

CASE STUDY: USER INPUT ON LABELS

by Josh Carroll, University of Minnesota

Introduction

Avoiding bias in user experience research is necessary for eliciting high-quality feedback from your target audience and is a difficult goal to accomplish. Jakob Nielsen recently addressed this topic in his Alertbox article “Card Sorting: Pushing Users Beyond Terminology Matches” (August 24, 2009). Nielsen observes that in card sorting, users sometimes group cards based on the way that the descriptions on the cards are written. He suggests using synonyms to describe concepts across sets of cards given to different users and writing the card description in non-parallel structures. Synonyms and non-parallel structures help users to avoid matching words from a card to a concept or category.

Another method for avoiding key-word matching bias in usability evaluations and card sorts is to elicit users’ input for naming concepts and terminology found on websites. Before conducting a series of card sorts, ask users to review planned content for various sections of a website and name the content. The goal is to find out how users think of and describe information that they would access on a website.

Methodological Problems

About a year ago I consulted for a web design team for College of Education and Human Development at the University of Minnesota. This team was at the beginning stages of a complete website redesign and wanted to do a card sorting study. One theme that constantly came into play was that users either didn’t understand the relevance of some the items they were asked to sort or that they misunderstood the working definitions. An example of both of these problems (there were several)

was with a term used to describe inter-disciplinary research cohorts in this college. When presented this particular term, users either said they did not understand its relevance to the website or grouped it with content with which it obviously did not belong. Then when explained the purpose and definition of the term, users said that they would not have expected the definition and suggested more concrete descriptions for that term.

After this project, my team and I began rethinking our card sort methodology. The question that we kept returning to was “why are we asking users to inform website structures but not giving them an unbiased way for informing website terminology and content labels?” A solution to this problem is to present users with descriptions of information they would find on a website and ask them to name that content. Using this methodology we were able to find out in users’ terms how they conceptualize the information on a website.

Concurrent to integrating a new labeling exercise into the card sorting methodology, we were moving our card sorting procedure to a web interface. Because of this we needed to determine if the labeling exercise we had developed could be adapted to a web interface.

Method

To compare face-to-face and web-based administering of the labeling activity, I again collaborated with the College of Education and Human Development web design team. The team developed 34 content descriptions representing information that they planned for discrete web pages within the College of Education and Human Development website. These descriptions were presented to the participants to label with link names they would expect to lead to the information. 15 participants completed the activity in a face-to-face sessions with a moderator and 38 participants completed the activity through a web application designed to elicit feedback about terminology, much like moderated method, but through the web. The labeling results from each part of the

study were entered into a spreadsheet for analysis. The spreadsheets from each portion of the study were compared to see if any differences in labeling exist between the two studies.

Results

15 potential users from the website's target audience completed the face-to-face labeling activity. In general, the users offered similar labels for most of the terminology. 20 of the 34 terms were given similar or the same labels by at least a majority of the users. Likewise, many of the user-generated labels were similar to the team-proposed labels, with the users' labels containing at least one similar word or form of a word.

38 potential users, of the 2706 recruited, completed the web-based labeling activity for a 1.4% response rate. Like the face-to-face activity, the web-based activity produced similar results. For the most part, users created labels that were close to the labels created by other users, but did not have the same continuity as the face-to-face responses. Another prime difference that emerged between the face-to-face and web-based data was that web respondents were more likely to provide longer labels as well as labels that seemed more abstract than their counterparts. Also, some web-based respondents indicated that they did not think that some content belonged on the College of Education and Human Development website, while no face-to-face respondents mentioned their dislike of any potential content.

An overall trend that did emerge for both stimuli was participants creating labels that used words the team used in their descriptions or creating labels that were similar to the descriptions.

Take Aways

The low response rate for the web-based portion was surprising. This could have been caused by a couple factors: 1) the incentive for

completing the web stimulus may not have been large enough to attract many participants and 2) the web application may not have appeared credible to the audience even though the University of Minnesota word mark was present.

The web-based study required significantly less time and work to prepare, administer and order results in a meaningful way. Based only on effort and time, the web-based labeling exercise would allow web designers to elicit feedback on content faster than the face-to-face method.

The issue of participants using labels similar to the descriptions they were presented is comparable to the bias in card sorting that emerged when Nielsen observed that participants categorized based on the language used to describe the items to sort. However, this trend was more prevalent in the face-to-face exercises. A few factors may have contributed to the face-to-face participants mimicking the descriptions in their labels. For one, moderator bias could have played a role. During the face-to-face exercises, participants often asked questions and looked for approval that they were completing the process correctly. While the moderator can take actions to eliminate bias, it is impossible to predict how a participant is going to respond instructions or answers to questions. Because the web-based participants did not have the option to ask clarifying questions about the procedure, this particular bias was mitigated. However, other types of biases could have crept into the web-based exercise such as participants self-editing their initial reaction to a description in favor of a label that they thought was more creative or what the design team was looking for.

The web-based labels displayed greater variety in vocabulary and overall construction. More diversity in responses gives the web design team better context and label choices for constructing the website's links and content labels. Some of the web-based responses were long, rambling

narratives that would never be used on a website. However, the lengthier labels provided more insight into participants' vocabulary and context.

A next step in improving the labeling exercise is to investigate how Nielsen's suggestions of using synonyms and non-parallel structures in the descriptions changes the nature of the labels created. Instead of presenting each participant with the same descriptions, use synonyms and non-parallel structures across the participants. Another option, that would take significant more time and effort to create, is to present the participants with a mocked up webpage representing a piece of content and have them label that instead of a short, text based description.