

Measuring Web Page Revisitation in Tabbed Browsing

Haimo Zhang, Shengdong Zhao

Computer Science, National University of Singapore

{z-haimo, zhaosd}@comp.nus.edu.sg

nus-hci.comp.nus.edu.sg

ABSTRACT

Browsing the web has been shown to be a highly recurrent activity. Aimed to optimize the browsing experience, extensive previous research has been carried out on users' revisitation behavior. However, the conventional definition for revisitation, which only considers page loading activities by monitoring http requests initiated by the browser, largely underestimates users' intended revisitation activities with *tabbed browsers*. Thus, we introduce a goal-oriented definition and a refined revisitation measurement based on page viewings in tabbed browsers. An empirical analysis of statistics taken from a client-side log study showed that although the overall revisitation rate remained relatively constant, tabbed browsing has introduced new behaviors warrant future investigations.

Author Keywords

Web revisitation, tabbed browsing, effective revisitation.

ACM Classification Keywords

H5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia – user issues.

General Terms

Human factors

INTRODUCTION

Browsing the web has been shown to be a highly recurrent activity [7]. Extensive research that has been conducted to understand user behavior and optimize the browsing experience focused on “web page revisitation”, a term that refers to the repeated visits to previously visited web pages [3,5,6,7].

Previous studies, conducted mostly between 1994 and 2000, were mostly on non-tabbed browsers [2,5,7] and revisitation was defined as “the repeated *loading* of a web page as identified by its URL” [7].

Following this definition, the revisitation rate is calculated using the number of repeated page loading events (the difference between the total number of page loading events

and the number of distinct URLs loaded, termed as the size of the URL vocabulary [7]) divided by the total number of page loading events.

While to some extent, tabbed browsing behavior can be achieved using multiple instances of non-tabbed browsers, but the pervasive usage of tabbed browsing behavior did not happen until major browsers adopted the tab feature. In 2009, Dubroy and Balakrishnan [3] conducted a study on user behaviors in Mozilla Firefox browser, highlighting the significance of tab usage in revisiting web pages. In 2010, Huang and White [4] investigated parallel browsing, specifically on how web searching tasks were performed with multiple tabs. Both papers adopted the conventional definition for revisitation and did not point out how repeated page visits in non-tabbed browsers were different from those in tabbed browsers.

Scope and Definition of Terms

From the user's perspective, the purpose of “visiting” a web page is to get information. Thus, we propose the term “effective revisitation” to describe the repetition of *obtaining information* from a web page as identified by its URL. Since the majority of web content is visual (rather than audio), we focus on revisitation of visual content in this paper.

A clear line needs to be drawn between the definition and the measurement of revisitation. Our proposed definition explains what revisitation is, rather than how it is measured. The conventional definition is somewhat misleading as it is actually a measurement of revisitations in non-tabbed browsers.

To understand why the conventional measurement for revisitation is insufficient under the tabbed browsing paradigm, several terms first need to be clarified:

- *Focused tab* or *current tab*. The current visible tab in a browser. In tabbed browsers, only the tab that is being displayed is visible. Non-tabbed browsers could be thought of as having only one tab, which is always in focus.
- *Background tabs*. They only exist in a tabbed browser. They contain opened pages, but unlike the focused tab, their content is not visible. All tabs except the focused tab are background tabs.
- *Page loading*. A page loading event is recorded whenever an http request is sent to the server.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

- *Page viewing.* A page is viewable when the tab containing it is in focus. Since it is difficult to know whether the user is actually viewing a web page unless eye-tracking mechanisms are used, we assume that the user is viewing a page whenever it is displayed in the focused tab.

Limitations of the Conventional Measurement for Revisitation

In any single instances of non-tabbed browsers (since non-tabbed browsers only displays one web page, it can also be regarded as single-tabbed browsers), all web pages are loaded into the only tab, which is always in focus; so they are displayed whenever they are loaded. Therefore, the conventional measurement for revisitation is able to determine the number of page viewings through the concurrence of page loading events and page displaying events, given the assumption that page displaying equals page viewing.

In tabbed browsers, however, a page loading is not always a page displaying. The conventional measurement cannot be used to determine the number of page viewings since it would introduce two types of errors:

- *Over-count of revisitation activities.* In tabbed browsers, a background tab can be closed without being viewed. This introduces three types of over-count errors that will be counted as revisitations:

Type 1. A previously loaded-but-**not**-viewed page is loaded again and viewed;

Type 2. A previously loaded-but-**not**-viewed page is loaded again but **not** viewed; and

Type 3. A previously loaded-and-viewed page is loaded again but **not** viewed.

In all of these cases, the same page was loaded several times but viewed not more than once. They are considered revisitations in the conventional definition but should be excluded under the new definition.

- *Under-count of revisitation activities.* When users switch to a tab to display its content, which has been viewed before, no additional loading events are triggered. The conventional definition does not consider this behavior as a revisitation; our proposed definition does.

Proposed Approach for Measuring Revisitations

To accurately measure revisitation in tabbed browsing, we propose to focus on page viewings instead of page loadings. There are two types of page viewing activities in tabbed browsers:

- *Loading-based page viewing.* A web page (uniquely represented by its URL) that is loaded into the focused tab is considered viewed.

- *Tab-switching-based page viewing.* A page in a background tab is considered viewed when the tab becomes the focused tab.

The *effective revisitation rate* (*EffRev%*) is calculated as the number of *effective revisits* (or repeated page viewings) (*#EffRev*) divided by the total *number of page viewings* (*#View*). The definitions of each term are listed below.

- *Number of page viewings* (*#View*). The total number of loading-based and tab-switching-based page viewings.

- *Number of effective revisits* (*#EffRev*). A subset of all page viewings. Effective revisits are viewings of previously viewed pages.

- *Effective revisitation rate* (*EffRev%*). The formula $\#EffRev/\#View$ calculates the effective revisitation rate based on page viewings.

Note that the difference between total number of page viewings and URL vocabulary size (number of distinct URL loads) does not equal the number of effective revisits, as some URLs could have been loaded but never viewed.

The new definition and measurement prompted the following research questions:

- Using the proposed viewing-based measurement, what is the revisitation rate under the tabbed browsing paradigm? How does it differ from the results using the conventional measurement?

- How significant are the over-count and under-count errors when the conventional method is used to measure revisitations in tabbed browsers?

- Has the frequency of users' repeated page visits changed in recent years?

To answer these questions, we carried out a one-month study of 20 participants and their browsing behaviors in tabbed browsers.

A STUDY OF REVISITATION BEHAVIOR

After examining the design of Dubroy and Balakrishnan's study on tabbed browsing behavior [3], we found that both their logger and study procedure can be used for our research on revisitation. We adopted those and changed only the interview questions to target our topic.

Participants and Duration

20 participants (7 females, age range 23-26, mean 24.1) from the university community took part in the one-month study. All participants use Internet daily on Microsoft Windows Operating Systems with Mozilla Firefox as the primary browser. Participants were instructed to browse the Internet as usual. No additional demonstration or suggestion of using tabs was given prior to or during the study.

Results and Analysis

A total of 235,707 browser events were captured from the 20 participants over one-month, among which there were 89,851 page loadings and 127,344 page viewings. We do not further distinguish the type of information users try to get (whether it is advertisement, news reading, etc.). While there are different ways to define a page, we adopted the

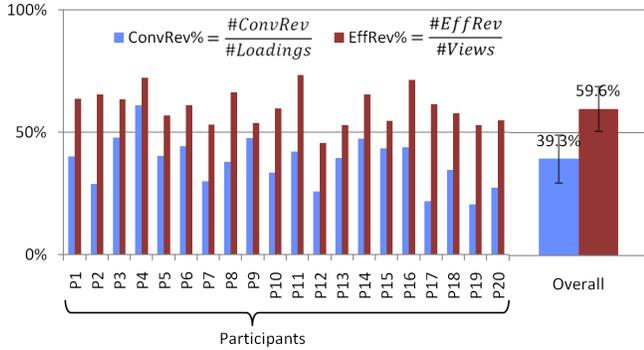


Figure 1. Significant Difference between Conventional (39.3%) and Effective (59.6%) Revisitation Rates

method of using full URL and the HTTP GET parameters to uniquely identify a web page [6].

Using the proposed viewing-based measurement, what is the revisitation rate under the tabbed browsing paradigm? How does it differ from the results using the conventional measurement?

The blue bar on the right end of figure 1 shows the overall conventional revisitation rate (39.3%, or 35,342 out of 89,851 loading events) based on page loadings, which was calculated using Tauscher and Greenberg’s method [7]. The red bar besides it shows the overall effective revisitation rate (59.6%, or 75,912 out of 127,344 page viewing events) using our proposed definition and measurement. The bars on the left of the figure show revisitation rates using the two measurements for each participant.

Our calculation shows that the conventional measurement largely underestimates the amount of revisitation activities under tabbed browsing ($T_{19} = 11.25, p < .001$). If we break down the effective revisitation of all 75,912 events into the two types of revisitation (loading-based revisitation and tab-switching-based revisitation), the former comprises 53.0% (40,221 events) while the latter comprises 47.0% (35,691 events). This shows that tab-switching-based revisitation, which was neglected in previous studies, is about as frequent as loading-based revisitation. It reinforces the point raised by Dubroy and Balakrishnan [3] that tab switching should be considered an important means of revisitation.

How significant are the over-count and under-count errors when the conventional method is used to measure revisitations in tabbed browsers?

There are a total of 4,135 over-count error events (11.7% of all conventional revisits). Of these, 3.9% (160) are type 1, 55.2% (2,283) are type 2 and 40.9% (1,692) are type 3.

There are 38,639 under-count error events (50.9% of all effective revisits), in which revisitation activities were done with tab switching alone.

These results show that conventional measurement suffers from both over-count and under-count errors. The under-count errors are much more than over-count errors, giving rise to the overall effect that conventional measurement largely underestimates revisitations in tabbed browsing.

Has the frequency of users’ repeated page visits changed in recent years?

We found four previous studies that explicitly addressed revisitation rate and present them chronologically alongside our study (table 1). The first three studies used non-tabbed browsers, while the last two studies used tabbed browsers. The study of Dubroy and Balakrishnan [3] is not included since it did not explicitly report revisitation rate.

Before the introduction of tabbed browsers, effective web revisitation activities were accurately measured as the loading of previously loaded pages. This was done in studies 1, 2 and 3, all of which have revisitation rates of above 50%.

Did the introduction of tabbed browsers change this rate? Study 4 suggested that there is a change: revisitation rate dropped to 43.7%. However, this may be misleading since the conventional measurement for revisitation was used in tabbed browsing (at least 15 out of the 25 participants in study 4 used tabbed browsers).

Our study shows that the effective revisitation rate has not dropped that dramatically (59.6% is lower than study 3’s 81%, but is comparable with study 1’s 61% and study 2’s 58%) even with the introduction of tabbed browsers. Due to lack of tab switching data from the previous studies, we are unable to calculate effective revisitation rates for study 4.

	Catledge & Pitkow [2]	Tauscher & Greenberg [7]	McKenzie & Cockburn [5]	Obendorf & Weinreich [6,8]	Our study
Date of Study	1994	1995-1996	1999-2000	2004-2005	2010
# Participants	107	23	17	25	20
Duration	21 days	42 days	119 days	52-195 days	31 days
Revisitation Rate	61%	58%	81%	43.7%	59.6% (Effective), 39.3% (Conventional)
Browsers Used	XMosaic	XMosaic 2.6	Netscape Navigator	Partly (15/25) <i>Firefox 1.0</i>	<i>Firefox 3.0</i>
# Page Loadings	31,134	19,000	83,411	137,272	80,811 (113,243 page views)
# Distinct URLs			17,242	65,643	49,450

Table 1. Comparisons between previous studies and our study (tabbed browsers are italicized)

While our study results suggest possible reasons behind the noticeable difference between study 4 and previous studies, comparison across studies should be cautiously interpreted to consider the differences in methods and other possible limitations. The purpose of this comparison is not to judge the previous studies, but to place our study in context and to highlight any significant difference to invite further investigations.

DISCUSSION, LIMITATION AND FUTURE DIRECTIONS

While our proposed method of measuring web page revisitation shed new light into research on the topic, it is, strictly speaking, still an approximation of actual revisitation behavior. It also has issues in overcounting and undercounting page revisits.

Our definition of revisitation is “the repetition of obtaining previously seen information,” but what accounts for “obtaining information” needs further clarification. Tab switching is one way of obtaining information from a web page contained in a tab. However, users may switch through tabs to find a previously opened tab or click on a tab by accident. In these cases, the tab switchings result in transient page views. Users do not actually intend to obtain information from these pages.

These tab switchings should be excluded from the calculation, and one way to do so is by excluding page views that are shorter than a certain time threshold. Table 2 tabulates number of page views, number of page revisits, and effective revisitation rates under different cut-off time thresholds.

As shown in table 2, although transient tab switchings increase the total number of page revisits, it is unlikely to significantly affect the overall revisitation rate. While applying a time threshold can help to remove some of the transient pages, determining the exact time threshold is difficult due to different scenarios and individual differences. Future research may use other methods (e.g., eye tracking) to more accurately tell whether the user is viewing the page to obtain information.

Undercounting may be caused by using multiple browser windows and revisiting non-visual contents of a page, e.g., radio stations. While these two cases happened rarely in our log study, future research in web page revisitation should take them into consideration.

Despite these limitations, we believe our method can be used as one of the important measures for future studies in

Threshold (seconds)	#page revisits	#page views	EffRev rate
0	75912	127344	59.6%
1	66359	113242	58.6%
2	56653	97516	58.1%
3	46625	81322	57.3%
4	40638	71269	57.0%

Table 2. Effective revisitation rates with varying cut-off threshold in seconds

web page revisitation. While it will be difficult to completely remove all issues in over and undercounting in any method of measurement, future researchers may consider reporting a percentage range, in which the actual revisitation rate resides with high probability (e.g., >95%).

In the future, our method of measurement could be further integrated into large scale analyses of web usage [1] or existing user behaviors such as back tracking [8], to deepen our understanding of tabbed browsing and to improve user experience.

CONCLUSION

We propose a goal-oriented definition and measurement for revisitation under the tabbed browsing paradigm. Our client-side log study shows that the conventional measurement for revisitation largely underestimates revisitation activities in tabbed browsing. Although the overall revisitation rate has remained relatively steady over the years, tabbed browsing has introduced new behaviors. In the future, these need to be taken into account in studies of web page revisitation and when optimizing the browsing and revisitation experience in tabbed browsers.

ACKNOWLEDGEMENT

We thank Xiaojuan Ma and anonymous reviewers for their constructive comments and feedback. We thank members of the NUS-HCI Lab for their support. This research is supported by National University of Singapore Academic Research Fund R-252-000-375-133.

REFERENCES

1. Adar E, Teevan J, and Dumais ST. Large scale analysis of web revisitation patterns. In *Proc. CHI 2008*, ACM Press (2008), 1197-1206.
2. Catledge, L.D., Pitkow, J.E. Characterizing browsing strategies in the World-Wide Web. In *Proc. WWW 1995*, ACM Press (1995), 1065-1073.
3. Dubroy, P. and Balakrishnan, R. A study of tabbed browsing among Mozilla firefox users. In *Proc. CHI 2010*, ACM Press (2010), 673-682.
4. Huang, J. and White, R. Parallel Browsing Behavior on the Web. In *Proc. HT 2010*, ACM Press (2010).
5. McKenzie, B. and Cockburn, A. An empirical analysis of web page revisitation. *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*
6. Obendorf, H., Weinreich, H., et al. Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In *Proc. CHI 2007*, ACM Press (2007), 597-606.
7. Tauscher, L. and Greenberg, S. Revisitation patterns in World Wide Web navigation. In *Proc. CHI 1997*, ACM Press (1997), 399-406.
8. Weinreich H, et al. Not quite the average: An empirical study of Web use. *ACM Trans. Web* 2008; 2(1): 1-31.