A Longitudinal Study of How Highlighting Web Content Change Affects People's Web Interactions

Jaime Teevan, Susan T. Dumais, Daniel J. Liebling
Microsoft Research
Redmond, WA, USA
{teevan, sdumais, danl}@microsoft.com

ABSTRACT

The Web is constantly changing, but most tools used to access Web content deal only with what can be captured at a single instance in time. As a result, Web users may not have a good understanding of the changes that occur. In this paper we show that making Web content change explicitly visible allows people to interact with the Web in new ways. We present a longitudinal study in which 30 people used a Web browser plug-in that caches visited pages and highlights text changes to those pages when revisited. We used a survey to capture their understanding of Web page change and their own revisitation patterns at the beginning of use and after one month. For a majority of the participants, we also logged their Web page visits and associated content change. Exposing change is more valuable to our participants than initially expected, making them aware of how dynamic content they visit is and changing their interactions with it.

Author Keywords
Web dynamics, longitudinal study, revisitation, re-finding.

ACM Classification Keywords
H5.4. Information interfaces and presentation: Hypertext/Hypermedia: User issues.

General Terms
Design, Human Factors, Measurement

INTRODUCTION
Web content changes regularly [4] and the content that people revisit is particularly likely to change [1]. While Web users have a general awareness of this fact, exactly how the content with which they interact changes may not be obvious to them. For example, people may recognize the ads change every time they issue the same query to a search engine, but they are less likely to realize that the results returned also often change [9]. Similarly, although people know that news sites change, they are probably unaware of exactly how and when those changes occur. Awareness of the details of change could cause people to use search engines and more general Web content differently.

In this paper we explore how making Web content change explicit affects how people understand that content and how they interact with it. We conducted a month-long study in which 30 people used a Web browser plug-in designed specifically to provide persistent, in-situ, self-referenced awareness of Web content change, as shown Figure 1. We show how the tool shaped people's revisitation patterns, awareness and use of dynamic Web content, and discuss the implications of our findings.

RELATED WORK
Researchers have studied how frequently Web pages change and by how much, and found that there are significant amounts of change. For example, Fetterly et al. [4] found that 35% of Web pages changed over 11 weeks. Adar et al. [1] sampled Web pages that people visited and found much higher rates of change (65% of pages changed over 5 weeks). They also reported that pages of different categories change at different rates.

Most Web tools do not explicitly deal with Web content change. Web browsers, for example, show only the current version of a page. Some systems like the Internet Archive (archive.org) provide access to historical versions of Web pages, and Web search engines provide a cached version of the pages they index. Other systems, such as the Firefox Update Scanner (updatescanner.mozdev.org), WebSite-Watcher (aignes.com), Change-Detect (changedetect.com), and browser enhancements explored by Kellar et al. [6] support monitoring and notification of Web page change. Jatowt et al. [5] discuss several ways historical information about Web pages could be used to enrich Web browsing.

Figure 1. An example of changes to the CHI 2010 page highlighted by the browser plug-in used by participants.

Question			Initial	Follow-up	*p*-value
How valuable {do you expect to, did you} find DiffIE to be?			2.83	3.20	0.04
How often do you revisit Web pages you have seen before?			4.03	4.17	0.11
When you revisit a page, how often is it to view new content?			3.63	3.77	0.15
What proportion of the pages that you visit change regularly?			3.33	3.60	0.04
How often do you notice changes in Web content you didn't expect to change?			2.63	3.07	< 0.01
How often do you expect the following page types to change?	Above average initial change	News pages	4.43	4.67	0.05
		Message boards, forums, newsgroups	4.27	4.21	0.41
		Search engine results	3.90	3.80	0.38
		Blogs you read	3.47	3.75	0.04
	Below average initial change	Pages with product information	2.93	2.76	0.24
		Wikipedia pages	2.90	2.67	0.04
		Company homepages	2.87	3.13	0.07
		Personal home pages of people you know	2.87	2.61	0.04
		Reference pages (dictionaries, yellow pages, maps)	2.67	2.59	0.29

**Table 1. Answers to a survey about participants' experience with changing Web content.
Participants took the survey before and after using a tool that exposes change for a month.**

A number of tools have been developed to allow people to view how a Web page has changed given a previous version [2, 3, 7]. While most of these systems require the user to actively request to view change-related information, Teevan et al.'s DiffIE system [10] provides constant awareness by highlighting changes when a person revisits a Web page. The research presented here builds on this previous work by exploring how such constant awareness changes the way people use and understand Web content.

METHODOLOGY
We conducted a study designed to capture people's Web experiences before, during, and after exposure to Web content change. In the study we asked 30 people to install DiffIE [10] and use it for their daily Web browsing. DiffIE is a simple Web browser plug-in that caches the content of the Web pages the user visits, and highlights any changes to the text content when the user returns to the Web page. An example of a highlighted Web page can be seen in Figure 1.

Prior to installation, participants completed a survey about:
- How valuable they expected the tool to be,
- their perception of their *revisitation* patterns, including how often they revisit Web pages and how often they revisit Web pages to see new content, and
- their perception of how the pages they visit *change*, including what proportion of Web pages they revisit change, how often they notice unexpected changes, and how often various types of pages change.

The exact questions asked are shown in Table 1. Ratings were given on a 5-point scale, with 1 indicating *none* or *never* and 5 indicating *all* or *always*. A month after installing the tool, the same 30 people answered the exact same questions. How the answers changed provides insight into how having changes highlighted modifies the way they understand Web content.

Seventeen of the 30 participants also opted in to having their interactions with the tool logged. We collected the URLs that they visited, whether they had previously visited the URL, whether the Web page content had changed, and by how much the content had changed if it had. After removing obvious instances of auto-refresh and visits to locally hosted URLs, this yielded approximately 25,000 page visits. A summary of what the data reveals about participants' revisitation patterns from the first week and last weeks of use can be found in Table 2. The overall revisitation frequency is consistent with prior work [8].

During the follow-up survey we also collected free form comments about users' positive and negative experiences with the tool, and asked about which situations they found the highlighting to be most useful for (Table 3). Combined with the logging information, this provides insight into why the observed changes in perception might have occurred.

All participants were Microsoft employees, and were recruited via an email to mailing lists focused on new and developing Internet technology. The individuals (26 male, 4 female) come from disciplines including engineering, consulting, and research, and represent seven countries. Significance is reported using one-tailed paired *t*-tests.

RESULTS
Tables 1 summarizes the initial and follow-up responses to the survey questions after using DiffIE for at least a month. During this time participants viewed thousands of Web pages with changes highlighted. Participants found the tool to be significantly more valuable than initially expected. Upon installation, the tool's value was estimated at 2.83; the follow-up survey revealed the value after extended use was 3.20 ($p < .05$). Here we explore potential reasons for this increase by looking at how the tool changed people's perceptions of and interactions with Web pages.

Web Page Revisitation
We begin by looking at how highlighting affected people's perceptions of their revisitation patterns, as well as their actual revisitation patterns. As can be seen in Table 1,

participants reported revisiting Web pages at the same frequency at during the first week as they did during the last (roughly "often"). They also reported that change to the revisited page's content motivated their revisitation behavior to the same degree prior to use as following.

Although participants' perceptions of their revisitation patterns were not significantly changed, we observe a marginally significant change in actual revisitation patterns (Table 2). During their first week using the tool and seeing changes highlighted, 39.4% of participants' page visits were revisits; during the last week 45.0% of their visits were revisits ($p = .06$). Being able to identify changes that have occurred to a page since it was last visited may motivate returning.

Web Page Change
To explore this further, we looked at how people's perception of and interaction with changing Web content changed during the study period. Here we observe significant differences in perception. When asked, "What proportion of the pages that you visit change regularly?" prior to having changes highlighted, people reported a value of 3.33, or "some." After a month, that value changed significantly ($p < .05$) to 3.60, or closer to "most".

There are several reasons why people may believe that a higher proportion of the Web pages they visit change. One is that by drawing their attention to change, the highlighting made them aware of the fact that the pages they already visit change more than expected. Evidence for this can be found in the significant change to participants' response to the question, "How often do you notice changes in Web content you didn't expect to change?" Initially they responded with an average value of 2.63, or between "rarely" and "sometimes." A month later, that response increased significantly ($p < .01$) to 3.07. By highlighting changes that have occurred, and drawing people's attention to unexpected changes, the tool may be making them aware of general trends in Web content change.

Participants may also perceive that they interact with pages that change more because they actually do visit more pages that change. To explore this, we turn to the log data, and find that over the study period people start to revisit more dynamic pages. During the first week with changes to Web content highlighted, 21.5% of the revisited pages changed on average, whereas during the last week 32.4% did ($p < 0.01$). Further, when a revisited page changed, it changed by a greater amount. During the first week, when a page changed, 6.2% of the content (in terms of DOM nodes) was different; during the last week, significantly more of the content changed (9.5%, $p < .05$).

Participants' free-text comments about their experiences with having change highlighted suggest that being able to see changes may drive visits to pages that change more often. For example, one participant reported refreshing a page to be able to quickly find what content on that page changed: "*Refreshing web pages while watching CEDIA*

	First week	Last week	p-value
% of visits that are revisits	39.4%	45.0%	0.06
% of revisits that have changed	21.5%	32.4%	< 0.01
% page changed given change	6.2%	9.5%	0.02

Table 2. Summary of log data collected by the tool.

announcements allowed me to quickly view the new content as well as new comments."

Web Page Change by Page Type
We also looked at how our participants' perception of Web page change by varied by page type. An interesting trend that emerged was the amplification of their beliefs of change frequency for certain types of pages. The page types that were initially thought to change a lot were later thought to change even more; likewise the pages initially thought to change a little were later thought to change even less. The page types thought to change more than average (top four types in Table 1) received an average initial score of 4.02, and an average follow up score of 4.11. In contrast, the pages that people thought changed less than average (bottom five types in Table 1) received an average initial score of 2.85, and a follow up score of 2.75.

This trend is illustrated by the four significant changes observed among the individual page types. News pages and blogs typically changes rapidly as new stories or posts are added. Participants stated that they found the news pages and blogs they read to change more after a month with highlighting (news pages: 4.43 to 4.67, $p < .05$; blogs: 3.47 to 3.75, $p < .05$). On the other hand, Wikipedia pages and personal homepages change relatively infrequently. These pages people experienced to change less at the end of a month having changes highlighted (Wikipedia pages: 2.90 to 2.67, $p < .04$; personal homepages 2.87 to 2.61, $p < .04$).

A possible explanation for this trend is that prior to having Web page change made explicit, our participants had a base assumption about the amount of change to Web content that commonly occurred. This base assumption may have varied some by page type, but not as much as it actually varies in the real world. It is possible that after having the Web content change exposed, they were then better able to differentiate page types that changed a lot from page types that changed little. In general, participants' intuitions of the relative frequency of change to different page types agreed with the empirical change rates found by Adar et al. [1], with the exception of personal pages.

Using Change
We also looked at what participants reported to be the most valuable types of Web behavior the highlighting of change enabled. Inspired by the different scenarios reported by Teevan et al. [10] for how people use Web content change, we asked participants to tell us how useful they found change highlighting to be in supporting different types of online behaviors. The results are reported in Table 3.

Use Scenarios	Rating
Finding unexpected but important changes	3.80
Drawing attention to expected changes (e.g., new blog entries)	3.14
Viewing edits made by yourself or by someone else at your request	3.10
Understanding that blocks of content change (or don't change) as a group	3.10
Seeing changes to activity counters (e.g., thread views, new comments)	3.10
Supporting serendipitous encounters	2.93
Noticing webpage features you hadn't noticed before	2.86
Monitoring changing content (e.g., weather, stock prices)	2.86
Finding unexpected but unimportant changes	2.80

Table 3. Value assigned by participants to different uses of the tool following a one month period of use.

The most valuable scenario was viewing unexpected but important changes. For example, one participant noted, "*Price changes for homes are being highlighted as they change which is really helpful.*" Two other valuable scenarios include viewing expected changes to web content ("*I really like when I hit a news site that I don't have to re-read content.*"), and viewing edits to a Web page ("*Changes made to a wiki were more easy to spot.*").

Two of the other most valuable scenarios relate to Web page content being used in new and different ways as a result of having content change exposed. Highlighting enabled participants to group blocks of content based on common change patterns, and made activity counters (e.g., counters of thread views or new comments) more useful because activity became visible. There was evidence in the comments, too, that highlighting helped participants understand page content differently. One participant said, "*I get surprised to see sections of web pages changing, which I had perceived as static.*" Another reported, "*I can easily find if there is anyone [who] re-edited their post.*" This would have been impossible without highlighting.

Monitoring content for change is a common Web activity [6]. Although it was not listed as one of the more valuable uses of highlighting, several participants mentioned the behavior in their free text comments. For example, one reported, "*There's a page I go to for monitoring the jobs submitted to a [computing] cluster, and DiffIE makes it easier to tell which jobs are new, which jobs have had status changes, etc.*" The discrepancy between the tool's value for monitoring in the data in Table 3 and the free text comments may be that highlighting change is not useful for traditional monitoring tasks like weather and stock prices, but is useful for monitoring information on pages not specifically designed to support monitoring behavior.

Participants did not seem to particularly value having changes that were not interesting to them highlighted. Serendipitous encounters and unexpected and unimportant changes were relatively less valuable when compared to other scenarios. In the comments, several participants mentioned that they did not like having unimportant changes highlighted. For example, one said, "*Some pages have date stamps. The date changes daily, and DiffIE points out the change. This is not valuable.*"

CONCLUSION

In this paper, we have looked at how people's perceptions of and interactions with changing Web content evolved during extended use of a tool that highlighted the changes. We found that people revisited more Web content as a result, and that the content they revisited changed more in a way that was easier to perceive and use. Further, for certain types of pages, participants' beliefs in the page's change frequency were strengthened.

Our study provides evidence that increasing the awareness of change on the Web, in a very simple manner, can affect revisitation patterns, information use, and people's conceptions of Web content. Some participants reported the highlighting had become an indispensible part of their browsing, and this is because it allowed them to interact in the dynamic information environment of the Web in an entirely new way.

REFERENCES

1. Adar, E., J. Teevan, S. T. Dumais, and J. L. Elsas. The Web changes everything: Understanding the dynamics of Web content. *WSDM 2009,* 282-291.

2. Borodin, Y., J. P. Bigham, R. Raman and I. V. Ramakrishnan. What's new? – Making Web page updates accessible. *ASSETS 2008*, 145-152.

3. Douglis, F., T. Ball, Y.Chen, and E. Koutsofios. The AT&T Internet Difference Engine: Tracking and Viewing Changes on the Web. *WWW 1998*, 27-44.

4. Fetterly, D., M. Manasse, M. Najork, and J. Wiener. A large-scale study of the evolution of Web pages. *WWW 2003*, 669-678.

5. Jatowt, A., Y. Kawai, H. Ohshima, and K. Tanaka. What can history tell us? Towards different models of interaction with document histories. *HT 2008*, 5-14.

6. Kellar, M., C. Watters, and K. M. Inkpen. An exploration of Web-based monitoring: Implications for design. *CHI 2007*, 377-386.

7. Liu, L., C. Pu, and W. Tang. WebCQ: Detecting and delivering information changes on the Web. *CIKM 2000*, 512-519.

8. Obendorf, H., H. Weinreich, E. Herder, and M. Mayer. Web page revisitation revisited: implications of a long-term click stream study of browser usage. *CHI 2007*, 597-606.

9. Selberg, E. and Etzioni, O. On the instability of Web search engines. *RIAO 2000*, 223-235.

10. Teevan, J., S. T. Dumais, D. J. Liebling and R. Hughes (2009). Changing the way people view changes on the Web. *UIST 2009,* 237-246.