WebViz that provide that selectively filtered a web server log, control bindings to graph attributes and also reissue of logged sequence of requests. WebViz, is a useful database utility, provide the user with the graphical information about document accesses and path taken by users through the database.

(vi) **WET (Web-Event Logging Technique):** Etgen et al. [22] proposed WET (Web-Event Logging Technique), WET is an automated usability testing technique that works by modifying every page on the server. It can automatically and remotely track user interactions. WET takes advantage of the event handling capabilities built into the Netscape and Microsoft browsers.

(vii) **WebQuilt:** WebQuilt [23] is developed by Hong et al. WebQuilt is a web logging and visualization analyzing tool which help to analyze the collected data from local and remote logs. The collected usage visualized in a zooming interface that shows which people viewed the particular web page

(ix) **Web log miner:** It coined by zaiane(1998), uses data mining and OLAP on treated and transformed web access files. Mining a web server log files.

(x) **Web mate:** The user profile is inferred training examples, proxy agent provides effective browsing and searching help.

(xi) **Web usage miner:** It exploits an innovative aggregated storage representation for the information in the web server log. It discovers patterns comprised of not necessarily adjacent events. Mining interesting navigation patterns in form graphs.

(xii) **i-miner:** To optimize the concurrent architecture of fuzzy clustering algorithm and fuzzy inference system to analyze the trends, pattern discovery and trend analysis from web usage data mining.

FUTURE SCOPE OF WEB USAGE MINING

Some of the future scope of web usage mining are:

- (i) Digital forensics investigations.
- (ii) Crime investigation
- (iii) Automated data cleaning
- (iv) Robot detection and filtering
- (v) Transaction identification
- (vi) Automated website design modification tools

CONCLUSION

The survey was performed on web usage mining its applications and tools available for web usage mining. In this survey we find standard processing steps followed by researchers for processing of web log files. We also listed the tools widely available for web usage mining. Applications of web mining is already covering a wide area and continuously growing day by day. Future scope of Web Usage Mining resides on Digital forensics investigations, Crime investigation, Automated data cleaning, Robot detection and filtering, Transaction identification etc.

REFERENCES

[01] Bharti Joshi, Ph.D., Suhasini Parvatikar, "Analysis of User Behavior through Web Usage Mining", International Journal of Computer Applications, International Conference on Advances in Science and Technology (ICAST-2014).

[2] Aditi Shrivastava, Nitin Shukla, "Extracting Knowledge from User Access Logs", International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012

[3] K. R. Suneetha, Dr. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File" IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.

[4] Amit Pratap Singh, Dr. R. C. Jain, "A Survey on Different Phases of Web Usage Mining for Anomaly User Behavior Investigation", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 3, May – June 2014.

[5] Lee U., Liu Z. and Cho J. (2005), "Automatic Identification of User Goals in Web Search", In Proc. 14th Int'l Conference on World Wide Web (WWW '05), Vol. 51, No. 3, pp. 391-400.

[6] G.D.Praveenkumar,R.Gayathri, "A Process of Web Usage Mining and Its Tools", International Journal of Advanced Research in Science, Engineering and Technology, Vol. 2, Issue 11, November 2015.

[7] Vijayashri Losarwar, Dr. Madhuri Joshi, "Data Preprocessing in Web Usage Mining", International Conference on Artificial Intelligence and Embedded Systems, 2012.

[8] Nanhay Singh, Achin Jain, Ram Shringar Raw, "COMPARISON ANALYSIS OF WEB USAGE MINING USING PATTERN RECOGNITION TECHNIQUES", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.4, July 2013.

[9] C. Sakthipriya, G. Srinaganya, Dr. J. G. R. Sathiaselvan, "An Analysis of Recent Trends and Challenges in Web Usage Mining Applications", International Journal of Computer Science and Mobile Computing, Vol. 4, Issue. 4, April 2015, pp.41 – 48.

[10] Mohd Helmy, Abd Wahab and Nik Shahidah, "Development of Web usage Mining Tools to Analyze the Web Server Logs using Artificial Intelligence Techniques", The 2nd National Intelligence Systems and Information Technology Symposium (ISITS 2007), October 30-31 2007, ITMA-UPM, Malaysia.

[11] K. R. Sunnetha and Dr. R. Krishnamoorthi, "Identifying User by Analyzing Web Server Access Log File", International Journal of Computer Science and Network Security (IJCSNS), Vol. 9, No. 4, 2009, pp. 327-332.

[12] Li. Chaofeng, "Research and Development of Data Preprocessing in Web Usage Mining", International Conference on Management and Engineering, 2006, pp. 1311-1315.

[13] R. Cooley, B. Mobasher and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Journal of Knowledge and Information Systems, Springer, 1999, Vol. 1, No. 1, pp. 1-27.

[14] Aarti M. Parekh, Anjali S. Patel, Sonal J. Parmar, Prof. Vaishali R. Patel, "Web usage Mining: Frequent Pattern Generation using Association Rule Mining and Clustering", International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 04, April-2015.

[15] E. L. Barse, H. Akan and K. E. Jonsson, "Synthesizing Test Data for Fraud Detection Systems", In proceedings of the 19th Annual Computer Security Applications Conference, December 8-12, 2003, pp. 384-394.

- [16] S. E. Salama, M. I. Marie, L. M. El-Fangary and Y. K. Helmy, "Web Anomaly Misuse Intrusion Detection Framework for SQL Injection Detection", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 3, No. 3, 2012, pp. 123-129
- [17] M. Eirinaki and M. Vazirgiannis Athens University of Economics and Business, "Web Mining for Web personalization," *ACM Transactions on Internet Technology*, 2005.
- [18] Dr.S. Vijiyarani and Ms. E. Suganya, "RESEARCH ISSUES IN WEB MINING", *International Journal of Computer-Aided Technologies (IJCAx)*, Vol.2, No.3, July 2015.
- [19] D. Antoniou, M. Paschou, E. Sourla, and A. Tsakalidis, "A Semantic Web Personalizing Technique The case of bursts in web visits," presented at IEEE Fourth International Conference on Semantic Computing, 2010.
- [20] Andrew J. Flanagin, Miriam J. Metzger, Rebekah Pure, Alex Markov, Ethan Hartsell, "Mitigating risk in eCommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention", *Springer Science+Business Media, Electron Commer Res* 14:1– 23, DOI 10.1007/s10660-014-9139-2, 2014.
- [21] J. Pitkow and Krishna K. Bharat, "WebViz : A Tool for World Wide Web Access Log Analysis", In First International WWW conference, 1994.
- [22] M. Etgen and J. Cantor, "What Does Getting WET (Web Event-Logging Tool) Mean for Web Usability?", In Fifth Human Factors and the Web Conference, 1999.
- [23] J. I. Hong and J. A. Landay, "Webquilt: A Framework for Capturing and Visualizing the Web Experience" In Proceedings of the International Conference on the World Wide Web (WWW'01), 2001, pp. 717–724.
- [24] Khushbu Patel, Anurag Punde, Kavita Namdev, Rudra Gupta, Mohit Vyas, DETAILED STUDY OF WEB MINING APPROACHES-A SURVEY, *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, Patel, 4(2): February, 2015.
- [25] V.Chitraa, Dr. Antony Selvdoss Davamani, "A Survey on Preprocessing Methods for Web Usage Data", *(IJCSIS) International Journal of Computer Science and Information Security*, Vol. 7, No. 3, 2010.
- [26] Vijayashri Losarwar, Dr. Madhuri Joshi , "Data Preprocessing in Web Usage Mining", *International Conference on Artificial Intelligence and Embedded Systems* , 2012 .