Evaluating the indicators of web resources operating in the ccTLD domain zone as a key component of national security

Yusif Sadıyev

Institute of Information Technology of ANAS, Baku, Azerbaijan
sadiyev.yusif@gmail.com

Abstract—**Currently, the number of web resources and Internet users is rapidly increasing as demand of time. This process generates Web Data, collected from web resources, within Big Data. Conducting multidisciplinary analytics of the web resource based on the collected data is an important issue for the country. Because the characterization of processes in the national domain zone of the country and predictive information depend on the analytical process mentioned. Web resource analytics not only help to generate statistical reports, but also help to detect situations that might be a threat to the National Information Security on time. It shows that analytics of web resources are a key component of the country's National Information Security.**

Keywords—web analytics, web content mining, web usage mining, web management, user profiling, e –Government.

I. INTRODUCTION

Analytics of web resources in the country code Top Level Domain (ccTLD) constitutes the basis of National Information Security, which is an important component of the national security of the country. It also allows analyzing citizens' online relationships with e-government.

Analytics of web - resources allows prepare reports on the following topics [2-4]:
• User behavior and profiles,
• The number of applications in the web-resource,
• Keywords sought by the user,
• Geography of visitors,
• Duration of users' stay on the web resource,
• Switch between pages, etc.

The e-government's web infrastructure is formed relying on the web resource of every agency, operating at public sector. Formed web infrastructure creates online interactive communication among organizations and citizens [1].

It should be carried out the analytics of web resources, in particular multi-purpose analytics in the process of establishing online relationships among citizens and governmental bodies. In this case, data collected in decision-making to improve the performance of web resources and increase effectiveness are used [5].

Detecting technical errors in web resources, operating in the e-government environment, identifying content-oriented errors and adapting to user requirements can be accomplished by effective web analytics. By using web analytics, it is possible to set user and visitor interest rates, identify new users' interests, and investigate the reasons of leaving the web site for regular users [6].

Timely identification of cases such as terror threatening security with the help of Web Content Mining and Web Usage Mining is now actual and relevant today [7].

The main task of the National Statistical System is to provide feedback between users and the e-government and to provide predictions by applying statistical mechanisms by archiving all the information collected. The National Statistical System is set up according to the national interests of the country and all analytical process is reflected there.

There are a large number of platforms involved in web analytics. Web sites of countries without a National Statistical System are analyzed by these platforms. In many cases, service policies of these platforms do not meet the national interests of the country. And analytics of web -resources, which is of critical importance for the country, is not accepted by taking into account the risk factor. As an example, we can show the e-government portal that offers online services to citizens. In this case, web analytics can now be regarded not only as a statistical reporting process, but also as an essential component of National Security: a key aspect of National Statistical System, one of the National Information Security goals.

It is known that, the users' behavior data is collected by sensors due to evaluate the performance of the web resources. The data collected allow for multi-purpose statistics. During multi-purpose statistics, it is possible to improve the work of the web - resource, assess security, conduct users' behavior map and other statistical reports. All of these can be considered as statistical mechanisms.

The first step in statistical mechanisms is the collection of initial information from web resources. This information is collected through sensors. Basically, two methods are used [8, 9]:

• Log Files – data is collected on server log files, and analytics is done with log analytics applications
• Page Tagging - is the collection of data through a web page code (script)

Both methods of data collection is used in the mechanisms of web resources statistics. Platforms, offering analytics services, are mostly used by Page Tagging. In the page tag method, the web page provides a "code" placed on the page.

It is a method of collecting, storing and processing of data, which is more relevant to the country's national interests, through Log files. However, big resource is needed in this case.

The page tag method is used widely. The missing key element of this method is to collect, store, and analyze data on an external server.

II. Assessment analysis of the indicators of web-resources and modern status

It is known that, web analytics of the web-resources are conducted by different platforms and these platforms use the page tag method [9, 10]. Sample: Google Anlytics, Yandex. Metrika, LiveInternet, etc. platforms.

However, these methods cannot be considered satisfactory for national interests to analyze the evaluation of the system of electronic services for citizens. And some of the subsystems of the web resources are closed systems.

In these cases, the collection and processing of data for the evaluation of the indicators of resources can be done by log files or the platforms providing analytical services are required to provide services based on national interests. It is important to note that to keep the information collected through the Page tag on the platform servers along with country's servers and use them in accordance with the National Information Security requirements is relevant.

As can be seen from Table 1, 33.8% of web resources do not use analytics services, and 66.7% use analytics services, the percentage of which is as follows:

Google Analytics	57.1%
Yandex.Metrica	5.3%
WordPress Jetpack	4.8%
LiveInternet	2.7%
Hotjar	2.1%
New Relic	1.8%
Matomo	1.3%
StatCounter	1.1%
Top.Mail.Ru	0.9%

TABLE 1: COMPARISON OF WEB-ANALYTICS PLATFORMS [11]

According to statistical figures, 178,329,0123 websites are active in the worldwide web and 66.7% of them, ie 1189420495 websites use analytics services. At this time, data collected for analytics creates Big Data. And their analysis comes to the problem of Big Data Analysis [12].

III. Conclusion

The experience of leading countries shows that analytics of web resources in the national domain zone of the country directly links to the national security issue. Leading countries with experience in this field build their own statistics system. Initial data is collected through the sensors placed in web-resource and statistics is carried out in accordance with the requirements of national interests.

The main direction of the National Statistical System is to maintain reliable the data collection from web resources, carry out analytics on existing experience and ensure the protection of the information integrity.

Analytics of web sites in ".az" domain is analyzed on different platforms because it has not got a National Statistical System, and web resources of strategic importance for the state cannot use these services for national interests. Thus, the services provided by web analytics are not used.

Acknowledgment

References

[1] Rasim M. Alguliev, Farhad F. Yusifov, "Some Actual Scientific And Theoretical Problems Of Formation Of The Electronic Government And The Perspectives Of Their Solutions " Journal of Problems of Information Society, pp. 5–7, April 2014.

[2] https://www.clicktale.com/academy/blog/a-brief-history-of-web-analytics/ (10.05.2018 09:35)

[3] Clifton, B.,"Advanced Web Metrics with Google Analytics", Second Edition, Wiley, New York, USA, 2008, page 531

[4] Google Analytics. http://www.google.com/analytics/. May 15, 2018.

[5] Kaushik A., "Web Analytics 2.0 - The Art of Online Accountability and Science of Customer Centricity", Wiley Publishing, Inc. 2010, 475.

[6] Rasim Alguliyev and Farhad Yusifov, "The Effective Mechanisms in Monitoring and Management of E-Government" Journal of Public Administration and Governance, 2015, vol.5, no2, pp. 58–60.

[7] Ramesh Yevale, Mayuri Dhage, Tejali Nalawade,.Trupti Kaule, "Unauthorized Terror Attack Tracking Using Web Usage Mining", Ramesh Yevale et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, page 1210-1212

[8] Agosti, M., Crivellari, F., Di Nunzio, G.M., "Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction", Data Mining and Knowledge Discovery 2012, 24(3), page 663–696.

[9] Cicchetti, A., Di Ruscio, D., Eramo, R., Maccarrone, F., Pierantonio, A., "beContent: A Model-Driven Platform for Designing and Maintaining Web Applications", pp. 518–522. Springer, Berlin, Heidelberg 2009, pp. 518-522

[10] Anthony Ferrini, Jakki J. Mohr, "Uses, Limitations, and Trends in Web Analytics. In Handbook of Research on Web Log Analysis", Chapter VII, 2009, pages 122 – 140.

[11] Xing, Guo, Fitzgerald, and Xu., "Google Analytics based Temporal-Geospatial Analysis for Web Management", A Case Study of a K-12 Online Resource Website. In International Journal of Information Science and Management, 2015, Vol. 13, No. 1, pages 87-106

[12] Ledford, J.L., Teixeira, J., Tyler, M.E., "Google analytics", John Wiley and Sons 2011, 434 pages.