Web Analytics and Metrics: A Survey

Lakhwinder Kumar
Dept. of Computer Science and Engineering
Guru Nanak Dev University(GNDU), Amritsar
contactlucky12329@ymail.com

Hardeep Singh
Dept. of Computer Science and Engineering
GNDU, Amritsar
hardeep_aph@rediffmail.com

Ramandeep Kaur
Dept. of Computer Science and Engineering
GNDU, Amritsar
ramansidhu1985@gmail.com

ABSTRACT

This is a survey paper which gives some different types of Web Analytics metrics and how the data is collected related to these metrics. As we know with the increasing need to meet customer preferences and to understand customer behavior, Web Analytics plays an important role to approach and fulfill the needs. The aim of the study is to contribute to the existing work by providing some of the key factors of change in the context of Web Analytics implementation and transition towards a data-driven analytics culture. Web Analytics has become a very important component of many web based system environment and helps in taking business decisions. This paper describes the process of Web Analytics, also the importance and the workings of summary for analytics. For the evaluation of web sites many tools and metrics are available. Basically these metrics are derived from the kind of response of the website users that tells the success of the website. Here we studied the information collected and the analysis done by various authors related to this collection. This paper will hopefully encourage a discussion and further study of Web Analytics.

Categories and Subject Descriptors

A.1 [**Introductory and Survey**]
D.2.8 [**Metrics**]: Complexity Measures, Performance measures
H.3.3 [**Information Search and retrieval**]

General Terms

Measurement, Performance. Design

Keywords

Web Analytics. web metrics, data collection.

ICACCI '12, August 03 - 05 2012, CHENNAI, India
Copyright 2012 ACM 978-1-4503-1196-0/12/08…$10.00.

1. INTRODUCTION

Web Analytics deals with the methods for measurement, data collection, data analysis and providing the related feedback on internet for the motive of understanding behavior of the customer using website. The benefit of studying behavior of the customer leads to optimize the usage of web site. As per nomenclature in [2] Web Analytics is the art and science of improving websites to increase their profitability by improving the customer's website experience. This is a science because it uses statistics, data mining techniques, report summaries, operations and a methodological process to process. It can be said as art because to improve websites there requires a deep level of creativity, imagination, analysis, balancing user-centric design, promotions, content, images and more.

The objective of Web Analytics is to provide the right direction to online users. It can be done by doing required and impactful changes in the web site. These changes are possible only when one can understand that what the customer wants. So this can be done by studying the behavior of customer i.e. how user works on website, how user will navigate from page to page and on which page the user stays more time. Web Analytics provides the ways to companies to adopt a customer-centric approach of business. Web Analytics and data management tools helps in many ways like to optimize the website, creativities for designing pages, changing the website material as well as the business. According to a survey by Forrester research [4] commissioned by Google more than 70% of the companies interviewed recognize that Web Analytics plays a significant role in the organization. To completely utilize Web Analytics, companies need to create a data-driven culture that thrives on making decisions based on statistical and analytics data rather than human feelings.

2. WEB ANALYTICS

There are mainly three fundamental pillars of Web Analytics. These can be said as methods used to enhancing the popularity of the website. These three pillars are data collection, data storage and data evaluation.

2.1 Data Collection

When decided to analyze the site visitor traffic then make it clear that exactly which data it is you are going to analyze. Proper implementation of web analytics helps to understand some basic questions like: From where online visitors come from? Which web

pages did they exit the site from? How long did they remain on the site?

This is explained in detail in section 2.6.3 "Web Analytics Process".

2.2 Data Storage

In this the storage of collected data will be analyzed i.e. analysis of possession of the data. These are of two kinds:

2.2.1 Internal Storage

To store all the collected data internally on a server, most important positive aspects of an internal storage is the data ownership. Possible costs factors comes in internal storage are hardware, software, licenses, staff, energy and security personnel as well as infrastructure.

2.2.2 External Storage & Software as a Service

Another option for websites is to store all the gathered data externally. Companies like Google offering to host data are also known as "Software as a Service" (SaaS) providers. Along with the lower costs, external storage offers additional positive aspects such as regular maintenances, software updates, software installation, technical assistance etc.

2.3 Data Evaluation

This is the last pillar of Web Analytics is the evaluation of collected and stored data. In order for Web Analytics users to have a fast overview of the most important information, proper metrics have to be chosen. Evaluation got easy by using a proper Key Performance Indicator (KPI) i.e. the basic metric on the basis of which evaluation is done.

2.4 Need of Web Analytics

The study of online user behavior and activities generates valuable marketing intelligence like:

- *Assurance*: - Measuring the performance of the website against targets.
- *Insight*: - Gaining clarity on user behaviors and needs, and how the site is meeting those needs.
- *Optimization*:-Ability to take a proper action to improve the website based on the results.

2.5 Some Basics about Web Analytics

A successful Web Analytics work requires a well-considered but simple framework that will help to address all the important aspects of the site measurement and reporting. Some important components of Web Analytics are:

- Set realistic objectives for the website.
- Set actionable and measurable KPIs.
- Implement measurement and reporting mechanisms.
- Continually improve the site through action.

2.6 Web Analytics Process

In Web Analytics process almost complete study has to be done starting from the goals, objectives of the customer till the website got implemented. In the entire process various steps involved e.g. setting up the goals, building effective metric to measure, gathering and studying the data for better approach etc. Actually it makes a cycle for the website optimization. The Web Analytics process includes the steps in Figure 1:

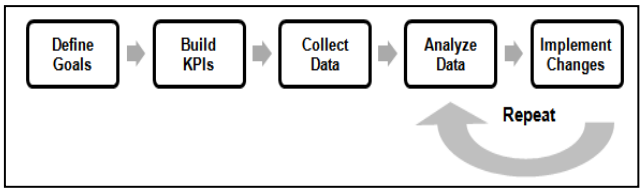

Figure 1. The Web Analytics Process.[2]

This process helps the holder of the website to measure the different costs in front of the profits made from the valuable visitors. The process helps to find out the behavior of the important customer visiting website routinely. In this one can optimize the website to improve the performance and profitability.

2.6.1 Goals

The goals helps to understand one important question that why should website exist? Website objectives are critical input that will help in recognizing the metrics that assist to measure the success. One of the key evolutionary trends in the last few years is the ability to measure the success no matter what the goals of the website are.

2.6.2 Defining Metrics

Measuring goal achievement can be done by creating KPIs that tells whether the website is getting closer to its objectives or not. There should be an action linked to each KPI proposed for a website. One important characteristic of a KPI is that it is highly adjustable. Good metrics should contain four attributes[5]:

- *Un-complex*: People with different backgrounds in a company from different departments can make decision easily by analyzing the metric.
- *Relevant*: The metric should be relevant i.e. it should properly shows the relevant result of the user.
- *Timely*: Great metrics must be provided promptly so that decision makers can make timely decisions. Even excellent metrics are useless if it takes a month to get information when industry changes every week.
- *Instantly Useful*: It is essential to understand what the KPI is, so that one can find the first blush of insights as soon as user looks on it.

2.6.3 Collecting Data

For the further analysis data should be saved. So data should be collected accurately and save on external database or local database for further analysis. Data collection is crucial to analysis results. Following are some ways to gather the information about the behavior of user:

2.6.3.1 Cookies & IP address

The cookies are the small text files have almost 4kb size. The cookies created on the user's computer when the user accessing

the website. Mainly two types of cookies are created; one is "session cookies" and the other "persistent cookies". While session cookies get deleted once the visitor leaves the website, persistent cookies remain on the visitors' computer. The reason why cookies are being used is to identify visitors. Each cookie holds a unique ID number, which makes it possible to differentiate between returning visitors and new visitors.

The IP address is a unique address, provided to each internet user. The IP address helps to identify the user that from where user is accessing the website. Mainly problem can be created if user is accessing the internet from a router e.g. in a company or hotspot. In this way the IP address of the router is accessed not the IP address of original user.

2.6.3.2 Logfile Analysis

The files required to host the website are on the web server. When visitors access and interact with the website, an entry into the log file is saved. The entry can contain information about the IP address of the user, browser type, operating system, accessed content, date and time of the access as well as the size of data sent. Every time a visitor to a website requests information from the server of the site registers this request in a log file.

2.6.3.3 Page Tagging

Page tagging is concerns to inserting a small JavaScript in every page of a website. This means that every time a visitor opens a page, this script is activated and the visitor information and actions are saved in a separate file.

2.6.3.4 Web Beacons

A common usage of web beacons is in tracking customer behavior across different websites. It tells how banner ads performing across multiple websites. Because the same server is collecting the data, reading the cookies and doing the tracking, it is quite easy for advertisers to track, anonymously, the same visitor across multiple sites or different visitors to the same site as explained in [2] [6].

2.6.3.5 Packet Sniffing

Although packet sniffing is very advanced in terms of technology, it is used mostly for multivariate testing. Its biggest advantage is that it need not tag pages; all the information goes through the packet sniffer (hardware) [2] [6].

2.6.4 Analyzing Data

To understand the customer behavior from the data, the analyst should follow a few initial steps. Any Web Analytics tool presents a summary report, a group of basic metrics that are available immediately after logging into the tool. The Figure 2 shows some metrics like:

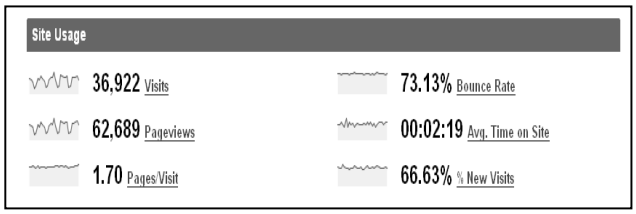

Figure 2. Google analytics basic model[2].

- *Visits*: The number of sessions on the website and number of times user interacted with the website.
- *Bounce Rate*: The percentage of single page view, visits (this metric can also have different definitions, such as a visit that last less than 5 seconds).
- *Page Views*: The number of pages that were requested in all visits.
- *Pages/Visit*: Number of pages seen in each visit.
- *Average Time*: Time on which the users stays on the web site.
- *% New Visits*: The count of the number of sessions which new users visited website first time.

2.7 Website Metrics

In section-2.6 discussed the Web Analytics process, which explains the complete step by step procedure for analysis of a website. Today there are many varieties of websites are present, to evaluate these websites some useful criteria is required. These criteria should meet three requirements. First, evaluation criteria should have sound theoretical backgrounds so that one makes it clear that no important aspect of website is missing. Second, make it sure that they are produce a reliable result regarding the outcomes of the websites. Third, evaluator can apply these criteria to the different types of websites. Some of important metrics to evaluate the website are listed as:-

2.7.1 Design and Usability

If there is poor website it can lead to loss in the use of website and its productivity. The design can be improve by taking the proper attention while designing the pages i.e. check properly the text elements, link elements, graphic elements and for each element there should be the proper formatting of each component.

2.7.2 Performance

Response time tells about the performance of a website. Response time represents the time (also known as an average time) that lasts between the issue of a request and the return of the data which is requested. If server is under full load i.e. processing a large number of requests, requests may take longer to complete as compare to the server were unloaded. Here the home page has the important factor.

2.7.3 Web Technology

The language or tools used to develop the website. A good technology have faster platform to work over the internet gives the better result. Some of the popular Web Technologies for

designing websites made their marks in the market, like: Microsoft Visual Studios, JAVA etc.

2.7.4 Content Updates

Updating the content regularly of the website i.e. remove the very old or unused data and update the new usable data into the website. If the website providing too old information it will lead to vanish the scope of the site. RSS is a format that allows users to find out about updates to the content of RSS-enabled websites.

2.7.5 Website Traffic

Website traffic used very efficiently to evaluate a website, several web traffics put themselves forward as candidates for inclusion in the study of unique visitors, page views, page hits, bounce rate and time spent on website.

2.7.6 Security

Implementing the security on the website for securing the information provided by the site i.e. avoiding hackers or any other type of the negative happenings which accessing website illegally or making the damage.

3. STUDY IN CONCEERN TO WEB ANALYTICS

With the appearance of web-based applications (e.g. ecommerce, e-learning, on-line education and information providing websites), analyzing and improving web communication has become essential in order to fulfill user expectations as told in [11]. Web Analytics is a part of web usage mining that has emerged in the corporate world, focusing on profitability in terms of how much money the web-site is making. Many metrics has been defined for judging the behavior of the user through the data collected from website in response to the user behavior. The website can be changed according to the requirement of the users, by just analyzing and operating the action against the feedback collected. Although mainly many websites have many similarities but their main objectives and techniques differ. This action implies a series of tasks as follows:

- *Data cleaning* is the process of removing all data recorded in web logs that are useless.
- *User identification* is the process of associating log entries to a connected user.
- *Session/visit identification* consisting of all logged contiguous actions performed by a user during a visit.
- *Data transformation* consists in calculating new attributes from existing ones, respectively, considering other sources of data such as structure or content information.
- *Data formatting* which is used for a better access. Solutions presented in use relational databases [12].

After the preprocessing phase, which can be very complex and time-consuming, pattern discovery techniques can be applied to data. There are many methods from various fields such as data mining, statistics, pattern recognition that can be applied on web usage data.

3.1 Web Analytics 2.0: Empowering Customer Centricity

In the paper [2] the author discussed the importance of additional sources of data to understand customer's behavior and the overall website performance as compared to competitors. Author explains that the click stream analysis is an important source to understand customer behavior. It explains only what is happening in the website. In his paper the focus on the "how much", the "why" and the "what else" user demands and get from the website. It is a new way to think about web data, a new way to think about new sources of data, which builds the complete picture of customer behavior in websites.

3.2 The Continuous Service Usage Intention in the Web Analytics Services

To predict continuous usage intention of the client companies, author adopted the three variables i.e. Satisfaction, Service Usage Period and Switching Costs, which have been considered as important in the areas of organization, strategy, and marketing. Results of the findings are:

First, a continuous usage intention in Web Analytics services was significantly and positively associated with satisfaction. This implies that Web Analytics services, satisfaction would be an important factor. Improving a service quality of Web Analytics helps service providers to holding their customers. Satisfaction is significantly associated with continuous service usage intention. When there are multiple service providing firms or when the service of the existing partner is unsatisfied then this firm may considered as to quit from the competition.

Second, the switching cost was not related to continuous service intention. Web Analytics service markets are not mature yet. However, it might find a clue for interpretation in that generally there is little investment on the transaction specific asset between two firms in the case of Web Analytics Services. In addition, the service usage period was not significantly related to continuous service intention. The main reason may be in the basis that the service usage period of the firms which participated in this survey was not more than two years. A maximum of two years may be not enough to affect continuous service intention.

Third, the two dimensions of service quality, reliability and empathy, were significantly associated with satisfaction while responsibility was not. Responsibility may be a mandatory or a prior condition in Web Analytics Services rather than service quality itself [6].

3.3 Web Analytics for Security Informatics

Web-based security informatics, and the specific challenges, by proposing a computational framework for exploring and analyzing the Web to realize three fundamental objectives in [13]:

- *Discovery:* Discovery of security-relevant data and information.

- *Situational Awareness:* Target situational awareness, typically by making near real-time inferences about phenomena of interest based on indirectly-related observables.
- *Predictive Analysis:* Predictive analysis, to include giving early warning for conflicts and crises and forming predictions concerning likely outcomes of emerging issues and contemplated interventions.

3.4 Acting on Analytics

It has been to identify key dimensions of successful implementation of change management towards a Web Analytics.

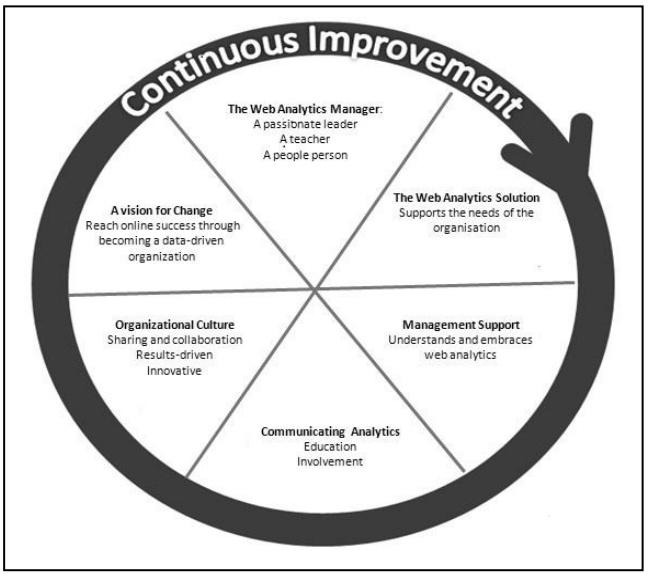

Figure 3. Key dimensions to successful change management in the context of Web Analytics [1].

The dimensions are presented in Figure 3 the representation should not be considered as a prescriptive model but as account of the change processes supporting a data-driven culture and as preliminary insight into the complexities of the subject [1].

Another point the change management that a successful change requires a group of talented, committed people responsible for the communication and implementation of the change. A key task in implementation is for the Web Analytics manager to continuously communicate the strategic importance of Web Analytics.

In this framework, communication is postulated as vital and strategic as it is said that employees will not participate in the change if they do not understand it. The results indicate that educating employees through tailored trainings and seminars, discussions and meetings is imperative for successful change achievement. Once the employees have understood the strategic value of Web Analytics they can get involved in the change. Distribution of critical insightful reports around the organization, educating employees how to use the Web Analytics tool or using testing as a persuasion tool, can raise the interest in analytics and encourage a data-driven thinking in the organization.

Furthermore, there is an important relationship between successful Web Analytics strategy and organizational culture. Hence for the booming communication of the change, it is important that the organizational culture supports sharing and collaboration. Finally, a change journey should be centered on the concept of continuous improvement. Embracing Web Analytics is a long process and thus executing the change too fast may turn to be de-motivating for employees. Dedicating time to the change and allowing the organization to which to analytically mature will yield better results.

3.5 Guidance Performance Indicator -Web Metrics for Information Driven Web Sites

In [14] the authors gave the information of various metrics and proposed a metric Guidance Performance Indicator (GPI). Particularly metrics for e-commerce sites based on transaction analysis are commonly available and well understood. In contrast to transaction based sites, the success of web sites geared toward information delivery is harder to quantify since there is no direct feedback of user intent. User feedback is only directly available on transactional web sites. They introduce a metric to measure the success of information driven web site in meeting its objective to deliver the desired information in a timely and usable fashion.

They propose to assign a value to each click based on the type of transition, duration and semantic distance. These values are then combined into a scoring model describing the success of a web site in meeting its objectives. By modeling desired user behavior patterns, the metric assigns positive as well as negative values according to the perceived success of a user session. They approach the challenge of making user feedback available without enquiring the user directly, by analyzing her behavior and visited content.

With help of the GPI metric a web site editor can discover important elements in the website structure and content, which influence user behavior. For example attractive content presentation, positive web site design or misleading navigation. Monitoring the development of the metric over time can reveal user acceptance of the web site and reactions to changes in content and design. The GPI can be applied to all kind of web sites without and can be adjusted to specific web site characteristics by fine tuning of the transition matrix.

4. CONCLUSION

This paper deals with the study of the Web Analytics, metrics used and also the some important steps before designing a website. Web Analytics services were significantly and positively associated with satisfaction. A good mechanism for controlling the quality of a website is the use of metrics. Though there are numerous general metrics are explained to measure web site usability and evaluate a website. In the paper a variety of methods are discussed to collect the information about the activities of the user, to interact with the collected information and ways to investigate this collected data.

Finally this paper explains how to implement the metrics, how to evaluate a website and collecting the information about behavior of

the user. Although this is a survey paper, in future we will present some practical metrics along with a proper framework and structure of the metrics.

5. REFERENCES

[1] Antoaneta Nikolaeva, "Acting on analytics," in A study of change management implementation in the context of Web Analytics, pp. 2-5, 2010.

[2] D. Waisberg and A. Kaushik, "Web Analytics 2.0: Empowering customer centricity," SEMj.org, vol. 2, no. 1, 2009.

[3] O. R. Zaiane, "Web usage mining for a better web-based learning," presented at the Proceedings of Conference on Advanced Technology for Education, pp. 60–64, 2001.

[4] Joe Stanhope, "The Forrester Wave™: Web Analytics, Q4 2011," Forrester Research, Inc. Reproduction Prohibited , pp. 3-6, October 2011.

[5] T. J. I. C. for Web Standards (JICWEBS), "Reporting standards website traffic", Auditing Bureau of Circulations electronic (ABCe), Saxon House, 211 High Street, Berkhamsted, Hertfordshire, UK, Report 1, January 2011. URL http://www.abc.org.uk/Corporate/AboutABC/documents/JICWEBSWebTraffic.pdf

[6] Jaesung Park, Kyungho Jung and Yunhee Lee, "The Continuous Service Usage Intention in the Web Analytics Services", Hawaii: IEEE 42nd International Conference on System Sciences, pp. 1-6, 2009.

[7] S. Dragos and R. Dragos, "WATEC: a Web Analytics Tool for Educational,Content," KEPT 2009, vol. Selected Papers, pp. 320–327, 2009.

[8] Brian Clifton, "Advanced Web Metrics with Google Analytics", Wiley Publishing, 2010.

[9] Sanda and Maria Dragos, "Why integrated E-learning Analytics are the best solution?", IEEE 2011, pp. 393-398.

[10] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: discovery and applications of usage patterns from web data," ACM SIGKDD Explorations Newsletter, vol. 1, no. 2, pp. 12–23, 2000.

[11] F. Facca and P. Lanzi, "Recent developments in web usage mining research," Data Warehousing and Knowledge Discovery, presented at the Proceedings of Springer Conferences of Computer Science in Heidelberg, pp. 140–150, vol. 2737, 2010. Kristin Glass and Richard Colbaugh, "Web Analytics for Security Informatics" , IEEE: European Intelligence and Security Informatics Conference, pp. 214-218, 2011.

[12] Catherine Dwyer and Starr Roxanne Hiltz, "Using Web Analytics to Measure the Activity in a Research-Oriented Online Community", presented at the Proceedings of the Tenth Americas Conference on Information Systems, New York, pp. 2668-2672, 2003.

[13] Kristin Glass and Richard Colbaugh, "Web Analytics for Security Informatics", IEEE: European Intelligence and Security Informatics Conference, pp. 214-219, 2011.

[14] Carsten Stolz, Maximilian Viermetz and Michal Skubacz, "Guidance Performance Indicator-Web Metrics for Information Driven Web Sites," IEEE: International Conference on Web Intelligence, pp. 186–192, 2005.

[15] N. C. Romano, C. Donovan, H. C. Chen and J. F. Nunamaker, "A methodology for analyzing Web-based qualitative data," Journal of Management Information Systems, Vol. 19 (4), pp. 213-246, 2003.

[16] B. J. Jansen and A. Spink, "How are we searching the world wide Web?: a comparison of nine search engine transaction logs," Information Processing and Management, Vol. 42 (1), pp. 248-263, 2006.

[17] L. Tooth, R. Ware, C. Bain, D. M. Purdie and A. Dobson, "Quality of reporting of observational longitudinal research," American Journal of Epidemiology, Vol. 161 (3), pp. 280-288, 2005.

[18] A. C. Vasconcelos, "The role of professional discourses in the organizational adaptation of information systems," International Journal of Information Management, Vol. 27 (4), pp. 279-293, 2007.

[19] M. Cutler and J. Sterne, "E-Metrics: Business Metrics For The New Economy", NetGenesis, Chicago, IL, 2000.

[20] A. Phippen, L. Sheppad and S. Furnell, "A practical evaluation of Web Analytics", Internet Research, Volume 14, Number 4, pp. 284-293.