A Tool for Supporting Web-Based Empirical Research: Providing a Basis for Web Design Guidelines

Jen Barrick laeren@u.washington.edu
Brandon Maust bmaust@u.washington.edu
Jan H. Spyridakis jansp@u.washington.edu
Matt Eliot mjeliot@u.washington.edu
Carolyn Wei cwei@u.washington.edu
Mary B. Evans evansmb@u.washington.edu
Kate Mobrand kmobrand@mobrand.com

University of Washington
Box 352195
Seattle, WA 98195-2195

Abstract

Current Web design guidelines are often based on little more than intuition or anecdotal evidence. When research is cited to support Web design guidelines, that research has frequently been conducted as part of a small usability study or a print-media study. Whether such sources can be validly generalized into standard Web design guidelines is questionable but the practice is widespread. Professional technical communicators may be interested in conducting experiments to determine what Web design elements most benefit their audiences, paying attention to how those users interact with their own computers at a place and time of their own choosing. Conducting such research, however, requires a level of technical expertise usually reserved for programmers. To address this problem, our research team has been developing a tool to aid non-programmers in conducting Web-based experiments of Web design features. The ultimate goal of such experiments would be the development of research-based Web design guidelines.

Keywords: website design, empirical research, research tools

1. Introduction

In the last ten years, the Internet has been transformed from a special-interest communication system to a mass communication medium with nearly 740 million users worldwide [1]. Along with the proliferation of Web sites has come the proliferation of Web design guidelines. However, these guidelines are rarely based on empirical research conducted with actual Web sites in naturalistic settings and with users functioning in their own environments. In fact, much of our current understanding of effective Web design comes from anecdotal evidence, usability testing, and automated evaluation studies, rather than from empirical research. Anecdotal evidence and observational studies may be strongly influenced by subjective bias. Usability testing relies on small numbers of participants in laboratory settings and is more strongly focused on immediate product development than hypothesis testing [2]. Automated studies are often better termed “automated evaluators” in that they evaluate existing Web sites or pages according to specified usability criteria rather than capturing live data relating to actual usage patterns [3].

To close the gap between current Web design guidelines and the research conducted to support them, our research team has been developing a tool to support the conduct of Web-based empirical studies of the effectiveness of various Web design features in naturalistic settings. To contextualize the need for this tool, the foundations of current Web design guidelines are briefly reviewed and some barriers to Internet research are discussed. Finally, our tool is described, centering on the benefits it provides to technical communicators interested in conducting Web-based experiments, to strengthen the research basis for Web design guidelines.
2. Background

The professional technical communicator is often at a loss when seeking Web design guidelines based on empirical, Web-based research. Many of the guidelines currently in vogue are the result of anecdotal evidence drawn from specific projects, making these guidelines incomplete or outright misleading [4]. Many, while claiming to cite empirical evidence, simply cite other guidelines, which in turn cite other guidelines. Thus the connections between Web design guidelines and the supporting research are often fragile.

A good attempt to connect Web design guidelines to empirical research is Research-based Web Design and Usability Guidelines [5]. The authors assembled 427 published sources in support of 187 extant Web design guidelines. These guidelines were aggregated into 17 categories with titles such as “Optimizing the User Experience,” “Headings, Titles, and Labels,” and “The Homepage.” Each guideline was rated according to the relative strength of the supporting evidence using a rating scheme consisting of five categories ranging from “Strong Research Support” to “Weak Expert Opinion Support.” Only 15% of the guidelines received the highest rating, “Strong Research Support,” a rating defined as equal to “cumulative and compelling, supporting research-based evidence...at least one formal, rigorous study with contextual validity...no known conflicting research-based findings...[and] expert opinions that agree with the research” (xxi).

Although 28 Web design guidelines with “strong research support” may seem like a sufficient basis for design, many of the cited studies refer to print-based research conducted long before the existence of computers or the Web. The extrapolation of print-based research findings to online media is understandable, but problematic, and has been criticized in the literature (e.g., [6], [4]). Further, many of the studies that are conducted online have been conducted as small usability studies with few participants (e.g., [7], [8]). Such approaches limit generalizability.

Design guidelines must be investigated via actual Web-based research if the results are to be applied to the Web. Conducting this type of research in a naturalistic setting through remote Web-based studies, rather than in usability labs, is vital because Internet usage is highly idiosyncratic [9], [10]. The artificial nature of usability settings may lead to data that do not accurately describe real-world Web behavior (for more information on Internet research, see [11], [12], [13]). And it is the idiosyncrasies and variety of users interacting with Web sites in the real world that must be incorporated into studies assessing Web design features.

3. Empirically testing Web design guidelines

Recognizing the limitations of current Web design guidelines, the professional technical communicator may wish to conduct Web-based studies to assess a variety of Web design elements. However, two major obstacles can hinder willingness and ability to carry out such studies: the development of multiple versions of a Web site, and recording, accessing, and interpreting study data.

3.1. Development of the experimental Web site

A useful approach to Web-based research of design principles is to conduct a comparative study of two or more versions of a Web site, varying a specific design element (such as link wording or heading frequency). However, creating the different versions of the Web site can be tedious and prone to inconsistencies and other errors, particularly for investigators without a great amount of technical expertise.

The development process can be greatly streamlined when a scripting language is used to vary the design element of interest, rather than creating multiple Web sites. By embedding code to manipulate design variables into a single copy of a Web site, version inconsistencies are eliminated and design changes are simplified. This approach requires, however, the assistance of a programmer.

3.2. Data access and interpretation

A commonly used approach for understanding the effects of Web site manipulation is to collect and analyze log file data. Web server log files support the conduct of a naturalistic study, reflecting participants’ habitual use patterns. However, server log file access may be limited—not all Web servers log data and not all server administrators are accessible or open to aiding researchers. Not everyone is fortunate enough to have full access to their Web server or, if they do have that access, to have the technical expertise necessary to ensure that data are logged properly.

Raw server log files, moreover, can be difficult to analyze. For example, standard log files track people by IP address, yet some Internet service providers assign IP addresses dynamically or mask several people behind what appears to be a single IP address, making differentiation of participant data difficult. Furthermore, server log files contain basic Web usage data, but often do not collect the particular data needed for a study, much less format that data in a manner accessible to those lacking great technical expertise. While some modifications can be made to server configurations, allowing for more useful log files, administrator access to
the server and advanced technical knowledge are both required. Regardless, even if these modifications are made, there is ultimately a limit to the type of data that can be collected using Web server logs (for more information on log file analysis, see [14], [15], [16]). To examine user behavior in meaningful ways, we need an expanded set of tools that can collect more granular data than what is available in standard server logs.

4. A tool to support Web-based empirical research

The obstacles described here have proved significant in the Web-based studies that our research team has conducted over the past few years (e.g., [17], [18], [19], [20], [21]). Our experiences have helped us focus on alleviating these issues with a tool that supports the conduct of Web-based studies. A recent study [17] serves as a good example to discuss the capabilities and functions of the tool. In this study, we investigated the effects of link wording and placement on browsing behavior and comprehension. We specifically sought to test the effects of three link wording conditions (generic, informative, and intriguing), in navigation bar and embedded links, on a variety of measures including user comprehension, perceptions, and behavior. The approach resulted in five experimental conditions (navigation bar-embedded links): (1) generic-generic, (2) generic-intriguing, (3) generic-informative, (4) intriguing-intriguing, and (5) informative-informative. We are still analyzing the data from this study, but of interest here is the tool we have been developing to support our work.

Although there were two programmers on our research team, we wanted a tool that could easily handle necessary design changes and that facilitated data access and analysis for all team members. We also wanted a tool that could be easily adapted to run future studies without direct programmer support.

Specifically, we sought to develop a tool that would:

- Support study instrumentation.
- Manipulate and deliver multiple versions of a Web site.
- Assign participants to study conditions.
- Gather Web site usage data and online questionnaire responses.
- Bypass raw server log file analysis.

4.1. Supporting study instrumentation

Instrumentation of Web-based research, particularly survey design and delivery when a variety of experimental conditions and datasets are under investigation, has proved difficult for both our team and the Internet research community at large [21], [17], [11]. Rather than relying on a cumbersome set of static HTML survey files, or a tool that could not apply changes throughout a survey, we opted to develop a method to dynamically generate online surveys from a set of source text files in a simple, defined format. If changes needed to be made to the survey, such as rewording an answer choice, moving questions around, or adding a new page of questions, any member of the team could easily edit the survey without needing to know how to edit an actual HTML survey form or having to fight with a recalcitrant survey generator tool.

4.2. Manipulating and delivering multiple versions of a Web site

We created a single experimental Web site that dynamically produced the five study conditions. PHP, a scripting language designed for use in Web programming, allowed us to generate all experimental conditions for the study from a single code source. For our specific interest in link wording, we inserted variables into the HTML code for the embedded links and navigation bar links. When a participant was assigned to a study condition, such as generic navigation bar links and informative embedded links, the dynamic page would automatically deliver the link wording that was appropriate for the participant’s assigned condition.

Although we were interested in links, PHP variables could have also handled other aspects of the page such as the background color, the displayed images, or any other design features that can be controlled by text in an HTML file. For example, in an earlier iteration of this study, we noticed that participants seemed to visit pages in the order these pages were listed in the navigation bar. We subsequently used PHP to randomly vary the order of the navigation bar sections, and the link order within these sections, for each participant.

4.3. Assigning participants to study conditions

We included a method that assigned incoming participants to study conditions sequentially or randomly depending on the sample size available for an experiment. Because participants come to the study Web site in an essentially random, self-selected order, both sequential and random methods should result in an equal and random assignment of participants across experimental conditions. The assignment process was transparent to participants.
4.4. Gathering Web site usage data and online questionnaire responses

Our method made it possible to collect specific, customized information about the participants’ behavior on the experimental Web site in addition to questionnaire responses. The measures we monitored were associated with individual users through the use of random ID codes entirely separate from their IP addresses. Some measures we collected about each participant included:

- **The specific pages visited.** Our tracking method allowed us to document an individual participant’s navigation path through the Web site, information that is extremely difficult to discern when using standard Web server log files and often impossible with large samples.

- **Page load times.** We were able to extrapolate the time spent on each page by looking at the load time of the next page.

- **The page each participant came from.** We were able to check whether the page that the participant’s request apparently originated from matched the last page they had actually visited. When there was a discrepancy, we inferred that participants had used their browser’s Back button or Forward button. This approach is similar to the approach detailed in [16], but is still subject to some problems because of browser caching.

- **The type of link clicked.** Our method allowed us to identify whether participants clicked on links tagged as navigation links, embedded text links, introductory page links, or survey page links. Given that the focus of this study included link placement, this information was crucial, yet such information would have been impossible to determine from standard log files.

- **Survey data.** Survey measures encompassed demographic and preference questions, comprehension questions (post-test), and open-ended comments. This data would have been entirely absent from standard server log files.

- **Passive data.** Passive data included such information as a participant’s assigned study condition, which Web browser he or she used, and what operating system the computer was running. Because all of the above information was linked through participants’ ID codes, we could easily match up different types of information, such as pages seen and comprehension, or condition and types of links clicked. An additional benefit of the method by which the scripts recorded data was that the page URLs did not contain excess information. The only clue to tech-savvy participants that any manipulation of the Web site was occurring was the .php file extension (as opposed to the long string of variables exposed in page URLs on sites such as amazon.com).

4.5. Bypassing raw server log file analysis

To avoid the difficult task of analyzing raw server log files, the tool directly exported data as a flat-file database that could be imported into a statistical program such as SPSS for analysis. Other scripts were developed to query the flat-file database for specific information such as comprehension results, number of participants who saw specific pages, or number of native versus non-native speakers. The results of these queries could be output either in a tabular form that could be easily viewed in a word processor or text editor, or as another text file, for further in-depth statistical analysis.

5. Conclusions and future directions

As the Internet becomes the mass communication medium of choice, the need for Web design guidelines based on rigorous empirical research will only increase. If we are to develop generalizable Web design guidelines applicable to a variety of users with different equipment and needs, we must move research from the usability lab into the broader arena of the natural Web.

The research tool that we are developing will support the conduct of such remote empirical studies. This tool will ultimately allow technical communicators who lack programming skills, or the time or desire necessary to learn them, to design and construct studies, create assessment instruments, manipulate Web design elements without constructing multiple Web sites, and collect, access, and analyze study results, all without relying on a programmer. We are currently seeking funding to support the further research and development needed to ensure that the tool’s final form provides high levels of accessibility to non-programmers while functioning as an effective aid for the naturalistic study of the effectiveness of Web design guidelines.

6. References


**About the Authors**

Jen Barrick is a Masters student in the Technical Japanese Program at the University of Washington, a sub-department of the Department of Technical Communication. She is interested in cross-cultural and computer-mediated communication and has a background in Web programming and computer technical support. She can be contacted by email at laeren@u.washington.edu.

Brandon Maust is majoring in English and Biochemistry at the University of Washington, and planning on pursuing a career in computational biology. His programming background is in system administration and Web-based applications. He can be contacted by email at bmaust@u.washington.edu.

Jan H. Spyridakis is a Professor in the Department of Technical Communication at the University of Washington. Her research focuses on document and screen design variables that affect comprehension and usability, cross-cultural audiences, and the refinement of research methods, particularly for Internet-based research. She can be contacted by email at jansp@u.washington.edu or by telephone at 206-685-1557.
Matt Eliot is a PhD student in the Department of Technical Communication at the University of Washington in Seattle. His research involves the development and use of accessible electronic and information technology, using ethnographic and other qualitative research methods. Matt can be contacted at mjeliot@u.washington.edu or by telephone at 206-683-1445.

Carolyn Wei is a PhD student in the Department of Technical Communication at the University of Washington. Her research focuses on adoption of information and communication technologies such as Internet and mobile devices in diverse settings. For her master's thesis from the University of Washington, she studied collaboration challenges within the distributed editorial team of a student computer science magazine. She can be contacted at cwei@u.washington.edu.

Mary B. Evans studies usability engineering and international technical communication. For her M.S. in technical communication from the University of Washington, she developed a set of empirically-supported Web design guidelines. She formerly served as a research analyst and Web developer in the National Oceanic and Atmospheric Administration, Seattle.

Kate Mobrand recently graduated from the Department of Technical Communication at the University of Washington in Seattle. For her master's thesis, she conducted a Web-based study on the effect of link wording on users' comprehension and perceptions of use. She currently owns and operates Mobrand Biometrics, Inc., an environmental consulting firm that develops analytical tools for application in fisheries resource management. Kate can be contacted at kmobrand@mobrand.com.