Effective Measurement of Website Engagement to Improve Website Content

Ashish Saxena¹, Vinod Mane², Ashwini Abhale³

¹ Department of Information Technology, D. Y. Patil College of Engineering, Akurdi, Pune- 411027, India

ABSTRACT

User engagement, which has been addressed as the emotional, cognitive and behavioral connection that exists between a user and a resource, is the result of trustworthy, quality, relevant and interesting content. In this paper, we evaluate one type of engagement, “stickiness” also referred to as site engagement, which is concerned about users “spending time” on a site. Certainly, the success of a site largely depends on itself, but also on how it is connected on the web and how web traffic arrives to it. This is particularly important for large online providers, such as AOL, MSN and Yahoo!, which offer a variety of content sites (e.g. news, sport, e-commerce) Social network sites (SNSs) are increasingly attracting the attention of academic and industry researchers intrigued by their affordances and reach. In the vein of other terms related to online system in which persons are deliberately capable of constructing an online image of self for example Matrimonial sites, online dating sites are compose of an key research framework for people investigating processes of friendship performance, self-presentation and impression management. While interacting online with people is now a day’s very common and effective way of enhancing the worth of websites used for business purposes. And due to these consequences, it becomes important for web site designers to understand the concept and its impact on the quality of website design.

Keywords: Site Engagement, Dwell Time, Security, Website Content.

I INTRODUCTION

Due to their extremely varied content, every website is typically studied and optimized one by one, as an example, by serving users the foremost relevant content in a beautiful and engaging manner, especially what relates to the layout and structure of the content. However, these massive on-line suppliers aim not solely to have interaction users with every individual website, however across all sites in their network, as sites will (and do) link to every alternative. As an example, if a website doesn't have any links on its pages to alternative sites of constant supplier, users can notice it troublesome to navigate to them, making associate engagement barrier [1]. Conversely, linking to relevant content of constant supplier can improve engagement.

On the opposite hand, users pay a lot of and a lot of of their on-line session multi-tasking, e.g. emailing, reading news, accessing a social network and usually navigating between sites. on-line multi-tasking has implications once viewing the network of web sites offered by on-line suppliers, as many of the supplier sites may be accessed throughout one session [3]. Therefore, website engagement ought to be examined not solely inside individual sites, however conjointly across sites, that is, the whole content supplier network. Since their introduction, social network sites (SNSs) like MySpace, Facebook,
Cyworld, and Bebo have attracted variant users, several of whom have integrated these sites into their daily practices. As of this writing, there are many SNSs, with numerous technological affordances, supporting a good variety of interests and practices [11]. Whereas their key technological options are fairly consistent, the cultures that emerge around SNSs are varied. Most sites support the upkeep of pre-existing social networks, however others facilitate strangers connect supported shared interests, affairs of state, or activities. Some sites cater to various audiences, whereas others attract individuals supported common language or shared racial, sexual, religious, or nationality-based identities. Sites conjointly vary within the extent to that they incorporate new infoand communication tools, like mobile property, blogging, and photo/video-sharing. students from disparate fields have examined SNSs so as to grasp the practices, implications, culture, and that means of the sites, furthermore as users’ engagement with them. Our goal is to showcase a number of the knowledge domain scholarship around these sites [2]. This paper investigates website engagement; by shaping a worldwide live of engagement that captures the impact sites wear the engagement on alternative sites inside constant on-line browsing session. Intuitively, our world live, that we have a tendency to name downstream engagement, measures the fraction of your time users spent on a content provider’s sites while not exploit out of the whole on-line session time [13].

II PROBLEM DEFINITION
2.1 Existing System
Due to their extremely varied content, every website is sometimes studied and optimized severally, for instance, by serving users the foremost relevant content in a pretty and attractive manner, specifically what relates to the layout and structure of the content [5]. However, these giant on-line suppliers aim not solely to have interaction users with every individual website, however across all sites in their network, as sites will (and do) link to every different. for instance, if a website doesn't have any links on its pages to different sites of a similar supplier, users can notice it tough to navigate to them, making associate degree engagement barrier. Conversely, linking to relevant content of a similar supplier can improve engagement [12].

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.

Therefore, the engagement of site should be examined across the individual websites i.e. the entire content provider, rather than examining within the individual websites.

III ROUTE TO SOLUTION
3.1 Planned System
The main contributions of this paper 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. We show that stylistic attributes, 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.

3.2 Benefits of Planned System

**Evaluate your content:** Engagement 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.

**IV LITERATURE SURVEY**

In the online industry, site engagement or “stickiness” is mostly measured through behavior measures aiming at assessing users’ depth of interaction with a site [4]. 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[7].

Several works investigated how users access the web and how they navigate between sites or web pages. From these and other studies, several user navigation models were developed, for example accounting for the usage of bookmarks, back buttons and teleportation [8]. 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[6].

Online behavior measures have been used for many years by the web-analytics community and Internet marketing re-search companies (e.g. com Score) [10]. 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 [9]. Furthermore, by varying specific aspects of the site, e.g. structure and layout, and assessing the effect on online behavior, these measures can provide implicit understanding on why users engage with the site. Our work extends online behavior 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 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.

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.

V IMPLEMENTATION MODULES

5.1 Module Description

5.1.1 Proposed work

This research work proposes the below measures 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 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.

VI SITE ENGAGEMENT MODULE

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 paper, 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, 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).

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: Namenode, Secondary Namenode, JobTracker.

1. The NameNode is the centerpiece 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 Namenode whole purpose is to have a checkpoint in HDFS. Its just a helper node for namenode. That’s why it also known as checkpoint node inside the community.

3. The JobTracker is the service within Hadoop that farms out MapReduce tasks to specific nodes in the cluster, ideally the nodes that have the data, or at least are in the same rack.

Slaves: Tasktracker, Datanodes.

1. A DataNode stores data in the HDFS. A functional filesystem has more than one DataNode, with data replicated across them.

2. A TaskTracker is a node in the cluster that accepts tasks - Map, Reduce and Shuffle operations - from a Job Tracker.

![Figure 3: User Engagement Analysis](image)

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.
REFERENCES

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.