

Analyzing Usability of a Beta-Version Web Site Through Server Logs, User Profile Data, and Online Questionnaire Responses

Laurie Kantner
Tec-Ed, Inc.
P.O. Box 1905
Ann Arbor, MI 48106
734-995-1010
laurie@teced.com

Larry Rusinsky
Tec-Ed, Inc.
P.O. Box 1905
Ann Arbor, MI 48106
734-995-1010
laurie@teced.com

ABSTRACT

To ensure “walk-up” usability of a custom search interface for a proprietary Web-based industrial product database, a publishing firm commissioned iterative usability studies. After two in-depth studies with individual participants, we decided to analyze beta-test sessions using automated data collection tools to learn from hundreds of beta-test users how successfully the product met their needs.

From web server log data, we reconstructed usage episodes to discover patterns and usability problems. Opinion data from online questionnaires and follow-up interviews provided qualitative information. User profile data from telerecruiting and online registration was also collected. The study provided convincing information for improving the web site.

INTRODUCTION

A publishing company that produces printed industrial product magazines has developed and released a Web-based database where users can supply detailed criteria to look up relevant product descriptions. After commissioning heuristic evaluation and usability testing of pre-alpha and alpha versions, the company expressed a strong desire to collect performance and opinion data from a large number of beta users. The usability team responded by recommending analysis of server log file data, online questionnaire responses, and user profile data.

To prepare for data collection, the usability team prepared a script for telerecruiting beta users; defined the log file data requirements; and developed the online questionnaires and the protocol for presenting them.

Within the three-month beta testing, the developers identified a two-week period that would provide a representative sample of data. Online questionnaire responses were collected for the whole three-month period, and profile data was collected for all recruited users. In addition, the telerecruiters conducted follow-up interviews with selected users to learn their opinions.

The usability team analyzed the tabulations of usage episode and user opinion data to draw conclusions about the usability of the web site. The user profile data confirmed that each beta-test user met the criteria we had specified. This method answered questions that an off-the-shelf log file analysis tool would not have addressed, making it useful for usability research, especially in conjunction with in-depth studies.

THE CASE STUDY

The publishing company's goal is to provide highly up-to-date, accessible industrial product information through a database on the World Wide Web. Anticipated users are engineers, purchasing staff, and others who specify or approve purchases of industrial products. To differentiate the site from existing Web-based industrial product search sites, the company chose a design consisting of a highly structured database that would allow very targeted searching.

Usability Objectives

Acceptable site usability meant meeting the following usability objectives:

- Users have first-time success searching the database. Success is defined as finding and viewing a narrowed list of product descriptions targeted to the user's specific requirements.
- Users can easily specify a search for the information they need.
- Users can easily move through search results to find the product description they want.
- Users meet their objectives in finding information.

The major user complaint the company sought to avoid was "too many irrelevant hits."

Site Design

The resulting Web-based database of industrial product information contains tens of thousands of industrial product descriptions organized into detailed product-type hierarchies and further indexed by many attributes. A product description contains data such as manufacturer contact information, materials used, product applications, industries in which used, and special properties of the product.

Users can navigate down product category trees (see Figure 1 for a sample list) or specify keywords in a search field. An additional layer of attributes enables users to filter their searches to products with the exact specifications they seek. During the beta-test period, users did not have to provide personal data until they had selected a detailed product description to view.

Figure 1: Sample Category Hierarchy

Categories (8 Total)

Assembly Systems: 1,013 Products

Bearings: 401 Products

Clutches: 601 Products

Conveyors: 1,401 Products

Drives: 471 Products

Feeders: 340 Products

Gear: 1,401 Products

Motion Control: 401 Products

Usability Evaluation History

Usability activities took place at key points in the product development cycle:

- When the search interface was first prototyped, the usability team performed two cycles of heuristic evaluation.
- After changes were made following the heuristic evaluations, the usability team performed a lab test of operational prototype software with representative users.
- After the search interface was modified to address lab test results and to incorporate professional graphic design, the usability team performed a prototype walkthrough of the new design.

These usability activities uncovered terminology and screen layout problems. However, the prototype walkthrough still did not evaluate how successfully users could search for products, because the product database was in development. The company next wanted to validate ease and success of searching using the operational user interface and a sample product database, with hundreds of beta-test users.

PROCESS FOR COLLECTING USER PROFILE DATA

The beta-test users needed to be people who would be likely to use the web site as a regular part of their jobs. To identify candidates, the publishing firm provided its circulation lists of individuals receiving industrial product magazines or inquiring about information within the magazines. Using these lists, which contained name and address information as well as profile information on a user's company (such as SIC code), a telerecruiting firm interviewed about 7,000 people, using a script developed by the usability team, to find people interested in participating in the beta test.

The script targeted candidates from the circulation lists who need to find industrial product information as part of their work, who would consider using the web to look up this information, who were interested in the product areas most heavily represented in the work-in-progress database, and who did not already work for a publisher of industrial product information. Over 2,000 people were recruited. The information collected during the recruiting process included name, title, company, web use (experience, home and/or work use, and frequency), information about the company's products and services, the candidate's product interests, and the candidate's role and responsibility for purchasing industrial products.

In a parallel effort, the user interface development team designed an online registration form that asked for name, title, address, job, and company information, plus a login ID. This information would be important for correlating user behavior in log file episodes and user responses to online questionnaires with user profile information.

PROCESS FOR COLLECTING OPINION DATA THROUGH ONLINE QUESTIONNAIRES

To evaluate thousands of usage sessions, we identified analysis of server logs as our best method of "observing" what users do. However, we also wanted to "hear" users comment about their experiences. To obtain this qualitative feedback, two activities took place: follow-up phone interviews and online questionnaires. The first activity took place as part of the telerecruiting follow-up, in which users who had not entered login IDs were contacted and asked if they in fact had used the site and, if so, their opinions of it.

Two online questionnaires were developed, one to be presented to the user when s/he first accessed a full product description, and the other when the user accessed his/her fourth full product description. Each questionnaire provided statements which respondents completed by selecting one of five provided answers. An example is:

The products found as a result of my search were:

- Precisely relevant to my search
- Mostly relevant
- Equally relevant and irrelevant
- Mostly not relevant
- Not at all relevant

The questionnaire data was stored in a spreadsheet for further analysis.

The data collected during the telerecruiting follow-up phone calls included whether users who had in fact accessed the site had found the information they were seeking, their opinion of the product categories, whether they found the site easy to use, and whether they intended to use it on a regular basis. This data was stored in a PC Access database for further analysis.

PROCESS FOR RECONSTRUCTING SERVER LOG DATA INTO EPISODES

The server log file, in a standard log file format, provided most but not all of the information the usability team had originally specified. Although an earlier test log file had included user ID information with session data, the file received for the analysis period did not contain user ID information if that information was in a cookie. (We later learned that the development team had stored cookie information in a parallel custom-formatted log file, but that file did not cover the entire usability analysis period.) Thus, for this study we were unable to correlate user profile information with user behavior to the extent we had originally intended.

Refining Sessions into Episodes

An episode was defined as a sequence of information exchanges believed to come from the same user. The site software assigned a session ID to a sequence of exchanges based on IP address and a maximum gap of 15 minutes between actions; in addition, a user's return to the home page signaled a new session.

However, when the usability team examined individual log file records, we found end-of-session interruptions in exchanges that seemed to belong to the same user with the same information need. We then refined this session data into episodes by adjusting two parameters: we raised the maximum gap between actions to 30 minutes, and a home page reentry did not signal a new session. These modifications resulted in fewer episodes than the automatically calculated number of sessions.

The usability team used standard PC database software and statistical packages to manipulate the log file data into episodes. (Log file tools are steadily improving and will eventually produce the desired information automatically.) The log file was reduced in size through stripping of graphics-access exchanges (which generated separate log file records) and simplifying of page URL notation.

One criticism of log file analysis for usability purposes is that the log file only records pages sent from the server, not client-side activity. However, in this study, most of the pages were dynamically generated on the server side, and home page accesses were always sent from the server rather than retrieved from cache, so all page activity could be traced in the log file. Examination of the log file revealed no discontinuity in page sequences, nor missing graphics accesses—both indicators of cache accesses.

Assumptions

It should be noted that equating IP addresses with individual users was a “safe” assumption for the beta test. However, when this site receives the high traffic expected in the future, or when evaluating any high-traffic site, this assumption will be risky. A single IP address could be used by different users within minutes of each other, or simultaneously by multiple users connected via proxy server. In those cases, we will collect extra information to correlate IP addresses with specific users, including browser version, user ID/cookie information, and session ID information.

Ensuring correlation between IP address and user would be especially important if we were analyzing the same user's behavior over time. However, this study was not intended to address repeat session rates (the beta-test database had limited content during the two-week log file analysis period).

THE VALUE OF THE RESULTS

The study yielded the following types of results:

- Where users entered the site
- How many episodes successfully reached each level of product information (the initial list of products and a full product description)
- The demographics of the users who reached a full product description
- The average number of pages accessed per episode

- Keywords entered for searching
- The relationship between the number of pages accessed per episode and the incidence of viewing full product descriptions
- How time between page accesses varied with the number of pages viewed per episode
- Comparisons between searching with product category hierarchies and with keywords

The most important finding from episode analysis was that searches using the product category hierarchy were statistically more successful in reaching full product descriptions than searches that did not use the hierarchy. Keyword searches did not achieve the same result. This finding would not have been convincing in a small study of only a few users, because the user's individual search goal would have been a confounding variable in deciding whether or not using hierarchies was in fact the problem.

Although standard log file tools can report important information for development and marketing purposes, current tools concentrate on per-page and per-visitor statistics. For example, the development team used a log file analysis tool to report the distribution of search methods used and the percentage of users that accessed full product descriptions; however, this tool did not report the relationship between these two characteristics, which the usability team was able to report using our tools. As stated before, log file analysis tools are improving, and some are beginning to report data that is immediately helpful for usability analysis.

The online questionnaire data provided useful data for usability improvements. The ratings information (such as rating the attractiveness of the home page) suggested areas where we could ask more targeted follow-up questions in telephone interviews. We did not include that step in this study but would strongly recommend it for future studies.

The follow-up telephone interviews for this study were strictly to ask users who had not logged in whether they had used the system and, if they had, their opinions of it. Free-form comments yielded the most useful information for usability improvements. We analyzed these comments in conjunction with textual information received in unsolicited email messages and in the online questionnaires to find common themes. This analysis yielded a myriad of suggestions for improvements, including the most common suggestion, "make the buttons more apparent." Again, we could design a very targeted follow-on study using those comments to explore usability problems in greater depth.

This study demonstrated that customized web-site log-file-analysis, performed in conjunction with automated collection of opinion and profile data, can be more useful for identifying usability improvements in web sites than using any of these single methods alone. Careful identification of the log file format is essential for having rich, meaningful data to analyze. The results can provide a solid basis for defining the focus of further usability study.

BACKGROUND OF THE PRESENTERS

Laurie Kantner has performed dozens of usability studies over the past 5 years, employing a variety of methods. In the past 2 years, she has specialized in Web studies, including navigation and downloading at mammoth sites and use of specialized search engines. She was an invited participant in the CHI '97 workshop, "Usability Testing WWW Sites." Laurie serves on the Editorial Board of *Common Ground*, the publication of the Usability Professionals' Association. Laurie co-presented "Multiple-User Testing: When One Person Can't See Everything" at the 1997 Usability Professionals' Association Conference. Her article on "Techniques for Managing a Usability Test" appeared in the September 1994 issue of the *IEEE Transactions on Professional Communication*.

Larry Rusinsky specializes in user interface evaluation and design and usability test design and administration. Recent projects include usability studies of Web sites for Netscape Communications, Sun Microsystems, Alta Vista, and Knight-Ridder, and user interface design for 3Com and Xerox. Larry worked previously for seven years in a development environment as product manager and user interface designer for bibliographic database management products, and he has a post-graduate degree in information science.