Old Cards, New Tricks: 
Applied Techniques in Card Sorting

William Hudson
Syntagm Ltd
10 Oxford Road
Abingdon, UK OX14 2DS
+44 1235 522859
whudson@acm.org

ABSTRACT
Card sorting is an extremely useful technique in the design of interactive systems. However, it is under-used in practice – often through a lack of understanding and the complexities of cluster analysis. This half-day, hands-on tutorial uses concrete examples taken from live web sites to guide participants through the analysis, design and execution of card sorting activities, particularly as they apply to web navigation. Specific topics presented include open and closed sorting, rapid data collection using bar codes, cluster analysis and extensions to traditional analyses using quality of fit metrics and measures of deviation.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Evaluation/Methodology; Theory and methods.

General Terms
Design, Human Factors.

Keywords
Card Sorting, Cluster Analysis, Dendrograms, Dendrograms, Topography, Navigation Design.

1. INTRODUCTION
(This brief paper provides a flavour of the content of the tutorial. It is based on a longer article written for ACM interactions [1]).

Card sorting is a knowledge elicitation technique often used by information architects, interaction designers and usability professionals to establish or assess the navigation hierarchy of a web site. The items are typically menu entries or hyperlinks while the groups are categories or headings. The process involves asking participants to sort items into meaningful groups. In open card sorts the number and names of groups are decided by each participant while in the closed card sorts these factors are fixed by the researcher in advance.

2. CARD SORTING ANALYSIS

© William Hudson, 2007
Published by the British Computer Society
Volume 2 Proceedings of the 21st BCS HCI Group Conference
HCI 2007, Lancaster University, 3-7 September 2007
Devina Ramduny-Ellis & Dorothy Rachovides (Editors)

Analysis of card sorting results range from simple counting of the number of times items were grouped together to the rather intimidating monothetic agglomerative cluster analysis (known simply as cluster analysis in most cases). Unfortunately, no single technique provides everything a researcher needs to know, especially if convincing evidence is needed to persuade colleagues or customers of the effectiveness of a proposed design.

2.1 Convincing Evidence
The evidence we need falls into three categories:

- Participants. Are these the right participants for our site? Are they all thinking about the items and their groupings in a similar way? Do they have a clear understanding of the card sorting task itself?
- Items. Are the item names well-understood by participants? Are there alternatives that should be considered – perhaps terms users are more familiar with?
- Groups. For closed card sorts, have we chosen the right number of groups and names for each? For open sorts, are participants largely in agreement about the number of groups needed? How well do participants feel the items fit into their groups?

Happily, the answer to this last question – how well participants feel the items fit into their groups – can also help us with many of the other issues listed. Coupled with a few data collection guidelines and alternative presentations of results, we can collect fairly comprehensive evidence about what is and what is not going to work in our navigation hierarchies.

2.2 Quality of Fit by Item
So let’s examine this last question in some more detail: How well do participants feel the items fit into their groups? It is possible to argue that this question is redundant; that the items must fit into their groups relatively well in any given set of results, because that is how the participant decided to group them. However, practical experience says otherwise. In many cases, participants place items into groups that are “good enough” but not necessarily ideal (satisficing versus maximizing behaviour). By asking participants to provide a “quality of fit” measure the cluster analysis can be extended to include the strength of relationship between items. Furthermore, analyzing quality of fit by participant or by item can give us useful insights into the some of the thought processes behind the sorting activity. For example, figure 1 shows quality of fit averaged by item for a closed card sort based on wines. The highest quality of fit in this example is

---

1 Quality of fit is indicated by participants on the following scale: fair (1), good (2) or perfect (3). If an item is omitted by a participant, its quality of fit is 0.
Claret while the lowest is Beaujolais. They are both red so it seems a little curious that the graph shows such disparity. The answer lies in the group names that were assigned in advance: full-bodied reds, dry whites and sparkling. Participants were happy to put both Claret and Beaujolais “full-bodied reds” even though they apparently knew Beaujolais was not full-bodied and despite instructions that they were to omit items they felt did not fit into any group.

In this case the quality of fit metric has allowed participants to express their dissatisfaction with Beaujolais (in the full-bodied reds group) while acknowledging their reluctance to discard the item entirely.

2.3 Alternatives to the Dendogram
Card sorting results are traditionally shown using a tree-like figure known as a dendogram. Unfortunately, in many applications, dendograms hide more information than they show. A simple alternative that is explored in the tutorial is a simple surface map of the underlying proximity matrix used for cluster analysis (each cell in the matrix is the frequency with which card pairs appeared in the same groups weighted by average quality of fit for the pair). An example, produced using Microsoft Excel, is shown in figure 2.

In this case the quality of fit metric has allowed participants to express their dissatisfaction with Beaujolais (in the full-bodied reds group) while acknowledging their reluctance to discard the item entirely.

The map shows three distinct groups with anomalies centred on Pinot Grigio. Further investigation brought to light a terminological confusion: participants were not generally familiar with Pinot Grigio (white) and were associating it with Pinot Noir (red). Consequently, participants grouped Pinot Grigio with red wines almost as frequently as they did white. While a hint of this problem would have been discernable in a traditional dendogram, it would not have been as dramatically obvious.

3. REFERENCES