

# Tracking the Client Internet Usage Data for Behavioral Analysis



Neetu Anand

**Abstract-**The objective of the study is to help the parents to track their child daily activities on computer, for which, there is a need to develop a Monitoring and Tracking System for computer and internet usage. To achieve the stated goals a framework was established and implemented. The interface was designed in such a way so as user will not get a sense that his or her activities are being traced. The database contain all the recorded logs to check over-all usage pattern over time, thus enabling parents to recall the past events and subsequently behavioral aspects of their child. In this study, Web usage mining for monitoring activity of kids has been addressed. The experiment was conducted on the dataset collected by using the developed software. The monitoring system (Systrack) has been compared with other available software's and it was found that Systrack was able to perform all the functions which other software's are capable of, in addition Systrack used data mining techniques to find the behavioural pattern of kids in different situation.

**Keywords:** Monitoring System, User Navigation Strategy, Log Data, Behavioral Analysis, Web Usage Mining.

## I. INTRODUCTION

Computers and the Internet have brought about an immense change in people's lives. The Web is no doubt a big space. In the last few years the World Wide Web has turn out to be the favorite area for creating Internet applications, thanks to its influential communiqué archetype focus on hypermedia contents and surfing, and to its open architectural measures which enable the amalgamation of diverse types of content and procedures [1]. During navigation through web, the browser requests numerous kinds of data from web servers and the result displayed timely on the screen, in even less than the seconds. Analyzing web usage is an important task just because of the obtainability of abundant and huge log files. There is a necessity to pre-process the data to change it, in a way so as it can be easily quarried for knowledge. It also necessitates the development of tools to assist a user to well capture the content of user sessions, interpret them and extract knowledge from them. When youngsters use home-based workstations instead of watching television and playing video games, it is mostly seen as constructive; but when they use computers keeping an eye on what they are doing is the new area of research.

By Considering parents' feedback indicate that they purchase computers for home and gain access to Internet just

only to give learning opportunities to their youngsters so that they can prepare themselves for the "information age". Even though they are more and more apprehensive about the impact of the web technology on their youngsters and are dissatisfied with a number of online activities their kids involve in-such as chatting, Internet browsing to download movies, games ,songs, images of celebrities, and various social networking sites etc. and many more. Parents usually belief that computer have many benefits, and even think about children without home computers to be at a drawback. By considering all these fact we prepared a supervising system which enables the parent to track many of their teenage activities that may be helpful in the analysis of their navigation pattern and trend. We proposed a new idea "child everyday behavior patterns", which categorize his or her very day actions based on their retrieval rate of various software or website at differed phase of a day. The remaining paper has been ordered as follows: Section 1 illustrates the motive behind the work and correlated work, Section 2 illustrates system monitoring, Section 3 introduces the WUM, Section 4 shows the methodology and procedure and section 5 concludes the work.

## II. MOTIVATION AND RELATED WORK

The objective of this study is to support the parent to keep track of their child daily activities on computer. Various advices that parents must implement to uncover their child daily routine actions and events are stated below:

- Superintend what kids are exploring and when and where they're going online-Maintain record on the list of websites visited and stuffs searched on the computer by reviewing computer and Internet browser's history.
- Be friend with your kids on Facebook, What apps or other social sites. According to recent studies, more than 80 percent of parents are friends on Facebook with their children to inspect their kids' interactions.
- Check online behavior offline. If you see your teen is using social website in an improper way, address it offline! Don't use their profile as a way to converse your concerns. As an alternative, take it as an alternative to dialog with your teen offline.
- Bring Internet use out from private room. When monitoring your teens' online behavior, it's important to stay positive and remind them that you are concerned as you care for them, the evidences specifies that the typically all parents keep an eye on the behavior of their child particularly during some time in a day. There are many issues which make it hard for parents to inspect and control children's computer Internet usage, these include:

Revised Manuscript Received on March 30, 2020.

\* Correspondence Author

Neetu Anand\*, Asst. Professor, Dept. of Computer Application, Maharaja Surajmal Institute of Technology College in Delhi, India. E-mail:neetuanand@msi-ggsip.org

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

- In this era of Wi-Fi and Internet, there is little possibility to keep an eye on child activity all the time. This is specifically relevant in case of mobile phones access during school or college time [4];
- The amount of time children and teenager devoted on events linked to internet;
- Teens' oppose to time bounds;
- Problem to avoid revealing of unacceptable matter;
- Teenagers' private control of details (for example by means of secret code, locks and personal browsing);
- Revolution of the Internet causes parents to face many problems in keeping pace with the social networking sites.

This paper discusses the investigation plan of adolescent computer and internet usage log file with the use of data mining and presents novel tools and techniques for addressing the problems. Event logs play a significant part in contemporary IT systems, since they are brilliant source of information for monitoring the system in an offline mode. Data mining tools and techniques are common choice for discovering the knowledge from event based log files, and the mining of patterns from those logs has been identified as a principal task in monitoring child activities. To tackle the above mentioned problem, this paper suggests inventive tools and techniques for log analysis.

The Internet changes so rapidly that technological measure is unlikely to be able to keep up. More efficacious and long-lasting deportments are those that build on the family, the society, edification and enfranchisement so that children and young people will make good choices and benefit from the generative power of the Internet insecure surroundings[7]. Our advice to parents is, if they have any apprehension sat all about their children visiting unacceptable websites or going online with Internet Prowlers, take the preventive measure that we suggest and employ Internet Monitoring software. You must have knowledge of your child involvement in any undesirable behavior, just as you would do with alcohol, drugs, schoolwork, friends contact details or driving privileges [8]. More than 50% of parents either do not have, or do not know if they have, software on their computer(s) that monitors where their youngster(s) go online and with whom they interact [9].

### III. SYSTEM MONITORING

The record sets mainly utilized for mining Web usage data, are gathered through the Web Servers, Proxy servers or client side. Client Side logs possess mainly exact report of user activities over internet. The foremost purpose of the client profiling is to understand and incorporate the client's habits, desires and detest patterns and choices for real-world applications. Log data accumulated by proxy servers, referred to as *proxy logs* or *proxy traces*, and comprises of documents about Web accesses by various users against numerous Websites.

The most important role of proxy servers or fire-wall is to work equally as a security measures to block an un- invited users or as a cache resource to reduce stream of network traffic by reusing their most newly obtained accounts. Their log files may contain various clients' records that were retrieving many web servers. In the log files, their client request records are interleaved in their received order. The logging method is instinctive and needs less interfering

compared to client-based logging. Its layout and design is reliant on the software used for logging. Its exactitude is however weakened by the client- level cache as some requests are not received by the proxy and are served from the most freshly retrieved files stored at the client computer. Web usage mining used mainly Server-level logs as a source for mining. Generally log files can be stored in any of these formats: Common Log Format (CLF), Extended Common Log Format (ECLF) format.

A number of browsers like Mozilla Firefox, IE, and chrome etc. can be modified or client side data is collected through JavaScript and Java applets. It requires support of users by authorizing the use of the Java Script and Java applets to collect data, or to voluntarily practice the improved browser. Client-side gathering is better than server-side pool because it lowers both the bot and session recognition difficulties. Keeping an eye on system is easy in case of client side monitoring comparative to server-side. The "next generation" of monitoring knows the key restrictions of server-side implications. As an alternative, current research has focused on augmenting software on the client machine to give a more gritty view of whatever are goings-on. To monitor the user actions on client side requires additional program installation with monitoring application to record all the actions of user. *Systrack* is a tool to visualize and look a tone's activities on individual computers. The interface was designed so as user will not get a sense that his or her activities are being traced. The database contain all the recorded logs to check over-all usage pattern over time, thus enabling parents to recall the past events and subsequently behavioral aspects of their child. Our goal was to help parents to answer three main questions about their child activity: What kind of things child do on computer? When do they do it? How all this advances over time?

#### A. Grabbing and Managing the Users' Movement

The two main challenges poses by Monitoring of users' action on a computer are:

- Client machine should not be overloaded by Data and data mustering must be done in an efficient manner. There are number of multifarious applications that can be monitored and this number even grows larger and larger if we take into account web applications.
- Data gathering must be non-intrusive, as users should not be disturbed or distracted at any time.

### IV. WEB MINING

WIC known as Web Intelligence Consortium described the term "Web Intelligence" resembles to "Research and development to explore the primary roles and impact of Artificial Intelligence and advanced Information Technology on the next generation of web-empowered systems, services, and activities." In this regard, [2] described Web usage mining (WUM) as the application of machine learning methods on data collected from the web for the automatic excavating of behavioral patterns of web users. This is the method to analyzing preferences of web user by means of analyzing usage pattern.



Discovering knowledge from raw data is one of the goals of data mining. Moreover data mining employ wide- ranging of techniques and tools. It does not depend on the internal physical procedures from which data is created. The mined knowledge can be better understood by means of tools which integrate various statistical methods, visualization techniques, and human factors. On the other hand, after discovering the pattern, machine learning algorithms can be used as a predictive modality. Successful applications on business and science have been performed such as credit scoring, weather prediction, revenue management, customer segmentation, and many others. Yet, it is not always clear how this knowledge is used in informative computing systems. Conventional method of mining data needs pre-processing of data and reformed before being use don log data. Various attempts have been made to enhance the characteristic of the resultant data. As soon as repository of web user behavior (Web Warehouse) is accessible [8], exclusive machine learning algorithms can be employed in order to extract patterns concerning the usage of the web site. In consequence, a number of applications can be instigating don adaptive web sites, such as recommender systems and revenue management marketing; among others. Web mining is the extension of data mining research [2] in the Web environment. It main aim is to inevitably discovering and extracting information from the Web. It is classified into three domains Content Mining, Structure Mining and Usage Mining shown in Fig1:

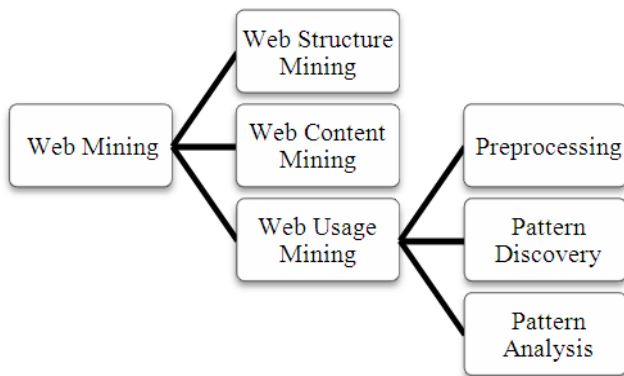


Fig. 1. Web Mining Classification [10]

**A. Web Content mining** encompasses various mining practices to extricate models from web contents like plain text, semi-structured documents (e.g., HTML or XML), structured documents (digital libraries), dynamic documents, hypermedia records. The extorted prototypes were employed to categorize web objects, to obtain keywords for use in information retrieval, to infer structure of semi-structured or unstructured objects. So Web Content Mining is defined as the way of extricating knowledge from the context of Web sites, for example contents of documents or their description.

**B. Web Structure Mining** makes use of the links and references in Web pages to deduce exciting information, such as identifying authorities- and hub-pages. Authorities are pages bearing vital content. Hubs are pages that act as resource lists, directing users to authorities.

Structure mining is further split into two categories

based on type of data used. A hyperlink is defined as a structural unit which connects web pages to different location, either within the same page or to a separate web page. Document Structure is based on the various HTML and XML tags within the page, the content within web pages can also be organized in a tree-structured format.

**C. Web Usage Mining**, also named as Web Log Mining, mines log files for finding attractive patterns in network traffic. The users' movements while surfing the Web is recorded in these sites' Web logs. The procedure of analyzing the user's surfing behavior is termed as Web usage mining Fig 2. The resources of data included in such an analysis might comprise of demographic data about the users, or data collected in web logs i.e. usage data.

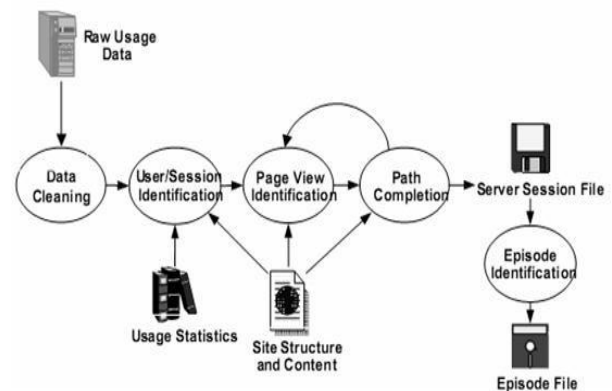


Fig. 2. Web usage mining process[12]

In this paper we mainly focus on collecting data for log files.[3] WUM can be viewed as a three-level procedure, involving data formulation, pattern finding and pattern assessment. In the first one, datasets gathered from the web are pre-processed for user identification, sessions finding, page views, and so on. The inserted data are chiefly the hits recorded in the web usage logs of the site, occasionally merged with other facts and figures such as registered user profiles, referrer's logs, cookies, etc. In the next one, statistical methods, along with data mining techniques (for example association rules, Regression, clustering, prediction, outlier detection and classification) are applied with the purpose of detecting significant patterns. These patterns are retained so that they can be investigated again in the last level of the web usage mining process[5]. New problems ensue in Web domain and novel techniques are needed for Web mining tasks. Web offers an astonishing prospect and confront to data mining.

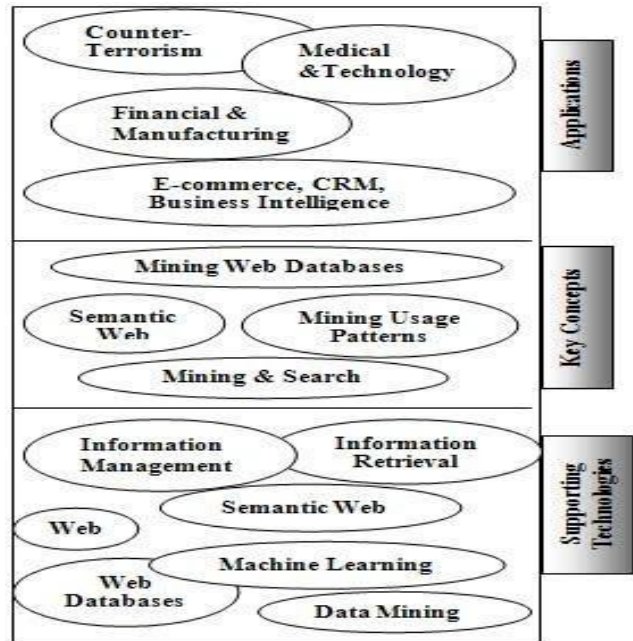
- The information on the web is very eclectic and comprehensive.
- The range of learning on web is enormous.
- Web information is entirely linked through hyperlinks.
- Various pages contain the same erudition in sometime different format.
- Material (Data and Info) of almost all types are available on web.

- Web is noisy and contains many contents on the same page.
- All types of services and assistances are available on the web.
- Web is essential and vibrant.
- Web is a simulated society.
- Web consists of open and deep web.

### D. Application Area of WUM

With the increase in interest of WUM its applications are growing day by day .Some of the area where it is used are described below [6]:

- Website Modification -WUM intensify the Web site pleasant appearance, by altering its content and organization according to the data mined from the users' behavior.
- Web Personalization- It is accomplished by matching the current navigation pattern of user with the previous web logs. These logs can be used to identify the typical browsing behavior of a user and subsequently to predict desired pages. Social sites use personal data to provide relevant advertisement for their users.
- Analysis of Clickstream Data-The massive volumes of click flow data, gets assimilated with
- Data from offline sources like transaction and demographic data and mined for other applications of business like e-CRM and others marketing programs.
- Web Caching-By revealing repeated retrieval methods of clients, desirable relationships can be detected to enhance the whole operation of potential retrieval. Knowledge regarding repeatedly retrieved pages can be exercised for caching.
- Commercial Communication- Pattern's generate after applying usage mining were used to gather information to enhance better relationship with customer in terms of investment, retention, and also in various others areas of business like sales, marketing and advertisement, cross sales etc.
- User Navigation Strategy- Usage characterization can also look into navigational strategy when browsing a particular site. It is also used to foresee user behavior while the user interacts with the Web. To understand the operational procedure, interface and use of the browser, pattern mining is very essential tool.
- Web Traffic Analysis-Web usage mining of patterns provides a key to understand Web traffic behavior, its performance and quality, which can be used to deal with policies on web caching, network transmission, load balancing, or data distribution.
- Infringement- Web usage mining is also beneficial for intrusion detection, fraud finding, and attempted break-ins to the system and to counter terrorism.
- Others Discipline-Web usage mining can be used in various fields of sciences such as Physical Sciences, Social Sciences, Engineering, Medicine, and Biotechnology.
- More usage of WUM are in CRM, eLearning, e- business, e-commerce, e-education, e-Newspapers, e- Government, Telecom, Digital Libraries, Manufacturing and financial planning.



**Fig. 3. Application area, key concept and supporting Technologies of WUM [11]**

### IV. DESIGN AND IMPLEMENTATION

To develop a Supervising System for client computer and internet usage, the Proposed Methodology is:

1. To build software to monitor all the actions executed on the kids computer that software must trace
    - Record all the recent files and folder opened by the user.
    - Record all the internet related activities like visited web-sites.
  2. A database is constructed having all the previous and current records, for this objective to achieve
    - A separate database is created for storing the record of usage of computer and internet by the kids.
    - A database is updated automatically after few seconds to record the further usage.
  3. Pre-processing on database is performed to clean, extract and refine the data to work on.
    - Data cleaning is the main task this step, the queries is to be applied to filter the database, to get the desired result.
  4. Automatically report generation
    - The Report is generated for parent to check the usage of computer and internet by their kids.
- To implement above mention methodology the algorithm is designed and executed for tracking the routinely activity of child on its personal computer. The various steps will be followed during the process:
- Make the program for capturing the usage record of kids.
  - The new record is added in the database(containing three files namely Récents files (computer usage data), Internet files (Internet usage data) and Image files (screen shots)) after every 4 s.
  - Data is cleaned from the database after every 5 days.

• Before the data is cleaned, data can be analysed for checking the user behavior in terms of its habit and interest. The detailed algorithm for the above mentioned steps is

*Algorithm*

```

Call form_load ( )
Call Getwindows Version ( )
Call Getuname ( )
Call Delete Record
Call Connection Open
Call Total Record
Delete 1 day old record of Recent file
Call Total Record
Call Connection
Closed
Call Connection open 1
Call Total Record 1
Delete 1 day old record of internet file
Call Connection
closed 1
Call Timer
Call RH
connection open
Fetching recent folder in file 1
Calling Total Record
Retrieving original path for 1st file in file1
Creating object to retrieve Created, Accessed and Modified
date of current file
Do loop
If path of the file matches with path in database and modified
date of file matches with the modified date in database then
f=1(i.e. data is already there in the record)
Else if f=0 and date modified of file matches with the
current date then stores record
Loop again for the 2nd file in file1 control
Call Total Record
Do Loop To print Records of Recent File in 1st path
Connection Closed
Call IH
Fetching "Title of the current open window and put it into
txt
If txt<> "New Tab" then
Then text2.text = txt
Now change event of text2.text been called as value of
"txt"comes in it .
Now call Add log and text2.text as its parameter.
Connection open 1
If value of strval="Internet"
Then fetching URL of open window and put it into text3.text
Call Total Record
Do Loop
If text3.txt=rs! Path
Then f=1
Else f=0
If f=0 and text3.text<>" " then Add Record
If value of strval="chrome, Mozilla,
etc" then

```

```

Call Total record 1
Do Loop
If strval= <>! Title browser f1=1
Else if f1=0 then add records
else if f1=1 then don't add records
Call TotalRecord1
Do loop to fetch databases records of internet files and put it
into I path
Call Connection closed 1
End
Call Timer2_Timer( )
To Capture Image of desktop
Call Generate Report1 ( )//report showing computer usage
Call Generate Report2 ( )//report showing internet usage
End
The module developed by using above said algorithm
will able to capture and store data of Recent files,
Internet file and desktop Screenshots .Comparison of
Monitoring software is shown in Table I

```

**Table I. Comparison of monitoring software**

Features	My Software (Systrack)	PC Pandora	Web Watcher
Price	Rs. 1000	Rs. 4272	Rs.6105
Monitor			
• Visited website	✓	✓	✓
• Computer operations	✓	✓	✓
Screenshots Recording	✓	✓	✓
Compatibility			
• Windows 7	✓	✓	✓
• Windows XP	✓	✓	✓
Built in Report	✓	✓	✓
• Pattern generation using Data Mining	✓	X	X

**V. CONCLUSION**

In this study, we presented a framework and algorithm to collect and analyze client Para data. The proposed algorithm was implemented using VB. The events on the client-side were collected and we exemplify user daily usage patterns across considerable software and websites. Methodology describes how algorithm helps to gather data. Future more, for cleaning data, SQL queries were applied along with the pre-processing algorithm. Based on these experiments, I affirm that the stated objectives were achieved by following the presented methodology. This paper has provided the initial foundation for research on monitoring and analyzing kid's activities. It has been deducted that computer and Internet use by the children should be supervised by parents in an intelligent manner. The main purpose in future is to build general behavior patterns from user's computer logs. What needs to be done in the future is these user daily behavior patterns can be used to build classification models, for numerous purposes.



We wish that this effort will accelerate supplementary research on leveraging previous happening and human intelligence to give valuable forecast about future events.

### REFERENCES

1. P Fraternal, Tools and approaches for developing data-intensive web application ACM Computing Surveys (CSUR), 1999.
2. Stavros Valsamidis1, Sotirios Kontogiannis1, Ioannis Kazanidis2, Theodosios Theodosiou2 and Alexandros Karakos1, "A Clustering
3. Methodology of Web Log Data for Learning Management Systems", ERIC 2012.
4. Vic Ciesielski and Anand Lalani, Data Mining of Web Access Logs From an Academic Web Site, ACM 2003.
5. [www.stopmedicineabuse.org](http://www.stopmedicineabuse.org)
6. Alexandros Kleftodimos, Georgios Evangelidis, An overview of Web Mining in Education, ACM 2013.
7. J.D. Vel'asquez and V. Palade. Adaptive web sites: A knowledge extraction from web data approach. IOS Press, Amsterdam, NL, 2008.
8. [www.internetociety.org](http://www.internetociety.org)
9. Michael Gough, Security and Privacy - How to monitor your children's use of the Internet Ketchum Global Research, Parents' Internet Monitoring Study
10. Maryam Jafari1, Farzad SoleymaniSabzchi1, Shahram Jamali2, "Extracting Users' Navigational Behavior from Web Log Data: a Survey"
12. [www.webology.org](http://www.webology.org)
13. <http://pubs.sciepub.com/ajss/3/2/3/>

### AUTHORS PROFILE



**Dr. Neetu Anand**, received her BSc. Degree from University of Delhi, her Master degree in Information Technology from GJU Hisar, Haryana. She did her PhD. in Computer Science from Lingaya's Vidyapeeth, Faridabad. Currently she is working in the Department of Computer Science, MSI, Affiliated to GGS Indraprastha University, Delhi. She has more than 19 years of teaching experience. She has published more than 25

research papers in reputed International and National Journals and conferences including IEEE, Elsevier and Springer. Her main area of research includes Web Mining, Cloud security and privacy, Big Data Analysis and Statistical Data Mining.