RESEARCH CHALLENGES IN WEB ANALYTICS – A STUDY

Mr. K. Suresh Kumar[1], Dr. S. Vijayarani[2]

Ph.D. Research Scholar, Associate Professor
Department of Computer Science, Bharathiar University, Coimbatore

---***---

Abstract

In today's digital world, an enormous volume of data is generated from the internet, social media platforms, and smartphones. Identifying the significant insights from these data needs in-depth analysis; hence data analytics has become vital. There are different domains have produced different kinds of data. It is necessary to apply the proper techniques, strategies, and effective algorithms for analyzing these data. This paper mainly focuses on web data analysis. Several numbers of websites are available which contain many web pages. Web Analytics is the progression of exploring web data to find significant knowledge. This paper will cover the basics of web analytics, usage, features, processes, and research challenges.

Key Words: Web Analytics, E-Commerce, Web Sites, Tools

1. Introduction

Web Analytics is a method and technology for the quantification, exploration, and visualization of web usage applications and websites. Web analytics is typically supported to examine the performance of the simple function of Hypertext Transfer Protocol (HTTP) traffic cataloging to a global suite for the procedure for tracking the data, analyzing data, and reporting the interpretations [2][4]. This process involves reviewing, reporting, and data tracking to quantify web activities together with the website use and their components like web pages, images, and videos [3][5][6]. It enables a business to retain customers, attract more visitors, and increase the money volume each customer spends. The web analytics industry and market are also growing with plenty of tools, platforms, jobs, and businesses. Further, it is not just a method for determining web traffic but it is an instrument for market and business research, assessment, and improving website effectiveness.

It also helps companies to extend the outcomes of traditional values or conduct marketing drives. It offers material about the website visitor count and the page views count; or creates user behavior profiles. Data gathered over web analytics may comprise traffic flow sources, referring websites, paths taken, page views, and transformation rates. This paper discusses the basic concepts of web analytics and various research issues.

The noteworthy tasks conducted during web analytics are collecting data from heterogeneous platforms, and data analysis of the data produced by the website visitors and the people who are networking with a website. Web analysis determines to measure customer activities, enhance the website's consumer capability; drift, and advance perceptions that support to increase the business objectives. Analysis of the Web is the procedural study of online/offline designs and trends. It is a system that is employed to accumulate, quantity, describe, and analyze website information. It is used to inspect the behavior of a website and improve its web usage data. It is a strategic method of collecting data and generating reports.

1.1 *Usage and Features*

The basic feature of web analytics is the collection and analysis of website usage data. Web analytics is utilized for various purposes in many numbers of applications. Some of the noteworthy applications are monitoring traffic on the web, optimization techniques for e-commerce sites, marketing/advertising, web development software, information system architecture, online website performance improvement, web-based courses/programs, etc. Some of the most important uses of web analytics [2][7][8] are:

Improving user experience and web application design on the website. The web analysis includes optimizing website information architecture, navigation, presentation and layout of the user interaction, and web content. It also benefits user awareness and reaction forecast and improves web application features. For example, the heat map with a graphical representation of data uses a system of color coding to represent different values and is most commonly used to show user behavior on specific web pages or webpage templates.

Enhance e-Commerce and improve e-CRM. Most companies analyze and report online website usage data to understand customer and visitor behavior patterns to

acquire and retain customers through stronger customer relationships to increase traffic and increase their revenue. Websites want to keep customers to encourage users to retain them and to make every visitor end with the completion of targeted action.

Tracking and measuring commercial campaigns. E-CRM is an Electronic customer relationship management application or Internet-based technology like electronic mail, chatting facility, discussion forums, websites, and additional stations to succeed in CRM objectives. It is well structured and processed in sales, customer services, and marketing. It tracks customers' history in real-time, creates and conserves a systematic data bank, and elevates customers in three aspects, they are maintenance, expansion, and attraction.

Identifying problems and improving the performance of website applications. The study shows a significant interconnection between page-load time and the likelihood of a user converting. Web analytics helps to address this issue. Page loading measures the average page visible time by internet browsers and environmental site is used to measure the performance. The actual time and ancient performance study allow for dynamic detection, investigation, and diagnosis of performance issues. Enhancements may start from normal image optimization to alteration of the termination date in the HTTP headers to strengthen the internet browsers to practice cached website content. A heat map might help to reveal website errors, such as when users click on buttons or images without links.

1.2 Limitations

Google search engine is not able to identify and track everything that takes place on a website. Information found in the reports contains only approximations [1][5][8].

Missing information

Several individuals will determine on the view of Google Analytics; they chunk cookies and JavaScript switch off and this is how Google trails the behavior of visitors. The above movement is not traced primarily to under-reporting.

Unwanted information

There are two foundations of undesirable facts that tip to over-reporting:

1. ***Internal Trafficking*** - The default user data provided through Global Capability Center (GCC) analytics comprises traffic from various users.

These data are considered unwanted but if data are concerned with the traffic to user pages by the universal community, will be essential to sort activities using the "External Traffic only" view.

2. ***Robots*** - The Google Analytics method is used to track user activity that excludes most of the robotics activity.

2. Web Analytics Tools

A group of web analytics tools have been developed and are available on the market. It aims to obtain quantitative and qualitative data in the decision-making process. The topmost important web analytics tools [9] [11] [12] are,

R and Python

R and Python are the top programming languages used in Data Analytics. R is an open source used for Statistics Analytics whereas Python is a high-level, interpreted language with easy syntax and dynamic semantics.

Microsoft Excel

It is a spreadsheet software program to organizes numbers and data with formulas and functions. It is one most popular tools for Data Analytics. Microsoft Excel provides operators with various features, for example, the distribution of workbooks, operational instantaneous teamwork, and adding figures to Excel.

Tableau

It is a market-based business Intelligence tool used to analyze and visualize data in an easy format. It allows us to work on the live dataset and spend more time on Data Analysis rather than Data Wrangling.

Rapid Miner

It is a Data Science and Machine Learning platform that analyses the collective impact of organization employees, expertise, and data. Rapid Miner is an objective for data processing, building Machine Learning models, and deployment.

KNIME

The acronym of KNIME is Konstanz Information Miner. KNIME is an open-source and free tool that is particularly used for the analysis of data, interpretation, and visualization and for providing an environment for integration for analytics on GUI-based workflow.

Power BI

It is an assortment of software facilities and connectors that effort together to turn our unrelated Microsoft product used for business analytics. It offers cooperative visualizations with business intelligence (BI) competencies, where end users can generate dashboards and reports on their own, without having to be determined by anybody.

Apache Spark

It is a unified analytics engine for large-scale data processing including built-in modules for SQL. It is a cluster computing structure that is open-source and is used for real-time processing. The main goal is to sustain rapidity in handing out excessive datasets in terms of the waiting time interval between running the data set and the queries.

QlikView

It is a leading Business Intelligence and Analytics Platform. It is designed for creating, reading, writing, and modifying Business Intelligence reports using QlikView. Its major role is to accelerate business value through data such as Data Integration, Data Literacy, and Data Analytics.

Talend

Talend is one of the ETL tools for data integration that is available in the market and is developed by Eclipse's graphical development environment. It lets to easily manage software solutions for data preparation, data quality, data integration, application integration, data management, and bid data.

Splunk

Data generated from machines, different websites, and various applications, this Splunk tool is used to perform different functions like searching, analyzing, and visualizing the results. It evolved products in various fields such as IT, Security, DevOps, and Analytics.

3. Web Analytics Process

There are six major steps of a web analytics process. This is presented in Figure 1.

> **Business Goals Setting:** This step is essential for describing the vital metrics that will govern the victory of the business industry and website.

> **Data Gathering:** This step collects the required information, statistics, and data about visitors to the website. This is done by using analytical tools.

> **Data Handling/Processing:** This step plays a vital role that converting the gathered raw data into significant ratios, KPIs, and other information.

> **Reporting the Data: The generated and/or processed results are d**isplayed in a user-friendly manner that can be easily understood by the users.

> **Online Strategy Development:** Building a strategy to augment the website understanding to meet business goals.

> **Experimenting**: Govern the finest approach to improve website performance by using various tests.
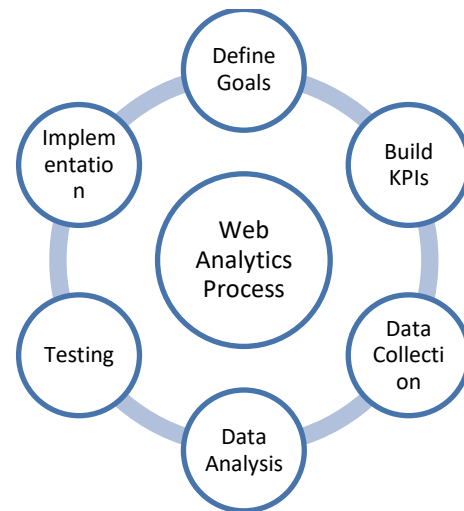

Figure 1. Web Analytics Process

4. Research Challenges

4.1 Data Privacy in Web Analysis

Data is undergoing a dramatic evolution. Users' personal data are collected by business industries, governments, and other organizations. The personal data contains the user's information, which may create a privacy issue because cookies, IP addresses, or login data are used to identify users. Web analytics tools are responsible for providing support to identify visitors' communications with websites. User data like select web pages, history of click behavior, location of access, browser settings, and access data must also be protected. Data privacy becomes

especially critical; if web data and metrics (including page views, visits, or conversion rates) are linked to sales data or personal data such as name, e-mail, postal address, or credit card number. Most countries have passed privacy or data protection acts to guarantee the correct and confidential handling of personal data.

However, if privacy policies do not conform with national or international privacy acts, if they are vaguely formulated, and if declarations regarding data collection are imprecise, incomplete, incorrect, or missing, trust and positivity of visitors towards websites can be disturbed and personal interactivity with websites or companies negatively influenced. In light of this, privacy policies, and declarations should not be understood as regulatory barriers, but rather as a bridge between visitors and websites. They are basics to establish long-term, profitable relationships between visitors and website operators, and between online customers and (electronic business) companies.

Research to study data collection and declaration on websites, besides reading and personally analyzing the privacy policies and declarations of all corporate websites, the tool Web Analytics Solution Profiler (WASP) was used to identify the web analytics systems used on every website. WASP is available as a plug-in for Mozilla Firefox and identifies more than 200 web analytics tools, using client-side data collection methods. Server-side data collection methods or other methods like reverse proxies or packet sniffing, cannot be analyzed by WASP and are not considered in this study. The privacy policies, terms, and disclaimers of the corporate websites were analyzed personally in this regard

> ➢ The declaration of the usage of cookies
> ➢ the collection of IP addresses
> ➢ The mention of the name of a web analytics tool.

4.2 *E-Commerce Sites Evaluation Using Google Analysis*

The success of an E-commerce site is related to how easy use it by the users. This research explores how to utilize Google Analytics software [10] [11] [12] in advanced web metrics. It is used to evaluate E-commerce sites' overall usability and identify future usability problems in several areas. E-commerce site managers can effectively improve the convenience of their websites. The research requests appropriate web metrics that are mandatory and useful for quickly indicating usability problems in the different areas and specific locations in an E-commerce site. However, this is very difficult to provide complete details about the particular issues that might be present on a webpage.

This research involved three E-commerce case studies. It compared the usability findings indicated by GA software to an analytical evaluation of the sites conducted by experts. To use GA software to track usage of the e-commerce sites it was necessary to install the required script on the company's websites. The site owners recognized crucial business practices for each spot and GA was set up to evaluate the usability of web pages come upon by users in finalizing these processes. The heuristic evaluation involved devising a set of comprehensive heuristics, specific to e-commerce websites. Detailed review and thorough investigation of the HCI literature have derived six different categories. They are internal search, content and design, navigation, customer service, architecture, and purchasing.

This research identified specific web metrics that can provide quick, easy, and cheap, indications of general potential usability problem areas on E-commerce websites. In some occasions, they can also be used to recognize difficulties on precise summons, however, not in all cases. By distinction, the outcomes exhibited that the experimental assessors were able to find complete explicit usability complications. The above results set a key for future research. The following stage is to progress a structure using GA as an initial phase in the course of ascertaining usability snags of an E-commerce website. Website methods including heuristic evaluation will also be used to identify the specific usability problems in the particular areas and pages on the website indicated by the web metrics. The goal is to provide a framework that enables specific usability problems to be identified quickly and cheaply by fully understanding the advantages or disadvantages of the various usability methods.

4.3 *Improve Online Access to Archival Resources*

According to the Web Analytics Association, [9] [10] Web analytics is "the measurement, collection, analysis, and reporting of Internet data for understanding and optimizing Web usage." In the past, it was also defined as "the study of the behavior of website visitors. Web analytics applications provide accurate and deep information about the actual use of online resources. Analytics software can also aid an archivist in apprehending how a certain resource type is being used.

In the early days' archives implemented a pilot project to test the utility of Google Analytics for measuring and analyzing the use of websites. The project provides advanced information about online use and demonstrates the concept of Web analytics. As the project progressed, it became apparent that it would allow us to identify usability blocks and improve the websites. The user uses

Web analytics as a method to implement principles of serviceable design by making small progressive changes to the existing website over several years.

Assessment of website use must be conducted against objectives that measure, in a statistically accurate way, whether the purposes or goals of the website are being achieved. The archives serve several complementary purposes. Specifically, it provides basic information about the archives (such as contact information and a general description of our holdings), promotes our programs and services, encourages users to find relevant descriptive information, and facilitates the use of online and offline materials. It would be tempting to analyze information regarding all of these goals. However, anyone conducting a user study should be reluctant to analyze a large number of data points, since many will likely be irrelevant to the task of understanding user behaviors and usability problems. It makes more sense to analyze closely defined data that speak to a well-defined hypothesis. The Web measurement literature notes that a successful analytics program will collect a limited amount of data and will seek to answer a few discrete and measurable questions based on an analysis of Web traffic over a defined period.

Web analytics software can be configured to provide specific types of information if you have precise questions in mind. Based on the user trends, it was most important to find out whether the website design encouraged or discouraged archival use and contacts with archival staff. Not only are such contacts reported to the administration as evidence of their use and effectiveness, but they also represent one way of measuring whether online services have an impact on users.

In this study, by analyzing the data around this specific question we can prove the study by identifying specific actions that take place to improve the website and make decisions in future services. In the future, to improve the data analysis in online access the archival resources, the main four questions must be followed by the typical users. The questions are:

1) Which parts of our website are most heavily used?

2) How do people reach our site?

3) What are the most popular searches on our site?

4) How do users navigate through our website (and in particular, which pages lead them to contact us)

5. Conclusion

This research aim is to increase the level of awareness of the intellectual and technical issues surrounding web analytics user data. Major attention is currently being paid to data science technology and government. In recent years have seen rapid growth in parallel and distributed computing systems as the backbone of the modern Internet-based information ecosystem. These systems have fueled search engines, E-Commerce, social networks, and online environments and provide the platform for data analysis issues. The main aim is to turn data into knowledge and to support effective decision-making important inferences derived [8] from the study are cause and effort and this should be derived based on the effort and cause of the relations. It is related to data variables, classes, samples, and groups. Generalizations are the strengths with simple structures, small strengths that can find certain general characteristics among themselves can be generalized. Data Reporting is organized data. Along with the results in the visualized form, the data should be in a report that is mined as a data report. The final report should contain the data along with research, methodology, and review of the final step of data analysis.

6. References

1. Chatham, B. (2005b). Web Analytics Cookies: Perception Versus Reality. Cambridge, MA: Forrester Research, Inc.

2. Chatham, B. (2005c). Web Analytics Data: The Truth Is Out There. Cambridge, MA: Forrester Research, Inc. Elliott.

3. Malacinski, A., Dominick, S., Hartrick, T.: Measuring Web Traffic, Part1, http://www.ibm.com / developerworks/ web/library/wa-mwt1.

4. McFadden, C.: Optimizing the Online Business Channel with Web Analytics, http://www.webanalytics association. org /en /art/?9.

5. Web Analytics Association, http://www.webanalytics association.

6. Kaushik, A.: Web Analytics, an Hour a Day. Wiley Publishing, Inc., Chichester (2007)

7. Peterson, E.: Web Analytics Demystified. Celilo Group Media and CafePress (2004)

8. A Visual History of Web Measurement, Web Site Measurement,
    http://www.websitemeasurement.com

9. Fang, W.: Using Google Analytics for Improving Library Website Content and Design: A Case Study. J. Library Philosophy and Practice, 1–17 (2007)

10. Haywood, M. E., & Mishra, A. (2019). Building a culture of business analytics: A marketing analytics exercise. The International Journal of Educational Management, 33(1), 86-97.

11. Agresti A (2013). Categorical Data Analysis, 3rd ed. John Wiley & Sons, Inc., Hoboken, NJ.

12. Alreck PL, Settle RB (2003). The Survey Research Handbook, 3rd ed. McGraw-Hill/Irwin.