

Automating the Evaluation of Web Site User Experience: Collecting Both Qualitative and Quantitative User Experience Data

Introduction

Web site user experience is a key component to any e-business performance management strategy. The challenge is applying an empirical approach that assesses the user experience as part of the effectiveness of any Web site, including e-business sites. Understanding user response cannot be understood merely by measuring response time and speeds and feeds, but requires the collection and analysis of data that truly represents user experience. In addition, you need a standardized, repeatable methodology to track user interaction with the Web site to determine where users go, what they look for, what they respond to, and generally what works and what doesn't.

This document will profile a new methodology that balances quantity of data with quality of data and provides a means to gather empirical usability and user experience data. Using this methodology will assist you in refining your Web site and increasing overall performance from a customer's perspective.

Balancing Quantity and Quality of Usability Data

When assessing Web user experience, two major questions arise: what group of users will be studied, and what kind of data will be collected? The answers to these questions are interdependent, as well as dependent on the resources available to the researcher. This interdependence usually results in a trade-off between the quantity of data collected and the quality of that data.

New technologies are emerging to assist researchers with data gathering and to help them balance data quantity versus data quality. Before considering some of those technologies and discussing methodologies, you have to make some initial decisions and set the ground rules to gather useful usability data.

The first step is to consider the required data properties, which will shape the criteria for the sample group. Generally, the users selected for a Web usability study should meet several criteria:

- **They should be representative of the population of interest.**

In order to generalize from the group you study to all site visitors, the sample group must be representative of the total population in terms of

demographics, "technographics" (which defines how users acquire and use technology), intentions, and experience. Selecting samples based on convenience or by paying panelists typically introduces bias not found in the population at large, which, if unaccounted for statistically, can negate the causal conclusions one might draw.

- **There should be no selection bias.** The method of selecting a sample from the population should not be related to success or the attainment of some outcome. For example, questioning or observing people after they complete a purchase creates selection bias because it ignores people who drop out of the process, which may be the group in most need of being studied.

- **There needs to be a large enough sampling to deliver meaningful conclusions.** To understand how a target population uses a Web site, you need a sufficiently large sample to draw statistically valid conclusions. The reactions of a small group will seem disproportionately important in the final analysis, where with a larger sample, the patterns observed have a statistical relationship to the patterns in the overall population.

When assessing Web user experience, two major questions arise: what group of users will be studied, and what kind of data will be collected?

Although the specific size of the sample can vary, you need to have a sample large enough to be representative of the overall target population, which gives you more support when prioritizing the importance of issues and determining how to make the right usability changes to improve performance.

- **There should be no measurement effects.** Observing participants doing an assigned task in an unnatural lab setting while being asked leading questions can create unnatural behaviors from which a researcher cannot draw valid conclusions. Data collection should be as unobtrusive and real-world as possible to avoid biased data.

WebEffective™ was developed to automate usability data collection and to strike the right balance between quantity of data collected and the quality of that data.

Along with the question of who to study is the question of what data to collect. Participants browsing a Web site can generate an immense amount of information. The data collected therefore must be filtered, but also provide sufficient detail to support the ensuing

analysis. Generally, usability data can be classified into three categories:

- **What did users see?** The content provided on the Web or in an e-business transaction is the raw material of user experience. Researchers must know what users saw so they understand the context of users' reactions. This is not trivial because Web sites are dynamic: promotional content changes over time, new functionality is added, pages are customized based on user preferences, ad banners rotate, and a single URL can render many different pages depending on visitor history and identity.

- **What did users do?** The complexity of Web sites allows users to take many paths to their desired destination. And, presumably, Web site designers have created paths they would like users to follow. Data collection must capture these paths in as much detail as possible so researchers understand how people transition between pages and how pieces of a site are experienced together.

The unit of analysis should be the user, not the Web page; understanding usability is about how people experience a Web site, not how a site experiences people. When user paths are aggregated instead of individual page hits, researchers can gain a better understanding of how to apply the user experience to their traffic reporting.

- **What did users think?** User attitude is continually shaped throughout their interaction with a Web site. You want to collect as much attitude data as you can for various reasons, including: providing interpretation of ambiguous behavior, such as long dwell time that can indicate interest or confusion; monitoring general attitude trends over time that result from marketing efforts and repeated exposure to the site; gathering user feedback and suggestions that can trigger direct action or further research. The process of capturing such data might interrupt the user experience and alter the user's attitude, which could corrupt the data and reduce its value. So, the method used to collect attitude data must be non-invasive and be incorporated into the user's experience at the site.

A New Approach to Gathering User Experience Data

Keeping in mind the constraints about the group of users you plan to study and the kind of data you need to collect, the next challenge is selecting the right methodology.

The desire for large, unbiased, representative samples suggests using automated methods such as log file analysis. However, the need for rich, contextually sensitive session data suggests usability lab testing. Thus, there is a dilemma of quantity versus quality: log files generate a larger quantity of data, whereas usability labs generate much richer data. Furthermore, each of these methods can produce serious flaws in its area of strength when used inappropriately or inconsistently by different researchers. The solution is in the middle where the two ends of the spectrum meet.

New data collection methods can help usability researchers capture large amounts of user experience data while maintaining qualitative richness. These solutions combine the best of both approaches with marginal sacrifices. The result is a more robust and standardized process to conduct consistent, reliable, actionable usability research.

Keynote WebEffective™ was developed to automate usability data collection and to strike the right balance between quantity of data collected and the quality of that data. As with other Keynote performance management solutions, WebEffective is deployed as a service using a network of servers positioned between actual site visitors and Web site servers. for analysis. By redirecting visitors to your Web site through Keynote servers, WebEffective can maintain its position unobtrusively monitoring and recording user interaction. WebEffective is able to see and capture all HTML content downloaded to the users' computer as well as all the upstream data requests sent to the Web server. WebEffective then reconstructs the data streams into actual visitor sessions

Balancing Data Quantity and Analysis Quality

A WebEffective user experience test contains a collection of sessions, and each session contains a collection of pages. Sessions can be analyzed individually at a very high level of detail, including VCR-style replay of the exact pages each visitor saw. Alternatively, session data can be aggregated and analyzed for patterns using analytic tools such as path models and page exit summaries. All analyses can be done after the data has been collected because Keynote preserves the data streams that constitute the original sessions. Thus, reconstructing and exploring usability data in depth is always possible when new insights occur. WebEffective's unique process of HTML redirection addresses critical issues in participant sampling. For example, a typical usability test implementation automatically redirects a group of visitors who come to a Web site home page through Keynote. The sample is selected randomly using a simple script with a random number generator.

Because the sample is drawn randomly from the overall Web visitor population, it is a perfectly representative sample of the total target audience. Making this assumption allows researchers to generalize from the sample to apply the results to real users.

Individuals within any sample will have differences that make them unique, and to the extent they are unique, they do not represent the group as whole. Hence, it is imperative to observe a sample that is sufficiently large that individual variations become statistical noise relative to the underlying pattern. As an extreme example, imagine doing a before-and-after test with just two participants, one in each "group." All of the variation in their behavior or stated opinions could be attributed to personal differences instead of changes to the Web site.

WebEffective is able to see and capture all HTML content downloaded to the users' computer as well as all the upstream data requests sent to the Web server.

Measurement effects also can be removed completely from the data collection process using WebEffective. Participants use their browsers as they normally do without any modifications to their computer system or browser. This ensures that their intentions at

the Web site and behavior are natural, not the result of special guidance or observation.

Depending on the degree of measurement precision required, several hundred observations may be necessary to make statistically valid comparisons between groups, with past observations, or with set standards and goals. Numbers of this magnitude are far beyond the time and financial resources available to most conventional usability labs. But, the automated session recording enabled by WebEffective gives researchers the hundreds of sessions needed to empirically estimate the importance of various Web design issues.

When collecting hundreds or thousands of sessions in a sample, the method of collection is crucial to the final quality of the data gathered for analysis. Partial tracking of where users go can be done by inserting beacons at various points in a Web site, but recording the actual session content with WebEffective explains how visitors transitioned between points and what they experienced in between. User experience researchers are also interested in what visitors think.

Getting this information requires interaction with the participant to ask detailed questions. Where this is typically a downfall of automated systems, WebEffective capitalizes on the user data recorded to trigger behavior-based questions. A variety of triggers can be specified such as page content, elapsed time of session, or number of pages into session. The questions that pop up are controlled by the researcher and do not require any HTML page creation. When site visitors answer these questions, their answers are recorded along with their full session information. During analysis, the researcher can see the distribution of responses and Investigate people who answered a particular way in more depth.

Conclusion

Automating user experience testing of any Web site is a difficult challenge at best, especially in light of the need to balance data quality and quantity: gathering rich data is essential to deriving meaning and understanding, and a sufficient quantity of data is essential to making findings valid and statistically significant. Technological innovation has begun to eliminate the need to sacrifice either of these important data characteristics. With its user experience assessment service WebEffective, Keynote has made significant strides in automating the collection of data that researchers need and in giving researchers the right tools to evaluate large quantities of exceptionally rich data. Such new data gathering techniques are ushering in a new approach to understanding user interaction with Web sites and the effectiveness of e-business systems.



1-800-KEYNOTE
(1-800-539-6683)
www.keynote.com