Measuring Actual Visitor Engagement in News Websites

Razib Iqbal
Missouri State University
Springfield, MO, USA
riqbal@missouristate.edu

Matthew Scott
Missouri State University
Springfield, MO, USA
matthew529@live.missouristate.edu

Tarah Cleveland
Valley City State University
Valley City, ND, USA
tarah.cleveland@vcsu.edu

ABSTRACT

As revenues from Internet advertising continue to grow, advertisers seek popular news websites for placing advertisements in an effort to maximize profits. An important measure of how well a website is performing or how attractive it is to the advertisers is how engaged the web visitors are with that website. During our background study, we explored articles covering metrics to measure online user activity and engagement. However, none of those proposed techniques address the need for advertisers and/or website owners to detect tab/application switching or to find the actual away time from a particular website. To address this need, in this paper, we propose two new metrics, *focus ratio* and *active ratio*, along with a proof of concept tool to track web visitor engagement more accurately. To the best of our knowledge, focus ratio and active ratio have not been proposed in the literature within the context of web analytics and visitor engagement.

CCS Concepts

CCS → Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

Keywords

Active ratio; focus ratio; online news; visitor engagement; web analytics.

1. INTRODUCTION

In the United States, Internet advertising revenues totaled nearly $42.8 billion in 2013 which is a 17% increase from the year 2012 [1]. With advertising being a large part of the revenue of websites, many tools have been made to track their effectiveness. These tools have been able to track many things such as page views, bounce rates, clicks, scrolls and many more metrics. Also, numerous methods have been proposed in the literature to measure user engagement (e.g. [2][3][5]), however, little has been done to assess the user engagement in terms of actual length of the user interaction. Therefore, our aim is to be able to track the actual interaction time of the web visitor from page load to when they finally close out of the page.

To better understand new methods for defining how web visitor engagement is measured when tabs or applications are switched, we briefly introduce the current definitions of user engagement. Authors in [9] described user engagement as the emotional, cognitive and behavioral connection that exists between a user and a resource. Therefore, we can say that higher user engagement refers to a website's ability to hold the attention of a visitor and/or motivate the visitor to participate in an activity online. According to [4], engagement refers to the degree of and depth of visitor interaction on a website against a defined set of goals. Therefore, the obvious question is - How do we measure visitor engagement? In [5], authors collect mouse tracking data to correlate with visual attention. They examine the cursor behavior to measure user engagement within a controlled environment. However, in reality, web visitors are in an uncontrolled environment where they can be away from the computer screen or looking at another application while a specific news website is open in the background.

INRA 2016 Workshop, Halifax, Canada.

Besides academic research, companies in the private sector are conducting research for marketers and suggest solutions for optimizing web engagement in an effort to maximize advertisement revenues. For example, in 2008, Forrester Research Inc. published the following five tools and technologies to measure engagement [7]: 1. A centralized customer data repository to establish a system of record. 2. Analytics tools to help understand customer behavior. 3. Brand monitoring tools to amplify customer feedback cross media channels. 4. Ethnographic research to expand the breadth of customer insight. 5. Measurement dashboards to share results with a wide audience.

However, measurement of actual visitor engagement lacks standardization and validated methods for measuring engagement, especially in an online context. The current web engagement metrics include a variety of measures, e.g. session duration (*duration index*), page views (*click index*), visits per visitor (*recency index*), conversion rates, customer satisfaction, brand index, interaction index, loyalty index, etc. [6]. Also, web user's activity can be tracked by means of teleporting, backpaging, hyperlinking, etc. These metrics give web advertisers a better indication of the sites for best placement of their advertisements (ads for short).

Our preliminary study has revealed that there are some fundamental problems with current measurements for web visitor engagement. For example, session duration which tracks visit time on a web page does not take into consideration the physical away time of the web visitor from the computer or when a visitor switches to a different tab or application. While advertisers commonly rely on a visitor's time spent looking at the website, traditional web analytics tools lack capability to measure this accurately. Therefore, in order to address the need for an efficient web engagement tool, in Section 3, we present two new metrics *focus ratio* and *active ratio* for accurately measuring session duration and web visitor engagement.

2. ENGAGEMENT METRICS AND TOOLS

User engagement is measured at large-scale through analytic tools assessing users' depth of interaction with a website, which include metrics such as dwell time [9] and clickthrough rate [10]. Authors in [8] explored user comments on web pages to devise a tool for measuring user engagement. Their initiative is focused on

human curiosity in online news engagement. In [9], authors adopted a framework for user engagement through researching existing information retrieval metrics, user engagement metrics, web analytics, and measures from immersion in gaming. Other web engagement metrics in [4] and [6] include session duration, page views per session, visits per visitor, loyalty, interaction, and brand index.

Session duration is a measure of the time a visitor spends on a webpage in a given session. By subtracting the last timestamp on a given session from the first timestamp on the session, one can calculate session duration [4]. Session duration neither takes into account the possibility of a web visitor leaving the computer with a webpage still loaded in the browser nor does it indicate if the visitor has changed tabs or switched applications. As long as the webpage is still loaded on the visitor's screen, the session duration will increase - whether the visitor is engaged in the site or not.

Page views can be measured by dividing the number of page views in a given period of time by the number of visits in a given period of time [4]. This considers the number of clicks as well as the total amount of information or content the visitor has viewed. The issue with using the number of page views as a metric alone is that a visitor may be familiar with the site, and go directly to the page he or she wishes to engage in. Whereas another visitor who is less familiar with the site may click through many pages before finding the material they wish to engage in. The second type of visitors would show a higher engagement based on the page views per session index, however, they may actually be much less engaged in the webpage's content.

Visits per visitor can be described as a measure of the loyalty, frequency, and recency of a visitor to a site over a period of time [4]. Visits per visitor is nearly a straight-forward count of the number of times a visitor has returned to a webpage. Loyalty index is used to describe visitors that have accounts with the website or that visit the website at least three times per week [6]. Both visits per visitor and loyalty index metrics use a count of the times a visitor has been on a particular webpage. These measures tend to be less cryptic than previous metrics such as duration (time elapsed) and page depth (number of clicks in a page). At the same time, recency index and loyalty index fail to differentiate between the visitors that are highly engaged but only visit the site once and the visitors that visit the site frequently, but are lowly engaged in the webpage's content.

Brand index measures the level of attention a web visitor is giving to a site's brand prior to landing on the site [6]. This is measured by examining incoming search phrases and comparing those phrases to brand keywords. If the incoming phrase matches the brand keyword or phrase, then the brand index measurement is used.

Interaction index is the engagement that occurs when a visitor is directly interacting with the site through commenting, posting in forums, and/or engaging in discussion boards [6]. Like the brand index, interaction index will score positively if the visitor completes any defined interactions during the session [4].

During our investigation phase, we set out to find resources that would serve as guidelines for implementing a new tool to accurately detect web visitor engagement. Initially, we found two methods: focus and blur. The focus method is triggered when a visitor focuses on an element. This focus is often measured by mouse movements or clicks [11]. The blur method is the counterpart of the focus method. This method is triggered when an element loses focus. The blur method is used to remove focus from an element. Both focus and blur methods are supported by Chrome, Internet Explorer, Firefox, Safari, and Opera web browsers. Through further investigation, we found that the focus and blur methods give a lot of false positives. For example, if a user displays a smaller window on top of the browser window, the browser window loses focus, so the blur is raised. However, the user is still able to see the browser. Relying on user activity also gives a lot of false positive as well. Consider a user watching a video where the user may not move her mouse, click, or produce any key strokes, but the user is still active on the web page.

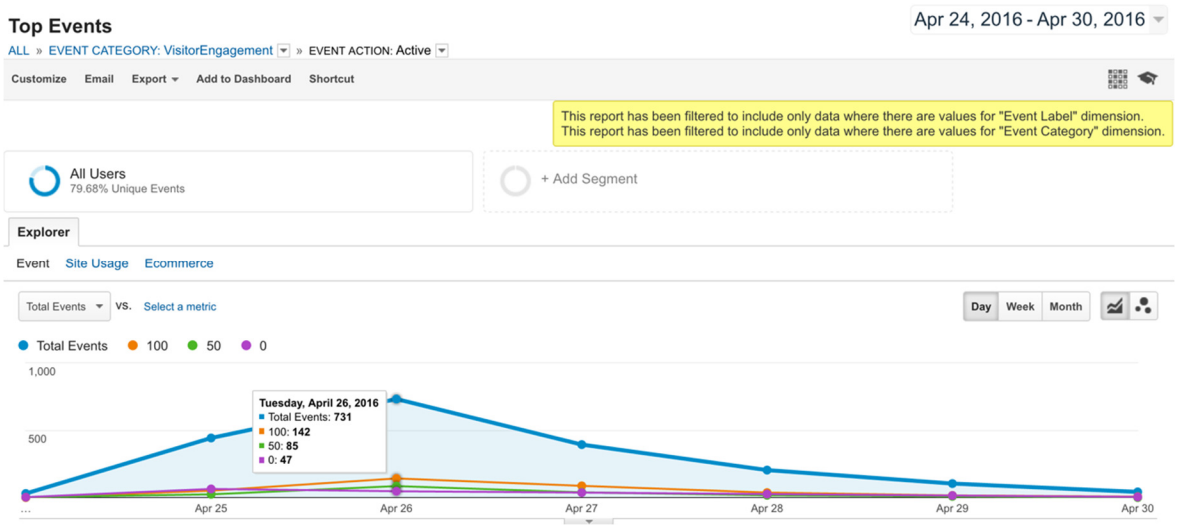
Finally, the Page Visibility API [12] is a way for website designers to determine the visibility status of a particular web page. It relies on the HTML5 specification. The Page Visibility API reveals when a webpage is visible or in focus. When a user switches between tabs or minimizes the current tab, the API triggers the visibility change event to notify the developers when the visibility state changes.

Of the metrics and tools discussed so far, it is safe to say that a lot of time and research has been plugged into pinpointing a definition of web engagement. The many tools and metrics we have reviewed in our background study suggest a vast knowledge of what the engaged visitor looks like compared to that of a less engaged visitor. However, none of the definitions have addressed the following questions - How can we determine if a visitor has switched tabs or applications? And, how long a user has actually been engaged with a particular webpage?

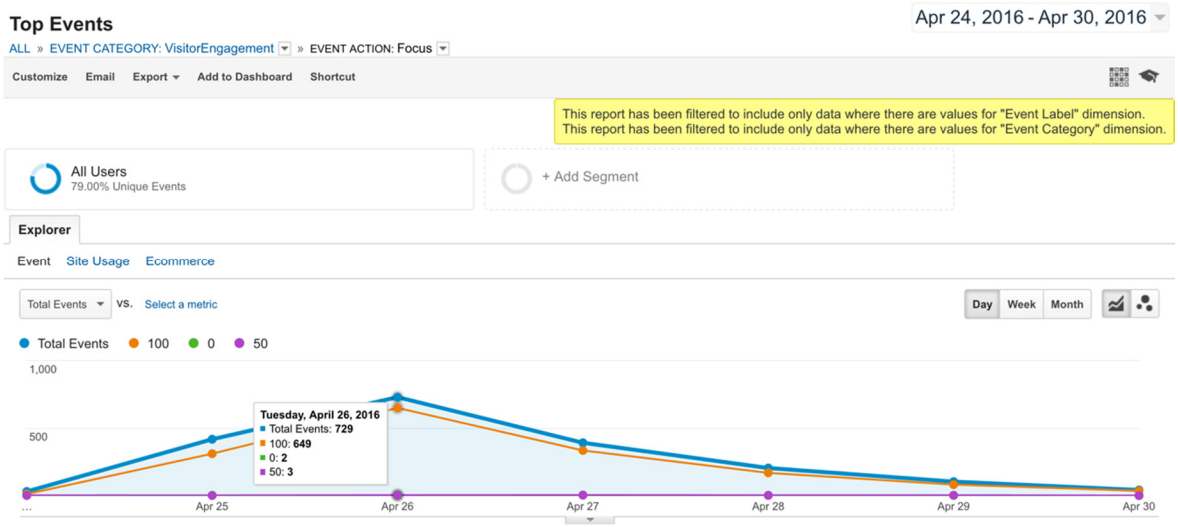
3. PROPOSED FOCUS RATIO AND ACTIVE RATIO

We define focus ratio as the difference between the time a webpage has been loaded in a browser and the time that page was actually visible in the active tab. For web advertisers, a higher focus ratio indicates a more attractive location for placing ads on. On the other hand, we define active ratio as the difference between the time a webpage is visible in the active tab and the user is actually considered viewing or interacting with that page. Both focus ratio and active ratio will help advertisers to determine the worthiness of placing an ad in a specific webpage at a relatively higher rate.

The benefit of tracking the ratio of active time a user is on the site compared with just seeing how long they are active is that we can check areas of our websites that might be causing web visitor to become inactive. If we have pages or news articles on our websites that have a very low active ratio, then we can take a look at these pages and see what it is about these pages or news contents that might be causing a user to leave the page. This is something we can measure with our proposed active and focus ratios instead of just simply tracking the average active time of that page. Also, having an average active ratio for each webpage on a news website, we will be able to find the contents that are the most appealing to the news readers. These will be the pages or contents that web visitors interact with the most, and thus we can promote these contents or categories to the prospective advertisers. Knowing which of our webpages are the most active will also make the ads more effective and better targeted. Therefore, high active and focus ratio pages will be a major selling points for ad spaces in an online news website. Showing these metrics to the potential ad space buyers or bidders will help us to be able to charge a fair price for that space.



Visitor Engagement events: *Active Ratio*



Visitor Engagement events: *Focus Ratio*

Figure 1. Visitor Engagement events sent to Google Analytics

4. PROOF OF CONCEPT TOOL FOR FOCUS RATIO AND ACTIVE RATIO

In this section, we describe our code snippets written in JavaScript and PHP to show how various times can be measured and stored to calculate and report the proposed focus and active ratios. The goal here is to show a tool that reports these new metrics, focus ratio and active ratio, and then provides an environment (in our case Google Analytics) where we can use these metrics to study the trends. To implement our concept, we used Riveted tool (*http://riveted.parsnip.io/*) for gathering user activity related data and to calculate the focus ratio and active ratio metrics. We then send this information to Google Analytics (*http://www.google.com/analytics/*).

4.1 Tracking active and hidden time

With the help of the Riveted tool, we add listeners for keystrokes, mouse click, mouse movement, scroll, page visibility for each of the web pages in order to detect user activity. When it detects any of these events it resets the *idleTimeout* value by calling the *clock()* function to increase the active time. In the *clock()* function of Riveted, *clockTime* value is being increased by one every second as long as the user is considered active. If the *clockTime* is greater than zero and a multiple of predefined report interval, then an event is sent to Google Analytics with the *clockTime*. *clockTime* value stops increasing when the user has not been active for the value in *idleTimeout*. So, if *idleTimeout* was set to 30, then after 30 seconds of inactivity *clock()* function would stop being called.

Page visibility listeners can detect if the page is the active tab in the browser. We added two functionalities (*hiddentTimeEvent()* and *totalIdleTime()*) to Riveted in order for it to keep track of how long the page has been hidden. The first function is called when the listener notices that the page is no longer visible, or becomes hidden. This starts a timer that calls the function *totalIdleTime()* every second and stores it in the variable *hiddenTime*. *totalIdleTime()*, when called, increases *hiddenTime* variable by one. If the user comes back to the page and it is visible again, then we kill the timer.

4.2 Tracking total time

Tracking the total time since the page has been loaded is needed for us to calculate the active ratio of the user. To keep track of total time, we added another two functions (*totalTime()* and *totalTimeEvent()*). We call the *totalTimeEvent()* as soon as Riveted is started, i.e. the webpage has been loaded. The function *totalTime()* increases *visitTime* value by one, and then calls the function *totalTimeEvent()*. The purpose of *totalTimeEvent()* is to wait one second before calling *totalTime()*. This effectively stores the total time since the page has been loaded, in seconds, in the variable *visitTime*.

4.3 Sending ratios to Google Analytics

Finally, we added a functionality to Riveted to calculate the ratios based on the times we are keeping track of, and then send those ratios to Google Analytics for further study of the webvisitor behavior. We achieve this with a function that we call once when Riveted is started. The first part of this function adds an event to the page that waits for the page to be closed. We then calculate the active ratio, and round it to a whole number, by taking the *clockTime* and dividing it by the *visitTime*. This gives us the percentage of time the user was active on that site. Next, we calculate the actual visible time to find the percentage of time the page was visible, and not hidden. We subtract the *hiddenTime* from *visitTime* to get the *visibleTime*. We then calculate focus ratio by taking the *visibleTime* and dividing it by the *visitTime*. This gives us the percentage of time the page was visible. The last part of the function sends these values to Google Analytics. We send an 'Event' with the category of 'VisitorEngagement', and label them as 'Active' or 'Focus' appropriately.

In Figure 1, we show the screen captures of the sample Google Analytics dashboard containing our visitor engagement data for the following website: *http://www.squidlessgames.com/* over a period of one week. In Figure 1.A., we show the total visitor engagement events reported versus how many users actually interacted with the website referring to an active ratio of 100%, 50%, and 0%. In Figure 1.B., we show how many users (out of those who were interacting with the website) departed right away (0%) or were active for the full (100%) and half (50%) of their visit duration to the website. These active and focus ratios are also available for the individual webpages. For example, in the above website we hosted two games: 'Floe' and 'Vocabulisitics'. Based on the collected user engagement data, we found that Floe has higher active and focus ratio compared to any other pages.

5. CONCLUSION

In this paper, we presented two new metrics and a proof of concept implementation to allow advertisers and/or news website owners to correctly detect when a visitor is physically away or changes applications in order to calculate the actual level of visitor engagement. We do not use any pre-determined concepts to characterize the user behavior or recommend news articles. However, focus and active ratio can be used to create an actual dwell time profiling of online news readers to help classify the news articles and rank news categories. This classification and rank information can be used by the news website owners to promote a specific category or news article to the advertisers in order to earn a higher revenue. Advertisers can also choose a specific ad space, or place a higher bid on a particular page or news category for their product promotion based on this information. Our tool detects user activity by tracking the keyboard and mouse activities as well as page visibility including tab switching in the browser. In the future, we plan to add capabilities to compare the ratios among different pages including a customizable dashboard to help users to get the most out of the information they are gathering. One major improvement will be to gather focus and active ratios for certain sections of a webpage to accurately determine user engagements in different sections of a webpage.

6. REFERENCES

[1] IAB Internet Advertising Revenue Report. Rep. Interactive Advertising Bureau (2014). Retrieved May 1, 2016 from http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2013.pdf

[2] Lagun, D. & Lalmas, M., (2016), Understanding and Measuring User Engagement and Attention in Online News Reading, *International Conference On Web Search and Data Mining (WSDM)*, San Francisco, USA.

[3] Barbieri, N., Silvestri, F., & Lalmas, M., (2016). Improving Post-Click User Engagement on Native Ads via Survival Analysis. *ACM international conference on World Wide Web (WWW)*, Switzerland.

[4] Peterson, E. T. & Carrabis, J. Measuring the Immeasurable: Visitor Engagement. Web Analytics Demystified, (2008). Retrieved May 1, 2016 from http://www.webanalyticsdemystified.com/downloads/Web_Analytics_Demystified_and_NextStage_Global_-_Measuring_the_Immeasurable_-_Visitor_Engagement.pdf

[5] Arapakis, I., Lalmas, M. & Valkanas, G., (2014). Understanding Within-Content Engagement through Pattern Analysis of Mouse Gestures, *ACM International Conference on Information and Knowledge Management (CIKM)*.

[6] Beckett, L. Getting beyond Just Pageviews: Philly.com's Seven-part Equation for Measuring, (2010). Retrieved May 1, 2016 from http://www.niemanlab.org/2010/10/getting-beyond-just-pageviews-philly-coms-seven-part-equation-for-measuring-online-engagement

[7] Haven, B., & Vittal, S. Five Tools and Technologies to Measure Engagement, (2008). Cambridge: Forrester Research Inc. Retrieved May 1, 2016 from https://www.adobe.com/enterprise/pdfs/five_tools_and_technologies_engagement.pdf

[8] Arapakis, I., et al. (2014). User engagement in online News: Under the scope of sentiment, interest, affect, and gaze. *Journal of the Association for Information Science and Technology*. 65: 1988–2005. doi: 10.1002/asi.23096

[9] Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011). Towards a science of user engagement (Position Paper), *WSDM Workshop on User Modelling for Web Applications*, Hong Kong.

[10] Chakrabarti, D., Agarwal, D., & Josifovski, V., (2008). Contextual advertising by combining relevance with click feedback, *ACM international conference on World Wide Web (WWW)*, New York, NY, USA, 417-426.

[11] Kantor, I. Focus/Blur: methods and events, (2011). (Oracle) JavaScript Tutorial. Retrieved May 1, 2016 from http://javascript.info/tutorial/focus-blur

[12] Mozilla Developer Network. Using the Page Visibility API. (2015). Retrieved May 1, 2016 from https://developer.mozilla.org/en-US/docs/Web/Guide/User_experience/Using_the_Page_Visibility_API