Web User Interaction Mining from Touch-Enabled Mobile Devices

Jeff Huang
Information School
University of Washington
hcir2012@jeffhuang.com

Abdigani Diriye
Human Computer Interaction Institute
Carnegie Mellon University
abdigani.diriye@gmail.com

ABSTRACT
Web services that thrive on mining user interaction data such as search engines can currently track clicks and mouse cursor activity on their Web pages. Cursor interaction mining has been shown to assist in user modeling and search result relevance, and is becoming another source of rich information that data scientists and search engineers can tap into. Due to the growing popularity of touch-enabled mobile devices, search systems may turn to tracking touch interactions in place of cursor interactions. However, unlike cursor interactions, touch interactions are difficult to record reliably and their coordinates have not been shown to relate to regions of user interest. A better approach may be to track the viewport coordinates instead, which the user must manipulate to view the content on a mobile device. These recorded viewport coordinates can potentially reveal what regions of the page interest users and to what degree. Using this information, search system can then improve the design of their pages or use this information in click models or learning to rank systems. In this position paper, we discuss some of the challenges faced in mining interaction data for new modes of interaction, and future research directions in this field.

INTRODUCTION
A recent survey puts mobile usage at 12.1% of all Web browsing in September 2012, near doubling from 6.7% a year ago [1]; and a talk by Google at a mobile convention in February 2011 notes "Roughly one in seven searches, even in the smaller categories, are happening on a mobile phone, ..."[15]. While search activity on desktops and laptops are still likely to be dominant for the near future, the growing portion of searches on mobile devices is becoming increasingly important. This evolution is an opportunity for Web search engines and other websites to begin using user interaction behavior on mobile devices for usability analysis and to inform their own design.

User interactions during Web search in a traditional cursor-enabled environment has been explored through tracking users' clicks [8] and more recently, through tracking users' mouse cursor movements. These movements include hovering, scrolling, opening links in new tabs, and higher-level behaviors such as following text with the cursor [6]. We can see a few of these interactions described in Table 1. The user interactions can be used to improve search [7] or understand user experiences resulting from Web page layout and ads [11]. But in touch-based interfaces such as tablets and smartphones, a cursor is not available to the user, and touch events have different meanings from cursor events. Additionally, viewing a Web page effectively on a small screen requires extensive panning and zooming, actions that can provide meaningful information.

	Interaction	Description
Cursor	Scrolling	Changes the page content being shown in the browser to a general region of interest. Often, this is vertical scrolling, which can mark what the user has or has not read.
	Highlight text	The selection of text on a page. This can identify terms or phrases of significance.
	Hover	Hovering is when the cursor idles over a region on the page. Hovering over search results may be interpreted as a signal that the user examined that result.
Touch	Pan	Changes of the visible content on a device. Panning towards a new area of the page shows user attention shifting from the previous region to the new one.
	Zoom	Magnifying or shrinking specific regions on a page, often performed by double tap or pinch gestures. Zoom can indicate degree of interest based on zoom level.

Table 1. Cursor and touch interactions that can be recorded by a website, and their potential usage for identifying content of interest to the user.

PRIOR WORK
Some online services and prior work relied on the assumption that eye-gaze is generally near where the cursor is pointing. For example, ClickTale, a leading Web analytics service remarks, "By aggregating the mouse movements of thousands of visitors on a webpage, we create a comprehensive, visual representation of what visitors

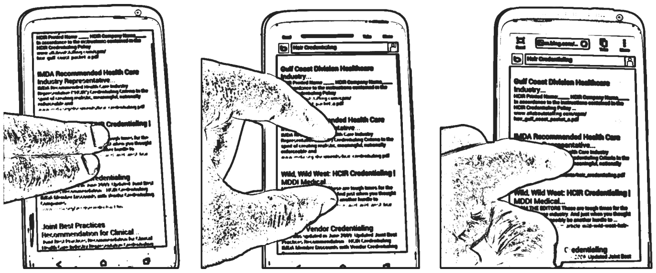

Figure 1. Touch interactions on a mobile device could potentially be recorded by the loaded Web site.

are looking at and focusing on within the page." Chen et al. performed a cursor-gaze correlation experiment and concluded, "..., a mouse saccade will move to a meaningful region and, in these cases, it is quite likely that the eye gaze is very close to the cursor." [5]. However, Huang et al. find that this notion of cursor-gaze alignment depends heavily on the intent behind the user's cursor behavior at a given moment [6]. As far as we are aware, there has been little work on the utility of user interactions in a touch-enabled environment—interactions which have the potential to affect Web search and other online websites. In this paper, we propose methods to best take advantage of these interactions, and note the challenges in pursuing this line of research.

Touch interactions have been used for a few applications. Leiva used touch interactions to make minor adjustments to the stylesheet of a Web page such as element and font sizes [10]. However, there has not yet been evidence that users benefited from these adjustments. Speicher goes beyond CSS metrics and tracks a greater portion of user interactions to adapt a Web page for mobile [14]. The users in Speicher's study found the adapted mobile Web interface to be better than the baseline page. Carta et al. record individual touch interaction events *touchstart*, *touchmove*, and *touchend* to incorporate into visual timelines [3]. Overall, these studies show some potential in using the touch interactions, but have not yet produced convincing practical applications.

TOUCH EVENTS IN PRACTICE

An immediate reaction may be to simply replace mouse cursor interactions with touch interactions for the same applications on the Web. Recording cursor coordinates becomes recording touch coordinates by changing *onmousemove* to *ontouchmove* [13]. The recorded data could include the centroid coordinates from the touches (Figure 1), along with timestamp. By doing so, the same applications follow—aggregating the coordinates result in heatmaps, and the individual interactions played back over time in an animation become replays. In fact, ClickTale has recently began testing a mobile version of their analytics service[1], explaining, "Businesses can now visualize whats working and what's not on their websites by seeing every swipe, pinch, tilt and tap."

[1] http://research.clicktale.com/ClickTale-Mobile-Beta.html

Challenges of Using Touch Events

Despite the potential in recording touch interaction events, using them in practice is problematic for a number of reasons.

The primary reason is that there is no evidence or rationale that the touched coordinates on the page relate to user interest or attention. While there is some justification for this in mouse cursor coordinates, touch-enabled devices do not have a cursor that users can possibly use as a marker to aid in reading text or mark interesting parts of the page. The "cursor" on a touch-enabled device is the user's finger and unlike the mouse cursor, it is not tracked when not clicking on the page. When users are performing gestures on the touchscreen, the area beneath a gesture such as pinching does not necessarily relate to the area the user wants to see more of; in fact, the user may be performing this gesture somewhere that does not obscure looking at the region of interest. The assumption that the touch coordinate is where the user's attention is focused is unfounded.

Additionally, current browsers do not reliably report touches. Speicher, who spent a considerable amount of time working with these events remarked, "When trying to recognise zooming gestures based on streams of touchdown, touchmove and touchup events using jQMultiTouch [a library specifically built for touch event tracking], we found out that the browsers used for testing showed different and partially unreliable behaviors in firing the corresponding touch events, ..." [14]. Although there is a W3C working draft on "Touch Events Specification", it is still up to the browser vendors to implement the specification. Additionally, the resolution in which the touch coordinates are recorded is variable; the W3C's specification offers, "the rate at which the user agent sends touchmove events is implementation-defined, and may depend on hardware capabilities and other implementation details."

Reproducing the resulting effect of a gesture is also difficult. When tracking a simple gesture such as a flick, browsers each record different points and many sample the gesture which makes it difficult to later reproduce the users' earlier action. Recreating different zoom levels from the pinch or double tap gestures is also difficult, since the zoomed level is device dependent and not all touch events trigger the JavaScript *ontouchmove* event.

FOCUSING ON THE VIEWPORT

Fortunately, the typically smaller screen size of touch-enabled devices, particularly for smartphones and smaller tablets, actually make one trackable feature more important. This feature is the shown area of the Web page at each moment, i.e., the viewport. Websites can record the change in this viewport, can therefore tell which parts of the page the user could examine, and for how long. On a large monitor, this information is not so useful since much of the page is visible on the screen without scrolling, commonly referred to as the region of the page "above the fold"; for example, on the author's

24-inch monitor, all 10 search results on Google and Bing show up in the browser window without needing to scroll. But on a smartphone for example, the user is limited by human visual ability even with high-resolution displays (Nebeling et al. find in a small survey that 0.4cm is the smallest font size that users can comfortably read [12]); hence, repositioning the viewport is a matter of necessity. This difference between desktop and mobile devices is further widened as people move to larger desktop monitors, while mobile screens remain small for portability. This makes tracking viewing area uninformative on desktop screens, but very useful on mobile devices, which are typically touch-enabled.

On small screens, users must move the viewing area left and right to different columns, and zoom in and out to alternate between a) seeing the overall layout of the whole page and b) enlarging the content of the page to be examined. Speicher [14] reports in a lab study that on smartphones, users zoomed into text by a factor of 2.73 on average in portrait mode and a factor of 1.95 on average in landscape mode. In tablets, users zoomed into text by a factor of 1.78 in portrait mode and a factor of 1.05 in landscape mode, so even on medium-sized screens, zooming can still be an important action. Even more strongly, the degree of zoom can be used as a measure of interest, since a lightly zoomed region of the page may still be hard to read, but by zooming in further, the user is disregarding the surrounding content in exchange for more visual detail in a small region on the Web page.

Figure 2. An illustrative example of a heatmap on a Web page generated from mobile device users' viewport data.

Applications of Viewport Data

A website that tracks the viewport can store the bounding boxes of these viewing areas and how long the user spent in each. These bounding boxes are easily recorded as they are relatively small in size (bytes), and aggregate well once collected. Generating heatmaps from the aggregated bounding boxes produces an easy visualization of which parts of the Web page the users focused on (Figure 2). When generating the heatmap, greater weight can be assigned to areas where there was a higher degree of zoom. Besides the information that users spent in a viewport, the act of moving the viewport away from an area can be useful information as well. A short dwell time may indicate that a user did not find an area on the page interesting after glancing at it, while a long dwell time indicates the user has read the contents in that region. It may also be useful to analyze which parts of the screen the user is often attending, since after zooming, the user is probably not equally likely to be looking at each part of the page.

The smaller viewports in touch-enabled devices are similar to an experiment by Lagun and Agichtein [9] who restrict the viewport of the user by blurring the page and unblurring only the search result hovered over by the mouse cursor. The experiment showed that having users intentionally shift their viewports enabled them to determine snippet attractiveness and help re-rank search results in a way that improved result ranking as measured by nDCG. These same applications can be developed when tracking the viewport on a device with smaller screen, substituting the unblurred portion of the page with the viewport on a small screen.

Finally, in information retrieval, where results are often presented as vertical lists, the viewport plays a key role. Search engines can determine which results are on the users' screens at any given time, and how much of the snippet they can see. Typically, Web search engines like Bing and Google can only present two or three search results at a time on a mobile device, and like ViewSer [9], touch-enabled smartphones can determine which results the user is looking at and how much time they spend examining the snippet. Using this information, websites can then use this information in traditional information retrieval models, such as searcher models [4] or learning to rank systems [2].

FUTURE RESEARCH DIRECTIONS

To leverage interaction data in touch-enabled devices, we propose exploring the following research agendas.

Categories of Touch Events

Touch events are a fairly recent and emerging mode of interaction data. This necessitates the need to carry out research that unpacks the different types of touch events available, and the contexts in which users use them. For example, a double tap to zoom may differ from a pinch to zoom gesture, despite similar end results. The double tap may be more likely to be aimed at the region of interest, while the pinch touches around that area. An investigation analyzing what event types are used in specific search contexts and information behaviors would enable us to move closer to understanding their value.

Utility of Touch Events

One of the first steps in leveraging touch events is understanding how events like zoom, scroll, swipe, pan, etc., correlate to factors like user attention, interest and relevance. Work by Huang [6] examined cursor events and hovers over different page regions on search engine results pages using activity logs from Bing; similar fundamental research needs to be performed for touch interaction. These interactions may be useful in determining environmental conditions such as the user's body posture or viewing angle, factors that may influence the type of content they may be interested in seeing.

Methods of Touch Interaction Data Collection

In understanding the usage context for touch events, work needs to be carried out to effectively collect, store, and mine touch event data that is able to reliably capture data from different browsers and display sizes. As discussed earlier, it is difficult to reproduce gestures because of the lack of standardized data polling and storage methods across different devices. Developments in this area would make it easier for search systems to leverage this data across a multitude of devices and browsers.

CONCLUSION

As a portion of online user activity such as Web search moves to touch-enabled mobile devices, online services will begin thinking about recording these user interactions as they do in traditional cursor-based systems such as desktops. While the initial reaction may be to simply replace cursor coordinates with touch coordinates, this is impractical because touch coordinates do not hold the same meaning as cursor coordinates. Touch events typically are region-free methods for navigating the page, and there is no rationale behind the assumption that users are attending to the specific touch coordinates. Additionally, technical difficulties in recording fine-grained touch coordinates prevent applicability of these events.

However, tracking the viewport coordinates can be tremendously useful in noticing where on the page a user is attending to, especially on small screens, where zooming and panning is a necessity. We believe this is the data that should be recorded and analyzed, and can potentially be used to improve the design of websites and search engines. To leverage this data, further work needs to be carried out for: 1) exploring the different categories of touch events; 2) understanding their empirical utility; and 3) developing browser- and device-independent methods to poll and store this data.

REFERENCES

1. Statcounter global stats: Mobile vs. desktop from sep 2011 to sep 2012. Retrieved September 21, 2012 from `http://gs.statcounter.com/#mobile_vs_desktop-ww-monthly-201109-201209`.
2. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of ICML*, pages 89–96, 2005.
3. T. Carta, F. Paternò, and V. Santana. Support for remote usability evaluation of web mobile applications. In *Proceedings of SIGDOC*, pages 129–136, 2011.
4. O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of WWW*, pages 1–10, 2009.
5. M. C. Chen, J. R. Anderson, and M. H. Sohn. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *CHI Extended Abstracts*, pages 281–282, 2001.
6. J. Huang, R. White, and G. Buscher. User see, user point: gaze and cursor alignment in web search. In *Proceedings of CHI*, pages 1341–1350, 2012.
7. J. Huang, R. W. White, and S. Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of CHI*, pages 1225–1234, 2011.
8. T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of KDD*, pages 133–142, 2002.
9. D. Lagun and E. Agichtein. Viewser: Enabling large-scale remote user studies of web search examination and interaction. In *Proceedings of SIGIR*, pages 365–374, 2011.
10. L. A. Leiva. Restyling website design via touch-based interactions. In *Proceedings of MobileHCI*, pages 599–604, 2011.
11. V. Navalpakkam and E. Churchill. Mouse tracking: measuring and predicting users' experience of web-based content. In *Proceedings of CHI*, pages 2963–2972, 2012.
12. M. Nebeling, F. Matulic, and M. C. Norrie. Metrics for the evaluation of news site content layout in large-screen contexts. In *Proceedings of CHI*, pages 1511–1520, 2011.
13. M. Nebeling and M. Norrie. jqmultitouch: lightweight toolkit and development framework for multi-touch/multi-device web interfaces. In *Proceedings of EICS*, pages 61–70, 2012.
14. M. Speicher. W3touch: Crowdsourced evaluation and adaptation of web interfaces for touch. Master's thesis, ETH Zürich, 2012.
15. J. Spero. The time for mobile is now. Presented at *thinkmobile with Google*, 2011.