

Real Time Visualization of DSpace Usage: An Integrated Framework for Libraries

Sukumar Mandal

Assistant Professor & Faculty in-Charge

Department of Library and Information Science, The University of Burdwan, West Bengal, India

E-Mail: sukumar.mandal5@gmail.com

(Received 24 December 2018; Accepted 16 January 2019; Available online 8 February 2019)

Abstract - Web analytics is also known as cloud analytics which helps to display the information regarding site metrics from massive data. This can be achieved in two ways via the user's web browser such as statistical and visualization of a particular digital web based interface. This paper explores high level digital archiving software by integration of DSpace with web analytics popular tool Piwik. The paper discusses to create a single window web analytic interface to represent the user log and present the DSpace Usage. For this the researchers select DSpace, an open source software as it is free and popular. They integrate DSpace software with Piwik web analytic tool. Piwik, a web analytic interface tool has been chosen due to its successful integration with DSpace and it is also a free web analytics tool. The whole integration is being performed in Ubuntu operating system due to its reliability and good rate of performance. After successful integration of DSpace with Piwik in Ubuntu Operating System, researchers found that there are multiple colourful link icons showing the different community, sub-community, collection and metadata. It will also display statistical visual graph showing unique visitors, average visit duration, page views, downloads, average generation time page views, unique page views total searches on your website, unique keywords downloads, etc. Through this web analytic interface, DSpace usage may be measured easily and the users' trend at present will be displayed to rectify any demerits behind this.

Keywords: Web Analytics, Real Time Access, Google Analytics, Institutional Repository (IR), DSpace, and Piwik

I. INTRODUCTION

Presently many new innovative products are being developed to meet the various challenges in the web environment. One of the challenges is to measure the level of satisfaction of user in the web environment efficiently and effectively. To meet this urgent need Web analytics study may be chosen through open source software to show real time access data. In this paper researchers has been chosen DSpace, an open source software in the field of digital archiving, developed by DuraSpace and it is popular among the open source software to build a repository for retaining academic and scholarly content of the document.

The researcher suggests an integrated framework for collecting and providing detailed log report regarding download metrics of users. This paper also introduces an activity of current analytical study. The relevant data were retrieved from DSpace Solr logs of an institutional repository and by using Piwik, free analytical software, the

data regarding user's choice and preferences were analyzed side by side. This study presents an evidence of log file analytics data appeared through Piwik, web analytic tool.

The study also supplies original real time counts of Content Downloads. This paper adopts a new framework to classify page views and download activity in which researchers proved how users data can be visualized easily. The visualization of real time access is very important to know the trend of users' choice and preferences and count the number of users at a time.

II. RELATED WORKS

Babu and O'Brien discussed on Web-OPAC interfaces and studied how Web-OPAC helps to improve web discovery (Babu and O'Brien, 2000). Niamh in this paper conducts a case study on Institute of Technology Library and reviews the responsive Web design project and at the end he opined that it is possible for a small academic library to effectively provide enriched Library Website using responsive design (Niamh Walker-Headon, 2016). Burby and Brown defined web analytics and elaborately discussed its version 4.0.

They also discussed various traits of web analytics, its application areas (Burby and Brown, 2007). Ganapathi & Zhang showed in his paper the art of data summarization and its applications in managing systems effectively (Ganapathi & Zhang, 2011). Rapoza stated the web analytics from new perspectives and discussed innovative features of web analytics and its pros and cons with a new look (Rapoza, 2010). Serov and others opined about current practice of web analytics (Serov & Others, 2016). Peter Adams presented Open Source Web Analytics framework, its architecture, modules and graphics (Open Source Web Analytics, 2017).

Bekavac and Praničević gave a comparative study on various web analytics tools and their impact to measure the performance of a business model. Accordingly, an overview of web analytics and web metrics tools is given, including their characteristics, main functionalities and available types (Bekavac and Praničević, 2015).

Schneider, Alexander in his paper tracked 15 years' user behavior on the use of websites and proved that the users

capabilities have matured now and they prefer open source projects (Schneider, 2015). The paper discusses how to enhance the library website using Google Analytics. In this paper it is found that how users use this site, how the website has been changed according to the ease of users and their impacts on the users (Web Analytics, 2017).

III. STATEMENT OF THE PROBLEM

The activity log of Users represents the human nature and their likings and disliking. The web analytics packages built into Information Repository (IR) software platform rely on log file analysis. Though the web analytics, researchers extract relevant data on users log through visual graph to provide real time access and to produce accurate download counts, times of opening and staying of pages, etc.

IV. OBJECTIVES OF THE STUDY

The objective of the study is to examine the trend of the users' preferences to use the content of the repository as real time situation. Apart from this the successful integration of DSpace with Piwik show the real time access metrics visually, herein lays the significance of the study.

V. METHODOLOGY

DSpace as open source software has been taken for creating IR. For this researchers integrate DSpace software with Piwik web analytic tool. The total integration will be done in the Ubuntu operating system. There are some reasons behind the choice of software, tool and operating system. DSpace is taken due to its popularity in creating Institutional Repository in the world and it is effectively and efficiently perform in the web environment. Piwik is chosen due to its successful integration with DSpace and it is also a free web analytics tool. The whole integration is being performed in Ubuntu operating system due to its reliability and good rate of performance.

VI. RESULTS AND DISCUSSION

At first some papers have been selected from various journals and the required data have been entered into the database for the development of Institutional Repository (IR). Through this IR, Title, author and issue date of papers will be searched by users at any time and full text and downloaded facility will be there in this repository. After successful entry of the data, the result will be displayed as in Fig-1.

Issue Date	Title	Author(s)
Jun-2017	A webometric analysis of National Libraries' websites in South Asia	Verma, Manoj Kumar; Brahma, Krishna
Mar-2017	Analysing the effects of cultural dimensions on the quality of library websites	Chakraborty, Sarthak
Mar-2017	Bibliometrics and scientometrics in India	Garg, K.C.; Tripathi, H.K.
Mar-2017	Scientometric study of books authored by Nobel laureate Amartya Sen	Sinha, Atanu Kumar

Fig. 1 Development of IR through DSpace

The integration of DSpace with Piwik, web analytics tool in Ubuntu 18.04 Operating System, researchers showed that there are multiple colourful link icons showing the different community, sub-community, collection and metadata. If the users go through different colourful link icons they will observe four things—date and time of opening the community, name of the community or sub-community and time of opening side by side how much time user stays on the community. In this way real time visitor statistics will be displayed through Piwik—web analytic interface, that is the number of visits and actions displayed in the last 30 minutes or last 24 hours. Moreover date-wise, month-wise user real

time access through visual graph will be displayed effectively in a single window interface of Piwik. The innovative feature of this interface is to show detailed usage statistics of IR. The statistical visual graph display unique visitors, average visit duration, visits bounced (i.e. left the website after one page), actions (i.e. page views, downloads, out links and internal site searches) per visit, average generation time page views, unique page views, total searches on your website, unique keywords downloads, unique downloads, unique out links, maximum actions in one visit, etc (See Fig-2).

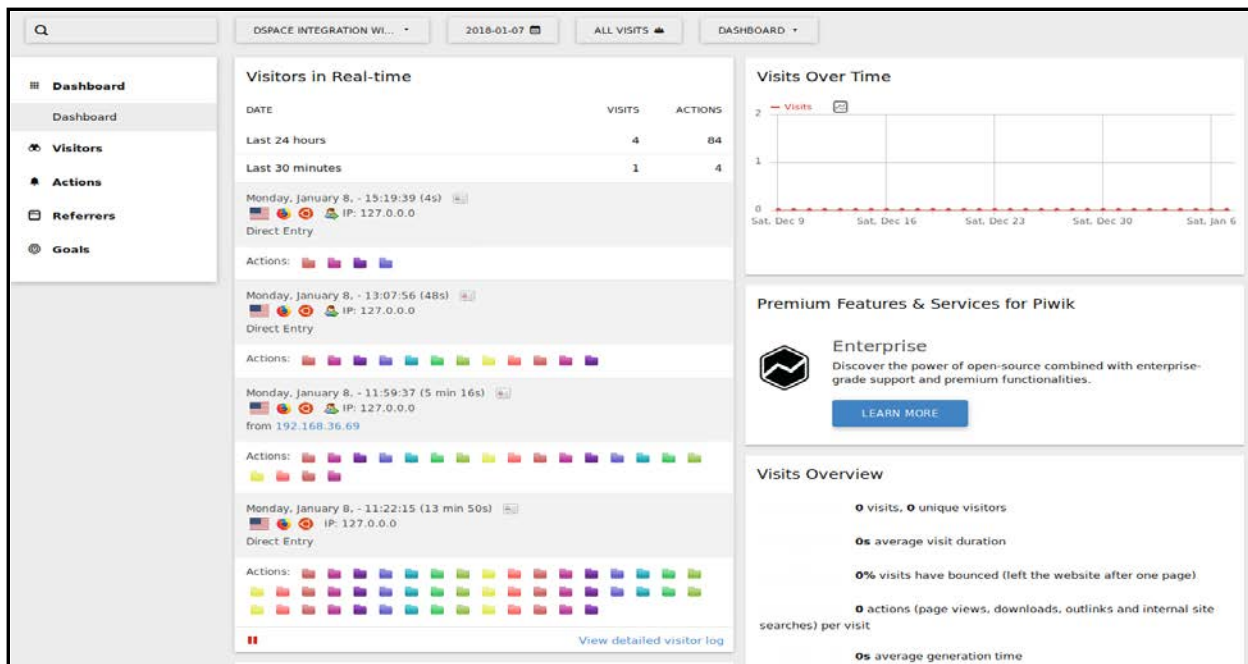


Fig. 2 Real time visualization of DSpace users in Piwik

VII. CONCLUSION

Web analytics log for IR are very much effective for tracking different activities of users like page views, real time access, number of item downloads, etc. The main feature is to ascertain the real time situation analysis of log file due to single window interface. In this case the report is really visible. Through this integration of DSpace with piwik in Ubuntu operating system, some unique innovative features like unique visitors, average visit duration, page views, downloads, time page views, unique page views, total searches on your website, unique keywords downloads, etc. have been observed. It is user friendly because the statistical graph will be displayed as per users' choice and the statistical data can be presented through pie chart, bar diagram, etc. In this case the actual result must come out to show the real time situation. Through this single window piwik interface, usage can be measured easily and the trend of present day users' will be displayed any time to show the state of the art reality.

REFERENCES

- [1] Babu, B. R. & Brien, A. (2000). Web OPAC interfaces: an overview. *The Electronic Library*, 18(5), 316-330.
- [2] Bekavac and Praničević (2015). Web analytics tools and web metrics tools: An overview and comparative analysis. *Croatian Operational Research Review*, 6, 373-386.
- [3] Burby, J., & Brown, A. (2007, August 16). Web Analytics Definitions - Version 4.0. Retrieved from <http://www.digitalanalytics.org/standards> (Accessed on 03.01.2018)
- [4] Dumbill, Edd. Data visualization. Retrieved from <https://www.fusioncharts.com>. (Accessed on 6th January 2018)
- [5] Ganapathi, A., & Zhang, S. (2011). Web Analytics and the Art of Data Summarization. In *Managing Large-scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques*, 6:1-6:9. New York, NY, USA: ACM. (Accessed on 03.01.2018)
- [6] Niamh, Walker-Headon. (2016). Responsive Web Site Development at the Library, Institute of Technology Tallaght: A Case Study. *Journal of Web Librarianship*, DOI: 10.1080/19322909.2016.1229147
- [7] OpenWebAnalytics. (2017). About Open Web Analytics. Retrieved from <http://www.Openwebanalytics.com> (Accessed on 05.01.2018)
- [8] Rapoza, J. (2010, December 2). Web Analytics: A New View. *Information Week*. Retrieved from <http://www.informationweek.com/web-analytics-a-new-view/d/d-id/1094560> (Accessed on 04.01.2018)
- [9] Schneider, Alexander (n.d.). Using Web Analytics Data to support Social Software Users Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.4012&rep=rep1>. (Accessed on 05.01.2018)
- [10] Serov, Igor, Leitner, Maria & Rinderle, Stefanie. (2016). Current Practice and Challenges of Data Use and Web Analytics in Online Participations. Retrieved from https://www.researchgate.net/publication/315788677Current_Practice_and_Challenges_of_Data_Use_and_Web_Analytics_in_Online_Participations (Accessed on 05.01.2018)
- [11] Web Analytics (2017). About Web analytics. Retrieved from http://digitalcommons.macalester.edu/cgi/view_content.cgi?article=1105&context=libtech_conf. (Accessed on 2.1.2018)