Scaling the Card Sort Method to over 500 Items: Restructuring the Google AdWords Help Center

Yelena Nakhimovsky

Google, Georgia Institute of Technology

1600 Amphitheatre Parkway

Mountain View, CA 94043, USA

yelena.nakhimovsky@gmail.com

Rudy Schusteritsch

Google

1600 Amphitheatre Parkway

Mountain View, CA 94043, USA

rudy@google.com

Kerry Rodden

Google

1600 Amphitheatre Parkway

Mountain View, CA 94043, USA

kerryr@google.com

Abstract

In this paper we describe the method we used to redesign the information architecture of the Google AdWords Help Center. We adapted the card sorting method to accommodate over 500 items, and produced a new Help Center structure that enables users to find information faster and with fewer errors, as verified in a formal experiment. Our process can be used as a model by those faced with organizing or reorganizing a large body of information, where conducting a traditional card sort would not be practical.

Keywords

Card sorting, usability methodology, online help, information architecture, redesign

ACM Classification Keywords

H.5.2 User Interfaces – *Evaluation/methodology, Graphical User Interfaces (GUI), Training, help, and documentation, User-centered design.*

Introduction

Thousands of advertisers worldwide use the Google AdWords interface to place ads on Google.com and Google Network sites. To assist advertisers, Google offers a Help Center as part of this interface. It contains information on how to resolve various issues that they might encounter while using the system, e.g. while signing up for an account, changing settings for their ad campaigns, or tracking the success of their advertising efforts. The Help Center also functions as a repository of specialized knowledge that many advertisers choose

183

Figure 1: A bird's eye view of one of the longer pages from the original Help Center (126 links to common questions, under the heading "My ads are live, now what?").

to draw upon to learn more about the AdWords system and how to optimize their advertising strategies.

Many new features have been added to Google AdWords over the last several years. Recently, it became clear that, with over 500 pages, the help content was starting to outgrow the information architecture (IA) of the Help Center and that a large-scale redesign would be necessary.

The specific goals for the redesign included making it easier for new users to browse available information, enabling returning users to locate previously-accessed information more quickly, and encouraging all users to seek answers to common and straight-forward questions through the Help Center rather than via a phone call or e-mail to AdWords Support.

Card sorting can be an effective technique to use in these situations: a number of users are provided with a set of cards, each listing one of the items intended for inclusion in the IA. By identifying patterns in how users choose to group items (and how they name and organize the various groups they create), an overall hierarchy can be derived. See [1] for more details on this method.

However, in the case of the AdWords Help Center, this traditional approach would have been problematic: for most users, sorting over 500 items would have been both overwhelming and extremely time-consuming. One possibility would have been to have the participants sort only a subset of the cards. Given the complexity of the content, however, it would have been difficult to ensure that this subset was representative. Also, since many areas of the Help Center are very

specialized, it would have been impractical for us to have average AdWords customers conduct a comprehensive card sort. Given these constraints, we had to radically modify the standard approach to card sorting for this project.

Our method
The process had four main stages. First, we carried out some initial research, to learn more about how the Help Center was being used. Then, we conducted two card sorting exercises: in the first, three usability analysts and an AdWords customer support representative (CSR) worked together to sort all of the cards. In the second, AdWords CSRs sorted a carefully-selected subset of the cards, to resolve ambiguities. Finally, we conducted an experimental evaluation, to test whether users could locate items more quickly using the redesigned information architecture.

Initial background research
We used a number of different sources to learn more about the typical context of Help Center usage:

▪ From search logs, we noted the most common queries, and those that did not return any results.

▪ From listening to phone calls to the support center, we learned more about the most common issues that AdWords users have, and the language that they tend to use when describing them.

▪ In informal interviews with CSRs, we identified circumstances under which they would and would not refer users to the Help Center, as well as the resources inside and outside of the Help Center that they considered to be the most effective for end users.

Figure 2: Sorting over 500 items during the first card sort.

This research led us to expand the original set of cards to include resources from outside the official Help Center, and also helped us to prioritize category names later in the process.

First card sort: usability analysts with expert consultant
The first card sort ("Round 1") was conducted by 4 Google employees, who worked together as a group to sort over 500 cards. Each card named either a frequently asked question (FAQ) from the Help Center, or a related resource from outside it (roughly 100 cards). The participants for Round 1 included two usability analysts who were moderately familiar with AdWords, one who had little familiarity, and an AdWords CSR who served as an expert to the others. During the card sort, the expert provided reality checks on the accuracy of the proposed category names, and participated in generating overarching labels for groups of categories.

The participants were instructed to work together to sort the cards into piles that seemed most logical to them, and to use a general "unsorted" pile to house all items that could not be categorized with reasonable certainty. They started by sorting into piles those that had the most obvious categorization (e.g., Billing) and agreed amongst themselves to use the "unsorted" pile if there was any doubt or disagreement about where to place an item. Duplicates were permitted; they were ultimately implemented by attaching a note marking a card as a duplicate, and listing the piles to which it was assigned. The process took several 1-2 hour sessions.

Second card sort: Customer Support Representatives
After documentation of Round 1 results (see "Documenting the card sorts"), the second card sort

("Round 2") was conducted. This round consisted of three independent but simultaneous card sorts, with 3 AdWords CSRs working quietly in the same room. To create a smaller set of cards for Round 2, we sampled 4-8 items from each category created during Round 1. To create each sample, we chose 2-4 items that seemed to closely match the given category, and 2-4 items that seemed to fit less well. We combined those representative cards with the "unsorted" pile from Round 1, to yield a total of 178 cards for Round 2.

CSRs were recruited as participants since it was still unreasonable to ask external AdWords users to sort 178 cards covering a broad range of specialized topics. The CSRs were asked to apply their direct and daily contact with AdWords customers, and consider how customers themselves might sort and label the items.

To alleviate the time pressure of the task, we also provided the participants with the proposed categories that resulted from Round 1. Round 2 participants were told that the categories were completely optional and that they were free to modify or ignore them. For these, cards of different colors were used to distinguish between the potential main categories, sub-categories, and sub-sub-categories. To encourage participants to generate new groups at any level (or to come up with their own hierarchy of category names), blank cards of every color were provided. Again, duplicates were permitted and marked accordingly. Round 2 was competed in a single 3-hour session.

Documenting the card sorts
We documented the card sorts using linked spreadsheets. One spreadsheet listed all items to be sorted, and assigned an identifier to each one. Both the

Figure 3: Documentation of the groupings and hierarchy identified during the first card sort.

item names and numbers were referenced by subsequent spreadsheets.

For Round 1, a second spreadsheet listed each item, grouped by proposed category, sub-category, and sub-sub-category names. Round 2 data was tracked in a third spreadsheet, using a separate worksheet for each of the three participants' results. Each spreadsheet included a field which indicated whether the given item was duplicated in other categories; this enabled more efficient analysis later on.

We then created a separate list of original category names and proposed names from Round 1 and Round 2, which allowed us identify the most problematic and most successful names.

Analysis of card sort data
The analysis consisted of a number of steps:

▪   Comparing the documentation from both rounds of card sorting, we identified discrepancies and redundancy in categorization of items. Items that were consistently placed in more than one category were noted, along with their associated categories; different organization structures were weighed against their potential for illuminating or misleading a user's understanding of how to use the product.

▪   We finalized the basis for a new hierarchy on a whiteboard, through merging the most consistent ideas and organization schemes from the card sorts.

▪   We applied this new IA to the full body of 500+ items. We used the documentation from the Round 1 card sort as a basis for the new document, since the file already accounted for every item, rather than an abbreviated set.

▪   We reordered sections for clarity, and sub-grouped as necessary, sometimes extrapolating from the logic shown in the participants' card sorts. Our guiding goal was to keep sub-groups of questions relatively small whenever possible (less than 10 items).

▪   We further organized each category and sub-category by ordering and grouping items to present basic or frequently accessed information towards the top of each group or subgroup, and non-FAQ material at the bottom.

▪   Iterating on the hierarchy, we altered its order and labels as patterns were identified.

▪   We then used information from the card sorts, previous AdWords Help Center activities, design conventions and previous usability work on Help Centers to inform the final editorial decisions regarding the proposed IA.

Experimental evaluation

Our next goal was to test whether the redesigned IA would make more sense to users than the old one, and in particular, whether it would enable them to find items more quickly. Because we had changed only the organization of the items, not their wording, we wanted an evaluation method that would let us test the organization in isolation from other factors.

Method
We created HTML mockups of both the original Help Center and the redesigned version, covering the home page and the different levels of each IA hierarchy. At the bottom level, the links led to a placeholder page. We then selected a set of items, and designed an experiment where users were asked to locate those items by browsing the hierarchy.

User	Session 1	Session 2
1	Original / Task set A	Redesign / Task set B
2	Redesign / Task set A	Original / Task set B
3	Original / Task set B	Redesign / Task set A
4	Redesign / Task set B	Original / Task set A
5	Original / Task set A	Redesign / Task set B
6	(etc)	(etc)

Figure 4: The experiment design, which was counterbalanced by both IA (original or redesign) and task set. Tasks were given in a random order within a set.

We aimed to choose a set of popular questions, taking into account the most frequently viewed pages and most frequent queries on the original site, as well as calls and e-mails to AdWords Support. We consolidated the lists from these sources and removed duplicates. In some cases the remaining questions were still similar to each other, so we decided to create two sets of questions, paired so that each question had a counterpart in the other set. We had to add some extra questions to ensure that each one had a counterpart, and ended up with 2 sets of 17 questions each.

To prevent users from simply scanning for the exact question (which they would not know in practice), we rephrased each task slightly. For example, the question "How do I delete or undelete an ad campaign?" was rephrased as "I'd like to delete a campaign. How can I do this?"

We wrote custom software to administer the experiment, implemented in HTML and PHP. Each user started from a home page which linked to separate pages for each task. On a task page, the user read the question text, then pressed a "Start" button, which opened the Help Center (original or redesigned, depending on the condition) in a new window. Users were asked to browse the Help Center (they were not allowed to use any search facility, e.g. the browser's search within page function), following links as necessary, until they felt they were done with the task, at which point they closed the Help Center window, and clicked a "Found it" button (if they found the question) or "Give up" (if they did not find it). They then followed a link to return to their home page, so they could move on to the next task. Clicks on any of the buttons were

recorded in a back-end mySQL database, with an associated timestamp.

We used a within-subjects design, so the participants did half of their tasks with the original Help Center IA and half with the redesigned one. The design was counterbalanced by IA and also by question set (see Figure 4). In addition, we conducted the study in two sessions, separating the first half and second half by at least two days, with the hope that the participants would forget most of what they had learned about the IA they used first.

Our participants were twelve existing AdWords customers, with varying levels of familiarity with AdWords and the Help Center. Ten of them participated remotely, using screen-sharing software that allowed them to access the mockups on our corporate network. Participants were asked to work silently, but for remote participants a phone line was kept open during the study, in case they had technical problems. Each session was scheduled for one hour, but most participants needed less time than this.

At the end of each session the participant filled in a questionnaire that consisted of a number of semantic differentials.

Results
Our main metrics were:

▪ **time to task completion**, i.e. the time from pressing the "Start" button to pressing either the

Figure 5: A bird's eye view of one of the longer pages in the redesigned Help Center (32 links under the heading "Text Ads").

"Found it" or "Give up" button[1]. The median was 37 seconds for the redesigned IA, compared to 50 seconds for the original one.

- **error rate**, i.e. the percentage of trials for which the user clicked "Found it" when in fact they had not found the required question (or "Give up" when they had). This was 10.7% (redesigned) compared to 20.3% (original).

- **"give up" rate**, i.e. the percentage of trials for which the user clicked "Give up" and had not found the required question. This was 8.0% (redesigned) compared to 16.0% (original).

We also carried out an Analysis of Variance (ANOVA) with time as the dependent variable[2], and participant ID, task number, session number, and IA (original or redesigned) as independent variables. This showed that users were significantly faster with the redesigned IA: $F(1,328) = 17.83$, $p < 0.001$.

We also examined the questionnaire results; all of the semantic differentials were more positive for the redesigned IA, but none of the differences were statistically significant.

One particularly useful outcome of the experiment was that it allowed us to identify the questions (9 out of 34)

---

[1] We chose to include times from "give up" trials in the analysis, instead of treating these as missing data, because the recorded time was a conservative estimate of the time the user would actually have taken to complete the task.

[2] In practice, the distribution of times was strongly positively skewed (i.e. with a long tail to the right), so we applied a log transform to make it closer to a normal distribution.

for which participants were slower with the new IA than with the old one. Subsequently, we reclassified these items and fixed any incidental usability problems with the new mockup that arose during the studies.

Conclusions

The experiment results show that our scaled version of the card sort method resulted in an improved information architecture for the Help Center site. Our method can be used as a model by those facing the task of organizing or reorganizing a large and complex body of information, where conducting a traditional card sort with only end-users would not be practical. One particular advantage of this method (compared to the alternative of sorting only a representative set of cards) is that every item is considered at least once, making it possible to discover and accommodate the items that should appear in multiple low-level categories.

Citations

[1] Courage, C. and Baxter, K. *Understanding your users: A practical guide to user requirements methods, tools and techniques.* Morgan Kaufmann, 2005.