Crowdsourced Web Site Evaluation with CrowdStudy

Michael Nebeling[1], Maximilian Speicher,
Michael Grossniklaus[2], and Moira C. Norrie[1]

[1] Institute of Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
[2] Computer Science Department, Portland State University
Portland, OR 97201, United States of America
{nebeling,norrie}@inf.ethz.ch, speichem@student.ethz.ch,
grossniklaus@cs.pdx.edu

Abstract. Many different automatic usability evaluation tools have been specifically developed for web sites and web-based services, but they usually cannot replace user testing. At the same time, traditional usability evaluation methods can be both expensive and time consuming. We will demonstrate CrowdStudy, a toolkit for crowdsourced testing of web interfaces that allows, not only to efficiently recruit larger amounts of test users, but also to evaluate web sites under many different conditions.

Keywords: Web usability, web site evaluation, crowdsourcing.

1 Introduction

Usability evaluation is an important topic in user interface design practice and research. Many different methods have been specifically developed for web sites [1], while the most prominent example is still user testing [2]. Existing techniques for remote usability evaluation use a range of different user activity tracking methods [3,4], but are frequently restricted to either server or client-side logging, which limits the kinds of information they can collect. Even worse is the fact that none of them specifically take the context of use into account. In contrast to previous approaches, our solution for web site evaluation integrates a simple notion of context-awareness, which is necessary to cater for the increased diversity of devices nowadays used for web browsing. Another shortcoming of existing usability evaluation tools is that essential tasks, such as subject recruitment including qualification tests, and many key aspects of remote usability testing, such as task distribution within or between subjects, are typically not even considered. Our solution is designed to integrate with existing crowdsourcing services such as Amazon Mechanical Turk[1], where requesters may, not only recruit a larger group of users in a short amount of time, but also control the

---

[1] http://mturk.com

conditions for each worker to be able to participate. This generally makes it a viable platform for conducting different kinds of online user studies [5], but has so far not been considered in existing usability evaluation tools.

In this paper, we present CrowdStudy, a web site evaluation toolkit that can be easily integrated with existing sites for remote usability testing and can also leverage services such as Mechanical Turk to advertise and facilitate online studies with the help of new users. CrowdStudy evolved out of our previous work on crowdsourcing web site adaptations for different devices [6], where it was first built to enable user evaluations for a specific set of tasks. However, we have now built several mechanisms into CrowdStudy that allow it to be configured with different metrics as well as different tasks to support a range of evaluation scenarios. As one example, we will show how CrowdStudy was used for evaluations on mobile touch phones, which is a new important scenario that poses a number of interesting challenges.

We start by presenting CrowdStudy in Section 2. This is followed by a description of our demo in Section 3 and concluding remarks.

2 CrowdStudy

CrowdStudy is designed to integrate with the typical client/server infrastructure of web applications. It introduces additional components for user activity tracking on the client-side and a context engine as well as data logging and crowdsourcing components on the server-side. Figure 1 shows the CrowdStudy architecture. Below we describe how each component contributes to supporting remote usability evaluations.

First note that clients can either be "normal" web site users or workers specifically recruited using crowdsourcing services. CrowdStudy can be configured with
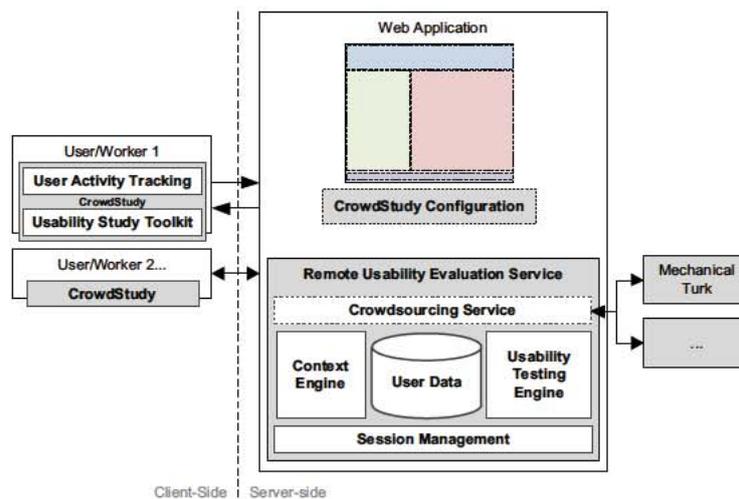
Fig. 1. CrowdStudy's architecture

multiple sets of tasks which are then distributed to users. Tasks may require that other tasks are completed first as well as user qualification in terms of the user's knowledge or skills based on a questionnaire or the client device characteristics based on programmatic checks. Each task may involve one or several web site components as well as navigation between pages. Using client-side scripting, some aspects of a task may be automated to allow users to focus on specific activities. For example, we provide tools that can automatically scroll to a component of interest and annotate and highlight certain parts that require interaction and attention of users. CrowdStudy then automatically tracks the user actions in terms of user interface events fired by the interaction components and also captures the context in terms of the screen size, window size and position in the page. The collected data is first buffered and cached locally before it is sent to the server-side at suitable intervals.

Looking at the server-side extensions, the first is a configuration for Crowd-Study and the second the CrowdStudy service which contains the usability testing engine responsible for assigning and distributing tasks to participants. We have built simple administrative tools for managing and running studies as well as evaluation tools for browsing and analysing the results generated by Crowd-Study. The analysis tools of CrowdStudy range from a simple response viewer that displays questionnaire data submitted by participants to a more advanced log analyser that allows evaluators to browse and visualise the data based on a combination of criteria. For example, CrowdStudy can visualise the touch data aggregated from several smartphone users to see how they interacted with the web site, e.g. where they have mis-clicked links and how much they had to zoom to counteract. CrowdStudy automatically takes care of the necessary user session management and stores all collected information in a database. Finally, as part of the crowdsourcing service component, we have implemented an interface to Mechanical Turk that can be used for recruiting external, paid crowd workers using Amazon's service. Other services could be integrated in the same way.

While most existing frameworks are proxy-based, our solution only requires that CrowdStudy is embedded in the web site using a single line of code similar to including JavaScript libraries like jQuery. User tasks may be based on simple textual descriptions that are displayed in an instruction box. Moreover, CrowdStudy supports custom tasks based on client-side scripting, but also provides a set of pre-defined task classes, e.g. for clicking certain links, navigating to parts of the page or reading text paragraphs and answering questions. In this case, the test designer only needs to mark corresponding web site elements with additional CSS classes that CrowdStudy will then automatically interpret and compile into a test scenario.

3 Demonstration

Our demonstration of CrowdStudy is based on a study we conducted for the Wikipedia web site. CrowdStudy was embedded in an example article and configured with the tasks shown in Figure 2 to assess the usability for smartphones
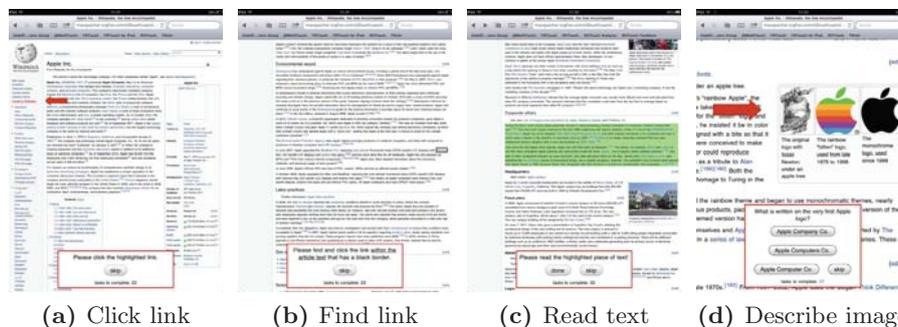
(a) Click link   **(b)** Find link   **(c)** Read text   **(d)** Describe image

Fig. 2. Set of tasks used in the evaluation of Wikipedia based on CrowdStudy

and tablets. Wikipedia was chosen for the test scenario since it provides both a desktop and mobile version with different layouts and features. We performed simple A/B testing by letting CrowdStudy assign one of the layouts to each participant and collected the data from 84 participants using a wide range of mobile devices and also different browsers.

CrowdStudy provided valuable insight into how users interact with such a text-heavy web page in terms of the optimal font size for reading, preferred device orientation with respect to different tasks (landscape was generally preferred on tablets and often used for reading on smartphones, where the other tasks were preferably done in portrait mode) as well as how they made use of multi-touch gestures for zooming and navigating within the page.

References

1. Matera, M., Rizzo, F., Carughi, G.: Web Usability: Principles and Evaluation Methods. Web Engineering (2006)
2. Insfran, E., Fernandez, A.: A Systematic Review of Usability Evaluation in Web Development. In: Hartmann, S., Zhou, X., Kirchberg, M. (eds.) WISE 2008. LNCS, vol. 5176, pp. 81–91. Springer, Heidelberg (2008)
3. Atterer, R., Wnuk, M., Schmidt, A.: Knowing the User's Every Move – User Activity Tracking for Website Usability Evaluation and Implicit Interaction. In: Proc. WWW (2006)
4. Hong, J.I., Heer, J., Waterson, S., Landay, J.A.: WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. TOIS 19(3) (2001)
5. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing User Studies With Mechanical Turk. In: Proc. CHI (2008)
6. Nebeling, M., Norrie, M.C.: Tools and Architectural Support for Crowdsourced Adaptation of Web Interfaces. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) ICWE 2011. LNCS, vol. 6757, pp. 243–257. Springer, Heidelberg (2011)