Implementation Of Web Analytics to Trace Visitors Transaction

Ms Kirti Gautam
M.Tech student, Department of CSE,
Jind Institute of Engineering and Technology, Jind (Haryana)

Gurdev Singh
Assistant Professor, Department of CSE,
Jind Institute of Engineering and Technology, Jind (Haryana)

Abstract: - Web Analytics is an enterprise-class based web analytics tool which provides a transparent view of website traffic & marketing effectiveness. Web Analytics has powerful & advance features that give rich insight into websites & improve website ROI (Return on Investment). This research paper consists of case study on Web Analytics, which demonstrates features Analytics provides & reports it generates. On basis of evaluation of web usage, web site owners or E-business marketers could enhance efficiency of marketing, & web traffics. This paper would also present limitations of Web Analytics & propose better approach to overcome issues.

I. INTRODUCTION

Web Mining

Web Usage Mining is application of data mining techniques to discover interesting usage patterns from Web data in order to understand & better serve needs of Web-based applications. Usage data captures identity or origin of Web users along within their browsing behavior at a Web site. Web usage mining itself could be classified further depending on kind of usage data considered:

- **Web Server Data**: user logs are collected by Web server. Typical data includes IP address, page reference & access time.
- **Application Server Data**: Commercial application servers had significant features to enable e-commerce applications to be built on top of them with little effort. A key feature of this is the ability to track various kinds of business events & log them in application server logs.
- **Application Level Data**: New kinds of events could be defined in an application, & logging could be turned on for them thus generating histories of these specially defined events. It must be noted that much end applications require a combination of one or more of techniques applied in categories above.

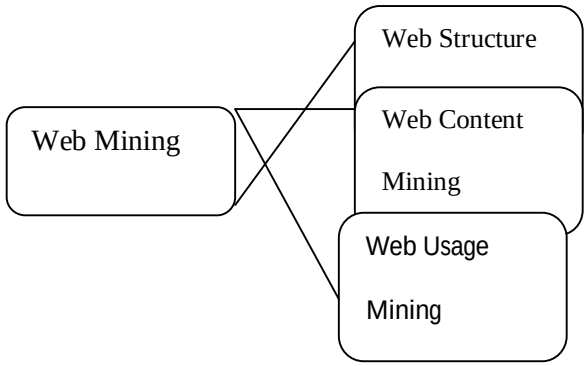


Fig 1 Web mining

Web Intelligence

Web Intelligence is the area of study & research of application of artificial intelligence & information technology on web within order to create next generation of products, services & frameworks based on internet. WI is a combination of digital analytics, which examines how website visitors view & interact within a site's pages & features, & business intelligence, which allows a corporation's management to use data on customer purchasing

patterns, demographics, & demand trends to make effective strategic decisions. As companies expand their reach into global marketplace, need to analyze how customers use company websites to learn about products & make buying decisions is becoming increasingly critical to survival & ultimate success.

II. METHOD FOR OBTAINING DATA FOR WEB ANALYSIS

There are two methods for obtaining web data:

A. Web Server Log Data

Web Server Logs give more accurate & comprehensive output, hence better insight in to website. They record various information from each request made to server. Server logs also provide important information about time & type of request, domain request, referring URL, 404 requests, broken links, & various other information. This way you could fix these problems to better accommodate your users & search crawlers. Logs are very detailed as they contain information such as Image requests, CSS files, JavaScript files, Search engine both traffic. In Server logs it is difficult to track at user level, little information about how visitors use site, requests from search engine bots could distort figures & they don't count views of pages which are cached.

B. Page Tag Data

Page tag-based web analytic systems obtain similar data as server logs. Additionally, they obtain information about client's browser, operating system, custom IDs & tags, as well as conversion data, unattainable from server logs.

III. RELATED WORK

According to Carrier & Povel (2003), general aspect of data mining consists of building of model from data. Each data mining technique could perform one or more of following types of data modeling like Association, Classification, Clustering, Forecasting, Regression, Sequence Discovery, Visualization. authors focus on point that choice of data mining techniques should be based on data characteristics & business requirements.

An Bhasin (2006) explains importance of data mining tools in extracting important information from existing data to enable better decision making throughout banking & retail industries. author further states that practical applications of data mining had been successful in areas like detecting frauds, predicting customer purchase behaviors, optimizing manufacturing process etc. Many retail industries are realizing that data mining could give them a competitive advantage. Data mining typically involves use of predictive modeling, forecasting & descriptive modeling techniques.

Faustina Johnson in June 2012 "Web Mining Techniques"

The Quest for knowledge had led to new discoveries & inventions. With emergence of World Wide Web, it became a hub for all these discoveries & inventions. Web browsers became a tool to make information available at our finger tips. As years passed World Wide Web became overloaded with information & it became hard to retrieve data according to need. Web mining came as a rescue for above problem. Web content mining is a subdivision under web mining. This paper deals with a study of different techniques & pattern of content mining & areas which had been influenced by content mining.

IV. WEB INTELLIGENCE BASED GOOGLE ANALYTICS

Web Intelligence based Google Analytics is a service offered by Google that generates detailed statistics about a website's traffic & traffic sources & measures conversions & sales. It's most widely used website statistics service. Basic service is free of charge & a premium version is available for a fee. Google Analytics could track visitors from all referrers, including search engines & social networks, direct visits & referring sites. It also tracks display advertising, pay-per-click networks, email marketing, & digital collateral such as links within PDF documents.

Features of Google Analytics

Integrated within Ad Words, users could now review online campaigns by tracking landing page quality & conversions (goals). Goals might include sales, lead generation, viewing a specific page, or downloading a particular file.

Google Analytics approach is to show high-level, dashboard -type data for casual user, & more in-depth data further into report set. Google Analytics analysis could identify poorly performing pages within techniques such as funnel visualization, where visitors came from (referrers), how long they stayed & their geographical position. It also provides more advanced features, including custom visitor segmentation. Google Analytics e-commerce reporting could track sales activity & performance. e-commerce reports show a site's transactions, revenue, & many other commerce-related metrics.

Dashboards give you a summary of many reports on a single page. Start within a dashboard within your most important performance indicators (your "Company KPIs"), then create detailed dashboards for other important topics like search engine optimization. Dashboards use drag-and-drop widgets for fast, easy customization.

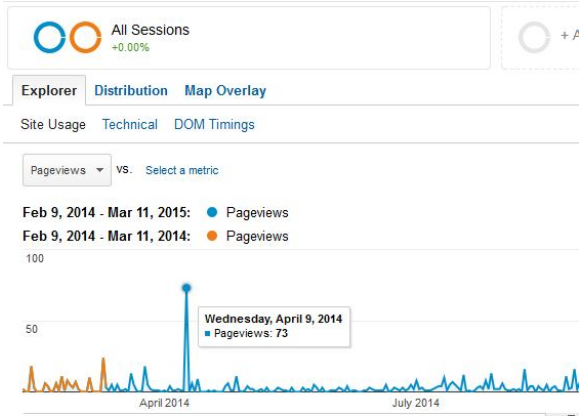


Fig.2 Map Overlay

Limitations of web analytics

1. Location of web user is traced according to countries only.
2. Data updating related to web browsing is updated after 24 hours to 36 hours.
3. Exact Ip Address of web user is not traced in this application because it changes when user is connected to network.

Advantages of web analytics

1. Repeated and unique visitors could be detected.
2. It is easy to detect Target Ratio point.
3. The number of pages viewed and the pages most visited could be checked.
4. The device is deteced whether it is PC or cell phone.
5. Bounce rate of web page is displayed.
6. Information has been displayed graphically.
7. Web Browser (ie. internet explorer,Firefox) are considered while taking readings.

V. PROPOSED WORK

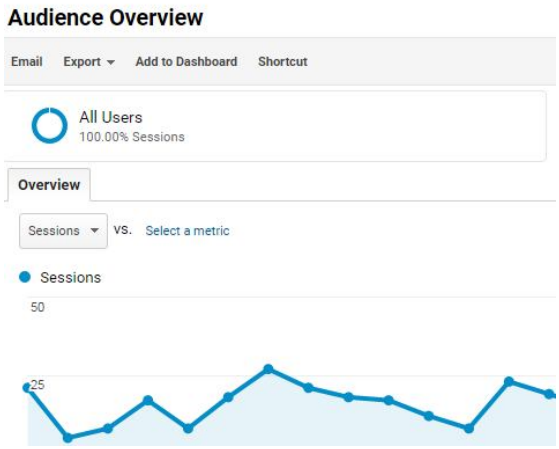
Web Intelligence is how to deal within uncertainty of information on wired & wireless Web. Adapting existing soft computing solutions, when appropriate for WI applications, incorporate a robust notion of learning it would scale to Web, adapt to individual user requirements, & personalize interfaces.

Algorithm

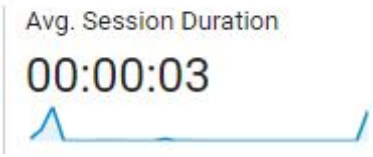
1. Get the keywords from web users.
2. Get the relevant links from database associated with keywords.
3. Trace IP address and login time of Visitor and store in database.
4. When user uses any application then API on application layer will capture application starting time.
5. When user closes application then application termination time is traced and stored in database.
6. When user switches to another web page then that would also be stored in database.
7. The Web Administrator would be able to trace how much time has been spent by user on web pages. How much time has been spent on application along with unique visitors and visitors who are visiting site multiple times with their web browser and operation system details?

SCREENSHOT OF WORKING SYSTEM

Here Repeated and unique visitors could be detected because Information is displayed graphically. It would be easy to detect Target Ratio point.



Here number of user and session of usage has been represented.



. The number of pages viewed and the pages most visited could be checked. Bounce rate of web page is displayed.

The device is detected whether it is PC or cell phone. Web Browser (ie. internet explorer, Firefox) are considered while taking readings.

Browser	Acquisition		
	Sessions ↓	% New Sessions	New Users
	354 % of Total: 100.00% (354)	25.99% Avg for View: 25.99% (0.00%)	92 % of Total: 100.00% (92)
1. google.com	135 (38.14%)	0.74%	1 (1.09%)
2. Safari	115 (32.49%)	2.61%	3 (3.26%)
3. Chrome	93 (26.27%)	91.40%	85 (92.39%)
4. Firefox	8 (2.26%)	25.00%	2 (2.17%)
5. Android Browser	1 (0.28%)	0.00%	0 (0.00%)
6. Opera	1 (0.28%)	100.00%	1 (1.09%)
7. YaBrowser	1 (0.28%)	0.00%	0 (0.00%)

VI. IMPLEMENTATION

Here we have taken keywords used by user in case of different websites. Following table represents the number of page views, Unique Visitor and Total Visitors.

SITE	Keywords	Page views	Unique Visitor	Total Visitors
1	8	50	31	73
2	12	40	20	54
3	14	55	33	63
4	7	65	40	77
5	9	51	32	76
6	11	35	30	62
7	6	25	20	34
8	5	20	15	27
9	4	15	12	19
10	7	34	21	30

Table 1 Show keyword with pages view unique total Visitors

We have created analytical chart of page views according to keywords

SITE	Keywords	Page views
1	8	50
2	12	40
3	14	55
4	7	65
5	9	51
6	11	35
7	6	25
8	5	20
9	4	15
10	7	34

Table 2 Site with keyword & pages view

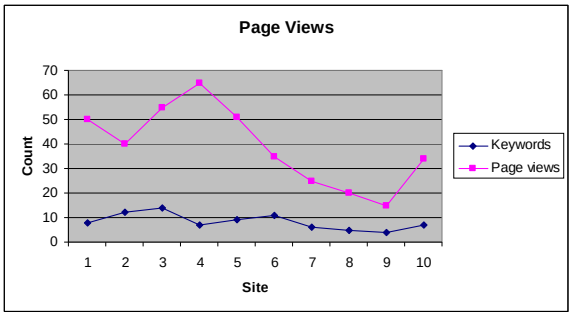


Fig 3 Chart in pages view

We have created analytical chart of Unique Visitors according to keywords

SITE	Keywords	Unique Visitor
1	8	31
2	12	20
3	14	33
4	7	40
5	9	32
6	11	30
7	6	20
8	5	15
9	4	12
10	7	21

Table 3 Site with keyword & unique visitor

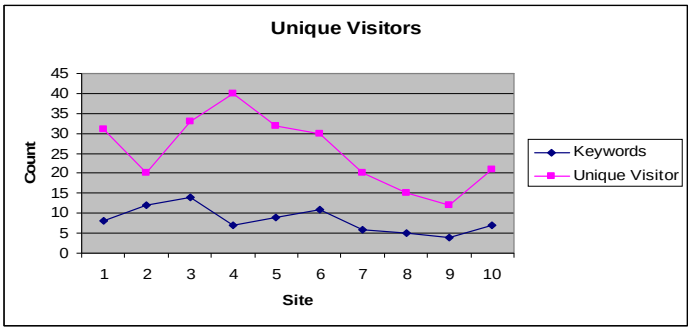


Fig 4 Chart in unique visitor

We have created analytical chart of Total Visitors according to keywords

SITE	Keywords	Total Visitors
1	8	73
2	12	54
3	14	63
4	7	77
5	9	76
6	11	62
7	6	34
8	5	27
9	4	19
10	7	30

Table 4 Total visitors

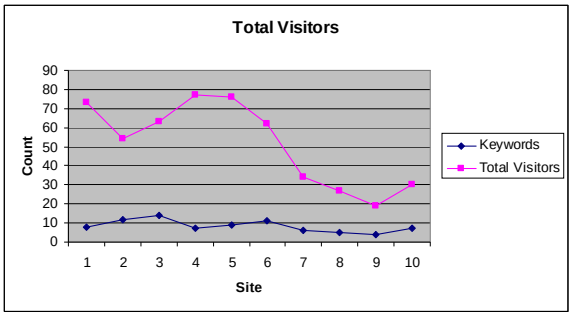


Fig 5 Show total visitor with keywords

We have created analytical chart of comparative analysis of unique visitors and total visitors according to keywords

SITE	Keywords	Unique Visitor	Total Visitors
1	8	31	73
2	12	20	54
3	14	33	63
4	7	40	77
5	9	32	76
6	11	30	62
7	6	20	34
8	5	15	27
9	4	12	19
10	7	21	30

Table 5 Comparison between unique and total visitors

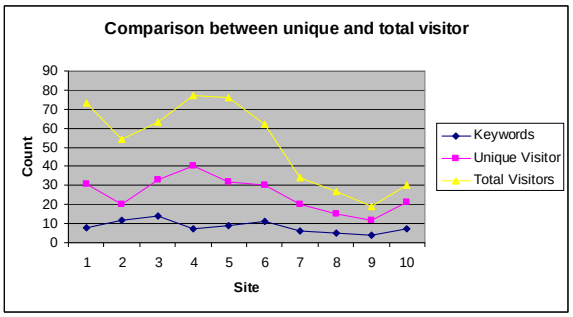


Fig 6 Comparison between unique and total visitors

We have created chart of time taken in LAN & Internet

SITE	Keywords	Page views	Time Taken in ms(LAN)	Time Taken in ms(Internet)
1	8	50	20	60
2	12	40	16	50
3	14	55	21	65
4	7	65	25	78
5	9	51	21	65
6	11	35	15	48
7	6	25	12	39
8	5	20	11	35
9	4	15	9	30
10	7	34	15	46

Table 6 Time taken in ms(LAN) within (Internet)

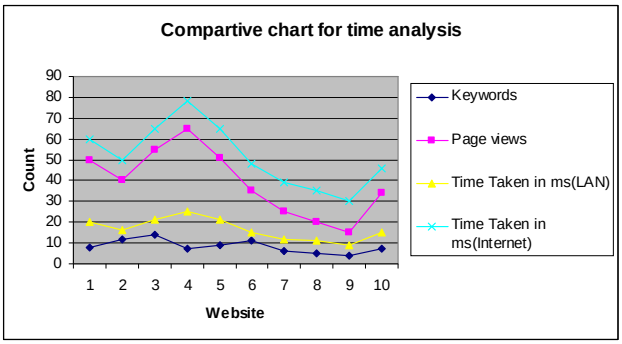


Fig 7 Comparative chart for time analysis

VII. CONCLUSION & FUTURE SCOPE

Current research within data mining mainly focuses on discovery algorithm & visualization techniques. There is a growing awareness that, within practice, it is easy to discover a huge number of patterns within a database where most of these patterns are actually obvious, redundant, & useless or uninteresting to user. To prevent user from being overwhelmed by a large number of uninteresting patterns, techniques are needed to identify only useful/interesting patterns & present them to user.

REFERENCE

[1] U. Fayyad & R. Uthurusamy, "Data mining & knowledge discovery within databases," *Common. ACM*, vol. 39, pp. 24–27, 1996.

[2] W. H. Inmon, "The data warehouse & data mining," *Common. ACM*, vol. 39, pp. 49–50, 1996.

[3] J. A. Major & D. R. Riedinger, "EFD—A hybrid knowledge Statisticalbased system for detection of fraud," *Int. J. Int ell. Syst.*, vol. 7, pp. 687–703, 1992.

[4] R. L. Blum, *Discovery & Representation of Causal Relationships From a Large Time-Oriented Clinical Database: RX Project*. New York: Springer-Verlag, 1982, vol. 19. of *Lecture Notes within Medical Informatics*.

[5] R. Heider, Troubleshooting CFM 56-3 Engines for Boeing 737—Using CBR & Data-Mining, Spinger-Verlag, New York, vol. 1168, pp. 512–523, 1996. *Lecture Notes within Computer Science*.

[6] U. Fayyad, D. Haussler, & P. Stolorz, "Mining scientific data," *Commun. ACM*, vol. 39, pp. 51–57, 1996.

[7] O. Etzioni, "The world-wide web: Quagmire or goldmine?," *Common. ACM*, vol. 39, pp. 65–68, 1996.

[8] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy, Eds., *Advances within Knowledge Discovery & Data Mining*. Menlo Park, CA: AAAI/MIT Press, 1996.

[9] "Special issue on knowledge discovery within data- & knowledge bases," *Int. J. Intel. Syst.*, vol. 7, 1992.

[10] Aggarwal K K, Singh Yogesh, Arvinder Kaur, Malhotra Ruchika (2006) Application of Neural Network for Predicting Maintainability Using Object- Oriented Metrics.Transaction on Engineering, Computing & Technology, Vol. 15.

[11] Aggarwal K K, Yogesh Singh (2008) Software Engineering: Program, Documentation & Operating Procedure. New Age International Publishers, 3rd edition.

[12] J. Furnkranz, J. Petrak, & R. Trappl, "Knowledge discovery within international conflict databases," *Applied Artificial Intelligence*, vol. 11, pp. 91–118, 1997.

[13] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

[14] S.N. Sivanandan, ‖Principals of Soft Computing‖

[15] Calvo R., Partridge M., & Jabri M., ―A Comparative Study of Principal Components Analysis Techniques‖.