A COMPARISON OF SUBJECT-BASED CLASSIFICATION STRATEGIES FOR ENHANCED USABILITY

Vicki Ahlstrom
Federal Aviation Administration
The National Airspace System Human Factors Group
Atlantic City International Airport, New Jersey

Collections of documents, such as technical notes, are often classified on websites using a set of keywords that describe general subject areas. These keywords are used as a set of links to help the user navigate to the information he or she is seeking. This subject-based classification is intended to help users find information. A poorly designed subject-based index can make it difficult or impossible for users to find the information they need, however, creating an effective subject-based index is not easy. There are many different classification strategies that can be used. Each of these strategies can result in a different set of key words/key phrases. Some of these strategies may result in an index that is more usable than others. This paper addresses the problem of how to create a subject index that facilitates usability. Several strategies are compared against user search data. The five strategies resulted in differences in search success rate and efficiency.

INTRODUCTION

It has become increasingly common for individuals or research groups to make collections of documents available on the Internet. Publishing documents on the Internet increases availability without significantly increasing cost.

The Federal Aviation Administration's (FAA) National Airspace System Human Factors Group (HFG), like many other research organizations, documents their research through published technical notes and reports. The technical notes and reports are available for download on the Internet through the HFG website (http://acb220.tc.faa.gov/). Prior to publishing the technical notes on the website, the number of people who were aware of, or had access to, the research conducted at the HFG were limited. Those desiring information about research conducted at the HFG had to request a hard copy of the document from the group manager. Project managers within the FAA were often unaware of research that could have benefited their project. By publishing the technical notes on the Internet, many more people have been made aware of the research and are able to benefit from it. Approximately 1,300 technical notes are downloaded every month from the website, compared to approximately five hard copies requested a month.

Making the technical notes available on the Internet is not sufficient, however. In order to maximize the benefit of the research done at the HFG laboratory, users need to be able to quickly and easily find what they need. In order to facilitate user information search, we structured the website so that users could look up technical notes by the author name, year, or by topic.

Structuring documents by topic is a common means of organizing information; however, one might ask why you would want to organize the technical notes by topic when the users have the ability to use a free text search to find technical notes. Ahlstrom and Allendoerfer, (2005) and Gremmett, (2003), found that most users will browse subject headings first, resorting to free text search only if the browsing fails. Thus, an effective topic index is vital for the users to find the information they need.

Creating an effective subject index is not easy. For users to successfully locate documents from a subject index it is necessary for them to find the right word to match the topic in which they are interested, however, many different words can be used to refer to the same topic. This is known as the vocabulary problem. The designer must accommodate a range of users, and it is unlikely that every user will pick the same word to refer to a particular topic when they are searching (Brown, 1995). If the designer of the index chooses terms that don't match the users' terms, it can be difficult or impossible for users to find the information that they need, with the user being unsuccessful in 80-90% of attempts (Furnas, Landauer, Gomez, & Dumais, 1983).

Thus the problem becomes, how do you create an effective subject-based index? There are two common paths for solving this problem: automatic indexing and manual indexing. Automated indexes can use complex algorithms to create and refine subject-based indexes. Most automatic indexing, however, is beyond the means of an average user or small group. Developers of small (less than 100 documents) document collections have limited time and resources, and need a quick and easy way to create a usable subject-based index. These developers may be restricted to using primarily manual methods for indexing, yet they have the same goals as larger groups – to make it easy for users to find the information they need.

We found five different strategies that could be used for creating a subject-based index within our budget. Four of these strategies were manual and one was automated. These five strategies were: designer-based subject index, author defined keywords, card sort, automated keyword generation based on frequency, and optimized keyword. Each of these strategies results in a different set of keywords or phrases for each document. The keywords or phrases can be used as the basis of a subject index. The five strategies are further described below:

Designer-Based Subject Index

A simple method for creating a subject-based index is to allow a subject matter expert to choose the subject categories, assuming the subject matter expert has some familiarity with the topic area. This technique requires significant effort on the part of a subject matter expert. The usability of the index is likely to vary depending on the individual subject matter expert.

Author Defined Keywords

A common practice is to require the author to provide a set of keywords or key phrases for a document. Each of the technical notes within the studied collection has a set of keywords chosen by the author of the technical note. Several authors reported that it only takes them a few minutes to generate the keywords for the document. The keywords can be consolidated into a master list and used as a basis of a subject–based index.

Card Sort

A strategy commonly used by human factors specialists for structuring information is a card sort. To create subject-based categorization using this technique, 5-10 participants are given index cards, each one containing an abstract, and asked to put the cards into as many or few categories as they would like, giving each category a name. Each of the category names can be used as a topic in the subject-based index. This can take a few days to complete, depending on the number of abstracts and the availability of participants. In our case, we used 8 participants; each participant took approximately 2 hours to complete the task.

Automated Keyword Creation Based On Frequency

A strategy for creating keywords automatically is to use software to identify significant words that occur with the highest frequency in the document. Words that occur with a high frequency are used for keywords. The key words were then used as the basis for topic headings. We used the automated function in Microsoft Word to create keywords for this index. This method is relatively easy and does not require much time. Microsoft Word will automatically generate keywords or keyphrases. One problem with this technique is that the software often produced words that were used frequently but did not summarize main points of the paper, for example, data, researchers, and participants. This was particularly problematic for the keyphrase generating function. We used the keyword function instead of the keyphrase function because we did not believe that the keyphrases that were generated for many of the papers were usable for topic headings to guide a search, for example, Table 1, method, and results.

Optimized Keyword

An optimized keyword index is based on a strategy suggested by Furnas et al. (1983). This strategy requires 4-6 participants. Each participant is asked to provide keywords/key phrases for each abstract. The top keywords, for each abstract across participants, would be considered the optimized keywords. Although people may have an extensive vocabulary, they tend to use a few words very frequently. Because the optimized keyword strategy requires different users to provide words for each abstract, it simulates various approaches to the categorization of individual abstracts.

In our study, five participants (different participants than those that participated in the usability study) were given index cards with the abstracts. Each card contained one abstract. The participants were asked to provide 4-6 keywords/key phrases for each of the abstracts. For each document, the researcher counted how often each keyword was used to refer to the document. The top keywords for each abstract across participants would be considered the optimized keywords (in our study, we defined the top keywords as those chosen by more than one participant). This method and the card sort method were the most time consuming of all the methods we tested.

The overall goal for this effort was to create a usable subject index for a relatively small (<100) collection of documents. Because the users could scroll through the lists of keywords, they were not expected to recall specific search terms or create an exact match for terms. Instead, recognition of key terms or phrases was sufficient. The operational definition we used to gauge usability is the number of users who were able to successfully retrieve a document.

Increasing the number of words cross-referenced to any given topic could increase the hit rate. In other words, any given topic could be cross-referenced with all the words that appear in a thesaurus for that term and every variation of that term. For example, auditory could be also coded as aural, acoustic, audio, and so forth. This can lead to an index that is of an unmanageable size and is difficult for users to browse. Because users may browse through the list of subjects, it would be beneficial to keep the number of subject topics high enough to allow for a high user success rate, without unnecessary subject topics. Ideally, to maximize efficiency, one would want to minimize unnecessary keywords. Thus, hit rate was not the only measure that interested us; we also were interested in looking at the efficiency of the list created by each technique.

METHOD

Initially, when we decided to create a new topic index for the website, we had a single subject matter expert create subject headings for the technical notes. This subject matter expert was knowledgeable about the content of the technical notes. This method is an example of the designer-based indexing strategy. After the initial index was completed, we wanted to evaluate the usability of the index.

To do so, we had seven human factors specialists participate in a usability evaluation. The specialists were taken from volunteers at the Research, Development, and Human Factors Lab at the William J Hughes Technical Center.

The HFG website is like many other websites in that it deals with a fairly restricted domain. The majority of the users who come to our website looking for information have some knowledge of both human factors and the rudimentary

structure of the National Airspace System. The participants we used for this task had about the same knowledge level as we would expect of typical users of our website.

Materials used for this part of the study included a set of 4 x 6 index cards, each with an abstract to a technical note printed on one side and a number on the other (for the researcher to use when collecting data). An Html mock up, presented on an Internet browser platform, was created using the index produced by the subject matter expert. The topics were listed in alphabetical order, with a link to each technical note represented by hyperlinked text underneath the appropriate subject heading.

Testing a search task is difficult. Users may come to the site looking for a general concept, yet simulating a general concept is difficult. We chose to use abstracts from the technical notes to simulate a general concept that the user may use when searching the site. This is a methodology developed in the library sciences and information retrieval area in which abstracts of documents are used as conceptual representations of the information contained within the document as a whole (Bates, 1977; Brown, 1995).

Each participant was given a set of 71 index cards containing technical note abstracts. The index cards were shuffled between sessions so that each participant received a random order of abstracts. The participant read through the abstract and then placed the index card face down and tried to find the title of the technical note that referred to the abstract. After the participant picked a card, the researcher noted the card number on a data collection sheet. The participants were asked to state out loud what keywords they were looking for as they scanned the index. The researcher recorded these keywords on the data collection sheet. When a participant found the item corresponding to the abstract, he or she clicked on the hyperlink. After the participant clicked on an item, the researcher returned the index to home position. The participant picked another card, and the process began again. In many web applications, the hyperlinks change color once a user has selected them. We set the test browser up so that when the researcher returned the index to the home position, there was no indication of which hyperlinks had been selected (all hyperlinks were blue). This was done to minimize the chance that previous selections would influence the current task.

As our goal was to optimize the usability of the subject-based index, we wondered if using another strategy to build the subject-based index would produce better results, allowing the user to find the information more quickly and efficiently. We had the list of keywords that were generated by the users when they were looking for the documents in the above mentioned usability study. We may assume that the keywords that they generated after reading the abstracts would be the same keywords they would use if they were visiting our website and browsing for that topic. Thus, we decided to evaluate the different indexing strategies by comparing the keywords generated by the users to the keyword subject headings generated by each of the index strategies. For example, keywords given by the users in the previous exercise were compared to the keywords given by the authors to determine the success rate that might be expected if the subject-based index based was created using keywords given by the authors. This was repeated so that each of the five strategies described above was used to create an index and the keywords generated by the users was compared to determine a hypothetical hit rate and efficiency rate for each index. We did the analysis for 19 of the 71 initial abstracts. The 19 abstracts were used because they were available in Microsoft Word format, which allowed us to try the automated keyword feature.

RESULTS

Optimizing the usability of a keyword index means that users have a high rate of success in finding the document they want. The success rate was computed for each of five different methods (see Figure 1). The optimized keyword method produced the highest rate of success and the card sort the lowest.

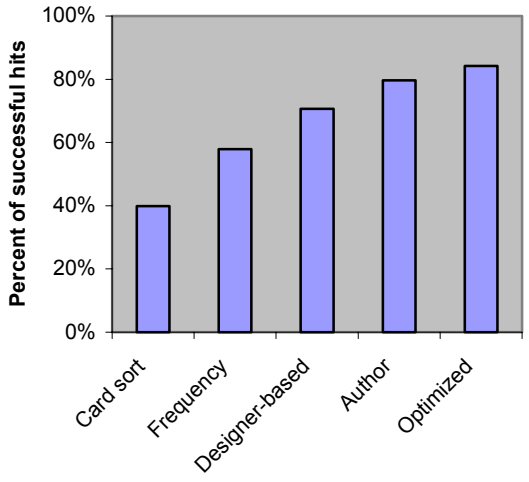

Figure 1. Percent of successful hits for the five different strategies.

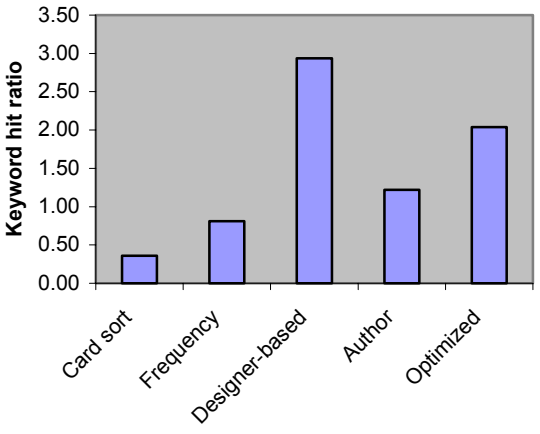

Figure 2. Efficiency of the five strategies.

One could argue that by including more terms in the index, the success rate could be increased, however, while it is important to include enough keywords to accommodate diverse users, it is impractical and inefficient to provide dozens of choices for each individual document. An optimal subject-based index maximizes the hit rate while minimizing the number of unnecessary keywords. Therefore, for each of the strategies, we computed an efficiency ratio equal to the number of overall hits (successful returns of information) over the total number of keywords generated by the index. [Total number of keywords in the index/total number of successful hits = keyword hit ratio]. Figure 2 shows the efficiency of each method, with a higher number indicating a higher rate of success per keyword. Lower numbers on the y-axis indicate that there are keywords present in the index that are not generating any hits. For indexes with the lower number, there are keywords that would be deleted without impacting the hit rate. The designer-based index and the optimized index had the highest hit per keyword ratio, thus they were the most efficient of the indexes. The least efficient indexes were the card sort and the frequency-based index.

When comparing the number of hits based on the web search terms, we found that using the keywords from the authors produced the most hits and the designer-based index created by the subject matter expert resulted in the least hits. The frequency index was the least efficient index when compared to free text search words, and the index created by keywords from the authors was the most efficient, followed closely by the designer based and optimized index which tied in efficiency. It is important to keep in mind that even though a keyword may have resulted in a "hit," we have no data from the free text searches about the relevancy of the hit. This is in contrast to the other set of data where the users were looking for documents related to specific abstracts.

DISCUSSION

The researcher designing a search task has a difficult dilemma. In order for the user to complete the task, the researcher needs to give the user something to search for, yet anything the researcher tells (or gives in writing) the user can potentially influence the vocabulary and resulting search terms chosen by the user. We were concerned that the specific vocabulary used in the abstracts, specifically the first sentence of the abstract, may have influenced the keywords used by the participants in the study. If this were the case, then the index resulting from the study may not be as usable as we would like.

To examine the extent to which the words used in the abstracts influenced the keywords chosen by the participants, we analyzed the first sentence of each of the abstracts and compared them against the keywords generated by the users. We found mixed results. Some of the keywords used by participants were exact words or phrases taken from the abstracts. For example, air traffic controllers have been known as many things over the years and different technical notes refer to them as controllers, air traffic controllers, certified professional controllers, and air traffic control specialists. We found that most of the participants tended to refer to the air traffic controllers using the same terminology that was in the most recent abstract that they read rather than referring to them consistently throughout the session. This data indicates the abstract may have had some influence on the keywords chosen. Thus, we felt it was important to compare the index to a broader vocabulary to maximize potential usability.

We compared keywords generated in free text searches of our web site to the indexes created by different methods. This represented a larger population of users and the wording of the abstracts would not influence the vocabulary used. By comparing the keywords entered in the free text search on the web, we also gained insight into how well the results might generalize to a broader population of users. We obtained the keywords and phrases entered into the search field on our website. We compared the vocabulary used in the free text searches with the vocabulary used in the indexes. The shortcoming of this strategy is that users are sometimes searching for something that does not exist on the site.

Additionally, even if we computed hit rates based on this data, we have no information on whether a hit was relevant to the user or not. It is difficult to directly compare free text search with a subject-based index because they are used for different purposes, thus the users may be drawing from two different vocabularies, one more specific than the other. Our feeling is that users are much more specific when using a free text keyword search than they would be if they are browsing a subject-based index. We suspect that users start general when browsing then become more specific. For example, a user might type in to a free text search "limitations of memory" but if searching a subject index may look for "memory," then browse the listings for "limitations."

Thus, rather than looking at the hit rate, we believe that free text search terms can be used to refine an index by comparing it to analogous terms. For example, in our indexes, air traffic controllers were referred to as air traffic controllers, controllers, certified professional controllers, etc. The dilemma is-which term to use? When there are multiple terms for a single concept, the designer can use the free text search vocabulary to identify which term may be more widely used. This strategy should be validated by additional usability studies.

There are several possible solutions for creating a subject-based index system. Each method results in a different set of keywords that can impact user success. This study compared the usability of the different index creation strategies available to the average designer. This allows designers with limited resources to create a usable subject-based index.

Although the card sort method provides information on which documents users considered related, the subject categories that come from our card sort do not appear to be sufficient to guide the users to the needed information. The card sort resulted in low hit rates and low efficiency. The card sort was also one of the most labor-intensive methods that we tried. It involved multiple users over several days. The results may have been influenced by the way we implemented the card sort. Some researchers who use this strategy have the participants work together to come up with

consensus topics. We chose to have each of the users generate categories individually rather than as a group because we wanted to capture the individual vocabulary differences. Although the users were not instructed to limit the categories that they used for the task, it appeared that the participants tried to group the documents into 10 or fewer major categories. Thus, the card sort strategy appears to be more effective when the goal is to create a small number of categories for a group of items. A subject-based index does not need to be this restrictive.

The optimized keyword strategy resulted in the highest hit rate and the second highest efficiency rate of all of the strategies used. This research shows that an optimized keyword index can be used to facilitate the usability of a subject-based classification system. This strategy can be helpful for users who have a small collection of documents to classify on a website. The drawback of this strategy is that it can be time consuming to get participants to read through the abstracts or articles and provide keywords.

Interestingly, a similar strategy is becoming popular on the Internet for categorizing websites. Sites such as Del.icio.us (http://del.icio.us) allow users to categorize links with their own keywords (called tags) and share those links with other users. The site lists the most popular tags for individual sites as input by actual users. This user-determined categorization has been called social bookmarking or folksonomy (Mathes, 2004).

Folksonomy is similar to the optimized keyword strategy in that multiple users are generating keywords for an item. As more users enter the categories, the top keywords evolve over time, similar to the optimization strategy.

In a future study, we would like to combine these two techniques to allow users of our website to enter keywords (tags) for each technical note. The top 3-4 keywords from actual users could then be used to categorize technical notes into a dynamic subject-based index. Essentially, this would be similar to conducting the optimization strategy using a much larger data set and a less contrived task.

REFERENCES

Ahlstrom, V. & Allendoerfer, K. (2005). Human factors usability study on a prototype portal. *South Jersey Human Factors Ergonomics Society Mini-Conference on Human Factors in Complex Sociotechnical Systems*. Atlantic City International Airport, NJ.

Bates, M. (1977). Factors affecting subject catalog search success. *Journal of the American Society for Information Science, 28*, 161-169.

Brown, M. (1995). By any other name: Accounting for failure in the naming of subject categories. *Library & Information Science Research, 17*, 347-385.

Furnas, G., Landauer, T., Gomez, L, & Dumais, S. (1983). The vocabulary problem in human-system communication: An analysis and a solution. *Bell System TechnicalJournal, 62*, 1753-1803.

Gremmett, P. (2003). *Utilizing a user's context to improve search results. CHI 2003-Best practices and future visions for search UIs - A workshop*. from http://home.earthlink.net/~searchworkshop/

Mathes, A. (2004). *Folksonomies- Cooperative classification and communication through shared metadata*. Retrieved May, 2005, from http://www.adammathes.com