Mining and Modelling Web User Engagement: A Survey on Academic Sites for Framework Establishment

Tarmo Robal[1], Ahto Kalja[1,2]
[1]Department of Computer Engineering, Tallinn University of Technology, Tallinn, Estonia
[2]Institute of Cybernetics at Tallinn University of Technology, Tallinn, Estonia

***Abstract*—The rapid development of information and communication technologies has made the Internet to become the main source of information, not to say a platform to share knowledge and information on a daily basis, consumed on various devices and platforms. This information consumption on the other hand produces data that can be used to measure user engagement into content and has been used for various purposes, e.g., for advertising and web personalization. The most common metric used has been click through rate. However, a simple click does not necessarily mean user engagement, although it can be one of the metrics applied. In this paper we focus on user engagement aspects and propose a framework to evaluate user engagement towards sites and provided web content. To demonstrate and prove the framework, we have conducted an extensive study on several websites in academic domain.**

I. INTRODUCTION

Today, the Internet provides us with an enormous ever-growing source of information and has become a crucial part of our everyday lives. We use it regularly to search for information, read news, communicate with friends and colleagues, buy theatre and movie tickets, goods from online stores, and use e-government services [12]. The raise of the importance of the Internet during the recent years with the rapid development of not only information and communication technologies but also e-governance and fostered shift to e-services, has made the Internet to become the main source of information, not to say a platform to obtain, share, and produce knowledge and information on a daily basis. The information provided over the Internet is consumed on various platforms through different devices and user interfaces, and the latter consumption process itself produces volumes of usage data in variable and complex forms – today known as big data.

Web user engagement (UE) has been mainly characterized by site traffic numbers, the latter being a strong driver for advertisement industry delivering users either a fixed set of web ads, or taking into account user background and context and providing personalized banners and links forwarding the visitor to a particular site based on his/her probable interests. However, to gain interest of companies and/or revenue from advertisement campaigns, the content provided to visitors has to be engaging, challenging them to return, and inviting to recommend the website to other possible visitors. Advertisers and investors on the other hand need to be sure that the return-on-investment (ROI) is beneficial for them. Although in the early phases of user engagement detection a simple click as a convenient marketing indicator was used, it has come evident that this does not describe user interest into the provided content and can easily be manipulated.

Various definitions have been provided for web user engagement; in the context of this paper we find the definition by Lehmann et al. [15] to describe it the best: "User engagement is the quality of the user experience that emphasizes the positive aspects of the interaction, and in particular the phenomena associated with being captivated by a web application, and so being motivated to use it".

User engagement has been measured in different ways starting with simple clicks, calculating time users stayed on page, bounce rate, number of page views, etc., whereas none of the abovementioned cannot be claimed to be 'bulletproof' on its own, yet when considered holistically they can provide an insight into user engagement. These aforementioned approaches are haunted by the problems like undetermined user distraction time, actual reading time, content acknowledgment and familiarization rate, which still have remained unsolved with today's technology. There is no way one could measure exactly how much time users spend reading web content besides being in parallel engaged with other activities like for example chatting on smartphone, sipping coffee, watching TV, and so forth. This is mainly because the World Wide Web lacks the opportunity which would allow users to express themselves in the sense of their information seeking and behavior without having their privacy invaded. Taking it further, it would be impossible to satisfy each and every person's informational needs based on one model. On the other hand, this would require enormous tracking abilities and would be in conflict with users' privacy that was one of the cornerstones of the Internet. Still, user studies in laboratory conditions, for instance eye-tracking studies during performing a particular set of tasks on the web, have provided researchers with valuable hints on how users interact with devices and user interfaces provided to them. The problem with this approach is that the interaction does not happen in its natural environment but instead in lab conditions.

To be able to reason about users' interest and engagement towards provided content in real life, systems need to collect data about users' activities, separate user sessions and store this data in repositories. In order to give meaning to collected data techniques of web mining and in particular web usage mining (WUM) are applied. WUM is defined as the application of data mining techniques to discover usage patterns from web data, in order to understand and better serve the needs of web-based applications [32]. Usually three main phases of WUM are applied: data preprocessing, pattern discovery, and pattern analysis. The collected and analyzed implicit user interaction data can be used to infer user engagement.

In our previous work we have used implicitly collected user interaction data to establish an interpretation for time spent on pages and researched the applicable time ranges based on anonymous user profiling on several websites [26], and using the latter in a methodology to identify popular portal pages [27] within a framework of web information systems (WIS) evaluation. We have also established a set of methods to derive user domain models and provide recommendations for anonymous ad-hoc web users [24], [25] by the means of web and user ontologies.

In this paper we continue our web studies based on implicit user interaction indicators derived from users' action logs, and focus on user engagement. We establish a three-phase framework to model and measure user engagement, the differences in usage and content engagement, and use several websites of academic domain to prove it. We see the novelty of our framework in the method of evaluating website, page and user engagement based only on implicit user interaction data and inferred evaluative ideal condition indicator values. In sense of users, this approach is non-invasive and does not disturb users' browsing behavior – thus our framework is based on usage of data collected in natural and real-life situation. The research is based on studying several conference and international research project websites of academic domain with different target user groups, providing a versatile and rich insight into the domain. The rest of the paper is organized as follows: Section 2 provides an overview of related works in the research area; Section 3 discusses issues of collecting and processing web usage data, whereas Section 4 addresses the framework. Section 5 discusses application of the framework and presents the results of our user engagement studies on academic websites. Finally, Section 6 draws conclusions.

II. RELATED WORKS

The research in the field is tightly connected to collecting user interaction data, web mining and web usage mining. Different methods of collecting data about web users have been explored in [2], [10], [16], [30], [31], [32] arguing whether explicit or implicit methods suit the best for the task. All in all, researchers tend to be in favor of implicit methods, as they are less invasive and intrusive. Web mining techniques have been explored by many researchers in [32], [17]. Over its application, user engagement studies are also besides WUM connected to user profiling and web personalization, both targeting to improve user experience and interaction in massive information space, and also establishing long-term user engagement.

Users' interest towards visited pages has been investigated by Hofgesang [11] noting that the majority of WUM researchers tend to apply a list of visited web pages and their order to express users' interest, not paying any attention on time spent on pages (TSP). He reasoned that the time users spend on a page is a significant indicator of relevance and interest in information retrieval (IR), human-computer interaction (HCI) and even e-learning, and should be exploited as a metric of importance – the more time users spend on it, the more important it is assumed to be. In [9] it was concluded that using TSP as an interest indicator helps to identify important usage patterns and that users are unable to accurately assess interest towards visited pages in less than 5 seconds. This is also consistent with the results of the eye-tracking study conducted by Pan and colleagues, which highlighted users to evaluate the importance of the information found on a visited webpage during the first few seconds [21]. Srivastava et al. investigated the effect of measuring TSP on client- and server-side, concluding that whenever possible server-side measuring should be applied because of the overhead on client side [32]. TSP is also indicated to be a good implicit indicator of user interest by Claypool and colleagues [5], whilst their study with a special web browser called 'The Curious Browser' through which they were able to track users' actions, reckoned mouse movements and clicks to be insufficient. The study of users' web navigation habits in [37] concluded that web browsing is a rapidly interactive activity, and even pages with plentiful information are viewed for a brief period of time only. The experiments in [37] showed that almost in half of the cases users were spending less than 12 seconds on a visited page – thus leaving the page before reading a substantial part of its contents.

Different engagement patterns have been explored by Lehmann and colleagues in [15], where they analyzed a large sample of user interaction data on 80 online sites and investigated user engagement in terms of popularity, activity and loyalty. In [39] users' site engagement, in particular inter-site engagement across Yahoo! Network sites was explored as a big data problem. The authors proposed a measure called downstream engagement showing the percentage of time spent within sites from the same provider in a contiguous fashion, namely provider session, from the total time spent online for a given site, showing that downstream engagement is different from commonly used dwell time (time spent on a site) as a measure of engagement. The novelty of their research relies in investigating the effect of the network of sites on site engagement. Drutsa et al. [8] proposed an approach to improve the sensitivity of user engagement metrics by predicting individual user future behavior using the data from the Yandex search engine for their research.

Researchers have also explored possible application of dwell time to deliver personalized user experience through a recommendation system, and thereby increase users' long term engagement [38]. Bian et al. [3] applied user engagement interpretation to improve online content optimization for personalization through more effective user segmentation and understanding of users' actions based on click events. Lee and colleagues [14] investigated web content engagement time in e-publication systems, such as e-magazines and other advertisement-rich publications, using an agent-based system. The aim of their work was to investigate whether engagement time could be used as a platform to evaluate e-publication content performance. They concluded that engagement time could be used as a tool like page views (click through rate) to evaluate quality of web content or e-publication, as normally users will spend more time on quality articles. Thomas et al. [35] were interested in whether click patterns, dwell times, and keyboard actions could correlate to "User Engagement Scale (UES); getting early promising results for their research. The UES defines user engagement through several attributes such as aesthetics, focused attention, novelty, perceived time etc [18], where UE is mainly measured through questionnaires.

Similarly, Aguiar and colleagues [1] investigated user engagement towards online video viewing, and established a model to predict how large portion will a user watch of a particular video. Their research was aimed on maximizing viewership retention, and as common was based on clickstream data. Clickstream data was also used by Chen and Su [4] to discover user's interest at e-commerce site through clustering, where they measured user interest through browsing path, frequency, time spent on a page and similarity to other users.

Rowe and Alani [29] investigated engagement dynamics across multiple social media platforms (Stack Overflow, Twitter, Facebook, etc.) and applied a machine-learning based approach for engagement prediction. Through their experiments they came to a conclusion that different features could have an opposite effect on engagement in different platforms, or across different non-random datasets from the same platform. In [36] Wang and colleagues investigated anonymous mobile messaging app Whisper where users communicate with random strangers. They showed that in this environment the ties between users and long-term user engagement are weak and users tend to be disengaging over time, and that stimulating user engagement might be of necessity. In Whisper this has been realized through push notifications.

User engagement plays a crucial role also in education. O'Brien and colleagues [19] investigated the relationship between user engagement and comprehension of varied academic reading environments. They noted that engaging systems do not necessarily produce better learning outcomes. Qiu and others [22] on the other hand investigated student engagement in MOOCs which have boomed the recent years, and the extent student learning behavior is predictable.

Understanding visitors and their behavior on the web has become very important with the objective to provide users the best user experience, information they are seeking for and get them engaged with a site, especially under diverse market conditions, where competitive web information systems may be available. It has also gained its value for content providers in sense of ROI and keeping users engaged to their sites. As shown, user engagement interests many researchers and covers variety of research fields from information technology, education to psychology, and researchers have different interests regarding user engagement.

III. DATA TO MEASURE USER ENGAGEMENT

A. Data Used to Measure User Engagement

Each interaction (click, scrolling, etc.) web users have with web user interface through a web browser window can be captured into log and used later to discover web usage patterns and users' preferences. With the era of big data it has become common to have some sort of a system to capture interaction data and an analytical tool to provide site owners statistics and insights about site, its usage and users. For most online analytical tools the trade is data for statistics. Probably the most popular in the category are the Google Analytics and Webtrends.

Data about WIS usage can be collected either explicitly or implicitly. Explicit data collection methods assume users to actively participate in data gathering, through evaluating websites, providing feedback, or participating in audience surveys – this active participation is also the main drawback of this approach, as generally users are unwilling to actively participate in such evaluations. Still, providing simple feedback in the form of likes or dislikes, e.g., the well-known 'Like'-button from Facebook, seems to work well enough. Implicit data collection methods on the other hand are hidden from end-users, less intrusive, require no effort from users, and do not disturb them during their normal browsing activities. This form of data collection can be organized either on server, proxy or client level using either web server logs, customized browsers, or special log systems, which utilize session based ID-s and cookies [32]. Implicit techniques enable to monitor accessed pages, time spent on pages, users' navigation traces, discover usage patterns and mine user profiles – thus they are very suitable for accumulating knowledge about users' behavior and provide insights into users' engagement towards WIS's.

There is no unique indicator of user engagement nor a methodology to measure it. Different approaches on various indicators are being applied to detect and measure users' interest towards a page, site, or group of sites, as previously discussed. Mostly users' interactions are captured site-wise based on identifiable user sessions. This implicit data however reflects only user actions as facts in time and not users' real intentions. Still, the captured data can be used to infer users' interest, predict their actions, and derive their engagement towards provided digital content.

Let us now turn to different indicators used to describe and evaluate user engagement either site- or page-wise in research and industry, and present short comments on those [3], [4], [6], [8], [14], [20]:

- Page requests expressed through users' click activity. These operations are important as they reflect the content that is of interest for users;

- In-page click activity reflecting interaction and events on current page, as not every click takes a user to another page;
- Bounces reflecting activity where a user leaves a page immediately after arriving, being an indicator of disengagement rather than engagement;
- Exit pages as the last page viewed in a session, reflecting the end of user browsing activities. Exit pages can be either an indicator of engagement or disengagement, depending on their type and content, and time spent on them by users;
- Scroll depth reflecting users' activity on a page and interest into page content, acting as an indicator of engagement. However, this can only be applied for pages that are scrollable due to their content length;
- Number of times website or its content page is shared on social networks. Not so strong indicator, as people tend to easily share anything, and also sites they actually might not be interested in at all. The same applies to 'Likes', which cannot be taken as a serious indicator of engagement.
- Time on site (dwell time) and time spent on page (TSP) are believed to be one of the most important reflectors of user engagement, based on the assumption that this time actually reflects active user commitment. However, this is not always the case, and the trouble with this indicator is user distraction and how to measure it.
- Pages per visit (user session) reflecting engagement towards a site; the more pages users view the more users are presumed to be engaged with the content. Yet, not only page view rate should be counted, but it should be considered together with TSP to apply this indicator properly.
- Sessions per user, which is an arguable indicator and depends a lot on a site type, e.g., online banking service provided only through one particular website, which users exploit only in case of necessity. Also, this indicator is prone to user detection on public websites, as users are free to switch devices and browsers, delete cookies and web storage, thus their revisits remain undetected, unless user identification via login is enforced.
- Ratio of returning users is definitely an indicator of engagement in competitive situations, where users are free to choose the site to obtain the information they are seeking for, e.g., news portals. However, this is not always the case, and for sites that have a monopolistic status, e.g., e-government portals, online banking services, e-health portals, users are forced to use the provided service as is; thereby this ratio on return merely reflects their necessity to use the service rather than their engagement towards it.

B. Capturing Indicator Data for the Study

In our research we have considered all of the aforementioned indicators (Section 3.1), except shares on social networks, as the sites under our studies do not include such functionality, and thus it is not captured either by the used log system [28].

The data used in our studies has been collected by a special log system we have developed to capture users' interactions in web information systems, regardless whether they are information portals, project websites or even e-learning systems. Our log system [28] is used in many of the WIS's developed and maintained at the Tallinn University of Technology (TUT). The system consists of two subsystems: log capturing and log analyzer system (Fig. 1). The log capturing unit handles server- and client-side data capturing using cookies, JavaScript and server-side scripting, and writing operations to the log system, whereas the analyzer unit prepares data for the repository and serves as a first stage data filter (e.g. identifying robot visits). The database layer runs on the MySQL platform. The main drivers to develop such a log system were the inability of web server logs to distinguish between user sessions [13] as they were designed to log each and every request, and data incompleteness issues with HTTP traffic logs [7], [17], [34]. The log system has well-served our research and has been updated with a web-service based approach allowing more precise capturing of user's interactions also on client-side.

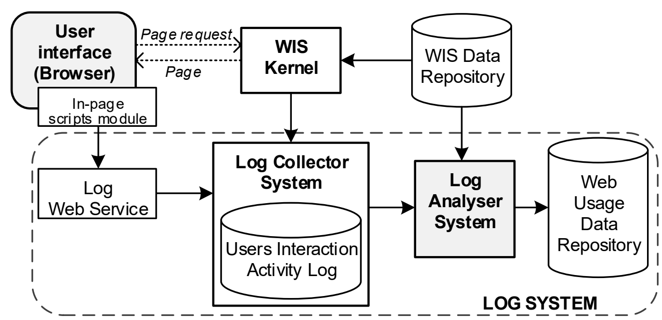

Fig. 1. General architecture of the log system and data capturing activities.

The log system (Fig. 1) works in the background of WIS and is transparent to visitors. The system uses several attributes to identify user sessions, such as embedded session identifiers, cookies and IP-address of the accessing host. In the produced log every access is coupled with timestamp and a variety of attributes describing users' activities, including:

- User session ID;
- IP and host of domain, where the request was made from;
- Page request;
- Browser and operating system information;
- Query method and full query string;
- Site referrer, if any;
- Operations performed during a session, including client-side UI events such as blur, focus, unload, click, etc.;
- Time to load and compose a page with a reference to server load during the page composition;
- Previous visit identifier and time, if available in cookie or web storage;

- Screen resolution as characteristic of viewing experience, and changes to viewing screen size during a visit.

We use this implicitly collected data stored in the Web Usage Data Repository as an input in our UE framework to measure user engagement and carry out experiments to prove our approach. The log system delivers necessary initial data that is sufficient enough to study user engagement aspects through the indicators previously described in Section 3 and draw conclusions on user engagement.

IV. MEASURING USER ENGAGEMENT

A. Framework Overview

Section 3 listed various indicators that can be and are used to measure user engagement towards websites or their content pages, and a method we used for data capturing. The main problem with most of the listed indicators is that alone they deliver little value and are error-prone to different conditions, whether it is a site specific issue (e.g., it does not make any sense nor is feasible to measure user engagement through returning users ratio on a monopolistic site), or a technology dependent restriction. Some of these conditions were outlined in our earlier discussion in Section 3.

Managing the evaluation and measuring user engagement requires a smart exploitation of these indicators of UE in an elaborated and exhaustive fashion, to make the data 'talk' and address effective user commitment. One should keep in mind that user engagement is more than just the sum of the values of its indicators.

Most of the approaches to measure user engagement look at some of the indicators of interest in some combination or separately, and try to specify user engagement based on these. For example, for years click through rate (number of page views) was applied as user engagement indicator to sell advertisements on websites. Yet, there are plenty of sites that still operate on this metric and do whatever to raise their click ratio. We believe that user engagement is far more complex, as is the human nature and behavior, and needs a systematic view and approach. With our framework on user engagement we have taken this path.

The build-up of our framework is based on a thorough look on the WIS itself, establishment of the ideal UE case of user engagement for a particular site and pages it consists of, and its comparative evaluation against the inferred ideal occasion. The framework addresses user engagement in two separate categories – site engagement and page engagement – through three phases (Fig. 2). The site engagement, as the name implies, addresses UE issues with a site in general, and measures user engagement towards the site as a whole. Page engagement on the other hand drills down to a page level and investigates UE for a particular page or a set of pages, and delivers engagement evaluation for the pages.

According to our framework and its methodology we firstly establish the objective and ideal UE case for a site and its content pages. This is done in *Phase 1*, on which the rest of model is set up. Secondly, in *Phase 2* we establish the average engagement model for the WIS based on collected implicit web usage data and collective intelligence reflected by its values in the Web Usage Data Repository (Fig. 1). This is the phase where we extensively apply web usage mining (WUM). These two phases set the necessary background system for the third phase, where user engagement on a particular user level can be evaluated contrasting the users' behavior to the collaborative averages and the ideal UE cases modelled in the two pervious phases. The *Phase 2* also addresses engagement issues with a site or its content pages. The third phase delivers important input for improving particular user engagement level and user experience through personalization and recommendations on content and keeping such users coming back. This is achieved though identifying such users via their current engagement level calculated based on their previous web behavior. Through this we target better personalization and thereby increase in long term UE, which also serves as a basis for improved revenue. Fig. 2 outlines the general architecture of the framework and indicators used to measure and evaluate site and page engagement.

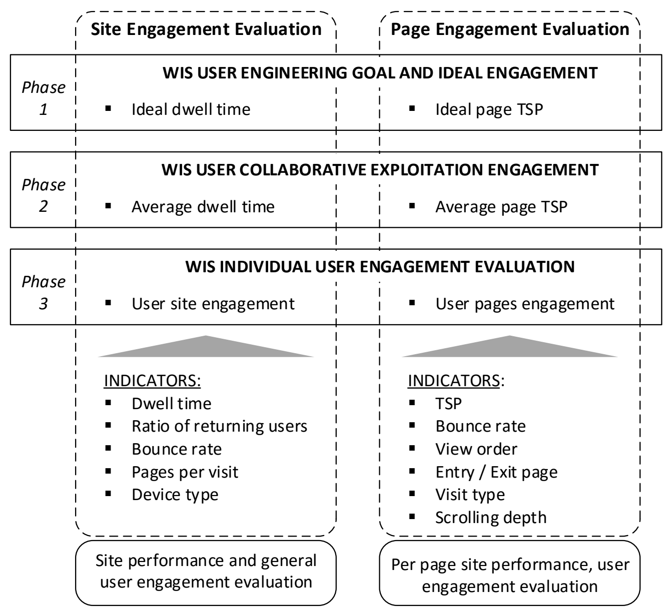

Fig. 2. General framework outline for site and user engagement measuring and evaluation.

In this paper we concentrate and limit our discussion on the first two phases of the framework, as the third phase is still an ongoing project. In the following subsections we will focus on user site engagement and user page engagement evaluation, and discuss the application of the framework.

B. Time Users Spend on Site and Page as Interest Indicator

Before continuing our discussion with site and page engagement measuring and evaluation, we would like to discuss time spent on page (TSP), its exploitation and limitations in our framework, as it is the most important and central metric of the framework.

TSP is the first and most trivial indicator of usability; more precisely it is an important indicator of user intention and interest towards viewed page and its content. It can be measured either on client side or on server side; nonetheless Srivastava et al. [32] have found in their research that client- and server-side measure of TSP is equal or in favor of server-side measuring because of the overhead on the client side.

The theoretical TSP is only the time user spent on particular page. Nevertheless, in reality we have to face also such factors as: first, the time spent for page generation on server, which usually has a marginal value and can easily be measured (in our previous studies we have found server-side composition time to be averagely 0.07–0.1 seconds [26]); second, the time spent on data transfer over network, which also under normal page load has a marginal value but technically troublesome to measure; and third, the effect of user distraction, i.e., user activities unrelated to site browsing, which adds uncertainty into the measured TSP time. To minimize this uncertainty, page focus and blur events can be detected, although this is no durable solution to eliminate user distraction from TSP. The actual user distraction remains immeasurable with today's technological capabilities without simultaneously conducted specific studies, e.g. eye tracking. Therefore, in practice TSP is found as the time between two page requests, and an upper limit is set to eliminate the impact of user distraction.

As our log model enables also capturing of page generation time, and page focus and blur events, we have used the TSP value calculated as given by (1), where t_i is the timestamp for a given operation, t_{gen} as time spent for page composition, t_f timestamp for page focus and t_b for page blur event. This formula also considers the fact that users might disable Javascript in their browser, and thereby it is impossible to capture the page focus-blur pairs. Relying only on summing these page focus-blur pair times, would not be an option to detect actual TSP.

$$t_{TSP} = t_{i+1} - t_i - t_{gen} - \sum(t_{f_{i+1}} - t_{b_i}) \qquad (1)$$

In order to be able to apply TSP properly, there are a few additional things to discuss. Firstly, after a page is loaded and presented to a user, it takes a moment for a user to evaluate its content. Research indicates that users are unable to accurately assess interest towards a page in less than 5 seconds [9], evaluating the importance of the presented information during the first few seconds [21]. Thus, approximately the first 5 seconds cannot reliably show user engagement in terms of content, as the user is still deciding over the appropriateness of the presented contents and its accordance to sought information.

Secondly, studies in [23] and [37] have shown average page view time as a measure of interest rate to be in between from 12 to 48 seconds. In [37] researchers highlighted that nearly in half of cases visitors were spending less than 12 seconds on viewed pages, and browsing for the next page before reading a substantial part of page contents. This conforms to the typical shallow behavior model of web users' towards content. In [11] it was shown that users tend to spend more time on pages that they have not visited before than on pages they have been already on. Thereby, we expect in the framework the actual TSPs to be lower than the ideal modelled TSP for a page and also lower in case of returning users in comparison to unique visits.

Hereby, in the framework we limit the appropriate applicable time spent on page as $TSP = [5 .. 2*TSP_{ideal}]$ seconds, where TSP_{ideal} is an approximated time to go through the content presented on a page. This approach of the ideal TSP in our method establishes a connection to actual content and presumable user commitment towards content in case of user interest. Values greatly exceeding these calculated TSP borders will be eliminated in the framework model.

The TSP_{ideal} can be calculated for text-based content articles and pages with video content, and has been done so in our framework and for the sites under study (Section 5). Studies on human reading performance have proven the reading speed for on screen content to be 180 words per minute (wpm), in contrast to 200 wpm from hardcopy for unfamiliar text [33]. This serves as a basis for calculating the TSP_{ideal} in our framework highlighting how much time in an ideal case would it take for users to read the full content article or look through the entire video content. A limitation of our framework at this point is that for other content types, e.g., graphics, the establishment of ideal TSP would need human evaluation preferably through a study with multiple participants, or keeping such pages out of the evaluation focus.

C. *Site Engagement*

Site engagement in our framework expresses the depth of user interaction on site. One of the major indicators for it is the time users contiguously spend on site – dwell time. Naturally, one would calculate time on site as the difference between the first and last page request time, as the TSP on exit page usually remains unknown and there is no way to determine when the user has left the page. However, we argue that this approach would deliver longer dwell time than the actual is, as it would count also user distraction. Thereby, to minimize the effect of user distraction, we propose that dwell time should be calculated over all visitor sessions finding the sum of TSPs over viewed pages during a session.

In our framework we address different types of dwell time: ideal time spent on site, which is the approximation of the time user should have spent going through all the content that was visited during a session, and actual dwell time. Both of these are based on actual site usage identified through web usage mining, and not on preset scenarios of navigation and their simulation. Thus, the ideal as well as the actual dwell time in our framework is expressed as given by (2), where t_{TSP} is correspondingly the ideal or the actual time user spent on page, and i covers all page views during user session, where user engagement is identified.

$$t_{dwell} = \sum_{i=0}^{n} t_{TSP}(i) \quad (2)$$

The method for site engagement evaluation consists of first finding the ideal engagement values, then the actual engagement, and providing evaluation through their comparison. For the evaluation we use the following indicators in our method:

- Ideal and actual dwell time,
- Ratio of returning and unique visitors,
- Bounce rate,
- Number of requests (clicks) per user session,
- User device type.

The dwell time is found according to (2) and expresses user session length with minimized user distraction effect. The ratio of returning and unique visitors is found only for sessions where actual user engagement is identified – in our method we do not count the sessions where no user engagement is present (e.g. bounce sessions); these sessions are identified and dropped from UE processing. It should also be noted that it is not possible to match all returning users as users are not bound to a specific device and browser and have the liberty to delete cookies and web storage from their devices, which results in lower level of returning visitors than the actual is. This is an inevitable limitation from technology at present time affecting also our approach. The bounce rate reflects sessions where no actual user engagement is present (recall the threshold values from Section 4.2) and the session consists of only bounce pages in comparison to sessions where some sort of user engagement is identifiable. Typically, as our experiments have shown, bounce sessions as well as bounce pages have a zero-like time values.

The number of operations per user session is counted in two categories: first the total number of operations, and secondly the number of operations with adequate TSP only. The ratio of the two latter values also highlights user engagement and is used as an indicator of site engagement in our framework.

Finally, user device type provides the segmentation over exploited device types and differences in UE on them – an important factor influencing user experience on today's era of smart and portable devices. These aforementioned indicators are used in our framework to characterize site engagement and provide evaluation for it. We will return to this with data from our experiments and study in academic domain in Section 5.

D. Page Engagement

Page engagement in our framework considers the level of user engagement to a particular page and its content. As with site engagement, one of the major indicators here is time, in particular time spent on page. However, there are several other indicators to follow in detecting user page engagement. In our framework we have used the following indicators to highlight user interest towards page content:

- Time spent on page (TSP),
- Bounce rate,
- Page view order,
- Entry/Exit page,
- Visit type (New/Frequent),
- Scrolling depth.

The time spent on page is found and calculated according to (1), and threshold limits discussed previously in this section are applied. The bounce rate in the method expresses the count of hits for a page where TSP < 5 seconds, thus users had decided to leave it before getting acquainted with the content, and thereby expressed disengagement. View order indicates whether a page was a mid-session request or any other type, to which in addition we explicitly also track entry and exit pages. The visit type is used to calculate the rate users revisit the page; whereas scrolling depth helps to identify whether users dug more into content and has a positive effect on user engagement determination. However, scrolling cannot be applied on all pages and is dependent on content length. Currently this has remained an open issue in our framework, and we have not yet established a method to solve it, leaving it as a future work. In addition, for each analyzed page its content type (text article, video page, graphics, etc.) is considered. Our method presumes this information is available in WIS or is prepared before applying the method for page engagement evaluation. The focus in the framework is on the level of user engagement as a comparison between the ideal expected and the actual highlighted engagement. The major outcome of phase 2 is the page-wise user engagement rate – commitment; which in our framework will further be used in phase 3 to evaluate a particular user's engagement level.

V. USER ENGAGEMENT STUDY ON ACADEMIC SITES

A. Studies Background

The aim of the studies described in this section is twofold: to prove our framework in practice on real data, and to investigate aspects of user engagement on academic and scientific websites. The study involves several conference and international research project websites for which we were in the managing role, had access to and a possibility to collect web usage log data from using our log system. These websites included 6 conference websites (hardware, software oriented, and/or educational conferences) and four research oriented websites (project websites and excellence centers). Jointly these 10 websites consisted of 279 different pages, 43% as project website pages and 57% of them as conference website pages. All the conference websites were of age over one year (rotating conferences with websites hosted at organizing institution TUT), and altogether delivered 61% of log records, whilst the project websites had a longer life span in average of 3 years (except one that had age of 3 months at the time of the studies) and provided 39% of records. Over this time the structure and layout of the websites remained the same, with only pages being added or removed as necessary. The collected log data to be analyzed consisted of approximately 800 000 records of page accesses.

On the collected raw data WUM was applied and data filtered to eliminate bogus requests. In addition the data was filtered to rule out accesses made from our department as in the role of administrator of these sites and therefore probably not reflecting the actual usage. The latter caused a lot of records to be dropped out of the analysis, which finally left us with only 37.6% of the original raw data set. After data filtering the ratio of conference and project websites data amounts changed and was correspondingly 65% from conference websites and 35% from research project websites.

The filtered data was further processed into several data objects such as session data, page view data, supplemented with page-specific data objects (e.g. content page type and raw content) from the WIS's the data was collected from, and calculations performed over several data objects according to the framework (e.g. session dwell time, TSP_{ideal} for each page, etc). In the remaining valid data of ca 36 000 user sessions the average rate of page request per session was 4.6 page views in a session. By this the data was prepared and ready for site and page engagement analyses.

B. Site Engagement Analyses

The site engagement analysis of the web usage log data from the 10 websites was carried out based on the indicators set for the framework – dwell time, ratio of returning users, bounce rate sessions, requests per visit, and device type. In addition we included site type (conference site, research project website) to investigate, if there are any deviations in the data on this parameter. The study was based on sessions and page views where actual user engagement could be identified as stated in the framework.

The analysis highlighted that out of the potential ideal time to deal with the content (ideal dwell time usage), visitors tend to use roughly only a third of it; returning visitors and visitors on mobile devices spend less time on content. The effective content engagement was clearly in favor of desktop platform users without any specific deviation in returning users or dependency on site type. The effective content engagement in our framework indicates the amount of time users spent on engaged content versus the time they could have spent in the ideal case. This provides an evaluation on users' engagement towards site. While taking a look at the rate of pages users showed up commitment in comparison to page requests in session (Table 1 rate of commitment pages), which is another measure of engagement in the framework, we discovered it to be clearly in favor of mobile users. This is probably due to the fact that mobile users, especially in the academic domain, tend to search for particular information rather than just browse around these sites under study, e.g., checking conference timetable or project news. The third measure of site engagement in our framework – site bounce rate indicating sessions where no user engagement was present, thus presenting user disengagement – revealed that mobile users are less prone to just browse around than desktop users, supposedly this is due to particular information needs and convenience of obtaining it. This higher engagement also supports the rate of commitment pages describing the amount of pages were user engagement was identified by the framework methods out of all viewed pages. Table 1 outlines the results of this site engagement analysis together with five site engagement measures of the framework.

TABLE 1. SITE ENGAGEMENT STUDY RESULTS

	Returning visitors	Bounce rate	Ideal dwell time usage	Effective content engagement	Commitment pages
Averages	31%	14%	29%	67%	47%
By category of study					
New visitor	na	19%	31%	69%	50%
Returning visitor	na	10%	27%	64%	45%
Desktop	37%	19%	30%	81%	36%
Mobile	24%	9%	28%	52%	59%
Conference	41%	13%	30%	69%	51%
Project	21%	16%	27%	64%	44%

In conjunction with site engagement we also studied the reading speed and its deviation from the implied value of 180 wpm [33], as the latter is a part of our framework. Based on the textual content length we calculated reading speed values (s_r) for pages, where user engagement was detected. The analysis of the calculated s_r values revealed an interesting fact that in average the deviation from this presumed reading speed is around 36%, and the average speed is actually lower at 135 wpm, whilst on mobile devices users tend to go through the content faster than on desktop devices. Table 2 presents the results of this analysis. It would be now interesting to test this on other sites, especially on sites not in the academic domain, e.g. news portals or social network sites and investigate the differences, as this deviation and lower reading speed for academic sites clearly describes more commitment into content of interest.

TABLE 2. READING SPEED ANALYSIS FOR PAGES WITH IDENTIFIED USER ENGAGEMENT

	Speed wpm	Deviation from 180wpm
Average	135	36%
By category of study		
New visitor	116	35%
Returning visitor	154	37%
Desktop	116	39%
Mobile	155	33%
Conference	150	37%
Project	120	35%

The analysis also confirmed the effectiveness of the exploitation of (2) in comparison to session length calculated as a difference between the first and last page request, as well as summing up just TSP's; thereby justifying the approach taken in our framework.

C. Page Engagement Analyses

The page engagement studies address users' interest and commitment towards the content on particular pages. The first phase (*Phase 1*) of the framework establishes the ideal case indicators whilst *Phase 2* delivers actual usage and its statistics based on web usage mining and allows to compute user commitment towards content on different pages. The third phase of the framework would be used to evaluate engagement level of a specific user, e.g. in the process of providing web personalization.

While analyzing users' commitment towards the content presented on the pages of the 10 websites under study, we first investigated user engagement towards presented content on the sites under study. The study outlined users' commitment page-wise over all sites. The commitment into page content was found based on the time spent on a particular page in respect to the ideal case found in Phase 1. Several threshold values were applied to rule out cases where no identifiable user engagement towards pages existed. Table 3 summarizes the results over all studied websites. These results do not contain any bounce pages and reflect only those page requests users made (i.e., excludes robots), and showed up interest towards presented content.

The results (Table 3) indicate that the overall users' engagement towards provided content is 57%; returning visitors tend to have higher engagement towards content, and the minimum commitment towards presented content is as low as 3% applicable for mid-session pages requested by new visitors. There is no indication of full commitment on exit pages. Single page views as visits where only one request was made during a session, thus the page is an entry and an exit point have the best user commitment rate. These are commonly pages users have accessed through a direct link, bookmark or search engine suggestion.

TABLE 3. INDICATED USER ENGAGEMENT THROUGH VISITOR AND PAGE REQUEST TYPE.

Visitor type		Indicated Commitment		
	Category	Min	Max	Average
New	Mid-session page	3%	153%	26%
	Exit page	62%	98%	80%
	Entry page	7%	112%	27%
	Single page view	75%	85%	79%
	Average	37%	112%	53%
Returning	Mid-session page	4%	163%	30%
	Exit page	63%	97%	80%
	Entry page	8%	114%	37%
	Single page view	75%	88%	82%
	Average	38%	115%	57%
Overall		40%	111%	57%

Secondly, we targeted the indicator of disengagement, namely bounce rate. Herein, in the framework a bounce page is declared as one that does not reflect user interest whereas users' leave the page without considering its contents. The thresholds highlighted in Section 4 are applied.

While the bounce rate in site engagement indicated the rate of sessions without any reflection of user interest towards site, the bounce rate in the page engagement category indicates the disinterest rate for each page in the WIS. In our framework the bounce page identification process is also bound to the ideal TSP (and therefore as well as content length), to rule out most of intermediate link pages which could otherwise mistakenly be identified as bounce pages as the time users spend on these pages is relatively short. For example, during the analysis a page for one of the research project websites came up with a bounce rate of 83%, being an entry page and with average visit time of 0.7 seconds (as a mid-session page this page had average visit time of 1.0 seconds and a bounce rate 43%, which is half of the entry page bounce rate). The count of words for this content page was 22 and the ideal TSP calculated was only 7 seconds. This is a good and rather extreme example of false positive bounce page. To tackle the problem of these false positives, we empirically introduced an additional threshold for bounce page identification at $0.3*TSP_{ideal}$, which is consistent also to the findings in the site engagement analysis (ideal dwell time usage ratio).

In identifying user disengagement through bounce pages the focus is on pages with long content (e.g., more than a few sentences to paragraphs of text) and high bounce rate. Evidently, issues with page commitment regarding bounce rate require human intervention and interpretation, and are to be dealt by site webmaster or content administrator, to find measures for lowering bounce rate through improving page content and user experience.

In our study out of the 10 websites and 279 pages they consisted of, 89% of pages had at least one request identified as a bounce; in average the bounce rate for pages where at least one bounce request was identified was 33%. Table 4 outlines the results of this analysis and shows the differences on whether the bounce page was an entry point or a mid-session page. As shown, for entry pages the decision to bounce away was made faster and the bounce rate is a bit higher (+7%) than on mid-session pages. The user commitment rate towards presented content based on the TSP ratio to ideal TSP is low in both cases for pages identified as bounce pages by the framework methods.

TABLE 4. USER COMMITMENT AND BOUNCE RATE ON BOUNCE PAGES

	AVG TSP	User Commitment	Bounce rate
Entry page	0.8 s	2.5%	36%
Session page	1.1 s	4.6%	29%
Average	1.0 s	3.6%	33%

VI. CONCLUSIONS

The Internet has become the main source of information and plays a crucial role in our everyday lives in obtaining knowledge and sharing information. Regardless of used devices and platforms, users expect the information to be delivered in a valuable form and are affected by the information consumption experience, the latter forming their attitudes towards provided web content, one of which is user engagement. User engagement is an important factor driving website success and feasibility to gain revenue, for example from advertisement industry. However, user engagement is not granted with site launch and efforts have to be made to gain, retain and improve it.

In this paper we have established a three-phase framework to measure and evaluate user engagement towards sites and towards pages they contain based on implicitly collected user behavior and interaction data, and comparative indicators of inferred ideal case values. While user engagement can be measured in different ways from interviews and questionnaires as the main explicit forms in contrast to tracking users' actions, our framework is focused only on data users produce during their uninterrupted browsing sessions and reaches out to combine this implicitly collected data with domain knowledge to evaluate user engagement level. The article, as a milestone of our ongoing work, focused on outlining the framework and proving the applicability of the first two phases of the framework. With these phases we have targeted the ability to evaluate site and page engagement on a collaborative level by the exploitation of web usage logs and web usage mining. The effectiveness of the proposed framework has been proved by analyses and experiments on 10 websites from academic domain. We have shown that implicitly collected user interaction data, supplemented with information from WIS itself, can be used to evaluate user engagement level towards site or its content pages. This has given us the assurance to continue our work with the third phase of the framework.

In the long run we see the framework to deliver user engagement evaluation on a single user level (third phase of the our framework) for application in the process of web personalization, where the economic effect is in better return of investment, better sales, and business success, with a common driver is user satisfaction. Regardless, whether there are competitive alternative websites or systems users might choose to use or not, we should be able to target our users and provide them with a content that is engaging to them. Better user experience and personalization delivers improved user engagement and leads to raise in site revenues.

As for the future work, we will continue our work with phase 3 of the framework and experiments in this category. This stage of the research and especially the experiments carried out have provided us with new ideas on improving the framework and for one of the final outcomes we see the establishment of a model for user engagement index identifying users' commitment as specific value on a preset scale.

ACKNOWLEDGMENTS

This research was partially supported by the Estonian Ministry of Research and Education institutional research grant no. IUT33-13 "Methods, tools and processes for software with strong guarantees".

REFERENCES

[1] Aguiar, E., S. Nagrecha and N. V. Chawla; "Predicting online video engagement using clickstreams," in *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, pp. 1-10, 2015.

[2] Al halabi, W. S., M. Kubat and M. Tapia; "Time Spent on a Web Page is Sufficient to Infer a User's Interest," in *Proceedings of the IASTED European Conference: internet and multimedia systems and applications (IMSA'07)*, ACTA Press, Anaheim, CA, USA, pp. 41-46, 2007.

[3] Bian, J., A. Dong, X. He, S. Reddy and Y. Chang; "User Action Interpretation for Online Content Optimization," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 25(9), pp. 2161-2174, 2013.

[4] Chen, L., and Q. Su; "Discovering user's interest at E-commerce site using clickstream data," in *10th International Conference on Service Systems and Service Management (ICSSSM),* pp. 124-129, 2013.

[5] Claypool, M., P. Le, M. Wased, D. Brown; "Implicit interest indicators," in *Proceedings of the 6th international conference on Intelligent user interfaces IUI'01*, ACM Press, New York, NY, USA, pp. 33-40, 2001.

[6] Colbert, M. and A. Boodoo; "Does 'Letting Go of the Words' Increase Engagement: a traffic study," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, ACM, New York, NY, USA, pp. 655-667, 2011.

[7] Davison, B.; "Web Traffic Logs: An Imperfect Resource for Evaluation," in *Proceedings of Ninth Annual Conference of the Internet Society (INET '99)*, 1999.

[8] Drutsa, A., G. Gusev and P. Serdyukov; "Future User Engagement Prediction and Its Application to Improve the Sensitivity of Online Experiments," in *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*, ACM, New York, NY, USA, pp. 256-266, 2015.

[9] Farzan, R. and P. Brusilovsky; "Social navigation support in e-learning: What are real footprints?," in *Proceedings of IJCAI'05 Workshop on Intelligent Techniques for Web Personalization*, pp. 49-56, 2005.

[10] Gauch, S., M. Speretta, A. Chandramouli and A. Micarelli; "User profiles for personalized information access," in *The adaptive web*, P. Brusilovsky, A. Kobsa, W. Nejdl Eds. LNCS, vol. 4321. Springer-Verlag, Berlin, Heidelberg, pp. 54-89, 2007.

[11] Hofgesang, P.I.; "Relevance of Time Spent on Web Pages," in *Proceedings of the Knowledge Discovery on the Web (KDD) ACM SIGKDD Workshop on Web Mining and Web Usage Analysis at WebKDD 2006*, ACM Press, 2006.

[12] Kalja, A., J. Põld, T. Robal and U. Vallner; "Modernization of the e-government in Estonia," in *PICMET'11: Proceedings Technology Management in the Energy-Smart World* [USB], eds.: D. F. Kocaoglu, T. R. Anderson, T. U. Daim, Portland, OR: PICMET, pp. 3151-3157, 2011.

[13] Kimball, R. and R. Margy; *The data warehouse toolkit: the complete guide to dimensional modelling*. John Wiley & Sons, 2002.

[14] Lee, R. S. T., J. N. K. Liu, K. S. Y. Yeung, A. H. L. Sin and D.T.F. Shum; "Agent-Based Web Content Engagement Time (WCET) Analyzer on e-Publication System," in *Ninth International Conference on Intelligent Systems Design and Applications (ISDA '09)*, pp. 67-72, 2009.

[15] Lehmann, J., M. Lalmas, E. Yom-Tov and G. Dupret; "Models of user engagement," in *Proceedings of the 20th International conference on User Modeling, Adaptation, and Personalization (UMAP'12)*, J. Masthoff, B. Mobasher, M. C. Desmarais, and R. Nkambou, Eds. Springer-Verlag, Berlin, Heidelberg, pp. 164-175, 2012.

[16] Middleton, S. E., N. R. Shadbolt and D.C. De Roure; "Capturing Interest through Inference and Visualization: Ontological User Profiling in Recommender Systems," in *Proceedings of the 2nd*

International Conference on Knowledge Capture, ACM, New York, NY, USA, pp. 62-69, 2003.

[17] Mobasher, B., R. Cooley and J. Srivastava; "Automatic Personalization Based on Web Usage Mining," *Communications of the ACM*, vol. 43, pp. 142-151, 2000.

[18] O'Brien, H. L. and E. G. Toms; "The development and evaluation of a survey to measure user engagement," *Journal of the American Society for Information Science and Technology*, vol 61, pp. 50-69, 2010.

[19] O'Brien, H. L., L. Freund, and R. Kopak; "Investigating the Role of User Engagement in Digital Reading Environments," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*, ACM, New York, NY, USA, pp. 71-80, 2016.

[20] Omniture whitepaper; "Metric definitions. Metrics for SiteCatalyst, and Discover", Retrieved 12/11/15 World Wide Web, http://www.webmetric.org/white_paper/metrics_definition.pdf

[21] Pan, B., H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner and J. K. Newman; "The determinants of web page viewing behavior: an eye-tracking study," in *ETRA '04: Proc. of the 2004 symposium on Eye tracking research & applications*, ACM, NY, USA, pp. 147-154, 2004.

[22] Qiu, J., J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang and Y. Xue; "Modeling and Predicting Learning Behavior in MOOCs," in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16)*, ACM, New York, NY, USA, pp. 93-102, 2016.

[23] Rafter, R. and B. Smyth; "Passive Profiling from Server Logs in an Online Recruitment Environment," in *Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalization (ITWP 2001)*, USA, pp. 35-41, 2001.

[24] Robal, T. and A. Kalja, "Conceptual web users' actions prediction for ontology-based browsing recommendations," in *Information Systems Development: Towards a Service Provision Society*, G.A. Papadopoulos, G. Wojtkowski, W. Wojtkowski, S. Wrycza and J. Zupancic, Eds. Springer-Verlag New York, pp. 121-129, 2009.

[25] Robal, T. and A. Kalja; "Applying user domain model to improve Web recommendations," in *Databases and Information Systems VII: Selected Papers from the Tenth International Baltic Conference, DB&IS 2012*, A. Caplinskas, G. Dzemyda, A. Lupeikiene, O. Vasilecas Eds. Frontiers in Artificial Intelligence and Applications, vol. 249, Amsterdam: IOS Press, pp. 118−131, 2013.

[26] Robal, T. and A. Kalja; "Learning from users for a better and personalized web experience," in *PICMET'12: Proceedings Technology Management for Emerging Technologies* [USB] eds.: D. F. Kocaoglu, T. R. Anderson, T. U. Daim, Portland, OR: PICMET, pp. 2179-2188, 2012.

[27] Robal, T. and A. Kalja; "Managing knowledge in Web portals for improved customer loyalty and satisfaction," in *PICMET'13: Proceedings Technology Management in the IT-Driven Services* [USB]. eds.: Kocaoglu, D. F. et al. Portland, OR: PICMET, pp. 1207-1216, 2013.

[28] Robal, T., A. Kalja and J. Põld; "Analysing the Web Log to Determine the Efficiency of Web Systems," in *Proc. of the 7th International Baltic Conference on Databases and Information Systems DB&IS'2006. Communications*, Vilnius: Technika, pp. 264-275, 2006.

[29] Rowe, M., H. Alani; "Mining and comparing engagement dynamics across multiple social media platforms," in *Proceedings of the 2014 ACM conference on Web science (WebSci '14)*, ACM, New York, NY, USA, pp. 229-238, 2014.

[30] Sarwar, B., J. Konstan, A. Borchers, J. Herlocker, B. Miller and J. Reidl; "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," in *Proceedings of ACM Conference on Computer Supported Collaborative Work (CSCW)*, Seattle, Washington, USA, pp. 345-354, 1998.

[31] Shapira, B., M. Taieb-Maimon and A. Moskowitz; "Study of Usefulness of Known and New Implicit Indicators and Their Optimal Combination for Accurate Inference of Users Interest," in *Proceedings of the 2006 ACM Symposium on Applied Computing (SAC '06)*, ACM, New York, NY, USA, pp. 1118-1119, 2006.

[32] Srivastava, J., R. Cooley, M. Deshpande and P. N. Tan; "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations*, vol. 1(2), pp. 12-23, 2000.

[33] Ziefle, M.; "Effects of display resolution on visual performance," *Human Factors*, vol. 40(4), pp. 555-568, 1998.

[34] Taghipour, N., A. Kardan and S. S. Ghidary; "Usage-Based Web Recommendations: A Reinforcement Learning Approach," in *Proceedings of the 2007 ACM conference on Recommender systems (RecSys'07)*, ACM, New York, NY, USA, pp. 113-120, 2007.

[35] Thomas, P., H. O'Brien and T. Rowlands; "Measuring Engagement with Online Forms," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*. ACM, New York, NY, USA, pp. 325-328, 2016.

[36] Wang, G., B. Wang, T. Wang, A. Nika, H. Zheng and B. Y. Zhao; "Whispers in the dark: analysis of an anonymous social network," in *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*, ACM, New York, NY, USA, pp. 137-150, 2014.

[37] Weinreich, H., H. Obendorf, E. Herder and M. Mayer; "Off the beaten tracks: exploring three aspects of web navigation," in *Proceedings of the 15th international conference on World Wide Web (WWW '06)*, ACM, New York, NY, USA, pp. 133-142, 2006.

[38] Yi, X., L. Hong, E. Zhong, N. Nan Liu and S. Rajan; "Beyond clicks: dwell time for personalization," in *Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14)*, ACM, New York, NY, USA, pp. 113-120, 2014.

[39] Yom-Tov, E., M. Lalmas, R. Baeza-Yates, G. Dupret, J. Lehmann and P. Donmez; "Measuring inter-site engagement," in *IEEE International Conference on Big Data*, pp. 228-236, 2013.