IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

A Report on Improving Website Content by Measuring Site Engagement

B Niket[1], Santosh Inchal[2], Arati S Gaikwad[3], Akarsh R Kapasi[4]

Department of Information Technology, D.Y Patil College of Engineering, Akurdi Pune, India[1, 2, 3,4]

Abstract: User engagement, which has been addressed as the arousing, cognitive and behavioral relation that exists between a user and a source, is the result of reliable, aspect, relevant and interesting content. In this Project, we evaluate one type of appointment, "adhesiveness" also known to as site engagement, which is apprehensive about users "spending time" on a site. The Site success depends on itself, but also it is dependent on how it is connected on the web and how the network traffic arrives towards it. This is of more concerned with large online providers, such as AOL, MSN and Yahoo websites!, which provides a mixture of content sites (e.g. news, sport, e-commerce) Social network sites (SNSs) and this has attracted the educational and engineering researchers intrigued by their affordances and reach. Like other online contexts in which individuals are knowingly able to construct an online representation of self such as online dating profiles and MUDS—SNSs constitute an important research context for scholars investigating processes of idea management, self-presentation, and friendship performance. The way of improving online interactivity is becoming a valuable way of communication quality of business web sites. As a result, it is important that web site designers understand the idea and how it affects the superiority of web site design.

Keywords: Site Engagement, Dwell time, security, Website Content.

1. INTRODUTION

Due to their highly different content, each site is usually studied and optimized separately, for example, by serving users the most relevant content in an attractive and attractive manner, in particular what relates to the layout and structure of the content. The main motto of large online providers is to only engage users with each individual site, but across all sites in their network, as sites can (and do) link to each other. For example, if a site does not have any links on its pages to other sites of the same provider, users will find it difficult to navigate to them, creating an engagement barrier. Conversely, linking to relevant content of the same provider will improve engagement.

On the other hand, users spend more and more of their online session multi-tasking, e.g. emailing, reading news, accessing a social network and generally navigating between sites. Online multi-tasking has implications when looking at the network of sites offered by online providers, as several of the provider sites can be accessed during a single session. Therefore, site engagement should be examined not only within individual sites, but also across sites, that is, the entire content provider network.

Since their introduction, social network sites (SNSs) such as MySpace, Face book, Cyworld, and Bebo have attracted millions of users, many of whom have integrated these sites into their daily practices. As of this writing, there are hundreds of SNSs, with various technological affordances, supporting a wide range of interests and practices. While their key technological features are fairly consistent, the cultures that emerge around SNSs are varied.

Most sites support the maintenance of pre-existing social networks, but others help strangers connect based on shared interests, political views, or activities. Some sites cater to diverse audiences, while others attract people based on common language or shared racial, sexual, religious, or nationality-based identities. Sites also vary in the extent to which they incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video-sharing Scholars from disparate fields have examined SNSs in order to understand the practices, implications, culture, and meaning of the sites, as well as users' engagement with them.

Our goal is to showcase some of the interdisciplinary scholarship around these sites. This report investigates site engagement, by defining a global measure of engagement that captures the effect sites have on the engagement on other sites within the same online browsing session. Intuitively, our global measure, which we name downstream engagement, measures the fraction of time users spent on a content provider's sites without leaving out of the entire online session time.

How to increase Engagement?

The ways to improve website engagement are:

1. Reduce page load times.
2. Consider Introducing a Brand Hero
3. Make it accessible with design and function.
4. Publish at All Hours
5. Let Fans Contribute
6. Always Link
7. Mention Others More

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

2. PROBLEM DEFINITION

2.1 Existing System

Due to their highly varied content, each site is usually studied and optimized separately, for example, by serving users the most relevant content in an attractive and enticing manner, in particular what relates to the layout and structure of the content. However, these large online providers aim not only to engage users with each individual site, but across all sites in their network, as sites can (and do) link to each other. Users are interested in site viewing if and only if the site is attractive and it depends on the webpage information also. For example, if a site does not have any links on its pages to other sites of the same provider, users will find it difficult to navigate to them, creating an engagement barrier. Conversely, linking to relevant content of the same provider will improve engagement.

2.1.1 Limitations of Existing System

1. If a site does not have any links on its pages to other sites of the same provider, users will find it difficult to navigate to them, creating an engagement barrier.
2. Online multi-tasking has implications when looking at the network of sites offered by online providers, as several of the provider sites can be accessed during a single session.
3. There is no strong site engagement in many of the websites.

Therefore, site engagement should be examined not only within individual sites, but also across sites, that is, the entire content provider.

3. ROUTE TO SOLUTION

3.1 Planned System

The main contributions of this project report are the following:

1. To the best of our knowledge, we introduce a new big (usage) data problem in the Web, as this is the first large-scale study that investigates the interactions between the different sites of a content provider and gives insight on site engagement.
2. The attribute with stylish nature are introduce, i.e., the elements of a web page, such as the links, tables, of each link can be used to predict and influence site engagement. This shows that web interconnections matter more than what some people may expect.
3. Speed Test is provided to the user on the website.

3.2 Benefits of Planned System

Evaluate your content. The users Engagement with the site can indicate that your content appeals to your audience – or doesn't. More in-depth analysis can reveal what types of followers are engaging the most and identify influencers and your biggest fans. Engagement metrics can show which networks are more valuable for your organization and where marketing teams should commit resources.

Boosts Website Traffic: Social engagement can increase exposure for your content, prompting more websites to link to it which in turn increases your website ranking. Study your Web analytics to see if rising social reach correlates with rising organic search traffic for your site. If it does, then drill deeper to learn which networks help the most. If it doesn't, re-evaluate your social media plans.

Generate sales leads: Although social media can bring visitors to your website, it doesn't sell. Your website must convert visitors to customers by offering relevant content that prompts them to stay on the site and return. It avoids the distraction from the current website and provides the user with live chat with the expert.
This design is made very responsive, so that the user can access from

4. LITERATURE SURVEY

4.1 Related work

In the online industry, site engagement or "stickiness" is mostly measured through behaviour measures aiming at assessing users' depth of interaction with a site. Widely-used measures include click-through rates, number of page views, time spent on a site (dwell time), how often users return to a site and number of users per month. Dwell time has proven to be a meaningful and robust measure of site engagement over the years; for example in the context of web search , where it is used to improve retrieval, Several white papers and reports contain studies on existing engagement measures and their usage , and proposals for a uniform measure of engagement based on several metrics

Several works carried out earlier shows that access to the web and navigation between the sites is shown by the user is different. From these and other studies, several user navigation models were developed, for example accounting for the usage of bookmarks, back buttons and teleportation. These models, based on formalisms such as branching processes, aimed to understand how users access sites and pages within them, and its effect on, for instance, site popularity, and loyalty to a site , but not the effect of the stylistics of a web page to the engagement of further web pages or sites.

Online behaviour measures have been used for many years by the web-analytics community and Internet marketing re-search companies (e.g. com Score). Because they are scalable to millions of users, they are commonly employed as a proxy for site engagement: the higher and the more frequent the usage, the more engaged the user. Although these measures cannot explicitly explain why users engage with a site, the fact that, for example, two million users choose to access a site daily is a strong indication of a high engagement with that site. Furthermore, by varying specific aspects of the site, e.g. structure and layout, and

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

assessing the effect on online behaviour, these measures can provide implicit understanding on why users engage with the site. Our work extends online behaviour measures with a measure defined to capture site engagement.
How long users spend on a provider set of sites (a provider network) from a given site is how we propose to measure site engagement. Previous work looking at aesthetics, accessibility and engagement based on dwell time include who showed that layout and textual features affect dwell time; and who showed that a combination of content and dynamic features (e.g. page size or time to download all URLs) had also an effect on page dwell time.
Following this line of work, this paper attempts to relate stylistic elements (e.g. layout and structure) of a web page, more particularly the main page of a site, and site engagement.

1. We show that stylistic attributes, i.e., the elements of a web page, such as the links, tables, of each site can be used to predict and influence site engagement. This shows that web interconnections matter more than what some people may expect.
2. Previous works include accessibility and engagement based on dwell time include who showed that layout and textual features affect dwell time; and who showed that a combination of content and dynamic features (e.g. page size or time to download all URLs) had also an effect on page dwell time. Following this line of work, this report attempts to relate stylistic elements (e.g. layout and structure) of a web page, more particularly the main page of a site, and site engagement.
It is a reasonable assumption that some nodes are likely to be deprived of secret keys by physical attacks [3]. So, a novel trust management scheme is necessary for secure and resilient wireless sensor networks.

5. IMPLEMENTATION MODULES

5.1 Module Description

5.1.1 Proposed work

In this report, we propose the following measure of site engagement: the total time spent on a contiguous sequence of provider sites from the next site until the end of the provider session, divided by the total remaining session time. We refer to this measure as downstream engagement.

By definition, if the next site in the session is not a provider site, the downstream engagement is zero. Intuitively, downstream engagement measures the fraction of time users spent on a content provider's sites, without leaving, out of the entire (remaining) time they had available to spend online (the total remaining session time). User data will be collected from a sample of users who gave their consent to provide browsing data through the Yahoo! toolbar. A total of 19.4M sessions were recorded from approximately 265,000 users.
To protect user privacy, no user identifiers were recorded, and only the top level domain was used in our analysis. We stripped the URLs to the last 3 components (for example mail.yahoo.com). This was also done to reduce the effect of sub-domains containing similar content (e.g., health in Yahoo! Lifestyle).

A site is an entity made of web pages put together to form a service. In the context of Yahoo! these include sites like Yahoo! News, Yahoo! Sports and Yahoo! Mail. Other examples include Face book (chat, apps) or Google (search, Gmail, scholar). We define a session as all the pages visited by a user within 30 minutes or less from the first interaction in the session. This definition captures over 95% of session boundaries.

Figure 1- User Engagement Measuring

The above figure illustrated the way in which the user engagement is measured in the form of cyclic way for good user engagement.

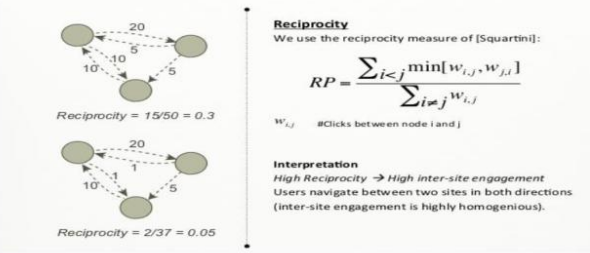

Figure 2- Site Engagement metrics

The above figure shows the site engagement metrics calculation when the user navigates from one website to other.

6. SITE ENGAGEMENT MODEL

A site is an entity made of web pages put together to form a service. In the context of Yahoo! these include sites like Yahoo! News, Yahoo! Sports and Yahoo! Mail. Other examples include Facebook (chat, apps) or Google (search, Gmail, scholar). We define a session as all the pages visited by a user within 30 minutes or less from the first interaction in the session. This definition captures over 95% of session boundaries. A provider session (in our case a Yahoo! session) corresponds to all contiguous pages of the provider sites visited within a session. Thus a session is composed of one or more provider sessions.

There are several ways we could measure site engagement. In this report, we propose the following measure of site engagement: the total time spent on a contiguous sequence

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

of provider sites from the next site until the end of the provider session, divided by the total remaining session time. We refer to this measure as downstream engagement. By definition, if the next site in the session is not a provider site, the downstream engagement is zero. Intuitively, down-stream engagement measures the fraction of time users spent on a content provider's sites, without leaving, out of the entire (remaining) time they had available to spend online (the total remaining session time).

Figure 3- Patterns of Site- Engagement

The above figure shows some of different websites which has the site engagements because of their content and response.

6.1 STYLISTIC ATTRIBUTES AND ENGAGEMENT

It is known that the style of a page is correlated with dwell time. In this section we test whether style attributes are correlated with engagement, by attempting to predict engagement (i.e. site engagement) given the attributes of a page. We used decision trees, graphs for classification, and 10-fold cross-validation to reduce the chance of an over-fit, for this purpose. We trained site-specific classifiers by setting the threshold for significant engagement as one standard deviation above the average engagement for each site separately. We include non-stylistic attributes in our prediction, to compare their effect to that of stylistic page attributes on downstream engagement. The average performance is reported.

6.2 FILE SYSTEM ANALYTICS:

HDFS works on Master-Slave technique:

Masters: Name node, Secondary Name node, Job Tracker.

1. The **Name Node** is the center piece of an HDFS file system. It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept. It does not store the data of these files itself.
2. Secondary Name node whole purpose is to have a checkpoint in HDFS. It's just a helper node for name node. That's why it also known as checkpoint node inside the community.
3. The **Job Tracker** is the service within Hadoop that farm out **Map Reduce** tasks to specific nodes in the cluster, ideally the nodes that have the data, or at least are in the same rack.

Slaves: Task tracker, Data nodes.

1. A **Data Node** stores data in the HDFS. A functional file system has more than one Data Node, with data replicated across them.
2. A **Task Tracker** is a node in the cluster that accepts tasks - Map, Reduce and Shuffle operations - from a Job Tracker.

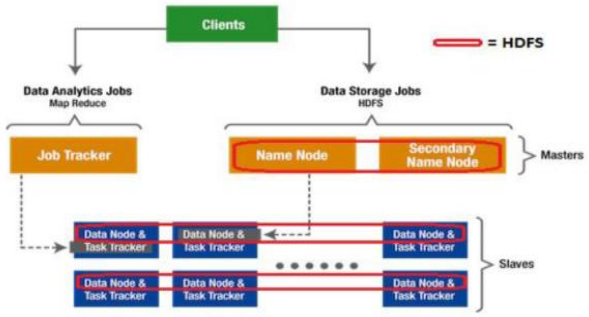

Figure-4: User engagement analytics

The above figure gives the view regarding the dwell time for the website which has good user engagement.

6.3 DWELL TIME

Dwell time is the time spent contiguously on a site, and is a popular measure of site engagement. It is often used as one of the official, standard measures by the web analytics community and Internet market research companies such as comScore. Dwell time per site, varies much from site to site, ranging from less than 5 seconds to more than 15 minutes. The sites with the shortest dwell time are e-commerce and sports sites. The highest average dwell time is recorded on leisure and home page country sites.

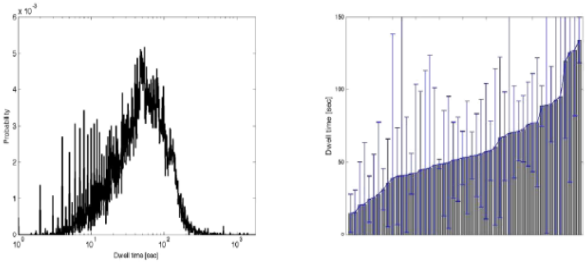

Figure- 5: Dwell time distribution (left) and Average dwell time per site

7. CONCLUSION

In this report, our primary focus was to analyse the activities of users on a particular website so as to measure the Engagement of the visitors and perform enhancements accordingly. We have prepared statistics of the user activities using Hadoop- Hive which would be used in further enhancement of the website. This report gives the brief review about the users engagement and also describes the content nature of the website for maximum user engagement.

DOI 10.17148/IJARCCE.2016.51182

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

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 Surviv al Analysis. ACM international conference on World Wide Web (WWW), Switzerland.

[4] Peterson, E. T. & Carrabis, J. Measuring the Immeasurable: Visitor Engagement. Web Analytics Demystifiedand_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]. H. OBrian, E. Toms, K. Kelloway, and E. Kelley. The development and evaluation of a survey to measure user engagement. JASIST 61(1):50-69, 2010.

[8]. E.T. Peterson and J. Carrabis. Measuring the immeasurable: Visitor engagement.Technical report, Web Analytics Demysti_ed, 2008.

[9]. R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpenning. A seasonal-trend decomposition procedure based on loss. Journal of O_cial Statistics, 6:373, 1990.

[10]. I. Dhillon, Y. Guan, and B. Kulis. A uni_ed view of kernel k-means, spectral clustering and graph cuts. Technical report, 2004.