User Engagement Analytics Based on Web Contents

Phoom Chokrasamesiri and Twittie Senivongse

Abstract User engagement is a relation of emotion, cognitive, and behavior between users and resources at a specific time or range of time. Measuring and analyzing web user engagement has been used by web developers as a means to gather feedback information from web users in order to understand their behavior and find ways to improve the websites. Many websites have been successful in using analytics tools since the information acquired by the tools helps, for example, to increase sales and the rate of returning to the websites. Most web analytics tools in the market focus on measuring engagement with the whole webpages, whereas the insight information about user behavior with respect to particular contents or areas within webpages is missing. However, such knowledge of web user engagement based on contents of the webpages would provide a deeper perspective on user behavior, compared to that based on the whole webpages. To fill this gap, we propose a set of web-content-based user engagement metrics that are adapted from existing web-page-based engagement metrics. In addition, the proposed metrics are accompanied by an analytics tool which the web developers can install on their websites to acquire deeper user engagement information.

Keywords User engagement · Web analytics · Web contents · Metric

1 Introduction

User Engagement is a relation of emotion, cognitive, and behavior between users and resources at a specific time or range of time [1]. In the past years, user engagement analytics on websites has been widely used by web developers to gather feedback from web users. In a survey of datanyze.com, 97.6 % of Alexa top one

P. Chokrasamesiri · T. Senivongse (✉)
Department of Computer Engineering, Faculty of Engineering,
Chulalongkorn University, Bangkok 10330, Thailand
e-mail: twittie.s@chula.ac.th

P. Chokrasamesiri
e-mail: phoom.c@student.chula.ac.th

R. Lee (ed.), *Computer and Information Science*,
Studies in Computational Intelligence 656, DOI 10.1007/978-3-319-40171-3_6

million websites use web analytics tools to analyze their user behavior data to obtain the insight of web usage [2]. The example of the analytics tools with the highest market share are Google Analytics, Google Universal Analytics, Yandex Metrica, comScore, and Quantcast, where Google Analytics occupies 45.5 % among these tools.

The analytics tools can provide web developers with the insight and feedback of web usage including what the users do on the websites, when they visit, how they interact with the websites, and many more. Once equipped with such information, the web developers can find ways to improve their web sites. For example, using an analytics tool, an e-commerce company [3] could identify the lost revenue due to high shopping cart abandonment rate and improved its website features to finally obtain an increase in checkout to payment page. Another case of a company [4], using an analytics tool to analyze which versions of its landing page should be used, could obtain an increase in homepage engagement and a boost to e-commerce conversion rate.

To analyze user engagement, a number of metrics have been developed to measure web usage and calculate user engagement. Click-Through Rate, Time Spent on a Site, Page Views, Return Rates, and Number of Unique Users are examples of popular metrics [5]. However, available analytics tools and metrics focus on analytics of user engagement with the whole webpages, rather than the web contents. Figure 1 depicts web contents, namely c1, c2, and c3, on a webpage, where Fig. 1a shows page-based engagement data that can be collected from users, i.e. mean visit to a page, mean hover over a page, and mean click to a page. However, the usage information regarding a particular content on a page is not known. If we can enhance page-based engagement metrics with content-based ones, i.e. mean visit to a con-

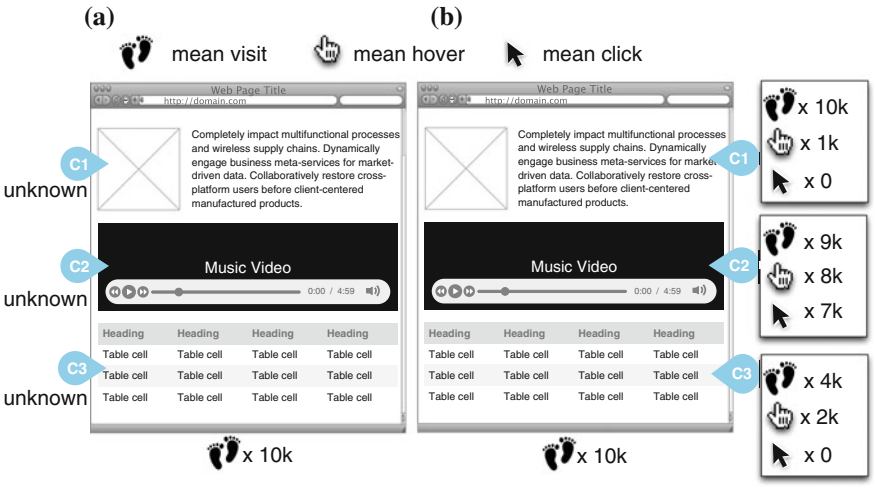

Fig. 1 Web usage analytics based on **a** webpage, **b** web contents on a page

tent, mean hover over a content, and mean click to a content as in Fig. 1b, web developers can obtain deeper understanding of how users interact with the contents and can improve their websites. This paper proposes a set of web-content-based user engagement metrics that are adapted from existing web-page-based engagement metrics. In addition, the proposed metrics are accompanied by an analytics tool which the web developers can install on their websites to acquire deeper user engagement information.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 explains the web contents model and content-based user engagement measurement method. In Sect. 4, user engagement metrics based on web contents are presented together with a case study. Section 5 describes the content-based user engagement analytics tool, and the paper concludes in Sect. 6.

2 Related Work

We focus the discussion about related work on user engagement metrics and a popular web analytics tool.

2.1 Web User Engagement Metrics: A Web Analytics Approach

Lehmann et al. [6] study models of user engagement with web applications which can vary as the way users engage with different applications can be very different. A set of metrics, grouped into popularity, activity, or loyalty categories, is used in their study as listed in Table 1. These metrics are widely used to measure how users engage with webpages or websites. This paper adapts from these metrics to devise a set of web-content-based engagement metrics.

2.2 Web User Engagement Metrics: A Physiological Approach

A physiological approach to user engagement measures the users involuntary body responses, e.g. eye tracking, mouse tracking, and facial expression analysis [7]. The measurement may involve specific hardware device, e.g. eye tracking, or only software, e.g. mouse movement detection. In this paper, we consider only the metrics that can be measured by software, i.e. mouse gestures, in particular. Huang et al. [8] examine the users' mouse cursor behavior on search engine results pages to help better design effective search systems. The mouse cursor behavior includes clicks, cursor movements, and hovers over different page regions. We adapt from the metrics in Table 2 which are used in their work.

Table 1 User engagement metrics based on web analytics [6]

Category	Metric	Description
Popularity (for a given time frame)	Users	Number of distinct users
	Visits	Number of visits (one user can visit more than once)
	Clicks (page views)	Number of page views (measured by number of clicks on link)
Activity	ClickDepth	Average number of page views per visit
	DwellTimeA	Average time per visit
Loyalty (for a given time frame)	ActiveDays	Number of days a user visited website
	ReturnRate	Number of times a user visited website
	DwellTimeL	Average time a user spends on website

Table 2 User engagement metrics based on physiological measure [8]

Metric	Description
Clickthrough rate	Average number of link clicked per visit
Hover rate	Average number of link hovered (mouse over link) per visit
Number of unclicked hover	Median of number of unclicks on link that is hovered
Maximum hover time	Time that the link is hovered
Cursor trail length	Distance that cursor moves (pixel)
Movement time	Time that cursor moves (pixel)
Cursor speed	Cursor trail length/Movement time

2.3 Web Analytics Tool

We discuss Google Analytics [9] as an example of the tools which can be installed on a website to track how users engage with the website. Table 3 shows metrics that are used.

Table 3 User engagement metrics used by Google Analytics [9]

Metric	Description
Session	Total number of sessions within the date range
Users	Users that have had at least 1 session within the selected date range
Pageviews	Total number of pages viewed
Unique pageviews	Number of sessions during which the specified page was viewed at least once
Pages/session	Average number of pages viewed during a session
Average page duration	Average length of a session
Average time on page	Average amount of time users spent viewing pages
Bounce rate	Percentage of single-page visits (i.e. visitors left your site from the entrance page without interacting with the page)
% New session	An estimate of the percentage of first time visits
Entrances	Number of times visitors entered your site through a specified page or set of pages
% Exit	(Number of exits)/(Number of pageviews) for the page or set of pages

Google Analytics visualizes the measurements as graphs and tables. Figure 2 depicts how Google Analytics, for example, shows pageviews of a single page in a graph and others in a table. This paper will use similar visualization techniques to report web-content-based user engagement information.

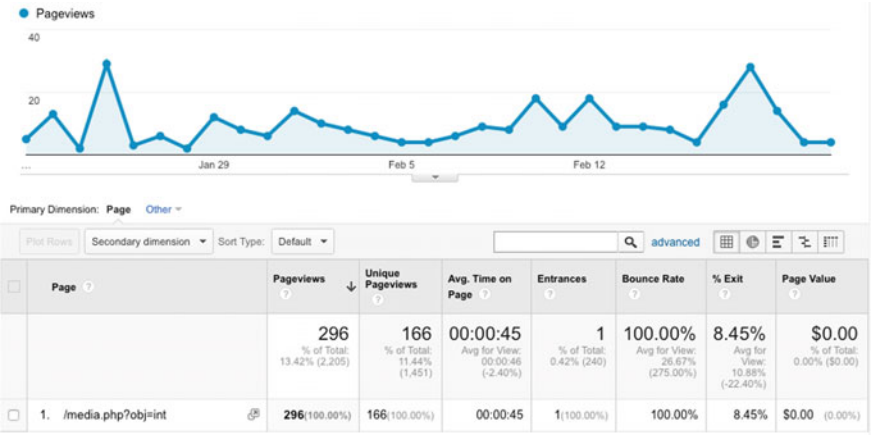

Fig. 2 Example of Google Analytics visualization

3 Web Contents Model and Content-Based Measurement

In this section, we give a definition of the web contents model, characteristics of web contents, and how user engagement is measured in this model.

3.1 Web Contents Model

The web contents model is depicted in Fig. 3. A website is a collection of webpages, where each webpage is composed of multiple containers. A container is a specific area on a particular webpage and can be nested inside another container. A web content is contained in a container, and can be seen on a web browser, e.g. text, image, sound, video etc. It can also be part of another content, e.g. text on an image.

3.2 Characteristics of Web Contents

Analyzing user engagement with web contents is different from that with webpages because of some unique characteristics of web contents as follows.

First, a web content can appear in different containers on different webpages at the same time. As shown in Fig. 4, a web content A can be in different containers on webpages 1 and 2 at the same time. We name a web content as cc or content in a container. That is, the same web content in different containers would be referred to

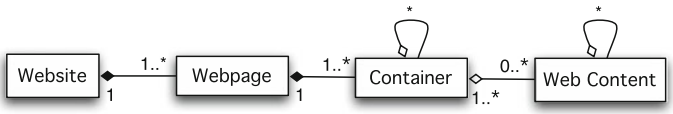

Fig. 3 Web contents model

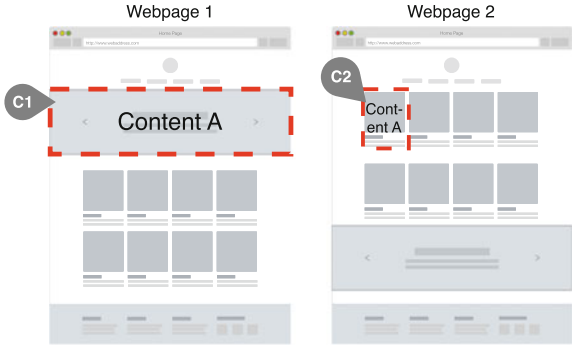

Fig. 4 A web content in different containers at the same time

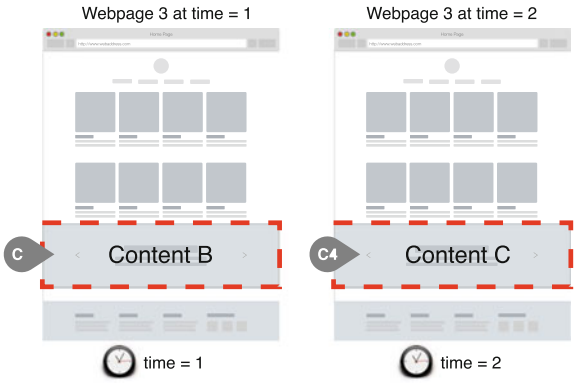

Fig. 5 Different web contents in the same container at different time

differently, e.g. the web content A in Fig. 4 is referred to as c1 and c2 with regard to its container on a webpage 1 and another one on a webpage 2 respectively.

Second, a container on a page can contain different web contents at different time. As shown in Fig. 5, a web content B contained in a container on a webpage 3 at time = 1 is referred to as c3. Later at time = 2, that container contains a different content C which will be referred to differently as c4.

Given these characteristics of web contents, we require a different set of metrics and tool for web-contents-based engagement analytics.

3.3 Measuring Content and Container Visit

Measuring visit to web contents and containers is different from measuring visit to webpages since the latter can be done when the page is loaded. However, for web contents and containers, they might be at the bottom of the page and cannot be seen by the user when the page is loaded. As shown in Fig. 6a, only the Content 1 is shown

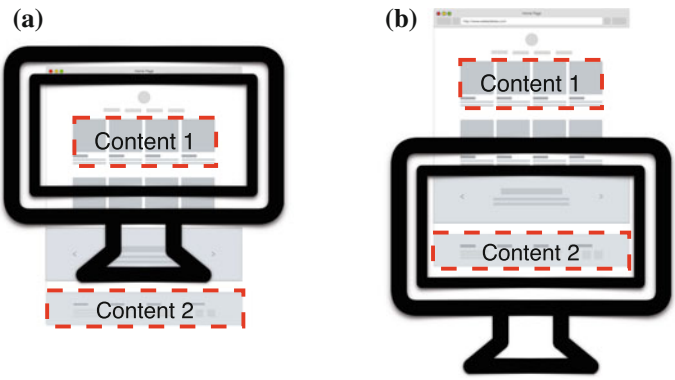

Fig. 6 Measuring content and container visit

in the viewport but the Content 2 is not because it is off the screen. In the case of measuring the visit to web contents and containers, the viewport of the screen will be used to determine whether the web contents and containers are visited. For example, Content 2 will be considered as visited when the page is scrolled down until shown in the viewport as in Fig. 6b.

4 User Engagement Metrics Based on Web Contents

The proposed user engagement metrics based on web contents are classified into three levels, i.e. low-level, high- level, and overall-level metrics. The low-level metrics are used to measure directly the behavior of the users on web contents. High-level metrics then are derived from the low-level ones. The overall-level metrics are further derived from the high- level metrics to obtain the overall engagement information. These three classes of metrics are adapted from several of those widely used engagement metrics for webpages discussed earlier in Sect. 2. Before we describe the proposed metrics, we first introduce a case study to which the metrics will be applied as an example.

4.1 Case Study

The example in Fig. 1 is used as a case study and revised with some information added as shown in Fig. 7. There are five CC or content in a container. c1, c2, and c3 are on page 1. At time = 1, c4 on page 2 has the same content (i.e. a music video A) as c2 on page 1. When time = 2, the container of c4 has its content changed to a music video B, and that location is then identified as c5. Table 4 summarizes the web content and container of each CC in the case study.

4.2 Low-Level User Engagement Metrics

Low-level user engagement metrics measure usage data directly from three sources, i.e. webpages, web contents, and containers. Since a content and a container share the same location on a webpage at the same time, we use the same low-level metrics for these two sources. Each metric is listed with its definition in Table 5.

Note that the metrics for webpage sources are taken from existing metrics but we enhance by proposing additional metrics for web contents and containers. Using the low-level metrics, we obtain the measurements of the case study in Tables 6 and 7.

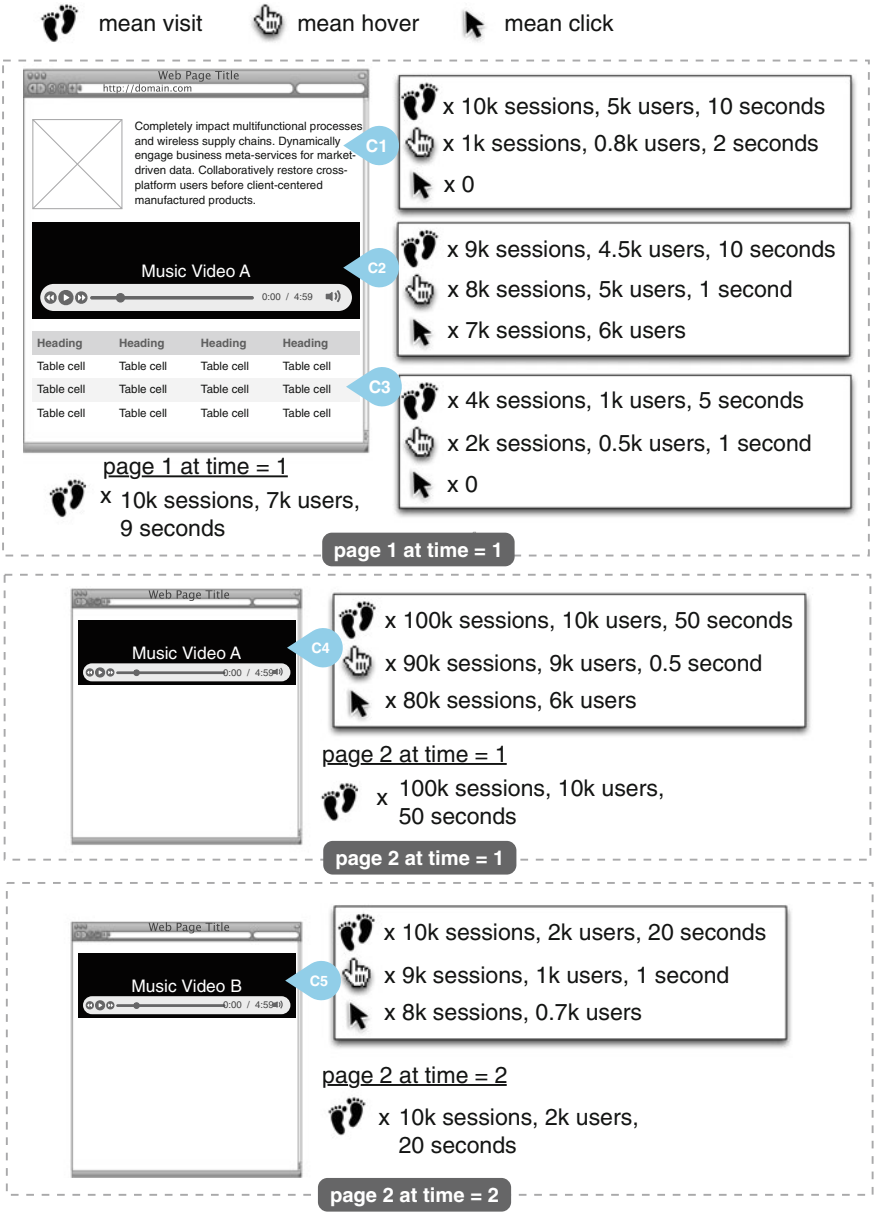

Fig. 7 Case study

Table 4 Content and container of each CC in case study

c1	c2	c3	c4	c5
Content 1[a]	Content 2[a]	Content 3[a]	Content 2	Content 4[a]
Container 1	Container 2	Container 3	Container 4	Container 4

[a] Content 1 = text, Content 2 = music video A, Content 3 = table, Content 4 = music video B

Table 5 Low-level user engagement metrics

Abbr.	Metric name	Definition
Webpage sources		
VS_{page}	Webpage Visit Session	Number of times users visited webpage for a given time frame
VU_{page}	Webpage Visit User	Number of distinct users visited webpage for a given time frame
VT_{page}	Webpage Visit Dwell Time	Total time that users visited webpage for a given time frame
Content/container sources[a]		
VS_C	C Visit Session	Number of times users visited C (counted when C appeared on screen) for a given time frame
VU_C	C Visit User	Number of distinct users visited C for a given time frame
VT_C	C Visit Dwell Time	Total time that users visited C (counted from the time C appeared on screen until C disappeared) for a given time frame
HS_C	C Hover Session	Number of times users hovered over C for a given time frame
HU_C	C Hover User	Number of distinct users hovered over C for a given time frame
HT_C	C Hover Time	Total time that users hovered over C for a given time frame
CS_C	C Click Session	Number of times users clicked on C for a given time frame
CU_C	C Click User	Number of distinct users clicked on C for a given time frame

[a]C stands for content or container

Table 6 Low-level measurements for webpage sources of case study

Abbr.	Metric name	Page 1	Page 2 at time = 1	Page 2 at time = 2
VS_{page}	Webpage Visit Session	10,000	100,000	10,000
VU_{page}	Webpage Visit User	7,000	10,000	2,000
VT_{page}	Webpage Visit Dwell Time	9	50	20

Table 7 Low-level measurements for content and container sources of case study

Abbr.	Metric name	c1	c2	c3	c4	c5
VS_C	C Visit Session	10,000	9,000	4,000	100,000	10,000
VU_C	C Visit User	5,000	4,500	1,000	10,000	2,000
VT_C	C Visit Dwell Time (seconds)	10	10	5	50	20
HS_C	C Hover Session	10,000	8,000	4,000	90,000	9,000
HU_C	C Hover User	800	5,000	500	9,000	1,000
HT_C	C Hover Time (seconds)	2	1	1	0.5	1
CS_C	C Click Session	0	7,000	0	80,000	8,000
CU_C	C Click User	0	6,000	0	6,000	700

4.3 High-Level User Engagement Metrics

High-level user engagement metrics are derived from low- level metrics and categorized by user behavior approaches, i.e. visit, click, and hover. Each metric is listed in Table 8 together with its definition. Table 9 shows the calculation results for the case study.

Table 8 High-level user engagement metrics

Abbr.	Metric name	Definition
Visit		
VSR_C	C Visit Session Rate	C Visit Session/C's Webpage Visit Session (VS_C/VS_{page})
VUR_C	C Visit User Rate	C Visit User/C's Webpage Visit User (VU_C/VU_{page})
VTR_C	C Visit Dwell Time Rate	C Visit Dwell Time /C's Webpage Visit Dwell Time (VT_C/VT_{page})
RR_C	Return Rate on C	C Visit User/Webpage Visit Session (VU_C/VS_{page})
Click		
SCR_C	C Session Clickthrough Rate	C Click Session/C's Webpage Visit Session (CS_C/VS_{page})
UCR_C	C User Clickthrough Rate	C Click User/C's Webpage Visit User (CU_C/VU_{page})
Hover		
SHR_C	C Session Hover Rate	C Hover Session/C's Webpage Visit Session (HS_C/VS_{page})
UHR_C	C User Hover Rate	C Hover User/C's Webpage Visit User (HS_C/VU_{page})
HTR_C	C Hover Time Rate	C Hover Time/C's Webpage Visit Dwell Time (HT_C/VT_{page})

High-level metrics can be used to determine user engagement with a web content in a container. However, to determine the overall user engagement with a web content that may appear in several containers as well as the overall user engagement with a container that may contain several contents over time, we need additional overall-level metrics.

4.4 Overall-Level User Engagement Metrics for Web Content

As depicted in Fig. 4, a web content may be contained in several containers across different webpages. To determine user engagement with this particular web content, we calculate an average engagement values of that web content over all containers in which it is contained:

$$Engagement_{content} = \frac{\sum_{i=1}^{n} Engagement_{content,i}}{n} \quad (1)$$

where $Engagement_{content}$ = overall user engagement with a web content at a particular time

n = number of containers in which that web content is contained, and

$Engagement_{content,i}$ = user engagement with that web content in the container i (obtained by using high-level metrics).

In the case study, c2 and c4 have the same content 2 (music video A). Using the engagement measurements for c2 and c4 in Table 9 as $Engagement_{content,i}$, we calculate the overall user engagement with the music video A at a particular time. In Table 10, the column "Content 2 at time = 1" lists the overall user engagement measurements for the music video A at time = 1. For example, $VSR_{content2} = (0.90 + 1.0)/2 = 0.95$ at time = 1.

Table 9 High-level measurements for content and container sources of case study

Abbr.	Metric name	c1	c2	c3	c4	c5
VSR_C	C Visit Session Rate	1.00	0.90	0.40	1.00	1.00
VUR_C	C Visit User Rate	0.71	0.64	0.14	1.00	1.00
VTR_C	C Visit Dwell Time Rate	1.11	1.11	0.56	1.00	0.40
RR_C	Return Rate on C	0.50	0.45	0.10	0.10	0.20
SCR_C	C Session Clickthrough Rate	0.00	0.70	0.00	0.80	0.80
UCR_C	C User Clickthrough Rate	0.00	0.86	0.00	0.60	0.35
SHR_C	C Session Hover Rate	0.00	0.70	0.00	0.80	0.80
UHR_C	C User Hover Rate	0.11	0.71	0.07	0.90	0.50
HTR_C	C Hover Time Rate	0.22	0.11	0.11	0.01	0.05

Table 10 Overall-level measurement of case study

Abbr.	Content				Container from time = 1 to time = 2			
	Time = 1		Time = 2					
	1	**2**	3	4	1	2	3	**4**
VSR_C	1.00	0.95	0.40	1.00	1.00	0.90	0.40	1.00
VUR_C	0.71	0.82	0.14	1.00	0.71	0.64	0.14	1.00
VTR_C	1.11	1.06	0.56	1.00	1.11	1.11	0.56	0.70
RR_C	0.50	0.28	0.10	0.10	0.50	0.45	0.10	0.15
SCR_C	0.00	0.75	0.00	0.80	0.00	0.70	0.00	0.80
UCR_C	0.00	0.73	0.00	0.60	0.00	0.86	0.00	0.48
SHR_C	0.00	0.75	0.00	0.80	0.00	0.70	0.00	0.80
UHR_C	0.11	0.81	0.07	0.90	0.11	0.71	0.07	0.70
HTR_C	0.22	0.06	0.11	0.01	0.22	0.11	0.11	0.03

4.5 Overall-Level User Engagement Metrics for Web Container

As depicted in Fig. 5, a container may contain several web contents at different time. It is useful to get an insight into how a particular container is engaged in order to design a container better or select suitable content for a container. To determine user engagement with a particular container over time, we calculate an average engagement values of that container over all contents contained in it:

$$Engagement_{container} = \frac{\sum_{i=1}^{t} Engagement_{container,i}}{t} \tag{2}$$

where $Engagement_{container}$ = overall user engagement with a web container over a range of time

t = number of times that user engagement with that container is determined (at a regular interval) over the time range, and

$Engagement_{container,i}$ = user engagement with that container measured at time i (obtained by using high-level metrics).

In the case study, c4 and c5 refer to the same container 4 (on webpage 2) with different contents (music video A and B) at different time. Using the engagement measurements for c4 and c5 in Table 9 as $Engagement_{container,i}$, we calculate the overall user engagement with this container 4 over a range of time. In Table 10, the column Container 4 from time = 1 to time = 2 lists the overall user engagement measurements for this container 4. For example, $VTR_{container4} = (1.0 + 0.4)/2 = 0.7$ over that time range.

5 Web-Contents-Based Analytics Tool

We develop a web-contents-based analytics tool that collects user interaction data, determine user engagements using the three classes of metrics, and visualize engagement information. The tool is implemented in PHP, with MySQL database. To install the tool on a website, a web developer has to register the website and generates jQuery code to put on the website. Once the code is installed, the tool will automatically collect necessary data, send to the server to process, and visualize the analytical results. The tool provides a dashboard by which the web developer can select particular web contents or containers (based on how the developers design the pages) and the engagement metrics that are of interest. Figure 8 shows an example of a graph displaying the UCR_C (User Clickthrough Rate) for the content 4 (music video B) at different time.

Fig. 8 Dashboard and visualization

6 Conclusion

The paper proposes an enhancement to existing web user engagement metrics by introducing additional engagement metrics that can take into account engagement with particular contents and areas on webpages. The analytical result is aimed to give an insight into user behavior and to help determine a way to improve website design by placing the right web contents at the right location. We are conducting a test on a commercial website that has installed the web-contents-based analytics tool in order to see if the site can benefit from the analytical results and improve their website design. Other future work includes improving the visualization of the tool and extending the set of engagement metrics.

References

1. Attfield, S., Kazai, G., Lalmas, M. et al.: Towards a science of user engagement (Position Paper). In: WSDM Workshop on User Modelling for Web Applications. http://www.dcs.gla.ac.uk/~mounia/Papers/engagement.pdf (2011). Accessed 9 April 2016
2. Datanyze: Analytics market share in the Alexa top 1M. https://www.datanyze.com/market-share/Analytics/ (2016). Accessed 9 April 2016
3. Google Analytics: Brian Gavin Diamonds sees 60 % increase in customer checkout with Google enhanced ecommerce. https://static.googleuserconent.com/media/www.google.com/en//intl/es%20_ALL/analytics/customers/pdfs/brian-gavin.pdf (2016). Accessed 9 April 2016
4. Google Analytics: With help from Periscopix, WBC employs advanced segments in Google Analytics to boost its e-commerce conversion rate by more than 12 %. https://static.googleusercontent.com/media/www.google.com/en//intl/es%20_ALL/analytics/customers/pdfs/wbc.pdf (2016). Accessed 9 April 2016
5. Lehmann, J., Lalmas, M., Baeza-Yates, R. et al.: Networked user engagement. In: Proceedings of the 1st Workshop on User Engagement Optimization (UEO'13) (2013)
6. Lehmann, J., Lalmas, M., Yom-Tov, E. et al.: Models of user engagement. In: Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization (UMAP'12) (2012)
7. Lalmas, M., O'Brien, H., Yom-Tov, E.: Measuring User Engagement. In: Marchionini, G. (ed.) Synthesis Lectures on Information Concepts, Retrieval and Services. Morgan & Claypool Publishers (2015)
8. Huang, J., White, R., Dumais, S.: No clicks, no problem: using cursor movements to understand and improve search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11) (2011)
9. Google Analytics: Google Analytics Solutions. https://www.google.com/intl/en/analytics/#?modal_active=none (2016). Accessed 9 April 2016