A Comparison of Card-sorting Analysis Methods

Nawaz, Ather

Document Version
Final published version

Published in:
APCHI '12. Proceedings of the 10th Asia Pacific Conference on Computer-Human Interaction

Publication date:
2012

A Comparison of Card-sorting Analysis Methods

Ather Nawaz
Copenhagen Business School
Howitzvej 60, DK-2000, Copenhagen, Denmark
An.itm@cbs.dk
+45 3815 2396

ABSTRACT

This study investigates how the choice of analysis method for card sorting studies affects the suggested information structure for websites. In the card sorting technique, a variety of methods are used to analyse the resulting data. The analysis of card sorting data helps user experience (UX) designers to discover the patterns in how users make classifications and thus to develop an optimal, user-centred website structure. During analysis, the recurrence of patterns of classification between users influences the resulting website structure. However, the algorithm used in the analysis influences the recurrent patterns found and thus has consequences for the resulting website design. This paper draws an attention to the choice of card sorting analysis and techniques and shows how it impacts the results. The research focuses on how the same data for card sorting can lead to different website structures by generating different set of classifications. It further explains how the agreement level between the users can change for similar data due to the choice of analysis.

Author Keywords

Card sorting; website structure; Method; Comparison; HCI; Classification

ACM Classification Keywords

H5.2. Information interfaces and presentation

INTRODUCTION

A number of studies of user-centred design (UCD) for websites use card sorting in the design, development and evaluation process of website structure. UCD approaches put the users of a website at the centre of the design, development and evaluation. Different approaches to evaluation, such as focus-groups, usability testing, cards sorting, participatory design, questionnaire and interviews are used as part of this process [1].

The choice of card sorting technique in usability studies has implications for the results of the resulting website's structure. The card sorting method is used to understand how users classify and structure website content. Data collected from multiple participants is compared between participants and with existing website structure. The comparison of the card sorting results between different participants is intended to achieve the best website structuring for a given domain of website. The domain of websites may include e-commerce websites, academic websites, healthcare websites or other such domains. The best structure of the websites is achieved by evaluating how users agree on structuring contents into categories. This users' intended structure and attributes for the website is compared to the existing content structure of websites.

APCHI '12, August 28–31, 2012, Matsue-city, Shimane, Japan.

Card Sorting and Analysis

The term card sorting applies to a wide variety of activities involving ordering, grouping and/or naming of objects or concepts. Card sorting is an established, intuitive method for understanding users' mental models of website structure. It is used frequently in software development, evaluation, and product design to understand the clustering of information and relationships between information from the users' perspective. Card sorting is used to group items into categories and to understand users' mental models of organization of website contents. In brief, in card sorting each card has a statement or product written on a card that relates to a page of the website, and these cards are then sorted by participants into relationships they find meaningful.

This paper argues that the choice of techniques and tools for card sorting has consequences for the ascertained website structure. In analyzing card sorting data, the data of multiple people is combined to determine an appropriate website structure. Thus, the data of multiple participants is analyzed in a variety of ways to come up with the aggregative sorting. Some studies use qualitative methods to analyze the data, looking for patterns in the sorts [2]. In this case, attention is paid to synonyms, concepts and themes in the sorting. Quantitative analyses for card sorting, on other hand, use different tools to interpret the users' sorting. These tools use algorithms such as cluster analysis and similarity matrixes to arrive at an interpretation [3]. A result can also be obtained by considering how far users place their cards from each other and how many steps are required to change one user's sorts to another user's sorts [4]. All such tools look for agreement in the patterns of uses' sorts. The most agreed-upon pattern is then used as the basis for the new website structure. In these sorts, there are number of cards for which users do not particularly agree on a specific placement. The choice of analysis for card sorting affects these cards most of all where users do not agree between each other.

There are two major types of card sorts used in most studies, the open card sort and the closed card sort. In an open card sorting study, users are presented with unsorted packs of index cards. They are asked to sort these cards according to their understanding and to label them. In a closed card sort, predefined groups are provided and users are asked to sort cards into these groups.

A problem which may arise during card sorting is that the choice of analysis and tool might impact on the resulting website structure. Most studies provide an analysis and visualisation of users' classifications that explains the agreement of users on the clustering of groups, but does not examine the logic used to conduct the analysis.

In fact, when determining the information structure of any website though card sorting, there is often considerable disagreement in the way users organize the cards into groups [2]. Despite their general similarities, users may vary in their mental models for organising concepts in a structure.

There is a need to understand the card sorting analysis and logics used in card sorting analysis because many of the websites determine their structure after conducting analysis of card sorting experiments. Still users find it hard to navigate on the websites despite adopting user-centred design approaches. The information architecture of websites represents the underlying structure that give a shape and meaning to their content [5]. Regarding the structure for navigation, the focus should therefore be given on users' view of the world for websites structure and understanding users' view of the world is vital to design optimal information structures of websites. The website structures are getting very large, and interaction is seriously limited by the available resources of the space of the screen. The users always look to get to the information quickly. A better understanding of how users conceptualize website structure can improve the quality of websites.

The perception of webpages' quality can also differ according to culture. Therefore, card sorting is also used to elicit cross-cultural perceptions of web page quality and structure [6] and to understand the attitude of different groups of users to a given system [7]. The use of different analysis in card sorting such as edit distance analysis, cluster analysis through similarity matrixes, and comparative analysis (i.e., thematic vs. taxonomic analysis) is common in the research studies of card sorting [2, 4, 8, 9]. Studies of website design use a variety of analysis for card sorting to come up with the user-centered structures for websites. Some of the studies conducted usability analysis of card sorting tools. However, few studies have conducted a comparison of the logic behind these tools used in card sorting.

Research Aims

This research paper aims to document how the choice of technique for card sorting has implications for the resulting website structure. The results produced through analyses of three techniques not only show different patterns of agreement by the users for the same data; but also different explanations of the data. This study shows that the choice of three methods of analysing data (analysis of edit distance, analysis of best merge method and analysis of actual merge method) has consequences for the resulting structure of the websites. These three analyses for card sorting are chosen because they are interesting from a research point of view. All of these techniques claim to determine an optimal solution for website structure in their own ways. Analysis of actual merge method (AMM) and best merge method (BMM) combine multiple card sorts into an aggregated card sort. The AMM and BMM are derived from cluster analysis. These two techniques are widely used in the industry to see the patterns of users' card sorting. AMM and BMM explain visual aspects of data with the analysis. Edit distance is used in academic circles to reflect on the variation in users' card sorts. It counts the difference between two sorts at a time and looks for one or more sorts that are central to all other user sorts.

The article is organized as follows. We begin by explaining why different structure matters. Different analyses for card sorting will be described afterwards. We then examine the data of 38 users through the analysis of edit distance, AMM and BMM. Finally we conduct a comparison on how data reveals different aspects of users' agreements. Then we will discuss the effect of number of users on the card sort and the threshold effect on the structure produced through card sorting.

RELATED WORK

Different Structures Matter

Different website structures matter to people's ability to navigate and find information. According to different structures, the contents of a website go into different levels of hierarchies. Different levels of hierarchies and locations of contents affect users' response time and success in finding information. Website structure becomes important when users look for information on a website at different levels of hierarchy. Allen investigated the effect of information depth on the response time and error rate at each hierarchical level of a website [10]. Response times became longer for searches deeper into the website, and users made more errors when the information to be retrieved was at deeper levels. Our previous study showed that there is some disagreement in how the users structure the contents of a website [2]. Further, websites often use different classification and navigation structures such as linear, tree, network, and global structures [11, 12].

Different structures matter because users have a tendency to perceive website structures in different ways. Users may perceive and group information in a thematic or taxonomic structure, for example, grouping items in a thematic classification are related to each other through a coherent story or situation [2]. In a thematic classification of a banana, monkey and panda, the two items banana and monkey go together via a classification based on eating habits and a coherent story of the situation of a monkey eating a banana. In a taxonomic structure, users classify items into groups according to the function or inferences drawn from the items in the group. The items are related to each other through higher level abstractions, or property [2]. Using previous example, in a taxonomic classification, monkey and panda go together in the same group because both are mammals.

Some studies have conducted analyses of card sort through comparison of different tools for card sorting [3, 5, 6]. Chaparro et. al. compared commercially available electronic card sort applications [3]. The study focused on user satisfaction, performance, usability and preference of card sorting tool. Results of the study indicated different preferences for the two user groups. Researchers who participated in the study preferred WebSort for creating and analysing the card sort. The end users preferred OpenSort for completing the card sort exercise. The study focused on the interface and functionalities of tools and did not look into the method which is used to conduct the analysis of card sorts. Katsanos et al. used semantic similarity between words, phrases and passages of user data to come up with an aggregative sort of webpages [5]. Katsanos et al. introduced a computational tool, autoCardSorter, which supports clustering of the web pages of a site. Petrie et al. investigated the difference between online card sorting and on-site card sorting [6]. Their study looked into the preferences between online card sort and offline card sort and found that online card sorting took significantly longer for non-native English speakers than native English speakers.

Most of studies which conducted analysis of card sorts did not look into the techniques and logics which are used in the card sort tools. Instead these studies tested the usability of tools, efficiency and effectiveness of users and preferences of user groups between online tools for card sorting and offline card sorts.

Card Sorting Analysis Work in Different Ways

There are different ways in which the card sorts of different participants can be compared in order to create an aggregative sort. Here we discuss some of the ways used to carry out this process.

A number of studies have used different techniques to analyse card sorts. Some of techniques examine the difference between the users' sorts. The University of Illinois at Chicago library redesigned their library website by conducting open card sorting studies and analysed the card sorts through factor analysis [13]. The study pointed out that qualitative analysis of data is also important in addition to Factor analysis. In the Katsanos et al. study the clustering during the design was built through taxonomical, statistical and hybrid techniques [5]. The taxonomical technique calculated the path length between two node-words. The taxonomical technique ensured a certain quality of the results because it involved human coding in the clustering of the words. It made it possible to model multiple synonym words. The statistical analysis used the probability of co-occurrences of captured text and clusters them together. The statistical analysis relies on machine learning of synonym words. The hybrid analysis combined the taxonomies of concepts with statistical properties of a text [5].

Petrie et. al. conducted a comparison of onsite card sorting data collection with the offsite data collection [6]. The onsite data collection was conducted through open card sort without using online tool for users' input. The offsite data collection used web portals and online card sorting tools for data collection. The outcome of the studies showed that the online version of data collection took a significantly longer time to complete than the onsite version. Kralish et al. compared card sorting results across Malaysian, Russian, British and German students [14]. The study used ranking of cards to come up with the final aggregative sort. The aggregative grouping was based upon the users' ranking of which information on the cards was most useful. Nawaz et al. conducted a qualitative analysis of card sorting to see if individual users grouped items according to a thematic classification or a taxonomic classification [2]. Martine and Rugg used co-occurrences matrixes to assess the similarity of webpage designs through card sorting [15]. The co-occurrence matrix shows how often a respondent places any two cards in the same group. Curran et al. investigated podiatrists' perceptions of expert systems in relation to their perceptions of other diagnostic of diseases through card sorting [7]. The study used multiple criteria to come up with an optimal sorting of expert systems. Petrie et al. used edit distance to see how users group items in a similar or different way [6].

Best Merge Method

The best merge method (BMM) is a technique based upon similarity matrixes and is the industry-standard. In brief, the similarity matrix counts the frequency of co-occurring pairs in the cluster [18]. Once all groups are broken into pairs, the method finds the most frequent pairs in all groups and constructs new groups out of those pairs. In other words, the best merger method accumulates the pairs of cards which are placed by the different users in the same group. The best pairs in the users' sorts are found and merged to form a group which is then assumed to be consistent.

Scenario 1:

X: [a, b, c] (1 group with 3 cards)

Y: [a], [b], [c] (3 groups with a card each)

Z: [a], [b], [c] (1group with 1 card)

Result of BMM = 1 x [a, b], 1 x [a, c], 1 x [b, c]

Scenario 2:

X: [a, b], [c]

Y: [a, c], [b]

Z: [a], [b, c]

Result of BMM = 1 x [a, b], 1 x [a, c], 1 x [b, c]

The pair reduction process in scenario one and scenario two has produced identical results for two different scenarios. The BMM only works by merging the pairs, so it does not reconstruct the original data.

Actual Merge Method

The actual merge method (AMM) works by looking into whole groups, rather than pairs, taking an inheritance perspective on information architecture and applying it to card sorting.

Scenario 3:

X: [a, b, c]

Y: [a, b], [c]

Z: [a, b], [c]

Result: of AMM = 1 x [a, b, c], 3 x [a, b]

$$\text{Disagreement \%} = \frac{\text{Avg. Neighbourhood Distance}}{\text{Total Number of cards}} \times 100$$

The AMM counts each instance of a complete group from every user. Each group with a non-zero score (a "real group") inherits the base score (i.e. before inheritance) of all superset groups. The group with the highest score is taken, and all conflicting groups are eliminated. The scores that the AMM analysis provides give an exact account of "X%" of users agree these should be grouped together.

Edit Distance

Edit distance is based upon a distance function that measures how far apart two card sorts are. The distance is considered to be the minimum number of stages required to convert one sort into another sort, where one stage consists of moving one card from one group to another group.

Consider the following example with two sorts A and B, both consisting of four groups of cards:

A= [A1, A2, A3, A4] and B = [B1, B2, B3, B4]

A1 [1; 2; 3] B1 [1; 2]

A2 [4; 5; 6] B2 [3; 4]

A3 [7; 8; 9] B3 [5; 6; 7]

A4 [ ] B4 [8; 9]

Sort A can be converted into sort B by moving items between groups. A minimum set of moves is as follows: move 3 from A1 to A4, 4 from A2 to A4 and 7 from A3 to A2 [4].

After the moves:

A1 [1; 2]

A2 [5; 6; 7]

A3 [8; 9]

A4 [3; 4]

Thus, the *'D'* function has a value of 3 because there were three moves needed to convert A into B. The most immediate application of the edit distance metric is for determining the similarity between two sorts. This is particularly useful when looking at sorts that use similar criteria [4] and is conducted through finding the "neighbourhood". A neighbourhood is a process of finding the sorts most closely related to a user's sort or to a websites' sort. Neighbourhood provides a measure of the dissimilarity between all sorts and shows which of the sorts is the closest to all the users; whereas AMM and BMM combine multiple card sorts into an aggregated card sort. In the end, all three sorts look to find an ideal user-centred website structure representation. If a single sort has many close neighbours, it may be part of a common theme in the overall data. Neighbourhood and edit distance are sometime mixed with each other. The edit distance is the method which explains the distance between two users and it uses neighbourhood as a way of analysing the distance between the sorts.

Among other methods, Hierarchical cluster analysis or cluster analysis is used to analyze card sorting. Hierarchical cluster analysis is an individual-directed method [16]. It is a method for assigning items into groups in such a way that the items whose themes are similar to each are grouped together. It focuses on the relationship between the individual items, and items can only appear in a single place in the hierarchy. Hierarchical cluster analysis is used in card sorting studies to see how different users group content. Hierarchical cluster analysis is best suited for data where a clear hierarchical organization already exists [17]. For example, plants are naturally organized into species, then genera, orders, etc. We focused mainly on Edit Distance, AMM and BMM because they are aligned together and can be explained with an approach and scenarios which are common in Edit Distance, AMM and BMM.

METHOD

In the first stage, data from users is collected onsite through open card sorts. The card sort data of users is analysed using edit distance, AMM and BMM to see how the users' structure provides different organisations of website structure. For the analysis of edit distance, we used UW Card Sort Analyzer[1], a Windows application. For the analysis of AMM and BMM, we used the web-based tool OptimalSort[2]. In the second stage of the analysis, we interpreted and compared the results collected with the three methods.

Procedure

Graduating students' organisation of website content was elicited using open card sort. Participants were asked to complete a background questionnaire regarding their computer use, internet use, language use on websites and educational background. They were later asked to perform open card sorts. The participants were given 15 minutes to sort the cards into groups. The time of 15 minutes was decided after conducting a pilot study with 5 participants.

Material

The participants in the study were provided with 37 2x2 cm cards with a home appliance's name mentioned on each. The cards represented the content taken from a local internet auctions website[3]. The information are organised on the website as shown in Table 1. We took contents from a local website which would most likely present the contents that are common in the studied group. The participants were asked to organise the cards in groups that made sense to them and were asked to write down a group name for specific groups of cards. The participants of the study were told that one card can be placed in one

[1] http://www.cs.washington.edu/research/edtech/CardSorts/

[2] http://www.optimalworkshop.com/

[3] http://www.lelong.com.my/

group and they can make as many groups as they would like to make.

Electronics and Appliances (7)	**Kitchen (7)**
White iron steamer	White oven glove
Home theatre system	Black non-stick paella pan
Karaoke system	Bamboo Chopsticks
Air conditioning	White Chop-board
Black VCR player	Kenwood Toaster
Calculator	Kenwood Hand mixer
Video camera	Bread maker
Electronics (7)	**Personal Accessories (6)**
Golden touchscreen-watch	Black micky mouse-necklace
White LED Clock	White Guitar Necklace
Black Analogue watch	Gold Locket
Metal Alarm Clock	Black sunglasses
Black Tablet PC	White Scratch proof-Bracelet
Apple iPad 2 white	Gold Swarovski Bracelet
iPod white	
Phone (5)	**Office (5)**
Black walkie-talkie	Stapler
Answering machine	Water dispenser
Pager	Window Curtains
VoIP-phone white	Computer desk
Fax machine	Hardcover file

Table 1. List of cards provided to the participants

Table 1 shows the list of items that were provided to the participants. All the cards were numbered randomly to be used for the analysis. Table 1 shows the items grouped as they are on the original auction website.

Participants

The participants were 38 undergraduate students at a Malaysian university. The call to participate in the study was advertised on the university notice board. We also applied the snowball method by asking each participant to recommend another participant for the study. Each participant received USD 5 to participate in the study. The study had twenty four (63%) female and fourteen (37%) male participants. Data were collected during summer 2011 following a standard protocol established for earlier study in Denmark and Pakistan [2].

RESULTS

Screening Criteria for selected Participants

During the screening of all participants to be used for the analysis of Edit Distance, AMM and BMM, we selected those participants whose classification was easily identifiable for top level group. This screening was conducted because each of tools could not handle multiple level groups. Among all participants, 40% of the participants (15 of 38) made only top-level groups, while the remaining 60 % (23 of 38) participants also made second level groups. We selected the 15 participants who made only top-level level groups because their data was easily identifiable for top-level groups. We further selected 10 participants from those who also made second level groups in one of the parent groups and treated all the cards in the second level groups under the parent group. We did not treat the cards of all those participants with second-level groups under the parent group because some of the participants made many second level groups under the parent group and treating all the cards in second level groups under the parent group would adversely affect the results. The selection of only 10 participants was conducted through qualitative analysis of the data to see that it would not affect the result by treating the cards in second-level groups under the parent group. With this screening we were left with 25 participants' data.

Analysis through Edit Distance

In order to evaluate the similarity and difference in the card sort data, we first used edit distance to see how much participants agreed with each other. On average, there were 14 moves taken to change one participant's sort into another participant's sort. Each of the sorts has a closest neighbour at a distance of 14 (SD $\pm$ 3.82) when comparing participants' sorts with each other. The analysis further shows that participants' disagreement was 38% from the original website sort.

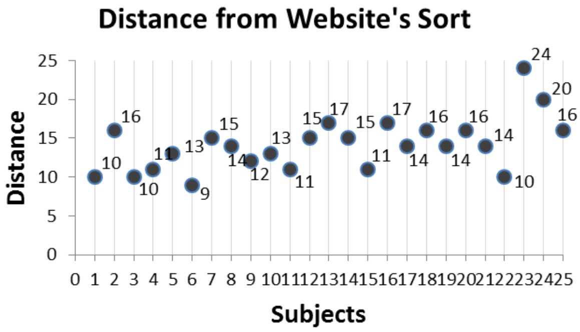


Figure 1. Distance of participants from original sort

Figure 1 shows the distance of the 25 participants' sorts from the original list as it was provided on the website and clearly shows that participants cluster the data considerably differently. The neighbourhood of edit distance provides a general understanding of whether the participants' sorts are close to the original sort as provided on the website. It can also provide information on how far apart each participant is from other participants. We performed analysis between the participants to see how close participants within the studied group were to one another.

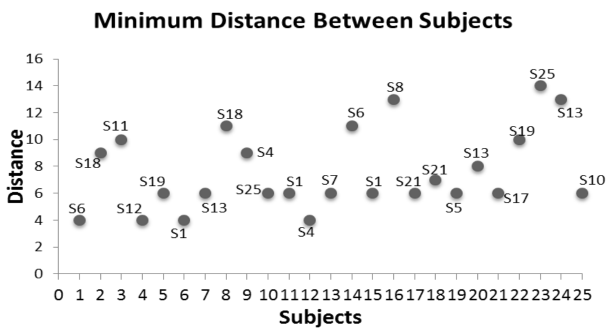


Figure 2. Minimum distance of the participants from another participant in the data

Figure 2 shows the minimal distances from each participant to another participant in terms of neighbourhood. Horizontal line shows each participant and vertical line shows the nearest participant in term of distance. The dots with annotations in figure 2, for example S6 stands for subject 6 or participant 6.It shows that participant one has a distance of 4 from participant 6 in the sort.

The analysis of neighbourhood shows each participant's closest participant of the study. On Average, each participant has a closest neighbour at a distance of 13 (SD $\pm$ 2.73). The average disagreement between the participants is quite high, calculated as 35%. The analysis between two participants about the closeness of a single participant with the nearest peer shows those participants who have similar way of clustering contents. On the other hands, it does not provide concrete information about the contents of the data which could be transformed into recommendations for the website structure.

Best Merge Method (BMM)

We performed an analysis using the best merge and actual merge methods. In both methods, we used a threshold of 60% agreement of items between participants, in keeping with Katsanos et al. [5]. A single card is included only if at least 60% of the participants have agreed to group it in the same group in their individual sorts.

Group 1 (7)	Group 2 (6)
Oven glove Bamboo Chopsticks Kenwood Toaster Non-stick paella pan Bread maker Kenwood Hand mixer Chopping-board	Sunglasses Guitar Necklace Locket Micky mouse necklace Scratch proof Bracelet Swarovski Bracelet
Group 3 (2)	**Group 4 (2)**
Apple iPad 2 white iPod white	Answering machine Fax machine
Group 5 (3)	**Group 6 (3)**
Hardcover file Calculator Stapler	Home theatre system Karaoke system Black VCR player
Group 7 (2)	**Group 8 (2)**
Touchscreen watch Analog watch	Metal Alarm Clock LED Clock
Group 9 (2)	
Computer desk Tablet PC	

Table 2. List of agreed groups by participant through BMM analysis

The group names for groups are suggested by participants and established by online tool1 used for AMM and BMM as: Group 1: *Gadget, Office and Entertainment;* Group 2: *Kitchen, home appliances, Kitchen items*; Group 3: *Accessories, personal accessories, Jewellery;* Group 4: *Clock, Others, Personal accessories;* Group 5: *Stationary, Office, and Office Appliance.*

The analysis of cards using BMM shows that for almost all of the cards (35 of 37, or 95%), participants agreed 60% and more for card placement in the same group. The participants clustered items in 5 groups with an average of 7 cards (M $\pm$ S) (7.0 $\pm$ 4.1) in a single group.

Table 2 shows the list of agreed groups by the participants through analysis of best merge method. The analysis of BMM shows that participant substantially agree and that there were only two cards on which the participants of the study did not agree 60% or higher.

Actual Merge Method (AMM)

The analysis of cards using AMM suggests that with an agreement of 60% for the cards where the cards have been grouped in the same group by all participants in their individual sorts, the participants do not agree greatly between each other and sort the items in groups with small numbers of cards. The participants agreed 60% and above for card placement in the same group for relatively fewer cards (29 of 37, or 78%) in comparison to the BMM (35 of 37, 95%). With an agreement of 60% and above about the grouping of cards in the similar group by individual participants and above for the cards, the AMM analysis showed that participants clustered the items in 9 groups with an average of 3 cards (M $\pm$ S) (3.2 $\pm$ 1.9) to a group.

Group 1 (13)	Group 2 (9)
Video camera Tablet PC Pager Computer desk VCR player Home theatre system Karaoke system Answering machine Fax machine VoIP-phone white walkie-talkie Apple iPad 2 white iPod white	Iron steamer Water dispenser Chopping-board White oven glove Bamboo Chopsticks Non-stick paella pan Kenwood Hand mixer Kenwood Toaster Bread maker
Group 3 (6)	**Group 4 (4)**
Sunglasses Guitar Necklace Gold Locket Scratch proof Bracelet Micky mouse necklace Swarovski Bracelet	Touchscreen watch Analog watch Metal Alarm Clock LED Clock
Group 5 (3)	
Calculator Hardcover file Stapler	

Table 3. List of agreed groups by participant through AMM analysis

The analysis of AMM indicates that participants did not agree as substantially with each other. The participants grouped most of the cards in fragmented groups and the agreement for 8 of 37 cards (Video camera, VoIP phone, Walkie-talkie, Pager, Window Curtain, Water dispenser, Air conditioner, iron Steamer) was below 60%.

Table 3 shows the list of agreed groups by the participants through analysis of actual merge method. The group names for groups are suggested by participants and established by online tool[1] as: Group 1: *Kitchen appliance, Kitchen Household, Kitchen*; Group 2: *Jewellery, Gold Accessory, Accessory* ; Group 3: *Apple*; Group 4: *Machine¸* Group 5: *Stationary, Office items, Personal;* Group 6: *Entertainment, Electrical Equipment, Living room appliances;* Group 7: *Watch;* Group 8: *Clock;* Group 9: *Study room, Computer Laptop*In analysing the same data with is used with two different methods (AMM and BMM). The results show different interpretation of how participants grouped items. Appendix A1 shows the comparison of AMM and BMM. The results show the variation in the agreement and the outcome of the users' sorts for website structure.

DISCUSSION

The comparison of card sorting analysis techniques revealed how the choice of technique can have an impact on the resulting structuring suggestions for a website. It also revealed that different techniques not only highlighted the different aspects of the data, but also confused the results for taking action and implementing a website structure. Secondly, in card sorting analysis, the eventual design of a website structure depends a great deal on the basis of structures created by the participants in a study. When different tools are used to analyse the data, the limitation created by the tools may potentially obscure or confuse insight into the users' sorts.

As the results of studies show, users tend to place information in different orders. Therefore the information classification on a websites should match the local users' way of perception for information classification. From usability aspect of website structure on the basis of card sorting, the structure of the websites should not merely come from the analysis of card sorting, but should be evaluated by subsequent usability testing.

The result of three techniques reveals that edit distance is slightly different from AMM and BMM. Edit distance provides a measure of the dissimilarity between all sorts and shows which of the sorts are the closest when compared with all other users. It points towards those user(s) who are central in card sorts, having the most in common with others. The analysis of AMM and BMM shows that it combines multiple card sorts into an aggregated card sort and approximates an agreement between different users on each card of the card sorting study.

The analysis of edit distance presented the data of users in two ways: a) comparing each of the users' sorts with other users' sorts and b) comparing each user's sorts with the original website's sort. To compare each of the users' sorts with other users' sorts provided an understanding of how some users were close to each other in terms of their mental models of the structure of the website. This information also highlighted the dissimilarity of users in their structuring approach and their disagreement as a whole but it did not highlighted each card's agreement level by the users as it could have been done through the visualisation of AMM and BMM. The neighbourhood did not provide the level of agreement of each item in the sort between different users of the study; it only showed the general level of agreement between users. Comparing users according to edit distance thus provided a general picture of the level of disagreement between users.

The minimum distance of a user from another user in the data indicated their level of agreement or disagreement, indicating the closeness of each user's sort. The analysis of edit distance was useful in understanding the impression of what extent the users were different in their structure from each other and to what extent the users were different in their structure from website structure. However, the interpretation of the results was difficult transform into a meaningful recommendation. The meaningful recommendation could not easily be determined because analysis of edit distance did not provide the contextual understanding of the result for each card. It was therefore difficult to translate information into meaningful representations which could be used to make decisions concerning website structure. Edit distance does not produce an aggregate categorization on the basis of multiple categorizations, but rather focuses more on the distances between users and websites.

The best merge method (BMM) looked for pairs in each user's sort and finally added up these pairs of sorts. The major issue with BMM was that it required reducing groups into pairs. If a user grouped [a, b, c] together, then BMM recognised it as if the user had placed [a, b] [a, c] and [b, c] together (i.e. 3 pairs). These pairs were then added up. This result was a fundamental loss of information, because once these pairs were added together, it became impossible to reconstruct the original data.

The actual merge method (AMM) looked for agreements in whole groups, rather than pairs, which made the natural disagreement promising when comparing it with BMM. AMM improved the result; it did not take the pair but it considered grouping together and showed it into a single group.

When comparing AMM with BMM, AMM not only take the pair but it also considered more than two items grouped together and showed into the group. On Surface, AMM did not show promising agreement between the users about their level of agreement. In reality, AMM provided a better picture of how users of the study agreed between each other and to what levels their agreement changed for each card because it not only looked for pairs but it considered more than two cards as a group if they were similar across different users' sorts.

The effect of threshold on BMM and AMM

The threshold of 60% agreement between users appeared to be an important factor in the resulting structure. The threshold of 60% and above explained that a single card was included only if at least 60% of the users agreed to group it in the same group in their individual categorizations. We wanted to see how the level of threshold affects the number of groups and average cards in a group if the threshold is changed in AMM and BMM. We also wanted to see at what level of threshold the structure produced and cards used by AMM becomes similar to the BMM.

	BMM	BMM	AMM	AMM
Threshold	60%	50%	60%	50%
Number of Groups	4	4	9	8
Avg. Cards in a Group	9	9	3	4
Card used	35 of 37	35 of 37	29 of 37	35 of 37

Table 4. A comparison of threshold of BMM to AMM

Table 4 shows the comparison of AMM and BMM at different thresholds and its impact on number of groups and average cards in a group. It shows that a decrease in threshold to 50% has no impact on BMM. Decreasing in the threshold for AMM to 50% changes in the number of groups and average cards contained in a group, although not greatly. In other words, even if the level of agreement between the users is decreased from 60% to 50%, it does not impact on the number of groups and average number of cards for BMM and AMM. It changes slightly for AMM when the threshold is decreased but it does not become equal to the number of groups for BMM if the threshold is decreased.

The Effect of Number of Users on Agreement

This study used 25 users in its investigation. The number of user may also have had an effect on the agreement levels, so we selected 5, 10, 15, and 20 users at random to generate AMM and BMM groupings and compared them to groupings generated on the basis of the data from all 25 users.

For BMM, a random subset of 20 and 15 users subsequently generated 4 groups with an average of 8.75 cards in a group, which is very close to the results achieved with 25 users. By selecting 10 users, the number of groups increased from 4 to 5. This suggests that in order to use results generated by BMM, the recruited users should be more than 15 to generate stabilized results.

For AMM, the random subset of 20 users generated 7 groups which varied from the groups generated by 25 users. By selecting 15 random users, the AMM results that the users made 6 groups. By selecting 10 random users, 8 groups were generated. This attentively indicates that in order to rely on results of AMM, more than 25 users are required for the study to generate stable results. This argument is aligned with the statement mentioned on the website of online tool providing company OptimalSort[4] which says that AMM is recommended for more than 30 participants and BMM is recommended if fewer than 30 participants are available.

Comparison of three analyses

Contrasting the three analyses, it seems that AMM provided a better understanding of the groupings determined by participants, which could be transformed into meaningful steps for a website's structure. The information visualisation of AMM and BMM provided a better understanding of the AMM and BMM analysis. The edit distance helped to understand the subjective distance of users from each other, although the information was difficult to leverage for specific decisions concerning structure.

When choosing between AMM and BMM, AMM seems to produce a larger number of groups in comparison to BMM. Such a difference in the number of groups explains the methodological issues with the choice of analysis for card sort. Appendix A1 shows the comparative scheme of AMM with BMM.

One of the implications of this study is that it is important to understand the methodological differences in each of the analysis when using them to construct website structures. Studies may have different requirements and this can affect the choice of analysis for card sorting. Researchers and practitioners need to conduct different analysis for card sorting. This would provide an overview of how these techniques and analyses of these techniques shape the results. The study indicates that information structure of the websites should also be evaluated by subsequent usability testing.

In one of the limitation of the study, the data for multiple level groups could not be handled by these analyses. Each of the three techniques could not deal with information in multiple groups. By not selecting the second level groups we introduced a fundamental loss of information which will have changed the outcome of the study. Tools to handle multiple level groups do not currently exist and therefore used here appear to be among the most suitable techniques available.

CONCLUSION

This study shows how the choice of analysis technique for a card sorting study can impact on the resulting information structure for a website by analysing the same data according to three techniques. It also suggests that the choice of analysis for card sorting has consequences for website designs because the agreement level for different methods varies for the same data and different method suggests different structures for web content. Finally, it also reveals that agreement levels for similar data changes if a different analysis for the same data is conducted. The study indicates that information structure of the websites should not merely come from the analysis of card sorting, but should be evaluated by subsequent usability testing.

The study concludes that it is important to understand the methodological issues for card sorts analysing tools. Card sorting tools have a great potential to use and understand users mental models because it can help to understand remote users view of information classification. However,

[4] http://www.optimalworkshop.com/help/

these benefits will only be realized if the card sorting applications visualization of analyses is understood by researchers and practitioners.

ACKNOWLEDGMENTS

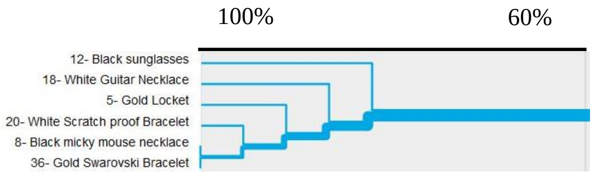
I would like to thank Morten Hertzum from Roskilde University, Denmark and Torkil Clemmensen from Copenhagen Business School, Demark whose valuable feedback made it possible to write this article. I would like to thank Dr. Alvin Yeo from University Malaysia Sarawak (UNIMAS) for his support in collecting data for this study. I would like to thank the participants in the study from the University of Malaysia Sarawak in Malaysia. Finally I would like to thank Optimal Workshop for providing the licencing to analyse the data for study

REFERENCES

1. Cairns, P. and A. Cox, *Research methods for human-computer interaction*2008: Cambridge University Press New York, NY, USA.
2. Nawaz, A., T. Clemmensen, and M. Hertzum. *Information Classification on University Websites: A Cross-Country Card Sort Study*. in *Information Systems Research Seminar in Scandinavia (IRIS)*. 2011. Turku.
3. Chaparro, B.S., V.D. Hinkle, and S.K. Riley, *The usability of computerized card sorting: A comparison of three applications by researchers and end users.* Journal of usability Studies, 2008. **4**(1): p. 31-48.
4. Deibel, K. and R. Anderson, *Using edit distance to analyze card sorts.* Expert Systems, 2005. **22**(3): p. 129-138.
5. Katsanos, C., N. Tselios, and N. Avouris, *Automated semantic elaboration of web site information architecture.* Interact. Comput., 2008. **20**(6): p. 535-544.
6. Petrie, H., et al., *Using card sorts for understanding website information architectures: technological, methodological and cultural issues*, in *INTERACT*2011, Springer: Lisbon, Portugal.
7. Curran, M., G. Rugg, and S. Corr, *Attitudes to expert systems: A card sort study.* The Foot, 2005. **15**(4): p. 190-197.
8. Capra, M.G. *Factor Analysis of Card Sort Data: An Alternative to Hierarchical Cluster Analysis*. in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2005
9. Shieh, J.C. and C.F. Liu, *The usability evaluation study of the university library websites.* Journal of Educational Media and Library Science, 2010. **47**(2): p. 163-198.
10. Allen, R., *Cognitive factors in the use of menus and trees: An experiment.* IEEE Journal on Selected Areas in Communications, 2002. **1**(2): p. 333-336.
11. Broughton, V., *Faceted classification as a basis for knowledge organization in a digital environment: the bliss bibliographic classification as a model for vocabulary management and the creation of multidimensional knowledge structures.* The New Review of Hypermedia and Multimedia, 2002. **7**(1): p. 67-102.
12. Morville, P. and L. Rosenfeld, *Information architecture for the world wide web*2006: O'Reilly Media, Inc.
13. Lewis, K. and P. Hepburn, *Open card sorting and factor analysis: a usability case study.* Electronic Library, The, 2010. **28**(3): p. 401-416.
14. Kralisch, A., A.W. Yeo, and N. Jali, *Linguistic and Cultural Differences in Information Categorization and Their Impact on Website Use*, in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 05*2006, IEEE Computer Society. p. 93.2.
15. Martine, G. and G. Rugg, *That site looks 88.46% familiar: quantifying similarity of Web page design.* Expert Systems, 2005. **22**(3): p. 115-120.
16. Johnson, D.E., *Applied multivariate methods for data analysts*1998: Duxbury Press Pacific Grove, CA.
17. Everitt, B., *Cluster analysis*1974: Wiley (New York).
18. Capra, M.G., *Factor Analysis of Card Sort Data: An Alternative to Hierarchical Cluster Analysis.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2005. **49**(5): p. 691-695.

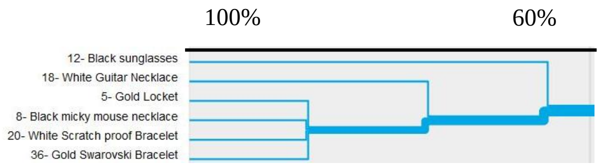
Appendix A

Best Merge Method (BMM)

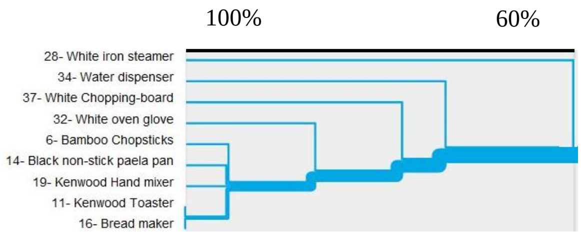


Grouping names: Accessories, Personal Accessories, Personal

Actual Merge Method (AMM)

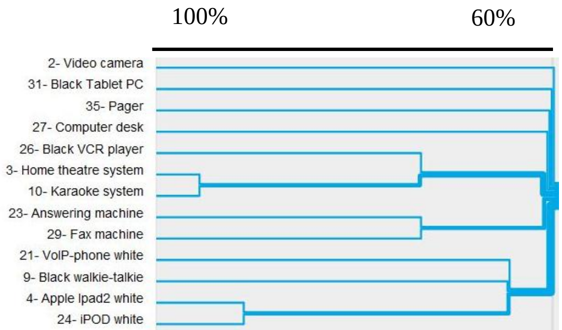


Grouping names: Accessories, Jewellery, gold Accessories

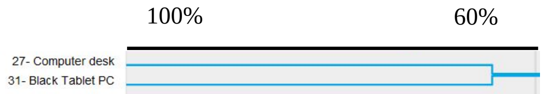


Grouping names: Kitchen, home, home items, communication, electrical, gadget

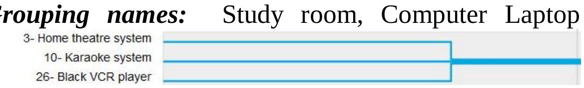
Grouping names: Mixer, Kitchen appliances, kitchen households



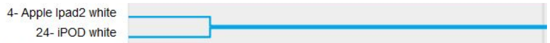
Grouping names: communication, electrical, gadget



Grouping names: Study room, Computer Laptop



Grouping names: Entertainment, Electrical Equipment, Living room appliances



Grouping names: Apple

Grouping names: Machine

A.1. A comparison between the results of Actual Merge method and Best Merge method