VisitorTrack: Accurate Data for Website Visits Analytics

Jakub Ďaďo[1], Erik Dzurňak[1], Tomáš Gaššo[1], Michal Vantúch[1], Dávid Spišák[2], Dávid Slezák[2], Peter Paška[2]

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`visitortrack.team@gmail.com`

We live in times, where Web provides many ways how to do a business. A trending way to make money on the Web is to create own website offering products to customers. Owners of websites should provide better functionality and better services to their customers then vast competing business to succeed. In order to discover whether number of incoming customers is increasing or decreasing you should track your website. By tracking visitors, one can discover how they react to changes or just simply monitor website visit rate. Tracking visitors is popular and there are many tools such as Google Analytics (GA), Piwik, etc. Visitor monitoring depends on quality data. The more accurate statistics are available the better actions could be taken.

In general, many of these tools use cookies to store data about visitors in browsers to be able to track them also when they come to the website another day. But there are few researches, for example[1] which says that 30% of Internet users are used to delete their cookies (up to 4 times/month) what causes anomalies in tracking, for example, "statistics for returning visitors". When returning visitor deletes his/her cookies, he appears as a new one so we are not able to recognize him/her as a returning visitor.

We propose a solution to the issue with cookie data. Our tool tracks activities of the users on the websites, creates a biometric track of each user and then tries to recognize him/her by comparing it with others stored in the database. Our main goal is to find out whether the user is on the website for the first time or he/she could be tagged as returning user.

We also provide data which are more accurate because we can merge duplicate records of the same visitor, who deleted his cookies together. For users who likes all their analysis and statistics in one place we can offer integration with GA. Based on our deduplication algorithm we can assign unique identifiers to the visitors and pass it to GA.

Our solution possesses some similar features to Piwik and GA. Those are analytical tools which collect data about visits using cookies and send them to statistics and graphs by JS APIs. Our system is different as we also use functions which are not dependant on cookies and unlike GA our solution diminishes problems with cross sessions and multiple web browsers.

---

[1] http://www.comscore.com/Insights/Blog/When-the-Cookie-Crumbles

Our approach consists of 3 main components: logger, server, client application. Logger serves as a data collector. It could be easily included on client's website by inserting just one line of JavaScript code. We tried to make this step as easy as possible to avoid client bothering. We gather data as coordinates of user behaviour on client website. It means that we track movement and events triggered on website (etc. click on button).

These data are processed once a day. Processing is based on creation of 10 features (in form of histograms), each of them for exact value obtained from data (i.e. Pause before clicking). Deduplication of visitors takes place after creating of histograms. We filter visitors by their browser data (browser, operation system, resolution, etc.) and then we choose visitors from this filtered collection based on their mouse usage biometrics. Deduplication is based on algorithm which compares histograms of visitors. Currently the accuracy of deduplication is 84%. Error of identifying new visitor as returning visitor is in the range of 0,45 - 0,55 and error of identifying returning visitor as new is in range of 0,15 - 0,25. These values are measured on an available dataset from browsing users. Our product are statistics based on data precomputed from algorithm. Client can view statistics in client application, which is personalized by his demands. Multiple websites could be registered and each website belongs to a unique API key, which must by passed to the logger. Client can display statistics for selected periods of time. We offer statistics such as returning visitors, visitor sessions per page and beyond that we offer heat maps for visualization of users' behavior.
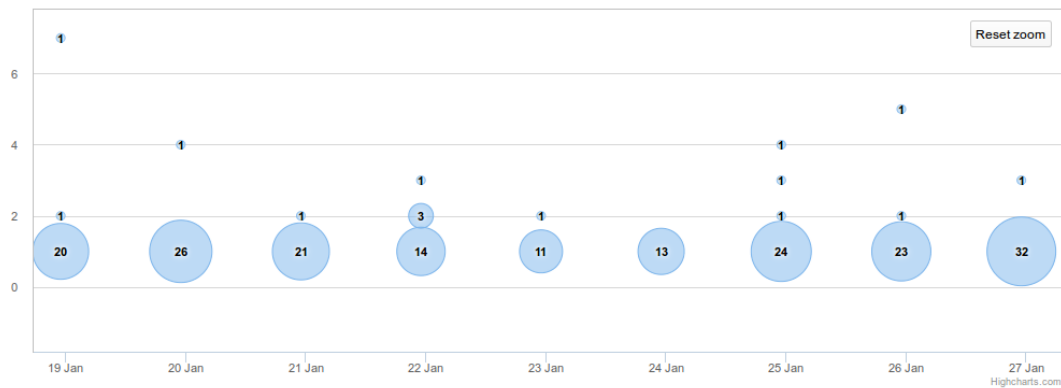

Figure 1 Graph – Sessions per user (y-axis "number of sessions"; x-axis "date")

References

[1] Daniela Chudá and Peter Krátky. 2015. Grouping instances in *k*NN for classification based on computer mouse features. In *Proceedings of the 16th International Conference on Computer Systems and Technologies* (CompSysTech '15), Boris Rachev and Angel Smrikarov (Eds.). ACM, New York, NY, USA, 214-220.

[2] Peter Eckersley, How unique is your web browser? Proceedings of the 10th international conference on Privacy enhancing technologies, p.1-18, July 21-23, 2010, Berlin, Germany

[3] E. Fettman, A Sweet Treat, But Users Delete: Cookies and Cookie Deletion in Google Analytics, 2015. [Online]. Available: https://www.e-nor.com/blog/google-analytics/cookies-and-cookie-deletion-in-google-analytics