How Many Users Are Enough for a Card-Sorting Study?

Tom Tullis - Fidelity Investments (tom.tullis@fmr.com)
Larry Wood - Brigham Young University (WoodL@byu.edu)

ABSTRACT
A study was conducted to assess the minimum number of participants needed for a card-sorting study. Similarity matrices and tree structures from various sample sizes were compared to those based on a set of 168 participants. Results indicate that reasonable structures are obtained from 20-30 participants.

Introduction

Card-sorting, either online or with actual cards, has become a very popular technique for helping to organize the elements of an information system in a way that makes sense to users of that system. It has become a standard tool in the toolbox for most usability professionals and information architects. Card-sorting has been used for designing mainframe menu systems (Tullis, 1985) and, more recently, for designing web sites (e.g., Frederickson-Mele, 1997; Tullis, 2003). A variety of computer-based tools are now available for conducting card-sorting exercises online and/or analyzing the data from manual or online card-sorting studies (e.g., EZSort, WebCAT, WebSort, Socratic CardSort, Classified, CardZort). However, one of the questions that has not been answered is how many users need to do the card-sorting exercise to get an accurate picture of how the information should be organized. The purpose of this study was to answer that question.

The Card-sorting Study

The study that these analyses are based on was conducted online at Fidelity Investments. The purpose of the card-sorting study was to determine how to organize the information for a redesign of the Intranet web site of our usability department. A total of 46 “cards” were used in the study, many of which represented services offered internally by the department, such as “Prototyping”, “Usability Testing”, “Wireless Design”, “Focus Groups”, and (somewhat circularly) “Card-sorting”. Some cards also represented general information about the department, such as “Who We Are”, “Where We Are”, and “Tour of the Usability Lab”.

The card-sorting study was conducted online on our company’s Intranet using the WebCAT tool. Users were presented with a list of the 46 “cards” in a random order. They would then drag a representation of each card into a region for each category that they wanted to create. Categories were not pre-defined; each user created and named their own categories. They could create as many or as few categories as desired. The exercise was complete when they had put every card into a category.

Employees of our company worldwide were invited to participate in the card-sorting study via an announcement in a daily message that is sent to all employees. The incentive to participate was entry in a drawing for a $50 gift check. A total of 172 employees participated in the card-sorting study. Four participants had to be dropped due to incomplete data, resulting in 168 complete card-sorts. For each participant, a file was created reflecting the cards that person grouped together and the names given to those groups. Each of these files can be converted to a similarity matrix showing all pair-wise combinations of cards, in which a pair that was grouped together received a similarity of “1” and a pair not grouped together received a similarity of “0”. Summing all of these individual similarity matrices resulted in an overall similarity matrix with entries ranging from 0 (if no one grouped that pair together) to 168 (if everyone grouped that pair together).

Data Analysis

These data were then analyzed using a modified version of WebSort to look at random sub-samples of different sizes from the full dataset of 168 participants.

The similarity matrix referred to above is the basis on which a statistical cluster analysis (Romberg, 1984) is performed, the result of which effectively "averages" the categorization cumulated across a set of participants. The resulting cluster analysis is then displayed as a hierarchical tree structure (known formally as a dendogram) on which organization of content and menu structures can be based.

The major goal of our research was to assess the degree of similarity of an organizational tree structure based on a sample of participants to a structure based on the full set of 168 participants in order to estimate the minimum number of participants needed to produce an *effective* organization. As a means to that end, correlation coefficients were calculated between the similarity matrices on which the cluster analysis was based. The assumption is that the more similar the trees, the higher should be the correlation between the similarity matrices on which they are based. Correlation coefficients between the sample similarity matrixes and the full similarity matrix were calculated for 10 samples each of sizes 2, 5, 8, 12, 15, 20, 30, 40, 50, 60, and 70 participants. A graph of the resulting mean correlation coefficients is shown in Figure 1.

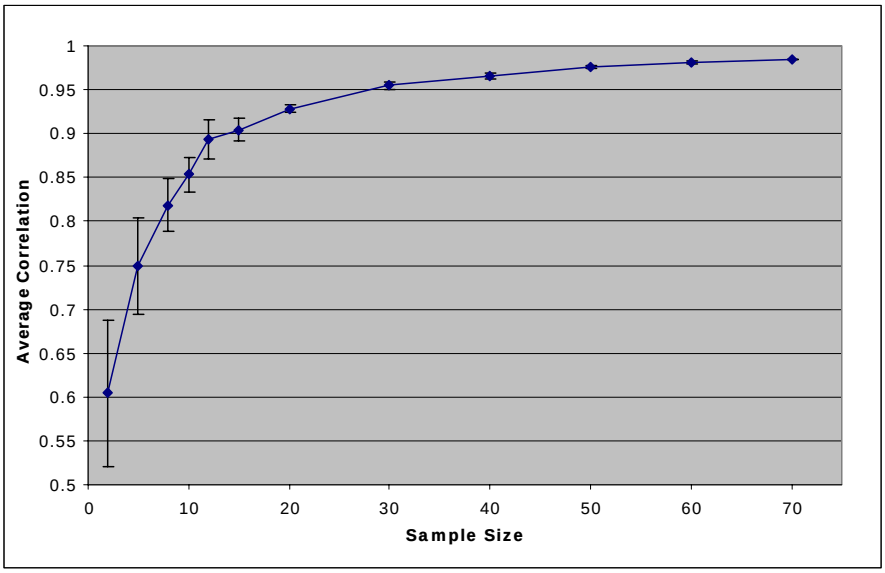

Figure 1. Correlation coefficients for various sample sizes, with error bars.

As shown in the graph, the relationship between the sample size and the average correlation is a negatively increasing function. Thus, the increase is more dramatic at the smaller sample sizes so that as the size increases beyond 20-30, there is little increase in the size of the correlation coefficient. Also note that the variance of the values, as indicated by the error bars, is much greater for the smaller samples.

An important question is how the function shown in Figure 1 relates to the similarity of the actual tree structures as a function of sample size. One practical implication is that the structures derived from sample sizes above 30 are very similar to that derived from the full set of participants, while those based on smaller sample sizes are increasingly different with smaller sample sizes. To the extent that this is true, it would have implications for determining the minimum number of users needed to obtain valid information.

To illustrate the kinds of differences in structure that occur with trees based on various sample sizes, trees are shown in Figures 3-6 of the Appendix for sample sizes of 40, 20, 15, 10, and 5, respectively. The tree based on all 168 participants is shown in Figure 2. The highlighted items in Figure 2 of the Appendix are the basic two-card clusters that are formed during a cluster analysis, based on cards that are most similar. A tree is then constructed by either combining these *base* clusters or adding individual cards to an existing base cluster iteratively until the tree is complete. As shown in Figure 2, there are 17 base clusters derived from the analysis using all 168 participants.

In Figures 3-7, one of the base clusters from the tree based on all 168 participants (card 18-*Remote usability testing* and card 21-*Portable usability testing*) is highlighted to indicate how the trees based on various sample sizes might differ from that based on all 168 participants. As shown in the figures, the cluster is still intact in the tree for N=40, whereas in those for N=20 through N=5, those two cards are separated by a greater distance as the sample size becomes smaller. Of course, this is not true of each tree for a given sample size, but does illustrate the trend.

To provide a more general idea of how tree structures differ as a function of sample size, five trees were generated from each sample size. For all 17 base clusters, the mean separation of the two cards in each base cluster was then calculated across the five samples. The results are listed in Table 1 for each base cluster as a function of sample size. Cluster separation was measured by counting the number of nodes separating the two cards in each base cluster. For example, referring to Figure 4, card 18 and card 21 are only separated by one node, whereas in Figure 6, they are separated by six nodes. A node was defined as the intersection of two branches. Thus, the number of nodes separating a pair consisted of the number of intersections that had to be crossed going up the tree from each pair until a common intersection was found.

Table 1. Mean separation of base clusters as a function of sample size

Base clusters from analysis of all 168 participants	Mean separation of base clusters across different sample sizes. Means are derived from five samples of each sample size				
	N = 5	N = 10	N = 15	N = 20	N = 40
42 Design_for_touch_screens 43 Design for voice_based	1.6	0,4	0.8	0.8	0.6
44 Design for elderly users 46 Design for blind users	2.8	0.4	1.2	1.4	0.8
29 Web Design Guide 40 Top 10 Web Design Mistakes	1.0	1.0	0	0	0
10 Prototyping 11 Card sorting	2.4	2.0	3	3.2	0.6
18 Remote usability testing 21 Portable usability testing	1.8	0.4	0.4	0.6	0.2
04 Usability checklist 41 Usability cycle	1.4	1.8	0.2	2.2	0.6
24 Web usability seminary 34 Sign up for usability studies	2.4	1.6	2.2	0	0.4
22 On-line help and documentation 23 Documentation samples	1.6	1.4	1.6	0	0
27 Usable Bits newsletter 28 Usable Bits archive	0.8	1.4	0	0.4	0
25 Study of the month 26 Study of the month results	0.6	0	0	0	0
12 User surveys 13 Focus groups	1.8	1.2	2.0	2.4	0.8
14 Expert reviews 36 Case studies	1.8	2.4	3.8	2.8	1.4
31 Eye-tracker_research 38 HID research	8	1.8	8.2	2.4	5.2
06 HID news 07 HID events	2.0	0.8	0.6	1.4	0.4
08 Who we are 45 Where we are	1.2	0.8	0.6	0.6	0.4
01 HID mission 20 Tour of HID lab	1.0	0.6	0.4	0.2	0.4
03 Site feedback 37 Customer_Testimonials	2.2	4.4	2.2	0.8	1.4
Mean separation across all base pairs	2.0	1.3	1.5	1.1	0.75
Mean % base pairs separated	69%	50%	45%	48%	35%

Conclusions

A general conclusion that can be drawn on the basis of this research is that it may not be cost effective to spend resources to gather information from more than 20-30 participants in a card-sorting study. However, it is important to note that even the trees based on the smallest sample sizes are probably closer to the one for all 168 participants than might be obtained from speculation by a designer who is not a potential user of the content or application for which the organization is being developed. As always, we must exercise appropriate caution in generalizing results from one study. Results will obviously differ as a function of the homogeneity of the participants in a sample and such things as the instructions given to the participants for the card-sorting task.

References

Frederickson-Mele, K. (1997) Usability Testing an Intranet Prototype Shell: A Case Study. CHI'97 Workshop on Usability Testing World-wide Web Sites. Retrieved 1/30/2004 from http://www.acm.org/sigchi/web/chi97testing/mele.htm.

Romesburg, C. H. (1984) *Cluster analysis for researchers*. Belmont, Calif. : Lifetime Learning Publications.

Tullis, T. S. (1985) Designing a Menu-based Interface to an Operating System. *Proceedings of CHI'85 Conference on Human Factors in Computing Systems*, San Francisco, CA, April 1985.

Tullis, T. S. (2003) Using Card-sorting Techniques to Organize Your Intranet. *Intranet Journal of Strategy and Management*, March 2003.

EZSort: http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410

WebCAT: http://zing.ncsl.nist.gov/WebTools/WebCAT/overview.html

WebSort: http://www.websort.net/

Socratic CardSort: http://www.sotech.com/main/eval.asp?pID=123

Classified: http://www.infodesign.com.au/usabilityresources/classified/

CardZort: http://condor.depaul.edu/~jtoro/cardzort/index.htm

Appendix - Trees from various sample sizes

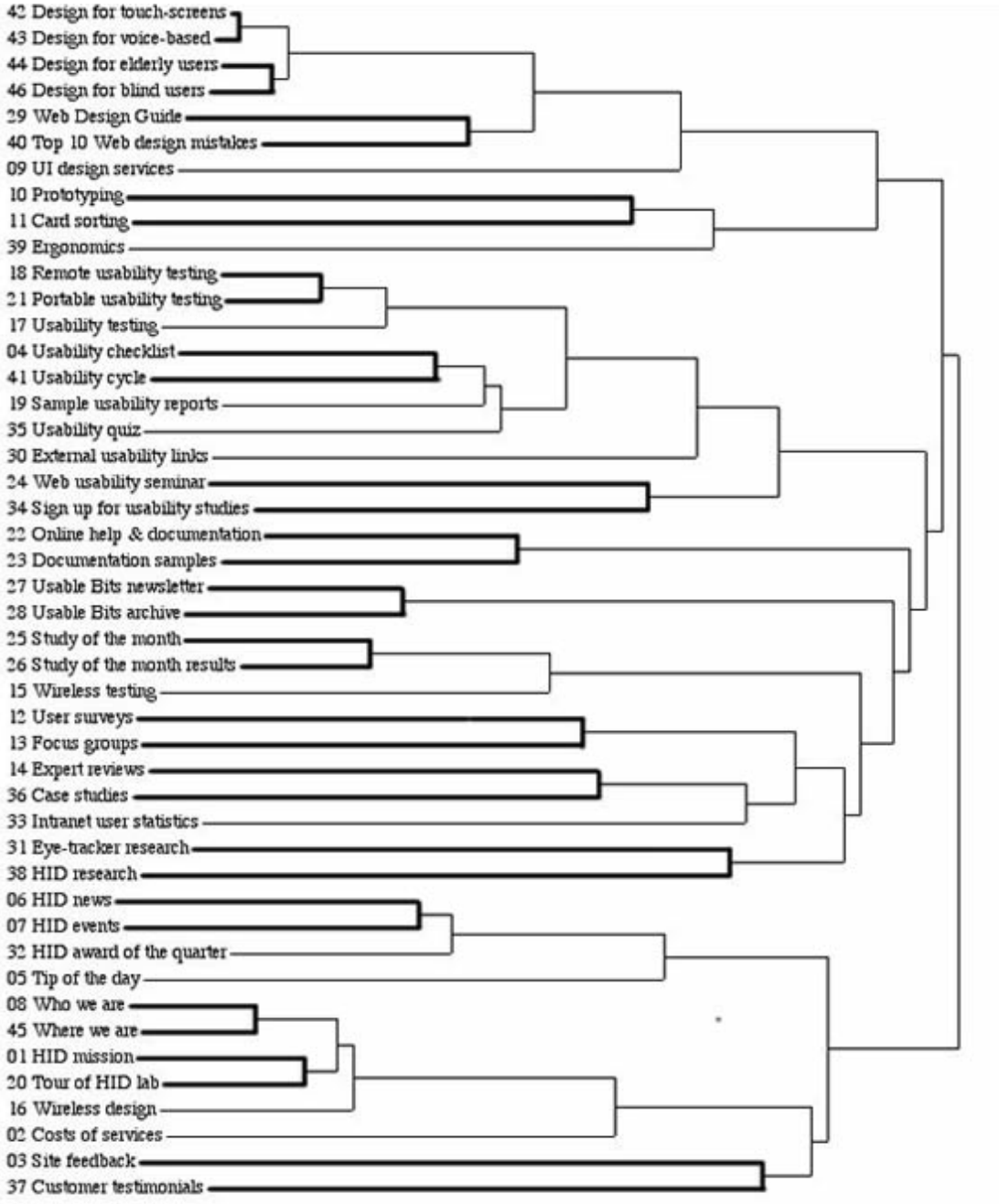

Figure 2. Tree based on all 168 participants with base clusters highlighted.

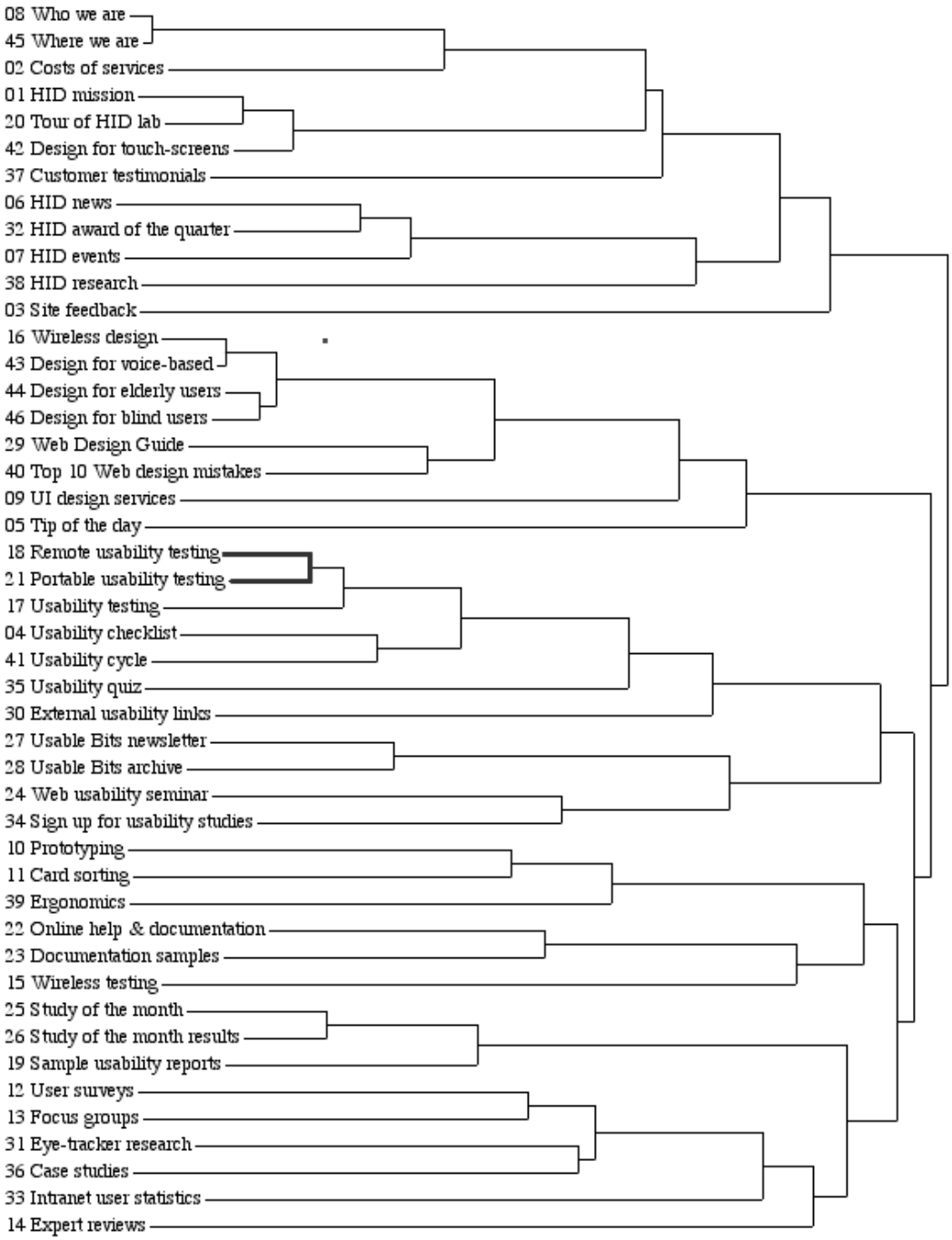

Figure 3 - A sample tree from sample size N=40

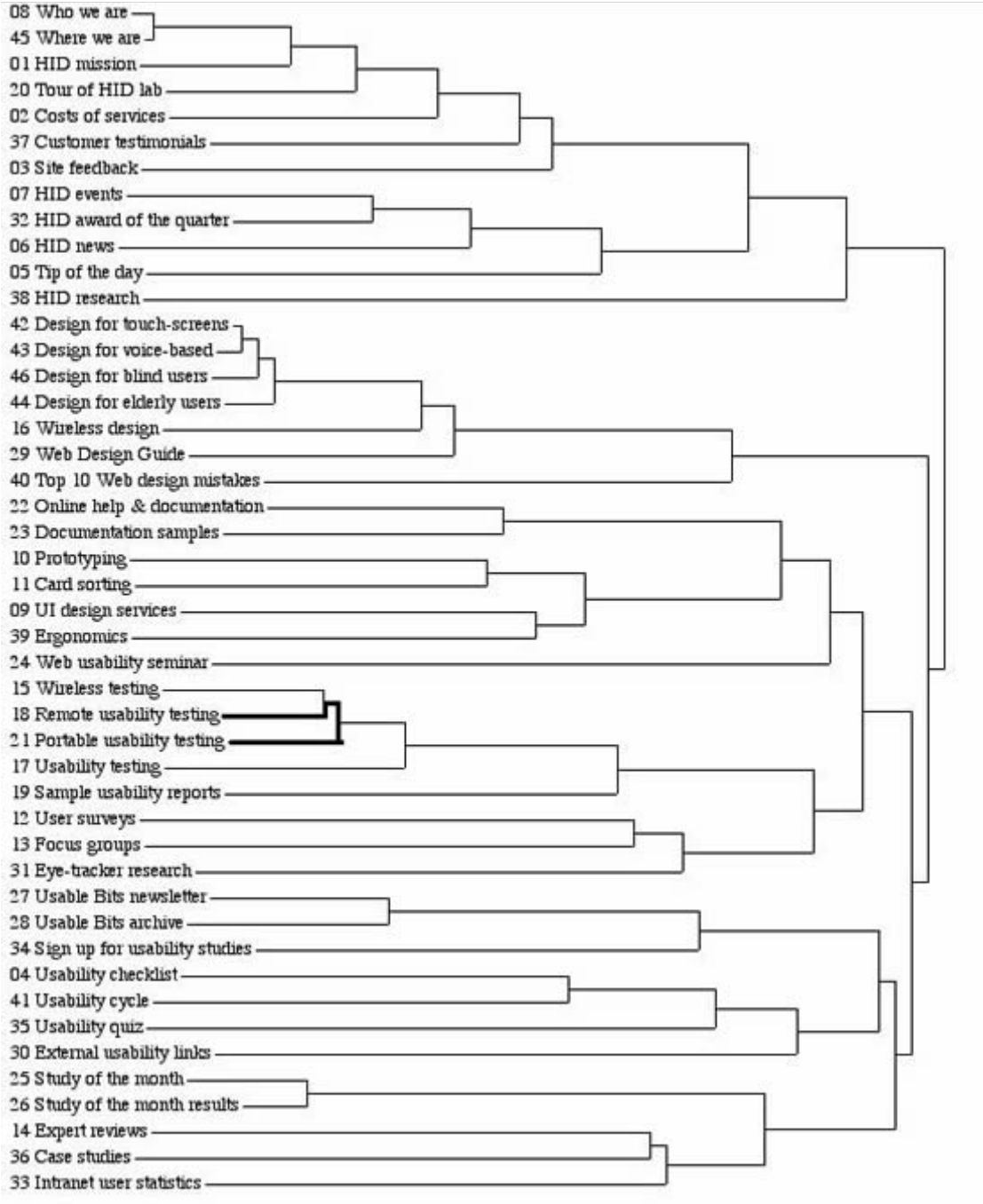

Figure 4 - A sample tree from sample size N=20

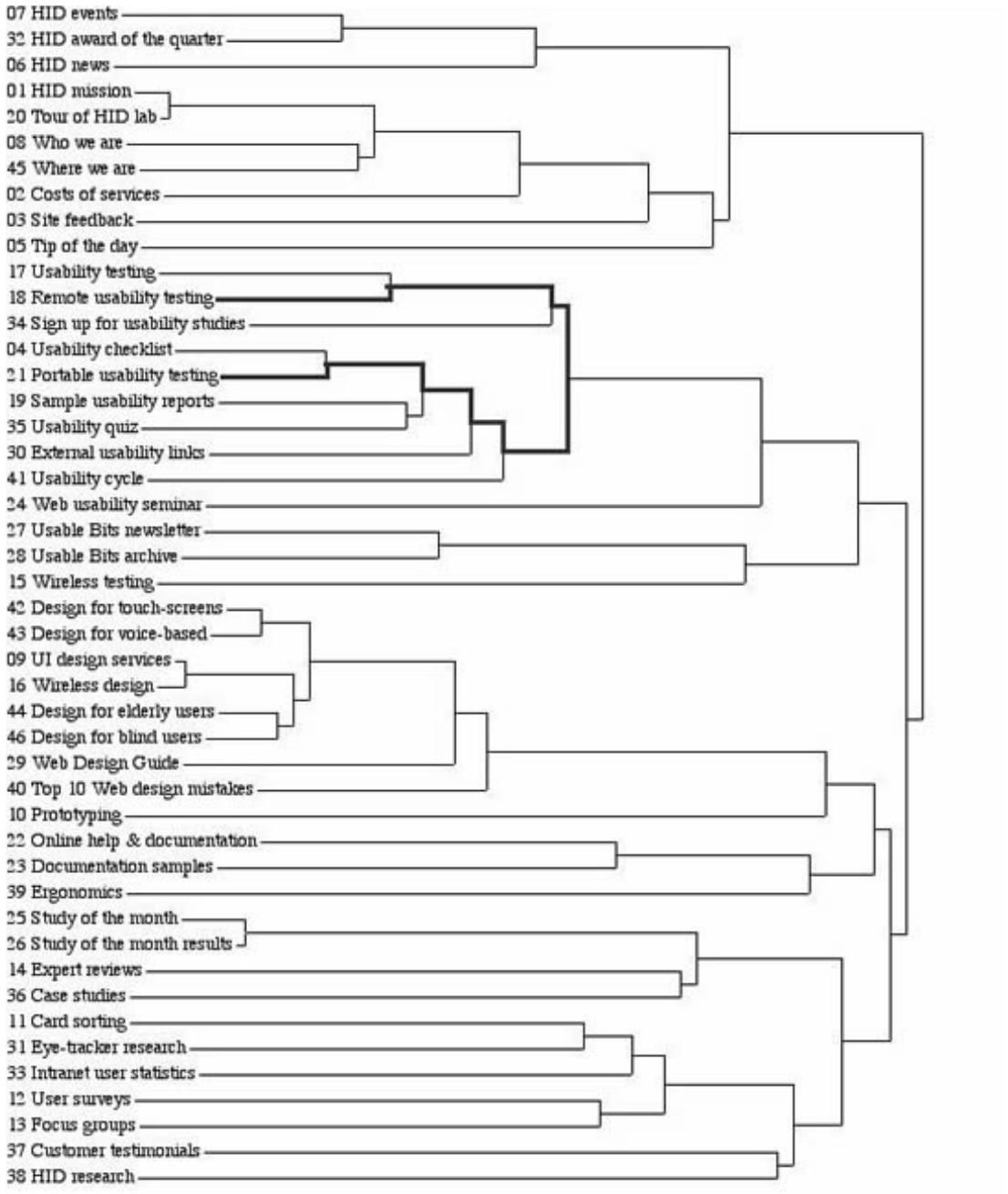

Figure 5 - A sample tree from sample size N=10

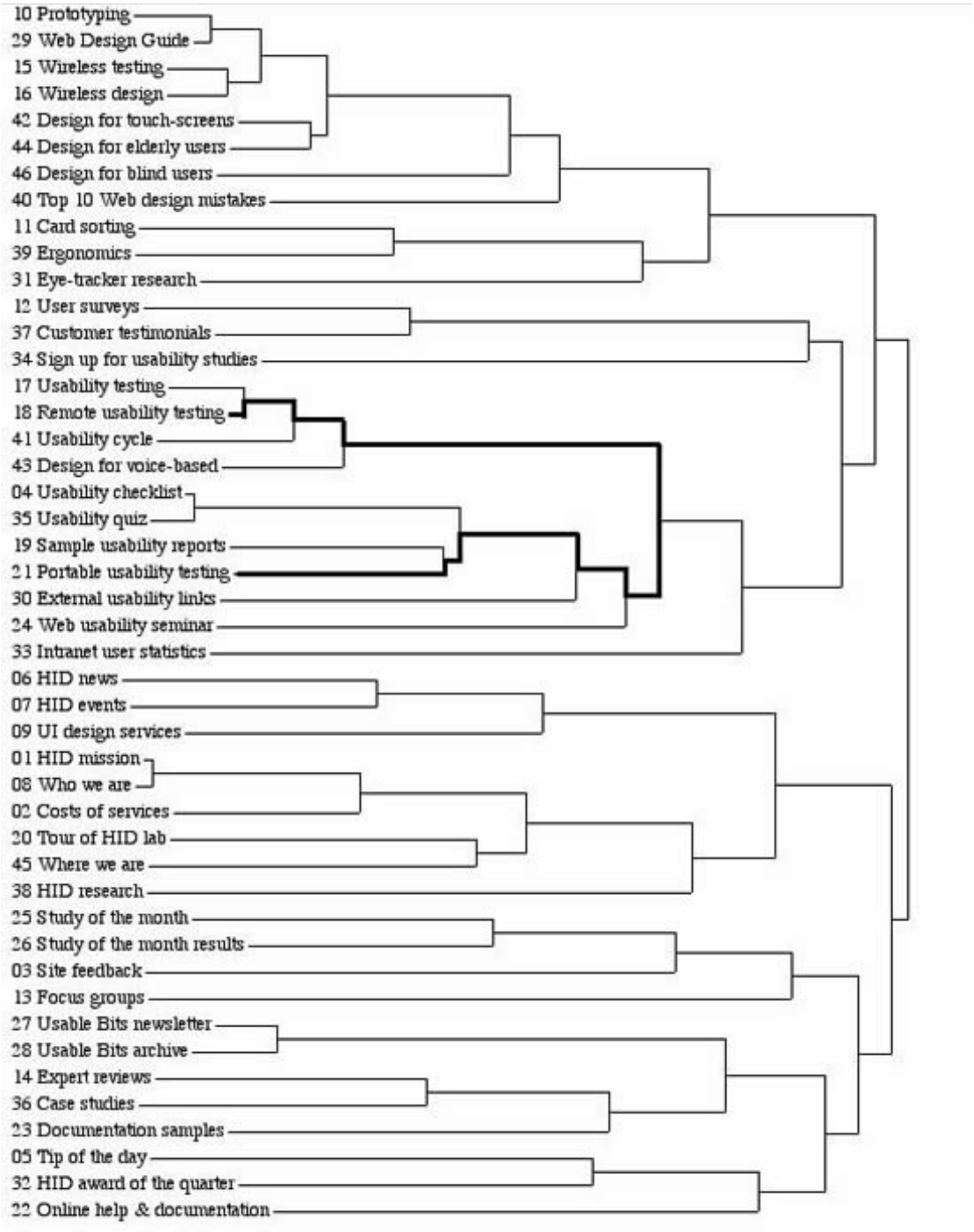

Figure 6 - A sample tree from sample size N=5