Automatic Capture, Representation, and Analysis of User Behavior

Sharon J. Laskowski
National Institute of Standards
and Technology (NIST)
Gaithersburg, MD USA
sharon.laskowski@nist.gov

James A. Landay
EECS Department
University of California
Berkeley, CA USA
landay@cs.berkeley.edu

Mike Lister
Netusability Limited
London, UK
mlister@netusability.com

Abstract
The goal of this workshop is to explore the implications of automated capture and analysis of user behaviors on HCI and UE research.

Keywords
Automation, evaluation, log analysis, validation, remote testing, usability

INTRODUCTION
With the advent of the Web and the refinement of instrumentation and monitoring tools, software user interactions are being captured on a much larger scale than ever before. Automated support for the capture, representation, and empirical analysis of user behavior is leading to new ways to evaluate usability and validate theories of human-computer interaction. It enables remote testing, allows testing with larger numbers of subjects, and motivates the development of tools for in-depth analysis. The data capture can take place in a formal experimental setting or on a deployed system.

The main questions are: can we leverage these capabilities to validate or change our models, to improve the user experience, and to change the user interfaces in products in measurably better ways? How will human-computer interaction (HCI) and usability engineering (UE) as bodies of knowledge and practice change? How *has* HCI/UE research and practice changed as new analysis, design, and evaluation methods have emerged and been adopted?

Specifically, a number of different approaches based on these methods have appeared in the research literature [3,7] and in commercial tools [1,5,6,9]. However, these have led to a number of unresolved issues under discussion in both the HCI and UE communities, such as how and when to apply these methods, when is remote, automated testing useful, and what can server logs provide.

GOALS
The goals of this workshop are to encourage discussion of these issues by the HCI and UE research communities, and, as a result, provide a foundation for a clearer understanding

and more systematic application of these methodologies. This will be accomplished by:

- Outlining the range of automated and semi-automated techniques for capturing and processing user behavior,
- Exploring key challenges to researchers in applying these techniques,
- Brainstorming approaches for evaluating and comparing methodologies, and
- Encouraging participants to contribute papers for a special issue of a journal or book summarizing the findings.

The survey by Hilbert and Redmiles [3] can be used as a starting point for identifying the range of techniques. This paper describes the range of data that can be collected and how the data can be applied to a set of usability indicators. However, one of their conclusions is "…that more work is needed in the area of transformation and data collection to ensure that useful information can be captured in the first place, before automated analysis techniques…can be expected to yield meaningful results." We expand on this observation in the next section.

CONTROVERSIAL/UNRESOLVED ISSUES
Here is a list of some of the issues that we have identified that need to be resolved to have confidence in automated and semi-automated techniques. This is not intended as a comprehensive list but to illustrate the kinds of issues we would like to see discussed at the workshop:

- Depth vs. breadth: semi-automated user testing usually implies analyzing large datasets. But, perhaps the need for automation is overstated and skilled testers should do user testing individually, at least in the case of evaluating a software application. However, for validating a theory automation can be valuable. See, for example, how the collection of eye tracker data and visualization to analyze the data are helpful in validating the CTVA-foraging theory of how a user interacts with focus+context information visualizations in [7].

- In general, can the behavioral data be used to support or disprove theories of behavior, such as ACT-R, Soar, EPIC, Activity Theory, Foraging Theory, Behaviorism, etc.?

- What is lost as compared to traditional usability testing methods? Is an onsite observer essential for a semantically deep description of the experience? When is an observer not essential? This is an especially important issue for testing web usability remotely. With Web-Quilt [4] and Enviz [6] it is assumed that some useful data will be collected without an observer, while Netusability's technology [5] and NetRaker's [1] technology captures some of the user reaction with a camera or chat window as part of the remote data capture. Server logs are used by some; they have severe limitations, but when are they useful?

- How many test participants are required? Recent results such as those in Spool [8] suggest that many more users than the traditionally accepted 5-8 are required under some circumstances when evaluating web sites. On the other hand, what do you do when you have a huge amount of data on user interaction? Can you mine it effectively?

- Is it feasible to develop one or more standard representations to allow data exchange and development of generic tools? For example, a standard format for user logs, such as that described in [2], will enhance interoperability among analysis tools. But, can we find a single format to cover a wide range of user testing methodologies? What can be done to support mapping of low-level system and user events into higher-level descriptions? What about capture of the part of the system's behavior that is apparent to the user, e.g., windows opening and closing, and the status of checkboxes? What context information needs to be captured?

- Can you effectively automate the tracking and managing of the web customer experience as in the approach described in [9]? This approach connects user satisfaction to click stream and page view user behavior for large sample populations. Can you infer the quality of the customer experience and how do you use this data for improving a web site? How does it compare to more traditional methods?

- Are there differences in data capture and analysis depending on whether data is supporting HCI research or usability evaluation?

- Which of the innumerable aspects of the user behavior should be captured? Mouse clicks? Eye gaze? Verbal self-reports and think-alouds? What can be done to automate the capture of this data?

- There are a number of technical issues relating to web-based applications. For example, can we abstract away from browser-specific event models? Is the Document Object Model (DOM) the answer? What is the best tool architecture for data capture? Server-side instrumentation, customized browsers on client-side, or proxies in between the two?

- What analysis and visualization tools are useful to researchers and usability engineers?

- Is there a methodology for benchmarking approaches so that there is some assurance that the automated tools are indeed measuring usability?

- What about privacy concerns? Should users' behavior ever be monitored without their explicit consent?

Disclaimer

NIST does not recommend or endorse commercial tools.

REFERENCES

[1] Chak, A., Usability Tools: A Useful Start, Web Techniques, August, 2000. Also available at: http://www.webtechniques.com/archives/2000/08/stratrevu/

[2] Cugini, J. and Laskowski, S., Design of a File Format for Logging Website Interaction, NIST SP 500-248, April, 2001, Also available at: http://www.itl.nist.gov/iad/vug/cugini/webmet/flud/design-paper.html

[3] Hilbert, D.M. and Redmiles, D.F., Extracting Usability Information from User Interface Events, in ACM Comput. Surv. 32, 4 (Dec. 2000), Pages 384 – 421. Also available at http://www.fxpal.com/people/hilbert/papers.html

[4] Hong, J.I., Heer, J., Waterson S., and Landay, J.A., WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. To appear in ACM Transactions on Information Systems. Also available at http://guir.cs.berkeley.edu/pubs/#webquilt

[5] Lister, M., Usability Testing Software for the Internet, in Extended Abstracts of CHI '01 (Seattle, WA, April 2001), ACM Press, 17-18.

[6] Moore, P., Conducting Experimental Research using Native Site Visitors, in IBM Make It Easy 2000 Conference at http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/1822

[7] Pirolli, P., Card, S.K., and Van Der Wege, M.M., Visual Information Foraging in a Focus + Context Visualization, in Proceedings of CHI '01 (Seattle, WA, April 2001), ACM Press, 506-513.

[8] Spool, J., Testing Web Sites: Five Users is Nowhere Near Enough, in Extended Abstracts of CHI '01 (Seattle, WA, April 2001), ACM Press, 285-286.

[9] Vividence white paper, The Vividence Approach and Methodology. Available at http://www.vividence.com/public/Research/methodology.htm