

Evaluation in the Trenches: Towards Rapid Evaluation

Sharon J. Laskowski, sharon.laskowski@nist.gov

Laura L. Downey, laura.downey@nist.gov

**The National Institute of Standards and Technology
Building 225, Room A216
Gaithersburg, Maryland 20899**

Introduction

At the National Institute of Standards and Technology (NIST), there are numerous Web sites maintained at many different organizational levels, each with varying objectives and user populations. This is not an uncommon situation, but it does present major challenges that affect evaluation and the subsequent usability of the Web sites. As HCI researchers, we have taken on the task of promoting usability within our organization. This, of course, includes Web sites. We are interested in developing evaluation methodologies that ensure usability of information systems within an organization.

Our in-house Web activities have provided the opportunity to explore and develop advanced evaluation techniques. We have performed several evaluations on NIST-related Web pages. These experiments have led us to some interesting conclusions and have also highlighted some key issues that need to be addressed in designing usable Web sites.

In this position paper, we outline four case studies of our attempts at encouraging and incorporating usability in Web interfaces. Next, we describe the web interface and organizational characteristics that shaped the evaluation methodology presented here. During the evaluations, we observed that, while typical evaluation approaches are useful, they are not sufficiently robust when applied to web interfaces. Based on these experiences, we have come to the conclusion that often one has no choice but to rely on techniques which support rapid, ongoing evaluation. We suggest, in this paper, some extensions to existing methods that support this "rapid evaluation."

Four Case Studies

The four case studies outlined below provide the basis for our discussion on the need for rapid evaluation. We believe they are representative of the Web pages of many organizations.

Case 1 was a home page for a large, diverse organization accessed by many users, both internal and external to the organization. We were asked to do a short evaluation to help improve the existing design and were presented with several alternatives. For our evaluation, we did a comparative analysis of alternatives using basic heuristic evaluation, an internal user walkthrough by evaluators who were part of the organization, and simulated external user walkthrough by evaluators with input from content-providers. Through this process, we were able to disambiguate terminology and overuse of graphics, remove centering, discover a better structure, remove long text, suggest more concise word usage, check for consistency, remove noise in terms of color and images, and encourage a higher fan-out of links. Some of our suggestions were not incorporated into the pages, for example, a more easily located position for

the "calendar of events" link. The major challenges in case study 1 were dealing with many levels of stakeholders, users and their expectations, and checking consistency across many levels of pages, when the page ownership is decentralized and the top level organization cannot exert control over those designs.

Case 2 was a virtual library top level page for a large, diverse organization. This also was an existing site with both external and internal users, with some pages only accessible by members of the organization. We were asked to do a short evaluation, "on-the-fly", to help improve the design. The evaluation consisted of one iteration using basic heuristic evaluation and an internal user walkthrough on a few pages by an evaluator who was part of the organization. This resulted in improved terminology and graphics, but the suggestions about color and text density were not taken. The major challenge was that the evaluation occurred at the end of the development process so any changes had to be minor, but with high impact.

Case 3 was a conceptual design for intranet access to sets of cross-referenced documents with the main purpose of document retrieval and viewing. The goal of this study was to provide design direction with usability in mind. Since we were not part of the design team, and there was no existing design to evaluate, we offered general guidelines and education, attempted to define the mental model for the site, and offered suggestions for organizing the information. The major challenge was how to provide helpful advice when there was no architecture or clearly defined user requirements yet developed on which to base an evaluation.

Case 4 was a clearing-house Web site designed to support a large number of standards developers and standards users requiring access to an information structure that included meta-information and complex searching capabilities. Since we were brought in at the conceptual stage, with only a domain analysis in place, but with the luxury of a small amount of funding (1 staff-year), we chose to provide a testbed to show the feasibility of a Web implementation. Our goals were to suggest an implementation strategy and apply usability principles from the start of the design. We were able to conduct several iterations of the evaluation process, since we had control of the testbed prototype. The evaluation included both heuristic methods and on-site walkthroughs with users at varying levels of expertise. In addition, we had access to a number of stakeholders from different organizational levels who had been involved in the concept development and domain analysis. These users evaluated the site from their perspective in a type of pluralistic, remote evaluation, which required some synthesis by these authors as primary evaluators. This remote evaluation encouraged a "buy-in" on the design across these multiple organizations. These methods were successful in that the prototype was well-received by the participating organizations, the testbed architecture was migrated to an operational system, and the testbed site received a LookSmart Editor's Choice Award. The major challenges were that, despite the extensive input of participating organizations, there was no architecture or clearly defined user functions. We had to use a tight development-evaluate-synthesize loop to illustrate our design concepts, architecture, and functions, then integrate comments very quickly. We also had to assess the impact of any design decision with respect to each of the stakeholders' organizational needs.

Note that in all four cases, while we saw a clear need for continued evaluation as the pages evolved, little or no resources were made available for this purpose. This emphasizes the importance of rapid evaluation under these constraints. In the following sections, we extrapolate from these case studies and present a general set of characteristics and an evaluation process. The

process is designed to identify the usability problems inherent in these types of Web applications even as they are dynamically changing.

Web Interface Characteristics

Our case studies reflect a particular set of characteristics associated with some Web sites. These include:

- Dynamic, evolving information that needs to be updated,
- Relationship-rich content and structure,
- Web sites that serve different communities with different objectives at the same time,
- Requirement to provide access to large amounts of information,
- Ubiquitous access, so that it is impossible to fully know your audience, and
- Evolving browser technology.

This complex nature of Web sites implies that the typical evaluation processes of formative, then summative evaluations, then release for use, is associated with high update costs, high levels of frustration for the users, or poor representation of the organization's information. Even when the cost of a new release is low, there is a requirement for rapid, ongoing evaluation, that the organization often ignores.

Organizational Characteristics

The organizations we dealt with had the following characteristics:

- Multiple levels of stakeholders,
- Lack of knowledge about the importance of usability and usability engineering, methodologies, and to some extent, system and software engineering practices,
- Difficulty in distinguishing prominent information and nonlinear ordering from organizational hierarchy,
- Designer and evaluator were often the same person,
- Limited access to domain experts and users,
- Scarce resources, and
- Tight time deadlines.

These characteristics then led to the need, in the course of an evaluation, to deal with such organizational issues as integrating different agendas and education about system engineering processes. We found that there was a disconnect between the ease of creating a web page and the difficulty in creating a well-designed, usable page. The organizational constraints also limited the analysis to small-scale evaluations, and caused a gap between the organization's understanding of

the content and the user perspective required by the interface.

Some Typical Usability Problems

The constraints imposed on the interface by the nature of this use of the web, by these types of organizations, leads to a number of typical problems, such as:

- Inconsistency of terminology and style across pages,
- Broken links,
- Mismatched mental models for information organization: organizational model does not match user models,
- Performance, often a byproduct of the mismatched models or misunderstanding of web capabilities, and
- Subsequent modifications that introduce new design problems and re-introduce previously eliminated problems.

The Need for a Rapid Evaluation Process

The above problems are not unique to the class of applications described here. However, they are more difficult to deal with because of the numerous web and organizational constraints. It is our contention that given the constraints, it is clear that what is required is an evaluation process that can be rapidly applied, not only at initial stages of design, but also during the evolution of pages over time.

In order to enable evaluation under these circumstances, the evaluation has to be an ongoing process. A quick turnaround at all times in the design and development process is required: fast "change and re-evaluate" is critical. Summative evaluation and formal usability lab studies take too much time and cost too much. The alternative is to construct a toolbox of rapid evaluation techniques that meet the needs of the web as a medium. In our case studies, we chose an ad hoc set of approaches that enabled rapid evaluation. We started with iterative expert review, a variation on heuristic evaluation. Because we knew the organization, we could do more than a generic heuristic evaluation: we made use of our knowledge of the organization and user populations -- this could be provided by a content-provider, or domain expert. We also adapted our evaluation process depending on the project objectives and resources. This adaptation was based on the stage of development, resources, time constraints, and receptiveness of developers and content providers. We wanted our advice to have some positive impact even if we were consulted for a very short time in the Web site design and development process. In general, the rapid evaluation needs to be an incremental and iterative approach.

Dealing with Some of the Usability Problems

In the context of the rapid evaluation process and usability problems described above, some of the following techniques can improve the usability of the design:

- Heuristic evaluation using standard best design practices to check for inconsistencies and obvious local design problems,

- Automated tools, such as browser compatibility checkers and spiders for identifying broken links and indexing the pages for searching by both designers and users,
- Task-based usage scenarios from different users' points of view to validate the information models and to test performance, and
- Recording of past designs, evaluations, and design decisions and justifications.

The main idea is to use whatever quick methods and guidelines are available to remove the major, but localized problems. Then, in order to find the more domain-specific, global problems, walkthroughs of various tasks from different points of view can be performed. These so-called "task-based scenarios" could be guided by a set of usability goals, some addressing general characteristics, such as identifying the email address of the webmaster, feedback mechanisms, and information location aids. Other goals that are domain-specific would involve subject matter experts identifying the major user classes, typical user information needs and associated query scenarios. This is based on the assumption that if a user can form a mental model of the structure of Web pages quickly, then it must be a usable structure.

Here is an example of a mental model disconnect. Recently, while searching an electronic magazine-related page, one of the authors of this paper was unable to find an email address or phone number for an editor. The information was at the bottom of the page after the advertisements, so a typical user would not think to scroll down that far. Instead, an email message was hesitantly sent to the technical webmaster to ask for this information.

Components of Rapid Evaluation

As we have seen, a number of component methodologies contribute to what we are terming "rapid evaluation." The extent to which the organization is willing to allocate resources to improve usability and the state of the Web site determines which components are applied. (Compare Case 2 with Case 4: more funding and early development stage means more impact.)

We believe that to be effective the components of rapid evaluation must achieve the following results:

- Evaluate and correct small problems with 1-2 days of analysis,
- Evaluate to identify major design problems and suggest corrections within 1 week, and
- provide on-going evaluation with very low level of support.

The components that we have identified so far that begin to meet these criteria are:

- Heuristic evaluation based on best practice guidelines and examples,
- Simulated expert and novice walkthroughs with task usage scenarios based on defined usability goals,
- Iterative expert review,
- Availability of information about domain, typical users, and cognitive task analysis,

- High degree of diplomacy and organizational savvy for evaluators, and
- Use of tools to facilitate maintenance of the Web site.

Issues and Challenges

While we have been very pragmatic in suggesting rapid evaluation as a means of identifying usability problems and improving a Web site, we readily admit that there are a number of issues and challenges that remain. First, we know that minor problems can be identified with our approach, but we do not know how well these rapid techniques can identify major design problems. Second, our walkthroughs depend on being able to design a set of usability criteria that can provide a minimum threshold for achieving usability and these might be difficult to generate. Third, we need rapid ways to measure improvements in design. Finally, we have just begun to use automated tools to maintain and test Web sites; there is much more we can do in this arena.

Sharon J. Laskowski

Dr. Sharon J. Laskowski is a computer scientist and group manager of the Visualization and Virtual Reality Group in the Information Technology Laboratory of the National Institute of Standards and Technology where she is currently investigating visualization for information retrieval and navigation and evaluation methods for intelligent collaboration and visualization. She has also participated in research and prototyping efforts focusing on usability and searching of World Wide Web-based virtual libraries. Previously, she conducted research and development in text analysis, information fusion, and plan recognition at the Artificial Intelligence Center of the MITRE Corporation. Dr. Laskowski received her PhD in computer science from Yale University.

Laura L. Downey

Laura L. Downey, is a computer scientist at the National Institute of Standards and Technology (NIST). Her research focus is on human-computer interaction with a special emphasis on usability engineering. She is committed to making technology work for people instead of the other way around. Currently she is investigating ways to make information retrieval (IR) interfaces more usable and developing techniques to evaluate visualization of IR results. Laura also designs and evaluates interfaces, conducts usability testing, develops evaluation methodologies and performs system analysis and system design.