UNIVERSITEIT TWENTE

Predictive value of card sorting on browsing performance and user satisfaction

Jan Sommer s1024833

8/24/2012

Supervisors:

Dr. Martin Schmettow

Dr. Matthijs Noordzij

Department:

GW-HFM

Abstract

During the last decennia website usability and user experience has become more and more important as people spend increasing time-spans on the internet. A wide range of information and services have become available digitally, which were formerly only provided physically. One such example is E-Government. In order to create usable and satisfying websites card sorting has often been used as a tool to evaluate website usability and inform website design. However, it has never been established how good a categorization task - e.g. card sorting- actually mirrors a searching task, which is what most people do browsing the web. In order to establish how good card sorting predicts browsing performance and user satisfaction a 50 participant experiment is conducted. Hereby, participants searched 5 different items on 5 different municipal web sites; one per site. Usage of the search function was not permitted. Selected items were chosen because of their wide range of difference to items in the mental model of users, as established by an earlier card sort. Items that were clustered like the mental model exhibited a small distance, while items which were clustered with other items than in the mental model presented a large distance. Repeated measures analysis showed that distance to the mental model is a bad, if any predictor of browsing performance as measured by time and clicks/path length until finding the target item. Additionally, it has no predictive value on user satisfaction. These results suggest that the method of card sorting has to be questioned in the context of website evaluation and design.

Contents

1.Introduction

Card sorting is a tool that helps us understand how people/ users structure knowledge and which categorization they would expect. Thereby, items are grouped into categories to find these underlying models of hierarchies and structures. Hereby, a variety of methods can be applied, most notably open or closed card-sorts. In open card-sorts participants are asked to group items into undefined groups whereas in closed card-sorts these categories are defined beforehand and participants are asked to fit items into them. Moreover, card sorting can be administered either physically or online as both modi seem to wield the same results (Bussolon et al.,2006). Furthermore, repeated sorts can be conducted in which participants are prompted to resort the items into different groups using another criterion. This can be repeated until the subjects are unable to sort any further. Either way, card sorting has come in handy in a lot of applications including psychiatry and clinical psychology (Heaton et al., 1993), sports and pedagogy (Reinhold, 1993), software development (e.g. McDonald et al., 1986), knowledge management (Shadbolt & Burton, 1990), market research (Dubois, 1949) or social development (Davies, 1996). It has also shown to be a simple and easily administered tool for finding out about the underlying mental models which users have of web sites (Gaffrey, 2010) It has generally been assumed that if a web site mirrors the mental model of users as approximated by card sorting it is also usable. Therefore, it has been repeatedly used as a tool in Usability tests:

In 1999, the MIT Libraries conducted an open card sort of their web site content in which students had to sort 106 cards. Thereafter a reverse category survey was conducted in which participants were asked in which out of five categories they would suspect to find a specific item from the web site. The results of these surveys were then used to group and label content, thereby redesigning the web site (Hennig, 2001).

In another study dating back to 1999 the Battleson and Weintrop conducted a card sort in which nine participants were asked to group 34 cards which depicted tasks commonly performed using the University of Buffalo's library web site. This study focused almost entirely on nomenclature rather than organization (Battleson & Weintrop, 2000).

In contrast to that the Cornell University used a card sorting study of its library web site with focus lying more on the organization of its help topics than on terminology. academic libraries to use the card sorting technique. This was done testing 12 users with 50 cards,

whose content was confined to the help section of the library. It was found that card sorting was a highly effective and valuable method for gathering user input on organizational groupings prior to total system design (Faiks & Hyland, 2000).

Later studies have often used card sorting as an evaluation tool for whole web sites and often these card sorts were part of iterative usability studies containing a variety of other methods of assessment such as focus groups, questionnaire surveys, heuristic evaluation, observation testing and label intuitiveness/category membership testing (e.g. Ebenezer, 2003; Turnbow et al., 2005).

Research on Usability and User Experience has grown in importance over the course of the last decennium. This is mainly because a lot of applications and services that were for formerly provided physically have become more and more digitalized and are nowadays often offered online. The question of how to test Usability and User experience as well as the question of limitations and definition of these concepts and tests has therefore risen in importance proportionally. Usability is generally understood as "the capability to be used by humans easily and effectively" (Shakel, 1991), "quality in use" (Bevan, 1995) or "the effectiveness, efficiency and satisfaction with which specific users can achieve goals in particular environments"(ISO, 1998). In his review of 180 Usability studies Hornbæk (2006) describes a variety of measures for assessing usability used in these studies. These fit into the categories of effectiveness, efficiency and satisfaction.

A concept closely related to satisfaction is user experience. It focuses on the hedonistic qualities of interactive products such as affect and aesthetics (Bargas-Avila & Hornbæk, 2006). Therefore it is often understood distinct from usability in that user experience is related to non-instrumental and hedonistic goals, whereas usability is associated with instrumental, task-oriented goals (Hassenzahl & Tractinsky, 2006). However, in practice non-instrumental and instrumental goals are often interwoven. For example did participants in one study which researched experience with riot!-an interactive play for voices- sometimes impose instrumental goals on themselves (Blythe et al., 2006). Note, that Hornbæk (2006) also added satisfaction to his list of measures of usability, while it is a prominent theme of User experience. Hassenzahl et al. (2000) noted that both hedonistic and ergonomic qualities can be assessed independently. However, they almost equally contribute to the appeal of the tested software prototypes.

Following these results Hassenzahl et al. (2003) invented a research tool which unifies User experience and Usability by measuring pragmatic as well as hedonistic qualities in interactive products named AttrakDiff. AttrakDiff measures four distinct dimensions of product evaluation: Pragmatic quality, hedonistic quality-identity, hedonistic quality-stimulation and attractiveness. Pragmatic quality refers to the usability of a product and is an indication of how good users were able to find their objectives. Hedonistic quality-Identity describes to what extend the web site lets the user identify with it. Hedonistic quality- Stimulation describes how novel, interesting and stimulating a web site's functions, content and presentation styles are. Attractiveness is a measure of perceived quality of a web site.

Furthermore did Katz and Byrne (2006) find that 60-90% of users use the menu and not the search function for finding information. Therefore a good hierarchical navigation structure seems to be a key feature in usability. A fitting hierarchy might be assessed by card sorting.

However, until today, card sorting studies have failed to assess how good a sorting task actually mirrors a finding task, which is what one typically does while browsing the internet/a web site. It therefore remains unclear how much value card sorting has as a basis for creating an information architecture (Hawley, 2008). Wood and Wood (2008) named two aspects which highlight the difference between sorting and finding; firstly, when searching for an item one has a specific need to fill (e.g. a mental description) which one wants to fill by finding something on the internet. Secondly, in finding tasks one does not have the entire context of all items that might be considered in grouping them. Another aspect, which might differentiate sorting from searching is path complexity. Research has shown that navigation path complexity plays an important role when measuring navigation performance (Melguizo et al., 2012). Navigation path complexity is a concept defined by Gwizdka and Spence (2006). They proposed that one could assess the concept by breaking it up into three components:

1. *Page complexity*: This is the complexity of navigation choices on each web site. It includes links on a webpage or its visual design.

2. *Page information assessment*: This is the degree of difficulty to judge the relevance of the information on a web site related to the goal information.

3. *Navigation path length:* This is the length of a navigation path. The longer a navigation task, the more navigation choices one has to make. If more navigation choices have to be made, more relevance judgments have to be made during search-tasks, which affects

navigation performance as measured by time performance, accuracy and lostness (Puerta Melguizo et al., 2006). These components, however, do not play a role when sorting items to fit a hierarchy, so that highly frequented items might end up in a long navigation path while sorted. Page complexity, too, is not especially considered when sorting items, which might then lead to a complex and even confusing navigation structure.

Because of all these differences it would be highly interesting to research if fit of a web site to a card sort can actually predict user performance and satisfaction in finding categorized items. If that were to hold true card sorting could provide an excellent tool for web site evaluation and redesign for various reasons. One of the most salient is surely its simplicity of administration: items in a stimulus set are given to a participant, who then sorts them into groups. These items can be given in any form, such as words, pictures or objects. Using other representations than words, cross-cultural evaluation of web sites could be easily conducted. In addition, card sorting requires no domain specific knowledge as it places no special cognitive burdens on research participants, such as time pressure or memory limitations (Fincher & Tenenberg, 2005). Furthermore, card sorting can elicit some of the semi-tacit knowledge that traditional interviews and questionnaires fail to access (Upchurch et al. ,2001). Besides, even a small sample of test subjects has shown to already wield significant results; that is it does not matter much if one lets 50 or 500 subjects sort cards (Wood & Wood, 2008).

Because of all these advantages card sorting offers, my research will focus on a few simple questions in assessing the evaluative quality of it. The first one being: Does card sorting have predictive value for finding items on web sites?

This question aims at how well a web site fits the mental model of users and if there is a correlation between performance of users and degree of match between the web site /search task and this mental model. The mental model of users is taken from a study assessing student's mental model of municipal web sites (Kowoll, 2010). In her study Kowoll let 27 participants categorize 70 items found on Enschede.nl. This categorization ultimately generated a matrix of Jaccard distances between all 70 items. Although common sense would expect a relationship between fit of a web site to this mental model and browsing performance an in-depth analysis is essential as categorization and search tasks could rely on completely different brain regions and cognitive functions thereby having no correlation at all. Browsing

performance is measured as time till completion and the quotient of dividing actual clicks done until finding the item by the minimal number of clicks necessary to find it.

The second question is if user experience correlates to the degree of match between web site and mental model. It is important to notice that user performance and satisfaction do not always correlate or co-occur. Sometimes users feel satisfied although web site design is complicated and sometimes users perform well but do not feel satisfied at all (Tullis & Albert, 2008). Not always is performance the most important variable in web site usage. For example, when choosing a travel-planning web site, users are strongly influenced by their satisfaction of a web site, while some more clicks (as a measure of performance) are quickly forgotten. (Tullis & Albert, 2008). One can easily imagine that satisfaction may be crucial for municipal web sites, too, as they aim at representing and advertising specific cities and informing their citizens.

Designing efficient municipal websites is of great importance and will become even more important as the total population of the world further rises as a means of not only informing citizens, but also handling requests and providing services which formerly had to be offered physically. This is called E-Government. Its main applications are government to citizen (G2C), government to business (G2B) and government to government (G2G) interactions. Thereby it mostly acts on two dimensions: informational and transactional. On the informational dimension the government informs about a variety of relevant facts, such as parking charges, locations of governmental buildings or parks. On the transactional dimension online forms can be filled in for e.g. filing a complaint, paying taxes or even voting online (e-voting). Through self-service and online-interaction these governmental processes can be streamlined. Prior work suggests that people prefer this sort of self-service over conventional, face-to-face service partly due to the considerable savings in time and effort, increased personal control and ease of use (Meuter et al., 2000). However, currently E-Government faces a variety of problems and limitations. For example, citizens may desire a high level of security for transactional E-Government services, but may subvert the use of security mechanisms due to its complexity (Princeton Survey Research Associates, 2002). This is mirrored in the fact that about 98% of the countries in the world have developed government Web sites, with less than one-third providing transactional services, such as online form submission (United Nations, 2010). Moreover, public satisfaction with E-Government is poor (Liu & Hu, 2012). Further research found that the usability of governmental websites in the UK - the second highest scoring country in the UN's global E-Government readiness report

2004- was generally quiet low (Ma & Zaphiris, 2003) . It is therefore essential to identify effective methods to design usable websites in order to improve E-Government and thereby government-citizen interaction overall. As mentioned above, card-sorting might turn out to be such a method.

2.Methods

2.1.Participants

Fifty students (24 female, 26 male) from the University of Twente in Enschede participated in this experiment. 49 of the participants were bachelor students of psychology or communication studies, with one being a master student (age range 18-26 , mean age 21,5). Furthermore, 19 of them were Dutch, whereas 31 were of German origin. Most of them were recruited through the participant-pool of the University which makes it mandatory for every bachelor student to participate in 15 hours worth of research. The rest was recruited word-of-mouth advertising. All participants showed reasonable command of the Dutch language as every student as every foreign student had to take a Dutch exam at the NT2 level 5 before being admitted to study at the UT. Furthermore, all participants filled in a form of informed consent.

2.2.Material and design

To create search tasks the 70 different items used by Kowoll (2010) were searched for on five municipal web sites (Enschede.nl, Nijmegen.nl, Utrecht.nl, DenHaag.nl, Amsterdam.nl). The web sites were selected because of their differing menu-structure, indicating a different web site structure. In order to later compare the mental model of municipal web sites with the web sites themselves the underlying structure of the five web sites with regard to the 70 items was analyzed via hierarchical cluster analysis. Only 36 of the 70 different items were present on every web site. Out of the 36 five items (taxation (belastingen), data of the population (bevolkingscijfers), parking licence (parkeervergunning), get married (trouwen) and shopping hours (winkeltijden)) were chosen, which varied the most in difference between web site design and mental model as established by card sorting. Dendrograms and heatmaps of both the mental model and the design model of the web sites helped in guiding the search for fitting items (see Appendix 4). The mental model was defined as the ideal structure of a municipal web site as measured by the open card sort of Kowoll (2010). An open card sort was be used as the categories established in closed sorts might not reflect actual categories that users would establish. Search tasks ranged from fitting the mental model nearly perfectly

to not fitting it at all. Differences between Items on web sites and the mental model were defined as the sum of the squared difference between the Jaccard distances each item has to every other in the mental model and the Jaccard distances each item has to every other item on the 5 web sites. This will be referred to as squared difference during the further course of this paper. Jaccard distances between items were established by the formula $J_\delta(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$ in order to establish a matrix of jaccard distances between all items. The sum of squared differences to the mental model was computed by the formula $\mathrm{SSD}(A_m, A_d) = \sum_{k=1}^{36}[(J_{\delta m}(A,B_k) - J_{\delta d}(A,B_k))^2]$, where $J_{\delta m}(A,B_k)$ is the Jaccard distance of a particular item A to another item B in the mental model and $J_{\delta d}(A,B_k)$ is the Jaccard distance of the same item A to the same particular item B in the design model. A low squared difference means that an item on a web site is nearly as far away to other items as it is in the mental model. If e.g. the item "tariffs (parkeertarieven)" on a web site fits the mental model that would mean that it is for example near "parking licence (parkeervergunning)" and "disabled parking (invalidenparkeeren)" but far away from "voluntary work (vrijwilligerswerk)" en "employment finding (arbeidsbemiddeling)". If the item "tariffs (parkeertarieven)" was nowhere to be found in the periphery of "parking licence (parkeervergunning)", "parking licence (parkeervergunning)" and "disabled parking (invalidenparkeeren)" it would have a large measure of distance. In that way, one difference was established for every item on every municipal web site, which made a total of 25 different tasks.

A within-subject design was used with five levels (each web site) and web site-design-difference as independent and performance time, clicks relative to the optimal path and User experience as dependent within-subject variable. Hereby, clicks were measured with help of the Morae Recorder ver.3.3 by TechSmith Corporation. The Morae Recorder is a tool, which lets one record everything that happens on the screen of a computer during a predefined time or task. Performance time was measured using a stopwatch. User experience was measured with AttrakDiff, a online Likert-scale based tool for product evaluation. Due to restrictions of this tool a two 20 participant study was conducted. Five different item-finding conditions were used so that no participant had to search for the same item twice, but all items would be searched for across all web sites equally often. Hindering repeated search of one item across more than 1 web site was done in order to ensure there were no learning effects between tasks. Participants were randomly assigned to the five conditions.

2.3.Procedure

At the beginning of each experiment the participant was asked to fill in the informed consent form. Thereafter a precise verbal explanation was given. Participants were told that they would have to search 5 different items on 5 different municipal web sites; one per site. They had to do it relying only mouse clicks, thus not using the obligatory search function every site has. This was done as I wanted to compare the hierarchical structures and as noted above 60-90% of users navigate through the menu rather than the search function (Katz & Bryne, 2006). Therefore searching without usage of the search-function seemed reasonable. Afterwards participants were given an example by the researcher ("Look for Infrastructure on Enschede.nl") to familiarize with the sort of task. Following the example the search tasks were given one after another. The search tasks included a short description of the item to be found. These were given in Dutch and English. An example of a description is

"Openings- en sluitingstijd van winkel, Openingstijden van winkelcentra, Shopping hours, opening hours" for "Shopping hours (winkeltijden)" (for further information about the items, see Appendix 2). If an Item was not found within 4,5 minutes time was stopped and the task was aborted due to temporal restrictions of the experiment. Furthermore, after completing tasks ,which had the lowest resp. the highest squared difference to the mental model in their condition, they were told to fill in the AttrakDiff list. 20 participants had to fill in the 2 web site evaluations. These evaluations were later used to see if there was any difference in rating between web sites that contain tasks that deviate little or much from the mental model. Familiarity with every web site was checked verbally before every experiment to prevent preexisting knowledge from interfering with performance results. None of the participants showed preexisting knowledge, but 6 participants said they had been to Enschede.nl before but were not really familiar with it or every looked something specific up on it. After the experiment participants were debriefed about the purpose of the research if they showed interest in it.

3.Results

3.1. Finding time

Finding time was measured as the time in seconds needed to find an item on a web site.

A repeated-measures analyses with 10 repeated measures over the 25 different tasks was conducted with the measures of difference of each combination and optimal path length (as measured in minimal clicks necessary to reach the goal information) as covariate. There was no main effect of the squared difference to the mental model to be found with F(1,23)= 0.72, p= 0.41, partial eta squared= 0.031. However, there was a significant effect found for length of the optimal path with F(1,23)= 6.7, p= 0.017, partial eta squared= 0.233.

The partial eta squared of resp. 0.031 and 0.233 means that the factor squared difference to the mental model by itself accounted for only 3% of the overall (effect+error) variance whereas the path lengths accounts for 23% of it.
Furthermore the post-hoc power analysis for difference to the mental model exhibited a quite low value, indicating that the possibility of rejecting the null hypothesis when the null hypothesis is false is also quite low (12,6%).

Table 1. Effects of squared difference to the mental model and path length on time until finding goal information

Tests of Between-Subjects Effects

Measure:Time

Transformed Variable:Average

Source	Type III Sum of Squareds	df	Mean Squared	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	58129,718	1	58129,718	2,138	,158	,089	2,138	,288
Difference	19085,457	1	19085,457	,702	,411	,031	,702	,126
Path length	182112,421	1	182112,421	6,699	,017	,233	6,699	,696
Error	598066,559	22	27184,844					

a. Computed using alpha = ,05

Mean times of the different tasks ranged from 13,5 seconds (for finding “taxation (belastingen)” on Amsterdam.nl, squared difference = 0,76) to 251 seconds (for finding “data of the population (bevolkingscijfers)” on DenHaag.nl, squared difference = 0,47). Additionally, one notices that out of 25 tasks 11 have a range of over 200 seconds, whereas only 3 combinations have a range of less than 100 seconds. This means that there was a huge difference in how good participants could find most items. One exception is “taxation (belastingen)” on Amsterdam.nl with a quite low range of 13, ranging from 6 to 18 seconds.

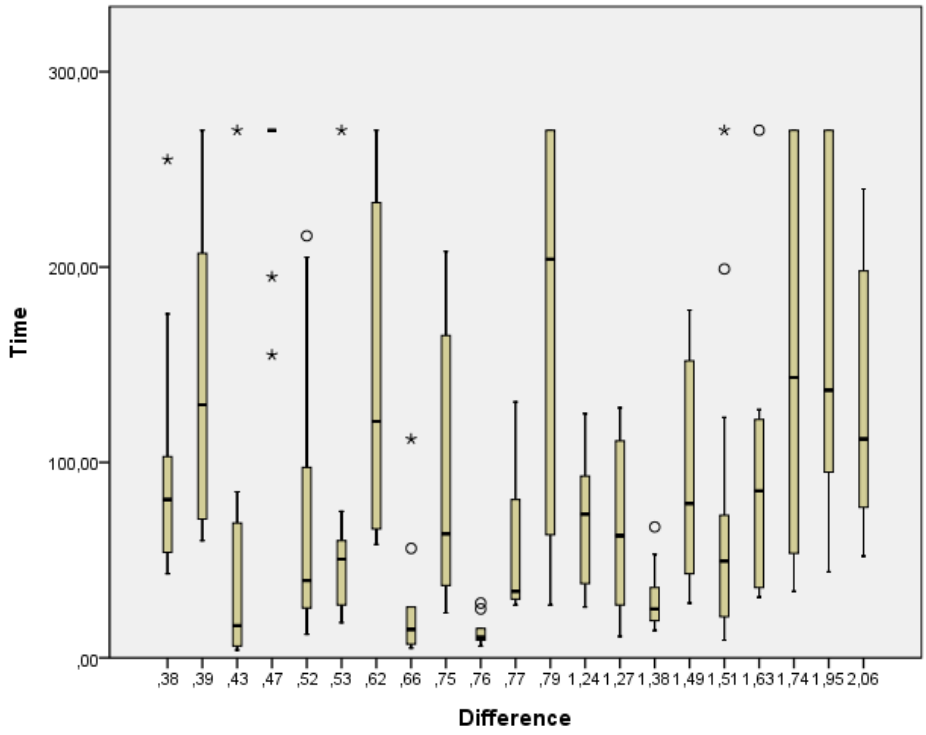


Figure 1. Time until finding goal information as related to squared difference to the mental model

Looking at figure 1 one notices that there are a lot of mild and extreme outliers which reflect the wide range of time scores for many tasks. Most notably hereby are the 2 extreme outliers of “data of the population (bevolkingscijfers)” on DenHaag.nl, squared difference to the mental model= 0,47. These are the only 2 scores which differ from 270 seconds on that task. The box-plot further demonstrates the wide range between the 25st and the 75st percentile often found in the time scores. Moreover, the maximal time scores for four different tasks lie within these percentiles (270 seconds for each task). These are “get married (trouwen)” on Enschede.nl (0,79 squared difference), “data of the population (bevolkingscijfers)” on DenHaag.nl (0,79 squared difference), “data of the population (bevolkingscijfers)” on Utrecht.nl (1,95 squared difference) and “shopping hours (winkeltijden)” on Enschede.nl (1,74 squared difference) (s. figure 1). This shows that a lot of participants could not find the information goals within 4,5 minutes.

Pairwise comparison of all items showed significantly higher values for "data of the population (bevolkingscijfers)" than for all other items ($p<0.05$) (s. figure 2). Moreover, pairwise comparison revealed significantly lower values for "taxation (belastingen)" than for all other items but “get married (trouwen)” ($p<0.05$). Further comparison showed that finding time values differed significantly between German and Dutch students ($p<0.05$).

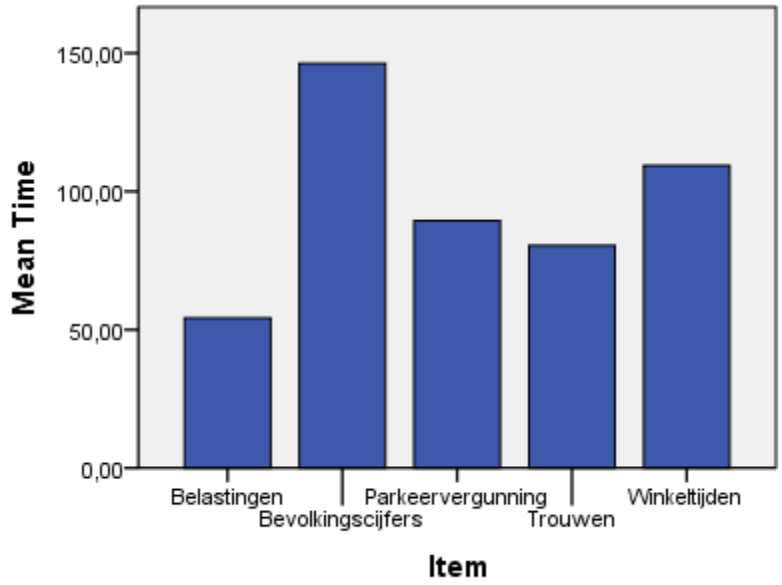


Figure 2. Mean time until finding goal information per item

Pairwise comparison of all tasks indicated that finding-time values were significantly higher for "data of the population (bevolkingscijfers)" on DenHaag.nl (0.47 squared difference) than for all other tasks (p=0.01) (s. figure 3).

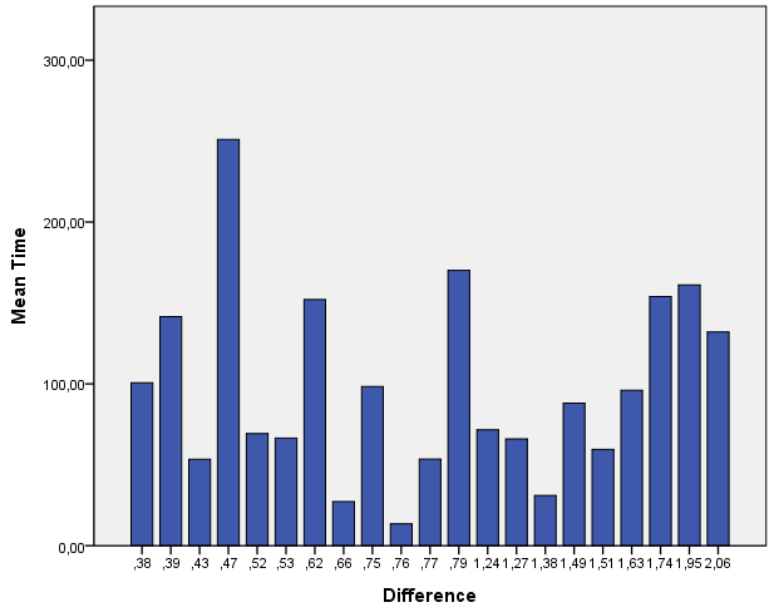


Figure 3. Mean time until finding goal information per squared difference to the mental model

3.2. Clicks

Clicks were measured as a quotient of actual clicks done until finding the item and the minimal number of clicks necessary to find it (the optimal path). A repeated-measures analysis with 10 repeated measures over the 25 different tasks was conducted with the measures of difference of each combination as covariate. There was no main effect of difference to the mental model to be found with F(1,23)= 2.69, p= 0.12, parietal eta squared = 0.105. The partial Eta squared being just .105 means that the squared difference to the mental model by itself accounted for only 10,5% of the overall (effect + error) variance. Post-hoc power analysis showed an observed power of 0.349 which means that the possibility of rejecting the null hypothesis when the null hypothesis is false is 34.9%. This, however, is no problem in this case as the null hypothesis is rejected due to insignificance of the squared difference to the mental model either way.

Table 2. Effect of squared difference to the mental model on Clicks/Optimal path

Tests of Between-Subjects Effects

Measure:Clicks

Transformed Variable:Average

Source	Type III Sum of Squareds	df	Mean Squared	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	942,280	1	942,280	31,611	,000	,579	31,611	1,000
Difference	80,212	1	80,212	2,691	,115	,105	2,691	,349
Error	685,601	23	29,809					

a. Computed using alpha = ,05

The mean quotient of clicks and optimal path ranged from 1,233 for “get married (trouwen)” on Nijmegen.nl and “shopping hours (winkeltijden)” on Utrecht.nl to 9,5 for “get married (trouwen)” on Enschede.nl. Ranges of quotients were rarely above 10 (only for “taxation (belastingen)” on Enschede.nl (23), “shopping hours (winkeltijden)” on Amsterdam.nl (10,33), “taxation (belastingen)” on DenHaag.nl (14), “get married (trouwen)” on Enschede.nl (20) and “get married (trouwen)” on Amsterdam.nl (11)). The high range for “get married (trouwen)” on Enschede.nl (0,53 squared difference) was due to one extreme outlier with a quotient of 24 (see figure 5). The lowest ranges of quotient scores (1) were exhibited by the tasks of finding “get married (trouwen)” on Nijmegen.nl and “shopping hours (winkeltijden)” on Utrecht.nl. For these tasks the minimal and maximal quotient were 1 and 2, which means that all participants found those items within two-times the clicks of the optimal path length.

Having said this, the minimal quotient for all tasks was almost always 1 with the exception of "shopping hours (winkeltijden)" on Amsterdam.nl (1,33), "parking licence (parkeervergunning)" on Nijmegen.nl (1,33), "data of the population (bevolkingscijfers)" on DenHaag.nl (3,0) and "parking licence (parkeervergunning)" on DenHaag.nl (1,5).
Pair wise comparison of the different tasks showed nearly no significance between different tasks but is showed a clear significant difference between "get married (trouwen)" on Enschede.nl (0,79 squared difference) and all other tasks (s. figure 4).

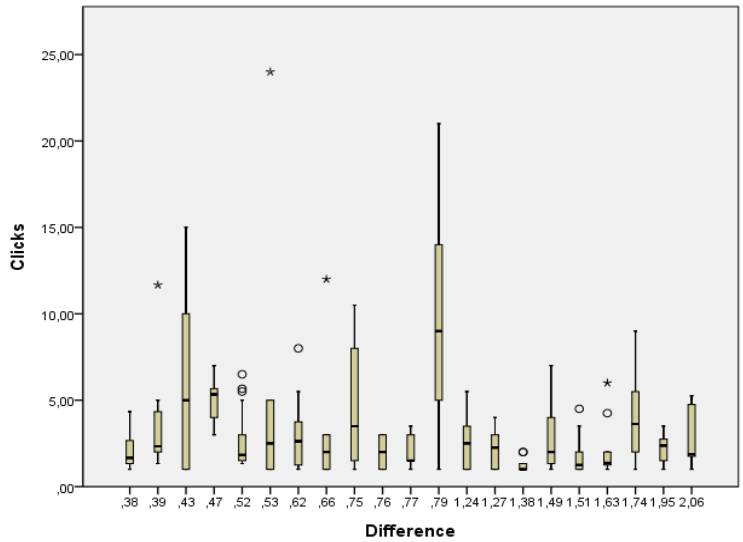


Figure 4. Clicks/Optimal path related to the squared difference to the mental model

3.3. User-Experience

In the first sample of 20 students web sites with tasks deviant from the mental model received higher ratings on the dimensions of pragmatic quality, hedonistic quality- identity, hedonistic quality- stimulation and attractiveness than web sites with tasks fitting the mental model. This difference in rating, however was statistically insignificant ($p>0.05$).

In the second sample of 20 students web sites with tasks fitting the mental model scored higher on the dimensions of pragmatic quality, hedonistic quality- identity and attractiveness than web sites with tasks deviant from the mental model. Of these differences only pragmatic quality was significant ($p<0.05$). Hedonistic quality- stimulation received slightly higher ratings on web sites with tasks defiant from the mental model than on web sites with tasks fitting the mental model. This difference was however statistically insignificant (note, that AttrakDiff does not give precise numbers, for more information see Appendix 5).

4.Discussion

The results show that card sorting has no or hardly any predictive value on browsing performance as measured by finding time and clicks/clicks on the optimal path. Therefore it seems that a categorization task does not sufficiently mirror a searching task on a web site. These findings let one reasonably doubt if card sorting is really fit for web site evaluation or even design as there seems to be no relation between proximity of a web site to a mental model and performance on that web site. Therefore other ways of reliable web site evaluation have to be found. One promising candidate might be a model called SNIF-ACT, which is short for Scent-based Navigation and Information Foraging in the ACT architecture (Fu & Pirolli, 2007). It has its basis in the Information foraging theory (Pirolli & Card, 1999), which in return grounds on the optimal foraging theory developed by ecologists in the 1960s (MacArthur & Pianka, 1966). Originally a model about searching for food (optimal foraging) it has now become a model about information search (information foraging). The latter theory tries to quantify the perceived relevance of a Web link to a User's goal. It assumes that users evaluate links sequentially and base their web-search decisions on previous and current assessment of the relevance of link texts to the information goal (Fu & Pirolli, 2007). Hereby, Information scent is one of the most important concepts. Much like animals rely on scents to indicate the chances of finding prey in current area and guide them to other promising patches, so do humans rely on various cues in the information environment to get similar answers. Human users estimate how much useful information they are likely to get on a given path, and after seeking information compare the actual outcome with their predictions. When the information scent stops getting stronger (i.e., when users no longer expect to find useful additional information), the users move to a different information source (Nielsen, 2004). This is very much applicable to my own research. Although the search task "data of the population (bevolkingscijfers)" on DenHaag.nl had a very low sum of squared differences to the mental model (0,47) it was not found within 4,5 minutes by all but 2 participants who were given this search task. As mentioned above this task showed significantly higher finding times than all other tasks. This could have been related to the headline leading to "data of the population (bevolkingscijfers)" on DenHaag.nl, which was "Actueel". This headline has no or hardly any semantic relation to "data of the population (bevolkingscijfers)" and therefore a very weak information scent. In contrast to that stands "data of the population (bevolkingscijfers)" on Enschede.nl with a mean search time of 39 seconds and a standard deviation of 28 seconds. Its search path "Enschede in Ontwikkeling-> Enschede in Cijfers->

Bevolkingscijfers" can be guessed to exert a much greater information scent than "Bewoners-> Actueel -> Feiten en Cijfers -> Kerncijfers Denhaag". Search paths for "data of the population (bevolkingscijfers)" on both Enschede.nl and DenHaag.nl are illustrated in figure 6. The U.S. Dept. of Health and Human Services' usability guidelines (2006) underline the importance of this finding by claiming that category labels should clearly reflect the items and information contained within the category. It is further stated that users will likely have difficulty understanding vague, generalized link labels, but will find specific, detailed links, and descriptors easier to use. Another guideline is of importance when looking at the search path of "data of the population (bevolkingscijfers)" on DenHaag.nl, namely that one has to scroll to get to the target information (as seen in figure 6, far right). The U.S. Dept. of Health and Human Services' usability guidelines, however, advice to use paging (navigating through links) instead.

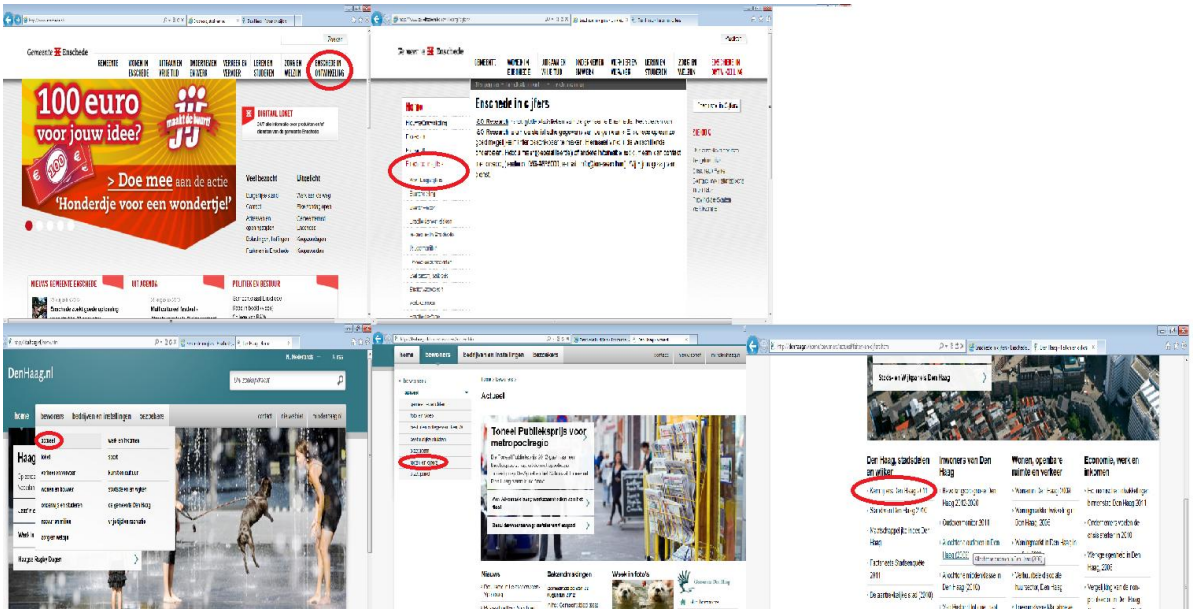


Figure 5: Search paths for "data of the population" on Enschede.nl (top) and DenHaag.nl (bottom)

Another result might be explained using Information foraging, namely the significant lower time values for finding "taxation (belastingen)" than for any other item. This particular item often had a direct link named "taxation (belastingen)" on the front page of the web sites (e.g. Enschede.nl, DenHaag.nl, Amsterdam.nl) or a semantically fitting path (Gemeente-> Belastingen en financiën-> Gemeentebelastingen) on Nijmegen.nl , Gemeentezaken-> Gemeentebelastingen) on Utrecht.nl). When participants did not find the direct link to "taxation (belastingen)" on the front page of Enschede.nl, DenHaag.nl or Amsterdam.nl, they often searched through categories such as "Ondernemen en werk" or "Bedrijven en

instellingen" to find the goal information. These categories are semantically highly related to "taxation (belastingen)" as one can only pay taxes if one works. This further underlines the importance of Information foraging in web search. However, for information scent to be a good lead to a target item, semantically fitting item categories are indispensable. Without proper grouping no fitting category names can be established, thereby decreasing information scent. In creating these semantically fitting item hierarchies card-sorting can be of great value after all.

The two examples that were listed (taxation (belastingen) and data of the population (bevolkingscijfers) on DenHaag.nl) highlight another candidate for web site evaluation, which is Navigation Path complexity. One aspect of it (Navigation path length) states that the longer a navigation task, the more navigation choices one has to make. If more navigation choices have to be made, more relevance judgments have to be made during search-tasks, which affects navigation performance as measured by time performance, accuracy and lostness (Melguizo, 2006). The item "taxation (belastingen)" was often only one click from the front page thereby having a very short navigation path. The item "data of the population (bevolkingscijfers)", which exerted significantly higher finding times than all other tasks often had relatively long paths. These were "Enschede in Ontwikkeling-> Enschede in Cijfers-> Data of the population (bevolkingscijfers)" on Enschede.nl, "Bewoners-> Actueel -> Feiten en cijfers -> Kerncijfers Denhaag" for DenHaag.nl, "Meer thema's->Stad in beeld-> Feiten en cijfers-> Amsterdam in cijfers-> Stand van de Bevolking" on Amsterdam.nl, "Gemeente-> Onderzoek en cijfers-> Cijfers-> Bevolking" on Nijmegen.nl and "Gemeentezaken-> Onderzoek en cijfers-> Bevolking, wonen en bouwen-> Bevolking" for Utrecht.nl. As one can see these navigation paths are significantly longer than those for e.g. "taxation (belastingen)". As seen in the result section, optimal path length was statistically significant for time. One can therefore conclude that the number of minimal clicks necessary to reach the goal information is a much better predictor of search performance than proximity to a mental model. Pairwise comparison shows that the difference in finding time was not significant between 1 or 2 minimal clicks necessary to reach the goal information, but it was significant between all other minimal path lengths. This is in accordance to the U.S. Dept. of Health and Human Services' usability guidelines (2006) which state that information should be available within two or three clicks.

In the context of E-Government there is another interesting finding to be reported:

Across all different websites, nearly each participant was given a task which they could not complete within the 4,5 minutes given. Seeing as the maximum number of clicks on the optimal path to reaching items was 4 this is a highly irritating outcome, indicating bad usability across all websites. In regard to the clicks/path length quotient significantly higher scores of Enschede.nl and DenHaag.nl to the other web sites were found. This can be explained due to the high quotient of "to marry" on Enschede.nl (9,5) and the generally high quotients on DenHaag.nl, with "taxes" being highest (5,9). Both of these items could only be found using the "digitaal loket" button, which was not embedded into the main menu. On all other websites items located in the "digitaal loket" section of the site could also be found in the main menu. This highlights the importance of making all information on a website also assessable through the main menu. This becomes most apparent when comparing "taxation" on DenHaag.nl and Amsterdam.nl. On Amsterdam.nl "taxation" could be found using the main menu whereas on DenHaag.nl it could not. Therefore, even though "taxation" could be found within one click and was located on the front page on both websites participants needed significantly more clicks ($p<0,01$) to find the item on DenHaag.nl.

Another interesting result was that Dutch students did significantly better than German students on the finding time and clicks/optimal path dimensions. This may firstly be due to difficulty with understanding the Dutch items (although an explanation was given in Dutch and English) and secondly due to not understanding the wording of the links on the path to that specific goal information. This should, however, have not been a problem due to all participant speaking Dutch at least on the NT2 level 5. However, as Dutch still was not the native language of the German participants it might have caused difficulties. This circumstance is closely related to another aspect of Navigation path complexity, namely Page information assessment. As noted in the Introduction this is the degree of difficulty to assess the relevance of information on a web site. When a web site is not given in a language one is highly familiar with an increase in difficulty of information assessment should be expected, thereby decreasing performance (Melguizo, 2006).

One result that has not been discussed so far is that in the second User-experience study web sites with tasks fitting the mental model were rated higher on the pragmatic quality dimension of the AttrakDiff product evaluation tool. It is a highly surprising result in so far as tasks fitting the mental model were neither finished faster nor with less clicks than tasks not fitting the model. With regard to these results, however, one has to notice that due to limitations of AttrakDiff only 20 subjects could fill in each of the two studies and the first one did not bear

any significant results at all. Therefore, it is highly probable that the significance in difference on the pragmatic quality dimension is only due to random fluctuations in the sample. Further complicating the issue users sometimes only needed about 5-10 seconds to reach the information goal on web sites they were asked to evaluate. It is therefore highly discussable if these evaluations can have any value at all. Sadly, AttrakDiff has no option of comparing or deleting several participants from the dataset, thereby giving no option to correct this possible error. Further studies might have to use other, more complex evaluation tools or just a more advanced version of AttrakDiff in order to get valid and reliable results.

There are still other implications of this research:

One is that the tool "card sorting" has to be questioned as it seems to wield no predictive value for browsing performance. If it has no evaluative value than its value for web site design is highly dubious as well. Therefore further research has to be done in order to confirm or debunk the findings of this experiment. It might be a good and valid tool to assess people's knowledge structures and mental models, but this has little value if it does not translate into real-world practice. Finding Items on web sites may rely on vastly different cognitive capacities and patterns than sorting items into groups and hierarchies. For designing web sites and evaluating them semantic Information scent and Navigation path complexity seem to wield much more promising results.

However, card sorting might only have little predictive value for municipal web sites or (more generally) web sites which users are not familiar with. As I mentioned earlier only 6 participants have ever been to one of the tested web sites. Card sorting may hold much more predictive value for web sites with which users are more familiar as they might have a rather elaborate mental model of that particular web site readily available. This might not be the case for unfamiliar web sites or unfamiliar information goals such as those tested for in this research as it is dubitable that students have a sufficient mental model of municipal web sites readily available, if they have an elaborated model of them at all.

5.References

Battleson, B. and Weintrop, J. (2000). University libraries web site nomenclature test using the card sort method: summary report presented to the University Libraries Web Access Team State University of New York at Buffalo", working paper, University Libraries, University of Buffalo, Buffalo. Retrieved Juli 21, 2012 from www.jkup.net/BuffaloNomenclatureTest-Spr2000.rtf

Bevis, N. (1995). Usability is quality of use. *Proceedings of UPA'98*, Washigton DC, 22-26 June 1998.

Bussolon, S., Russi, B. and Del Missier, F. (2006). Online card sorting: as good as the paper version. *Thirteenth European Conference on Cognitive Ergonomics*. Zürich, Switzerland, 20-22 September, 2006

Davies, R. (1996). Hierarchical card sorting: A tool for qualitative research. Retrieved Juli 22, 2012 from http://www.mande.co.uk/docs/hierarch.htm.

Dubois, Cornelius (1949). The Card-Sorting or Psychophysical Interview. *In Public Opinion Quarterly*, 13 (4) pp. 619-628

Ebenezer, C. (2003). Usability evaluation of an NHS web site, *Health Information and Libraries Journal*, Vol. 20 No. 3, pp. 134-42.

Faiks, A. and Hyland, N. (2000). Gaining user insight: a case study illustrating the card sort technique, *College & Research Libraries*, Vol. 61 No. 4, pp. 349-57.

Fincher, S. and Tenenberg, J. (2005) "Making sense of card sorting data," *Expert Systems*, 22(3)

Fu, W.-T. and P. Pirolli (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction* 22: 355-412.

Gaffney, G. (2000). What is card sorting? www.infodesign.com.au

Gwizdka, J. and Spence, I. (2006). What can searching behavior tell us about the difficulty of information tasks? A study of web navigation. In: *Proceedings of the 69th annual*

meeting of the American Society for Information Science and Technology (ASIS&T), 43, Austin, TX.

Hawley, M. (2008). Extending Card-Sorting Techniques to Inform the Design of Web Site Hierarchies. Retrieved Juli15, 2012 from http://www.uxmatters.com/mt/archives/2008/10/extending-card-sorting-techniques-to-inform-the-design-of-web-site-hierarchies.php

Heaton, R. K., Chelune, G. J., Kay, G. G., & Curtiss, G. (1993).*Wisconsin card sorting test.* Odessa: Psychological Assessment Resources

Hennig, N. (2001), "Card-sorting usability tests of the MIT libraries' web site: categories from the user's point of view", in Campbell, N. (Ed.), *Usability Assessment of Library-Related Web Sites: Methods and Case Studies*, LITA, Chicago, IL, pp. 88-99.

ISO 9241-11 (1998) *Ergonomic requirements for office work with visual display terminals (VDTs)* -- Part 11: Guidance on usability.

Katz, M. A. and Byrne, M. D. (2003). Effects of Scent and Breadth on Use of Site-Specific Search on E-Commerce Web Sites. *ACM Transactions on Computer-Human Interaction*, Vol. 10, No. 3, September 2003, Pages 198–220.

Kowoll, S. (2010). Mental Models of municipal web sites: A card sorting study

Ma, H. & Zaphiris, P. (2003). The Usability and Content Accessibility of the E-Government in the UK, *Universal Access in HCI,* 760-764

MacArthur, R. H. and Pianka, E. R. (1966). On the optimal use of a patchy environment. *American Naturalist*, 100

McDonald, J. E., Dearholt, D. W., Paap, K. R., & Schvaneveldt, R. W. (1986). A formal interface design methodology based on user knowledge. *ACM SIGCHI Bulletin*, 17(4), 285–290.

Melguizo, M.C.P., Vidya, U., Van Oostendorp, H.(2012). Seeking information online: the influence of menu type, navigation path complexity and spatial ability on information gathering tasks. *Behaviour & IT* 31(1): 59-70

Meuter, M.L., Ostrom, A.L., Roundtree, R.I., Bitner, M.J., 2000. Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of Marketing* 64 (3), 50–64.

Pirolli, P. and Card, S. K. (1999). Information Foraging. *Psychological Review 106*(4): 643-675.

Princeton Survey Research Associates, (2002). A Matter of Trust: What Users Want from Web Sites, Research Report.

Puerta Melguizo, M.C ., Lemmert, V.R., and van Oostendorp, H, (2006). Lostness, mental models and performance. In: *V.P. Guerrero Bote, ed. Current research in information sciences and technologies: multidisciplinary approaches to global information systems*, 256–260.

Reinhold, D. (1993). Erfassung der innerpsychischen Prozesse (Bewegungsvorstellung) mit Hilfe der Kartenlegetechnik. [Recording the psychological processes (notion of body movement) using the card sort technique.] In V. Lippens (Ed.), Forschungsproblem:

Subjektive Theorien. Zur Innensicht in Lern– und Optimierungsprozessen, *Berichte und Materialien des Bundesinstituts fuer Sportwissenschaft 10/93*. Koeln: Sport und Buch Strauß

Shackel, B. (1991). Usability - Context, Framework, Definition, Design and Evaluation. In B. Shackel, & S. Richardson, (Eds.), *Human Factors for Informatics Usability* (p. 21-37). Cambridge:University Press.

Shadbolt, N., & Burton, M. (1990). Knowledge elicitation. In *J. R. Wilson & E. N. Coreltt (Eds.), Evaluation of human work: A practical ergonomics methodology* (pp. 321–345). London: Taylor & Francis.

Smith, P. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers*, 8 (4): 365-381.

Tullis, T., Albert, B. (2008). *Measuring the User Experience*. (p. 48). Burlington: Elsevier.

Turnbow, D., Kasianovitz, K., Snyder, L., Gilbert, D. and Yamamoto, D. (2005). Usability testing for web redesign: a UCLA case study, *OCLC Systems & Services*, Vol. 21 No. 3, pp. 226-34.

United Nations, (2010). United Nations E-Government Survey 2010: Leveraging Egovernment at a Time of Financial and Economic Crisis.

Upchurch, L., Rugg, G., And Kitchenham, B. (2001). Using card sorts to elicit Web page quality attributes, *IEEE Software*, 84–89.

U.S. Dept. of Health and Human Services (2006). *The Research-Based Web Design & Usability Guidelines*, Enlarged/Expanded edition. Washington: U.S. Government Printing Office

Wood, J and Wood, L. (2008). Card Sorting: Current Practices and Beyond. *Journal of Usability Studies*, 4 (1), 1---6. Web. March 8, 2012

Appendix 1: Geïnformeerde toestemming

GEÏNFORMEERDE TOESTEMMING

Ik, …….. (*naam proefpersoon)*

Stem toe mee te doen aan een onderzoek dat uitgevoerd wordt door

Jan Sommer

Ik ben me ervan bewust dat deelname aan dit onderzoek geheel vrijwillig is. Ik kan mijn medewerking op elk tijdstip stopzetten en de gegevens verkregen uit dit onderzoek terugkrijgen, laten verwijderen uit de database, of laten vernietigen.

De volgende punten zijn aan mij uitgelegd:

1. Het doel van dit onderzoek is uit te vinden in hoeverre de structurele afwijking tussen een web site en het mentale model van users de performance en experience van users tijdens search-tasks beinvloedt.
   Deelname aan dit onderzoek zal meer inzicht geven omtrent voorspellende waarde van card sorting m.b.t. web site design.
2. Er zal aan mij gevraagd worden om verschillende items op 5 verschillende web sites te zoeken. Dit zal zonder hulp van de zoek-functie gebeuren.
   Het hele onderzoek zal ongeveer 30 minuten duren. Aan het einde van het onderzoek zal de onderzoeker uitleggen waar het onderzoek over ging.
3. Er behoort geen stress of ongemak voort te vloeien uit deelname aan dit onderzoek.
4. De gegevens verkregen uit dit onderzoek zullen anoniem verwerkt worden en kunnen daarom niet bekend gemaakt worden op een individueel identificeerbare manier.
5. De onderzoeker zal alle verdere vragen over dit onderzoek beantwoorden, nu of gedurende het verdere verloop van het onderzoek.

Handtekening onderzoeker: …… Datum: ……..

Handtekening proefpersoon: …… Datum: ……..

Appendix 2: Itempool

Items	Description Dutch	Description English	Nr
Algemene Projecten	Informatie over projecten in havengebieden, luchthaven, …	Projects	1
Arbeidsbemiddeling	Informatie over arbeidsomstandigheden, onderzoek arbeidsinpassing en sociale werkvoorzieningen	Employment-finding	2
Armoedebeleid	Informatie over de armoedebestrijding met bijzondere focus op de nota "Niemand aan de kant" en de website Geldkompas.nl	Poverty policy	3
Begraafplaatsen	Verschillende details over alle begraafplaatsen en de te betalen tarieven ervoor, informatie over verschillende soorten graven en een contactmogelijkheid	Cemetery	4
Belastingen	Lijst van taxation (belastingen) die betaald moeten worden en toevoegende informatie erover	Taxation	5
Bereikbaarheid	Wegwerkzaamheden, bereikbaarheid van de stad met de auto, trein, bus en fiets	Accessibility	6
Bescherming Milieu	Informatie over afvalregels, gevelreiniging, milieuvergunning en natuurgebieden	Protection of environment	7
Bevolkingscijfers	Informatie over ontwikkeling en stand van de bevolking van de gemeente, bevolkingsgegevens zijn in de buurtmonitor en in het statistisch jaarboek te vinden	Data of the population	8
Bezwaar en beroep	Digitaal een bezwaarschrift indienen. Verder regeling voor klachten en eventuele Rechtshulp	Complaints	9
Bioscoop	Informatie over bioscopen, adres en link naar bioscoop-website	Cinema	10

Bomen	Informatie en regels om hout kappen en te herplanten, kapvergunning en opsomming van bijzondere en monumentale bomen	Trees	11
(Ver)bouwen,slopen, verbeteren	Informatie over bodemsanering, rioolaansluiting en interne verbouwingen. Verder bouwverordening en de mogelijkheid een onbewoonbaarverklaring in te dienen	Have one´s house altered, rebuilt, renovated, converted, demolition	12
Brandpreventie	Informatie over barbecueregels, alarmnummer, sirene en vluchtplan	Fire precaution	13
Burgerinitiatief	Recht van burgers om zelf onderwerpen op de agenda van de gemeentepolitiek te plaatsen: woon- en leefklimaat in de straat, de wijk of het stadsdeel verbeteren	Petition	14
College van B & W	College van Burgemeester en Wethouders van de gemeente, overzicht van de wekelijkse besluitenlijsten	Board	15
Crisisbeheersing	Omschrijving van hulpverleningsdiensten, informatie over terrorismebestrijding, rampenplan, risicoplan	Crisis management, crisis control	16
Fiets- en wandelroutes	Korte omschrijving van de bestaande fiets- en wandelroutes en verwijzingen naar nuttige websites	(Bi)cycle routes, trails	17
Fietsparkeren	Adres en openingstijden van rijwielstallingen	(Bi)cycle lock-up (bi)cycle park	18
Geboorte	Informatie over de erkenning van een pasgeboren kind, aangifte doen van het kind bij de gemeente en de geboorteakte	Birth	19
Gehandicaptenzorg	Nuttige informatie over gehandicaptenvervoer, rolstoelen, thuiszorg en woningaanpassing voor gehandicapten	Health care for disabled people	20
Geld- en schuldproblemen	Informatie over hulp bij schulden, langdurigheidtoeslag, verklaring omtrent inkomen en vermogen, link naar	Debts	21

	Geldkompas.nl		
Gemeenteraad	Algemene informatie over het hoogste orgaan van de gemeente, lijst van alle raadsleden, - commissies, agenda`s en vergaderstukken	College	22
Gezondheid	Informatie over gezondheidsvoorlichting en geestelijke gezondheidszorg. Verder reizigersadvisering en contactmogelijkheiden van hulpdiensten	Health-promotion, advise travellers, helpline	23
Heffingen	Lijst van heffingen zoals rioolheffing etc	Charge, impost, duty, tax	24
Huisafval	Algemene informatie over afval, zoals glas, papier, klein chemisch afval, textielinzameling. Standpunten van afvalcontainers	Domestic/ household waste	25
Hulp bij het huishouden	Informatie voor klanten, hulpverlenende krachten en zorgaanbieders, verder informatie over mogelijke hulp, contactgegevens en een aanvraag-formuluier	Aid, assistence in housekeeping	26
(Ver)huren van een woning	Informatie over huurtoeslag en het aanbod Huurwoningen	Rent flat	27
Invalidenparkeren	Parkeerkaarten en -plaatsen voor invaliden	Disabled parking	28
Koopzondagen/Koopavonden	Opsomming van koopzondagen- en avonden binnen en buiten de binnenstad en openingstijden	Sunday opening hours, evening opening hours	29
Kopen van een woning	Informatie omtrent het aanbod nieuwbouwwoningen, bijvoorbeeld kaarten en plattegronden, informatie over nieuwbouwprojecten en koopsubsidie	Buy flat	30
Leerplicht	Officiële maatregelen rond het onderwerp Leerplicht	Compulsory education	31
Leren en studeren	Informatie en advertentie van alle	Studying	32

	onderwijsfaciliteiten en links ernaartoe		
Musea	Overzicht van musea, adres en openingstijden	Museum, art gallery	33
Nationaliteit	Informatie over het bewijs van Nederlanderschap, problemen met discriminatie, mogelijke inburgering in Enschede en het proces van naturalisatie	Nationality	34
Natuurprojecten in de stad	Opsomming van bijzondere beleefroutes- en plekken, Natuureducatie-acties (bijv. de schoolnatuurtuin voor basisschoolen)	Nature projects	35
Onderwijs	Informatie over leerlingenvervoer, vakanties en een opsomming van alle onderwijsinstellingen	Education	36
Onderzoekspanel	Leden van dit onderzoekspanel ontvangen een aantal keren per jaar via e-mail een uitnodiging om een online vragenlijst in te vullen	Panel, consortium	37
Ontevredenheid woonomgeving	Informatie over burenrechten, regels omtrent het houden van honden en hondenvoorzieningen, mogelijkheid om klachten en meldingen te maken en ongediertebestrijding	Dissatisfaction with environment/ housing conditions	38
Ontheffingen	Informatie over ontheffing geluid, parkeerontheffing en stookontheffing	Exemption	39
Openbaar vervoer	Informatie over de bus, busroutes, link naar Busbedrijf	Public transport	40
Over de stad	Korte historie van de stad en bijzondere gebeurtenissen uit de geschiedenis	About the city	41
Overlijden	Informatie over het aangifte doen van een overleden persoon, de overlijdensakte en mogelijke begraafplaatsen	Death, deceased	42
Parkeergelegenheden	Opsomming van parkeergarages en een plattegrond met parkeerlocaties	Parking facilities, parking space/ accommodation	43
Parkeertarieven	Prijsoverzicht voor abonnementen, overzicht van vergunningstarieven	Tariff	44
Parkeervergunning	Omschrijving van een parking licence	Parking	45

	(parkeervergunning), Aanvraagmogelijkheden en benodigdheden	licence/alicense parking permit	
Parken	Informatie en beschrijving van parken en een plattegrond daarvan	Parks	46
Recreatie en tourisme	Informatie en opsomming van kinderboerderijen, speelplekken in openbare ruimtes en speeltuinen	Recreation, leisure & tourism	47
Regels voor evenementen	Informatie over het beleid en de vergunningen van evenementen	Event-rules	48
Reisdocumenten	Informatie over het aanvragen en hebben van een ID en/of paspoort en een contactadres	Travel documents	49
Rekenkamer	Onafhankelijk orgaan dat toezicht houdt op de Rijksfinanciën	Auditor`s office	50
Rijbewijs	Opsomming van verschillende soorten Rijbewijzen	Driver´s license	51
Scheiden	Informatie over de echtscheidingsakte en de afloop van een echtscheiding	Divorce	52
Sport voor speciale groepen	Informatie over sportaanbiedingen voor gehandicapten, scholen, ouderen, allochtone vrouwen en meisjes	Sport for special groups	53
Spoortvoorzieningen	Informatie over ondersteuning sportverenigingen, sportinstellingen, sportaccommodaties	Facilities of sport	54
Stadsarchief	Geschiedenis van het archief, openingstijden, tarieven en een beschrijving hoe het precies in zijn werk gaat.	Archive	55
Stadsdelen	Informatie over stadsdelen, actuele berichten, projecten, cijfers en feiten	Districts in town	56
Trouwen	Informatie over huwelijk en geregistreerd partnerschap en huwelijks-/ partnerschapsregistratieakte	Get married	57
Uitkering	Informatie over arbeidsongeschiktheid en uitkering, (bijzondere) bijstand, bijstand voor zelfstandigen, hulp bij betaald werk zoeken, uitkering bij tijdelijke ziekte, zorgtoeslag	Remittance	58

Vergunningen	Informatie over o.a. aanlegvergunning, bouwvergunning, evenementenvergunning, inritvergunning, kapvergunning, milieuvergunning	Permission	59
Verkeer	Informatie over gladheidbestrijding, verkeerslichten en verkeersmaatregel	Traffic	60
Verkeersveiligheid	Informatie en omschrijving van verkeersprojecten, aanpak maatregelen en gedragsprojecten	Road/traffic safety	61
Verkiezingen	Informatie over de uitslagen van de verkiezingen in de gemeente, over Europese, gemeenteraadsverkiezingen en verkiezingen van de Provinciale Staten	Election	62
Visie en ambitie van de gemeente	Informatie over concrete doelstellingen voor de toekomst binnen de stad	Vision and pursuit	63
Vrijwilligerswerk	Vraag en aanbod van vrijwilligers, informatie over vrijwilligersbeleid, contactgegevens van het bijhorende servicepunt	Voluntary/volunteer work	64
Water	Informatie over de maatregeling van afvalwater, grondwater en grondwateroverlast	Water	65
Weer	Weerbericht met meetgegevens van de regio	Weather	66
Wet Maatschappelijke Ondersteuning	Algemene informatie over de Wet Maatschappelijke Ondersteuning en de hierbij belangrijke politieke agenda. Bovendien zijn er pdf-documenten over de wetsartikelen te vinden.	Social Support Act	67
Winkeltijden	Openingstijden van winkelcentra	Shopping hours, opening hours	68
Wonen met zorg op maat	Informatie over woonvoorzieningen met zorg en begleiding, locaties, projecten en omgevingsfactoren zijn hier opgesomd	Living accomodatios for people in need	69
Woningcoöperaties	Opsomming van woningcoöperaties	Collaboration of lessors	70

Appendix 3: SPSS Syntax

TIME:

Finding Time repeated- measures:

```
DATASET ACTIVATE DataSet1.
GET
  FILE='W:\Thesis Jan Sommer\spss bestände Documents\alle umgedreht.sav'.
DATASET NAME DataSet3 WINDOW=FRONT.
GLM var001 var002 var003 var004 var005 var006 var007 var008 var009 var010 WITH Difference pathlenghts
  /WSFACTOR=RPT 10 Polynomial
  /MEASURE=Time
  /METHOD=SSTYPE(3)
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=RPT
  /DESIGN=Difference pathlenghts.
```

Descriptives Difference-Time:

```
GET
  FILE='W:\Thesis Jan Sommer\spss bestände\Long-ausgefuellt1.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
MEANS TABLES=Time BY Difference
  /CELLS MEAN COUNT STDDEV MIN MAX RANGE.
```

Pairwise comparison tasks:

```
UNIANOVA Time BY Task
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(Task) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=Task.
```

Pairwise comparison nationality:

```
UNIANOVA Time BY Nationality
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(Nationality) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=Nationality.
```

Pairwise comparison task-website combination:

```
UNIANOVA Time BY Difference
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(Difference) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE
```

```
  /CRITERIA=ALPHA(.05)
  /DESIGN=Difference.
```

Clicks:

Repeated measures clicks:

```
GET
  ='W:\Thesis Jan Sommer\spss bestände \umgedreht clicks.sav'.
DATASET NAME DataSet2 WINDOW=FRONT.
GLM var001 var002 var003 var004 var005 var006 var007 var008 var009 var010 WITH VAR00001
  /WSFACTOR=RPT 10 Polynomial
  /MEASURE=Clicks
  /METHOD=SSTYPE(3)
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=RPT
  /DESIGN=VAR00001.
```

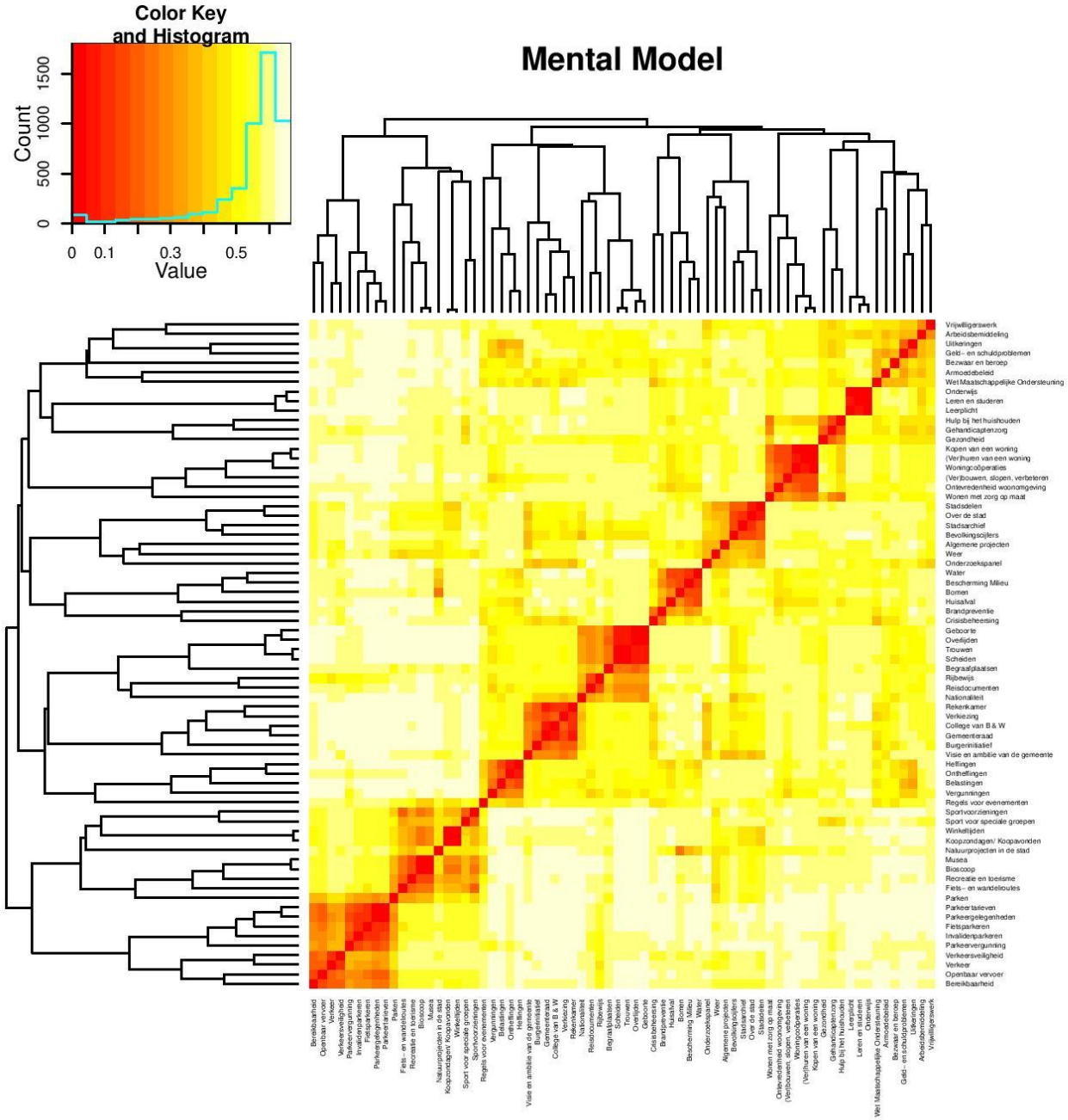
Descriptive statistics task-website combinations:

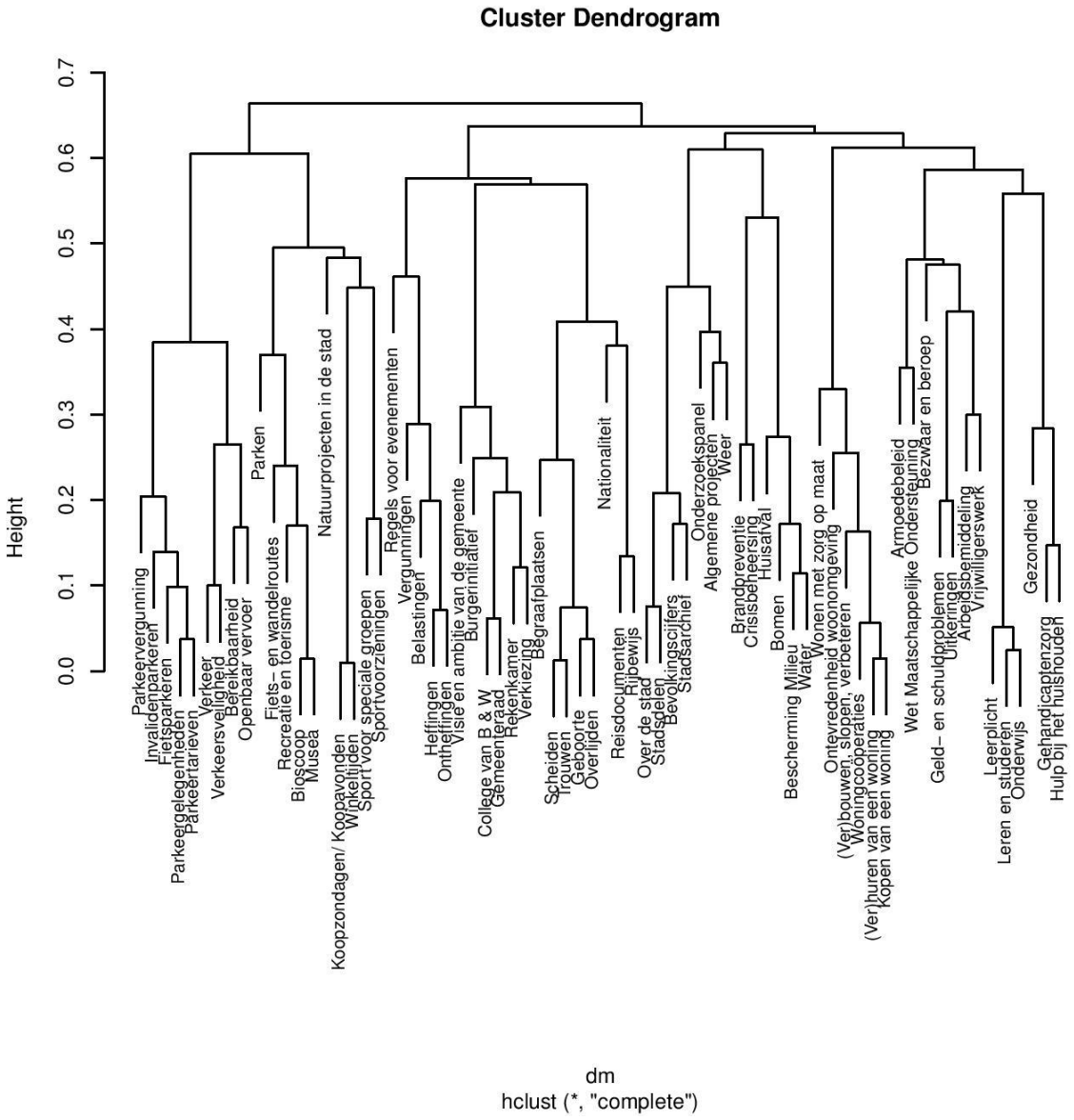
```
GET
  ='W:\Thesis Jan Sommer\spss bestände\Long-ausgefuellt1.sav'.
DATASET ACTIVATE DataSet1.
MEANS TABLES=Clicks BY Difference
  /CELLS MEAN COUNT STDDEV MIN MAX RANGE.
```

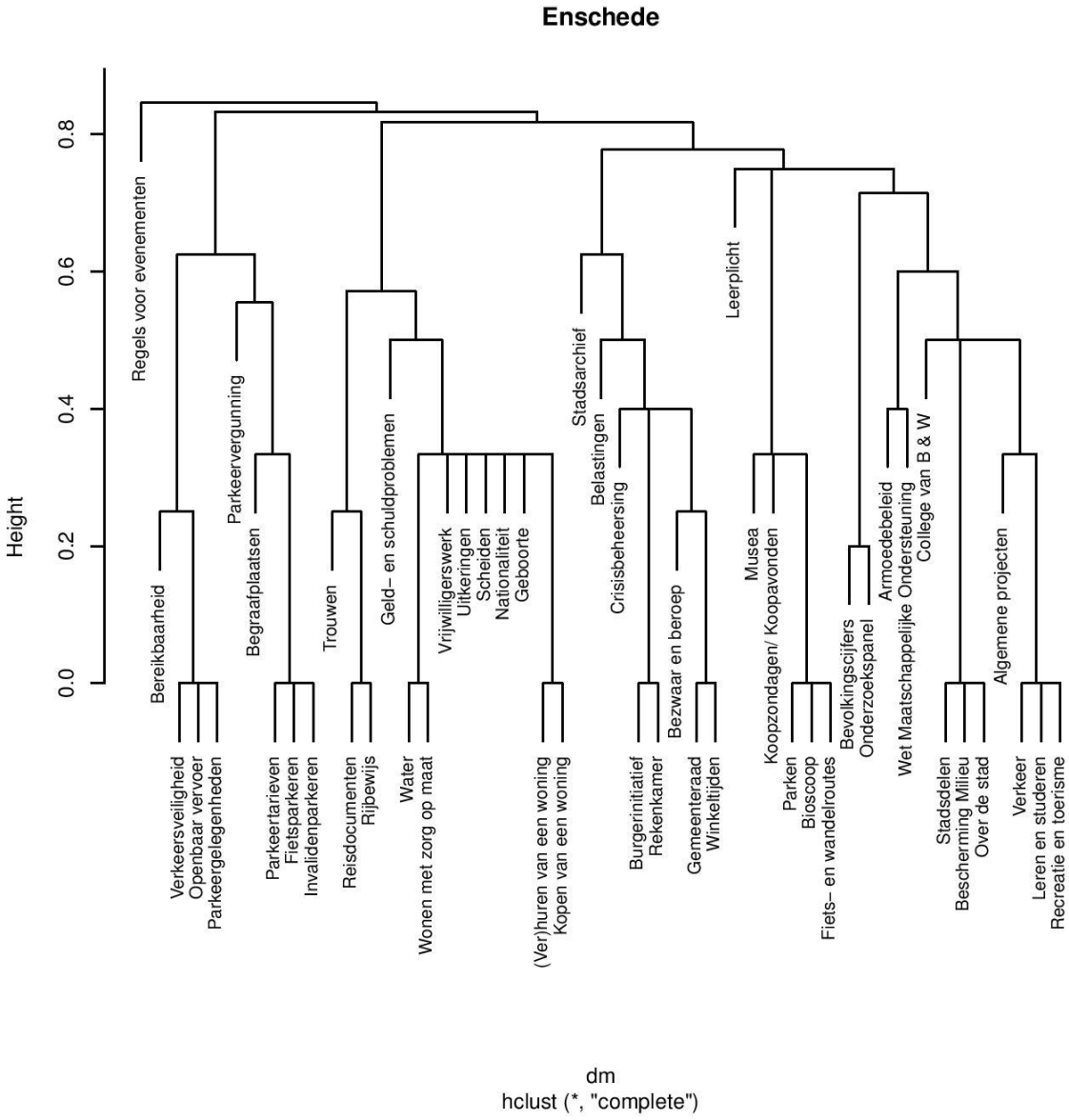
Pairwise comparison of task-website combinations:

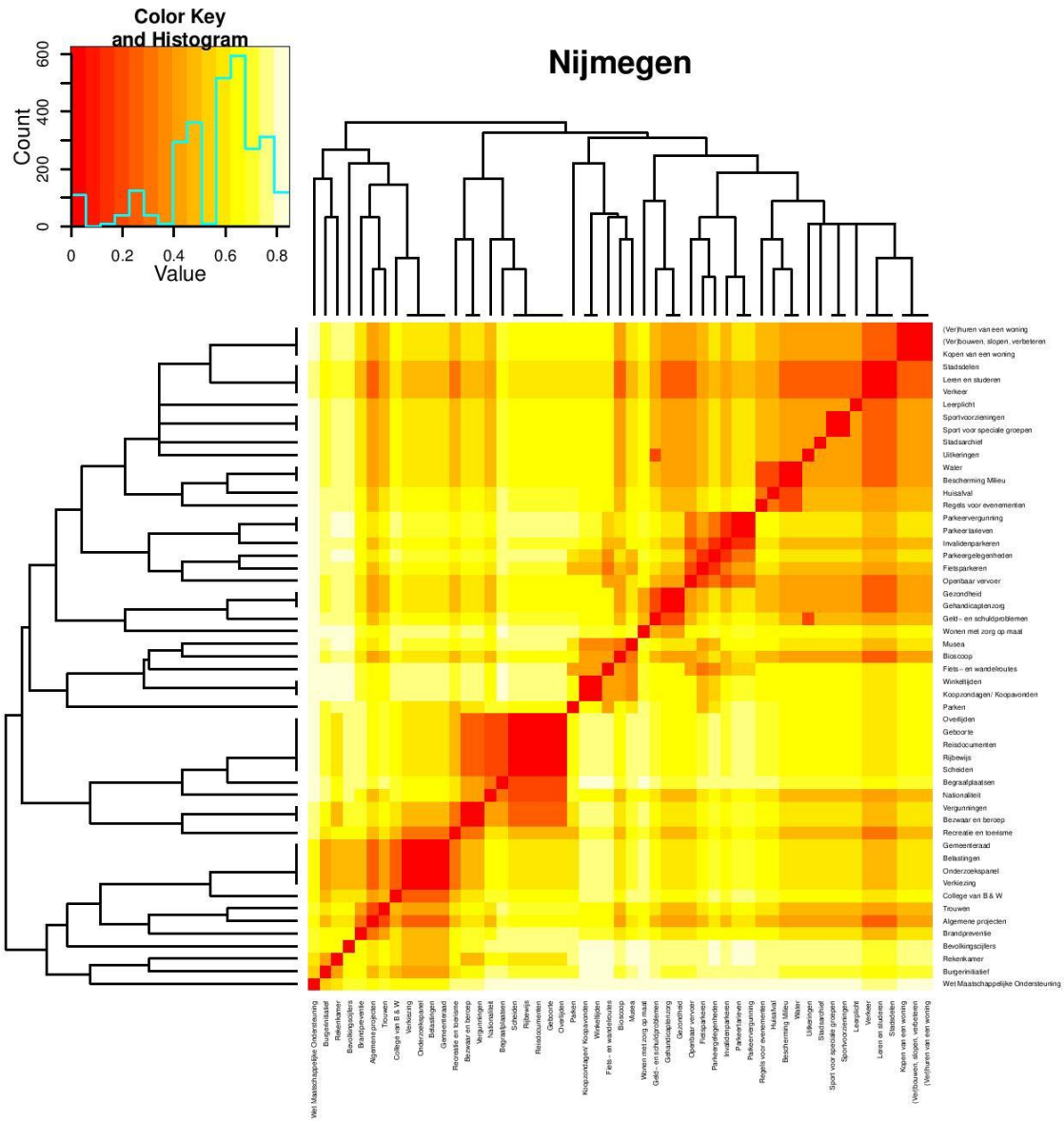
```
UNIANOVA Clicks BY Difference
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(Difference) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=Difference.
```

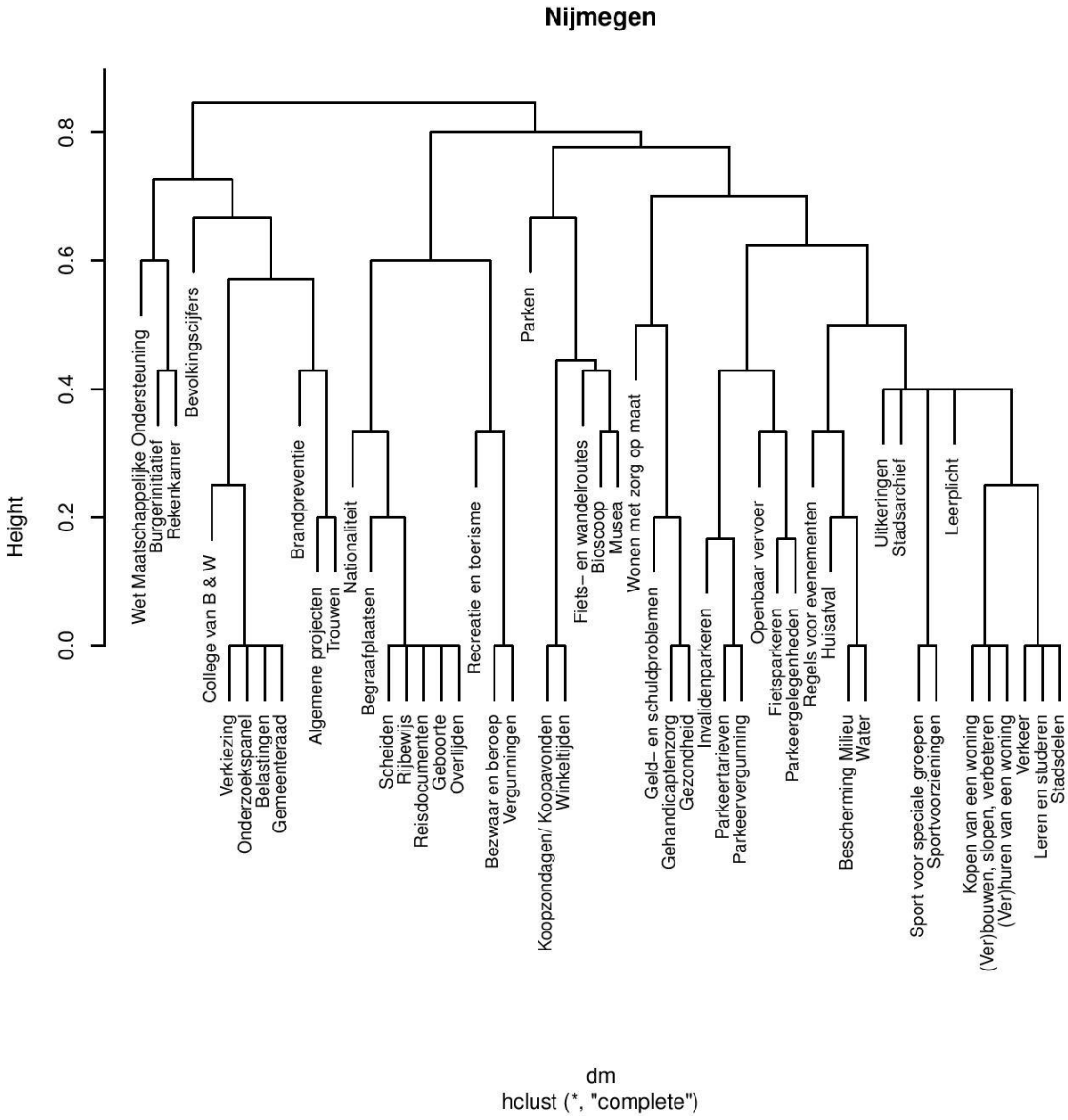
Appendix 4: Dendro- and Histograms

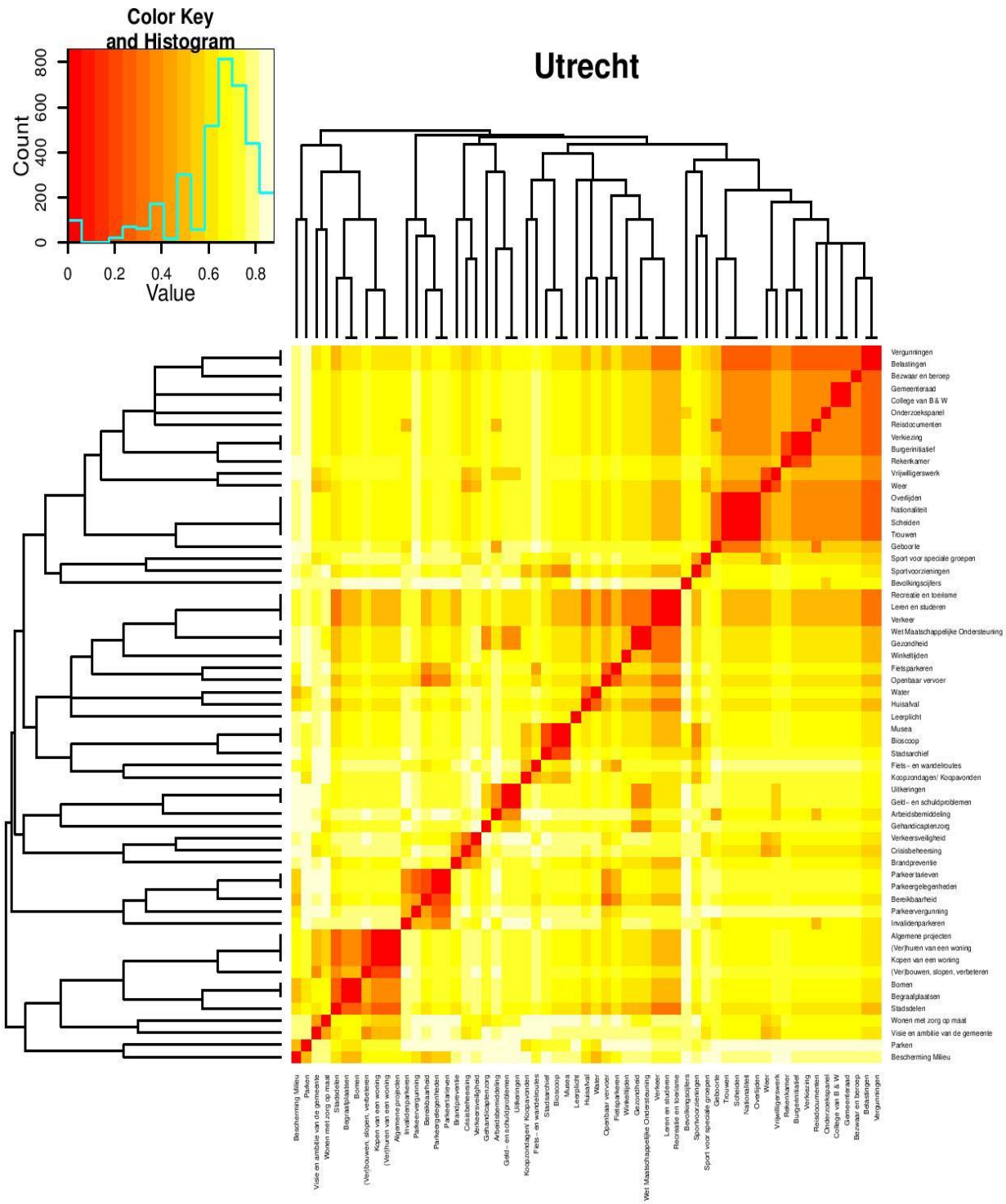


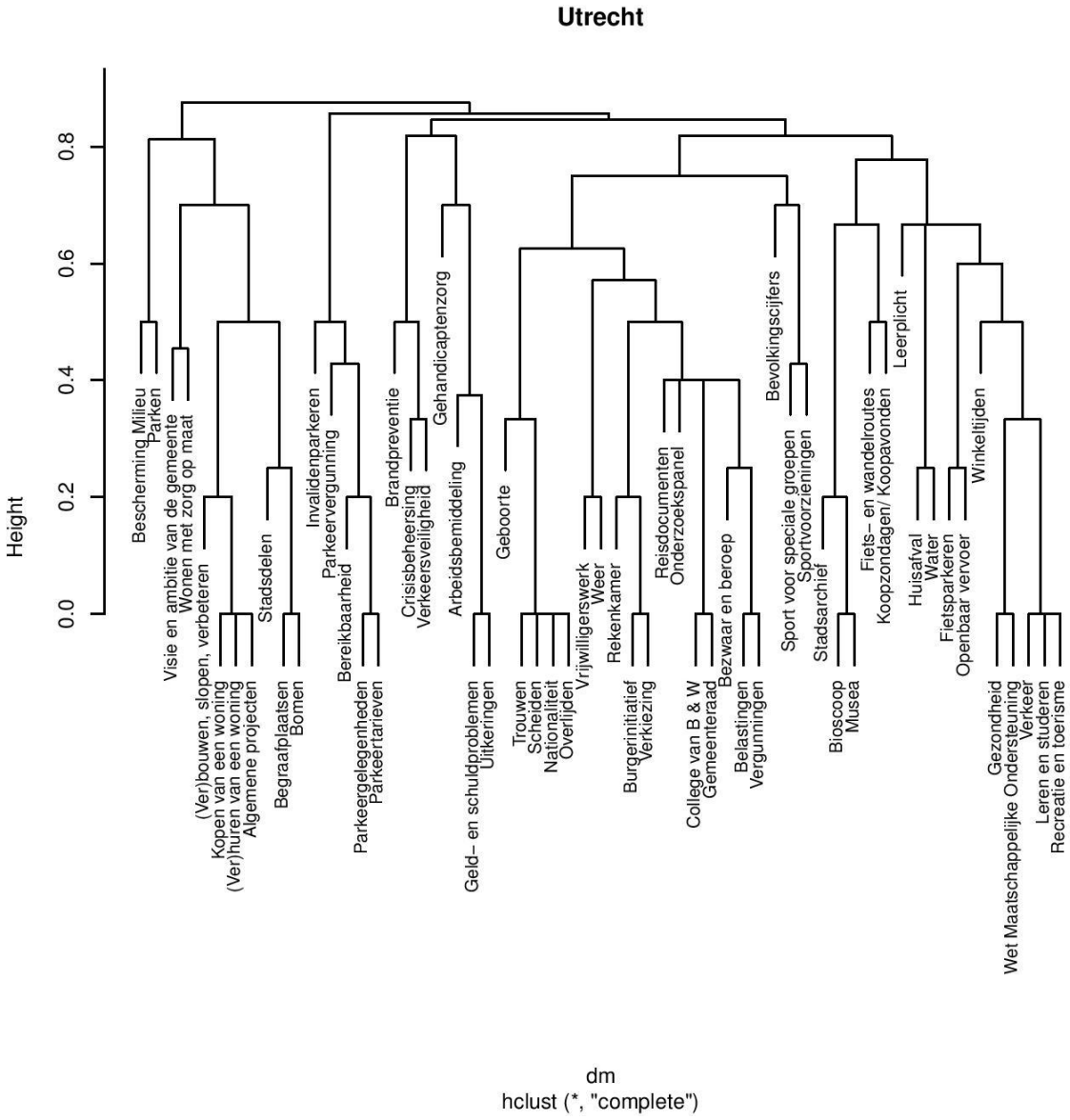

Cluster Dendrogram
Height
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Parkeervergunning
Invalidenparkeren
Fietsparkeren
Parkeergelegenheden
Parkeertarieven
Verkeer
Verkeersveiligheid
Bereikbaarheid
Openbaar vervoer
Parken
Fiets- en wandelroutes
Recreatie en toerisme
Bioscoop
Musea
Natuurprojecten in de stad
Koopzondagen/ Koopavonden
Winkelijden
Sportvoorzieningen
Regels voor evenementen
Vergunningen
Belastingen
Heffingen
Ontheffingen
Visie en ambitie van de gemeente
Burgerinitiatief
College van B & W
Gemeenteraad
Rekenkamer
Verkiezing
Begraafplaatsen
Scheiden
Trouwen
Geboorte
Overlijden
Nationaliteit
Reisdocumenten
Rijbewijs
Over de stad
Stadsgeleding
Bevolkingscijfers
Stadsarchief
Onderzoekspanel
Algemene projecten
Weer
Brandpreventie
Crisisbeheersing
Huisafval
Bomen
Milieu
Water
Bescherming
Wonen met zorg op maat
Ontevredenheid woonomgeving
(Ver)bouwen, slopen, verbeteren
Woningcooperaties
(Ver)huren van een woning
Kopen van een woning
Armoedebeleid
Wet Maatschappelijke Ondersteuning
Bezwaar en beroep
Geld- en schuldproblemen
Uitkeringen
Arbeidsbemiddeling
Vrijwilligerswerk
Leerplicht
Leren en studeren
Onderwijs
Gezondheid
Gehandicaptenzorg
Hulp bij het huishouden
dm
hclust (*, "complete")

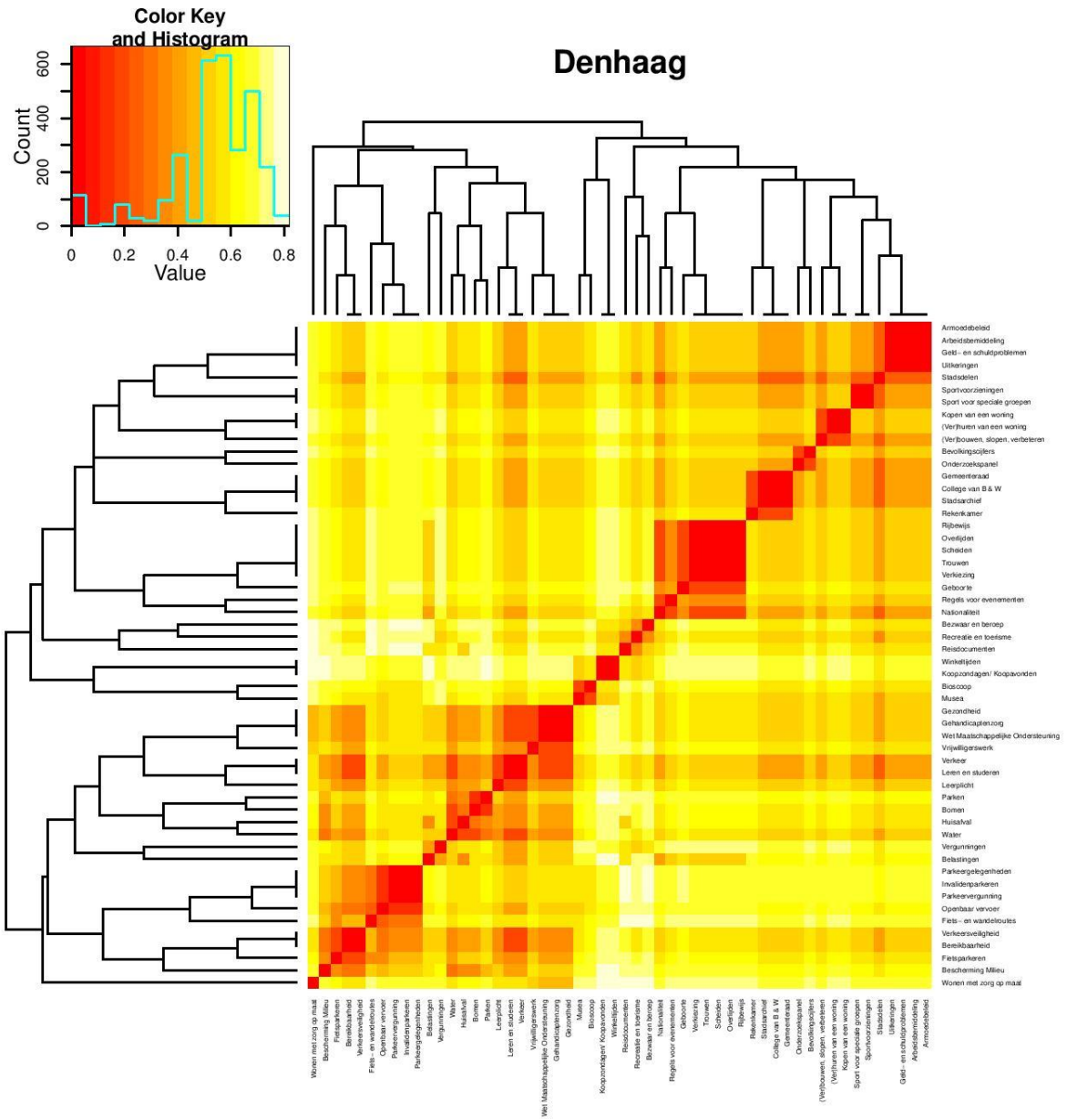

Enschede
Height
0.0
0.2
0.4
0.6
0.8
Regels voor evenementen
Bereikbaarheid
Verkeersveiligheid
Openbaar vervoer
Parkeergelegenheden
Parkeervergunning
Begraafplaatsen
Parkeertarieven
Fietsparkeren
Invalidenparkeren
Trouwen
Reisdocumenten
Rijbewijs
Geld– en schuldproblemen
Water
Wonen met zorg op maat
Vrijwilligerswerk
Uitkeringen
Scheiden
Nationaliteit
Geboorte
(Ver)huren van een woning
Kopen van een woning
Stadsarchief
Belastingen
Crisisbeheersing
Burgerinitiatief
Rekenkamer
Bezwaar en beroep
Gemeenteraad
Winkeltijden
Leerplicht
Musea
Koopzondagen/ Koopavonden
Parken
Bioscoop
Fiets– en wandelroutes
Bevolkingscijfers
Onderzoekspanel
Armoedebeleid
Wet Maatschappelijke Ondersteuning
College van B & W
Stadsdelen
Bescherming Milieu
Over de stad
Algemene projecten
Verkeer
Leren en studeren
Recreatie en toerisme
dm
hclust (*, "complete")

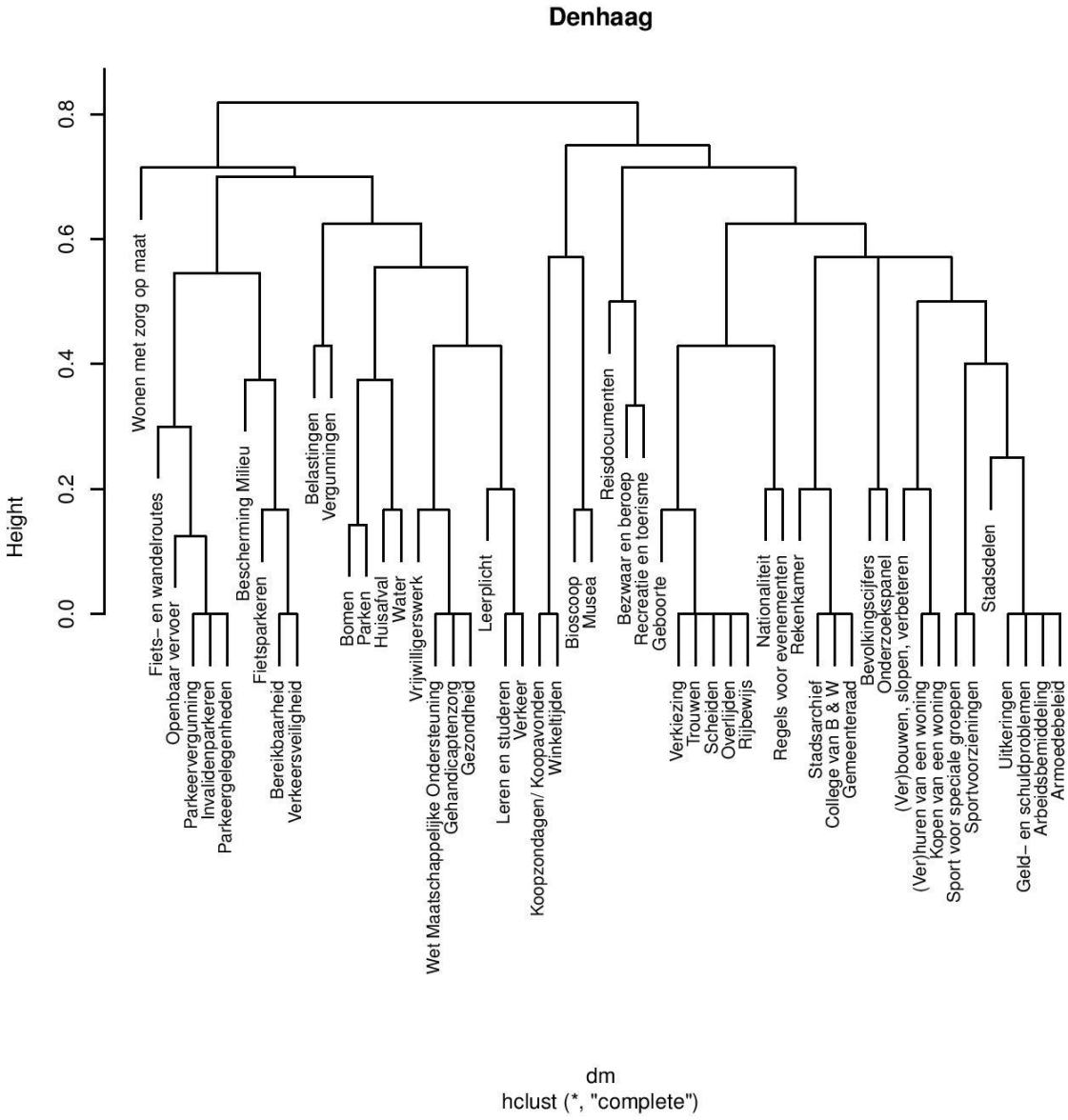

Color Key and Histogram
Count
Value
Nijmegen
(Verhuren van een woning
(Verbouwen, slopen, verbeteren
Kopen van een woning
Stadsdelen
Leren en studeren
Verkeer
Leerplicht
Sportvoorzieningen
Sport voor speciale groepen
Stadsarchief
Uitkeringen
Water
Bescherming Milieu
Huisafval
Regels voor evenementen
Parkeervergunning
Parkeertarieven
Invalidenparkeren
Parkeergelegenheden
Fietsparkeren
Openbaar vervoer
Gezondheid
Gehandicaptenzorg
Geld- en schuldproblemen
Wonen met zorg op maat
Musea
Bioscoop
Fiets- en wandelroutes
Winkeltijden
Koopzondagen/ Koopavonden
Parken
Overlijden
Geboorte
Reisdocumenten
Rijbewijs
Scheiden
Begraafplaatsen
Nationaliteit
Vergunningen
Bezwaar en beroep
Recreatie en toerisme
Gemeenteraad
Belastingen
Onderzoekspanel
Verkiezing
College van B & W
Trouwen
Algemene projecten
Brandpreventie
Bevolkingscijfers
Rekenkamer
Burgerinitiatief
Wet Maatschappelijke Ondersteuning

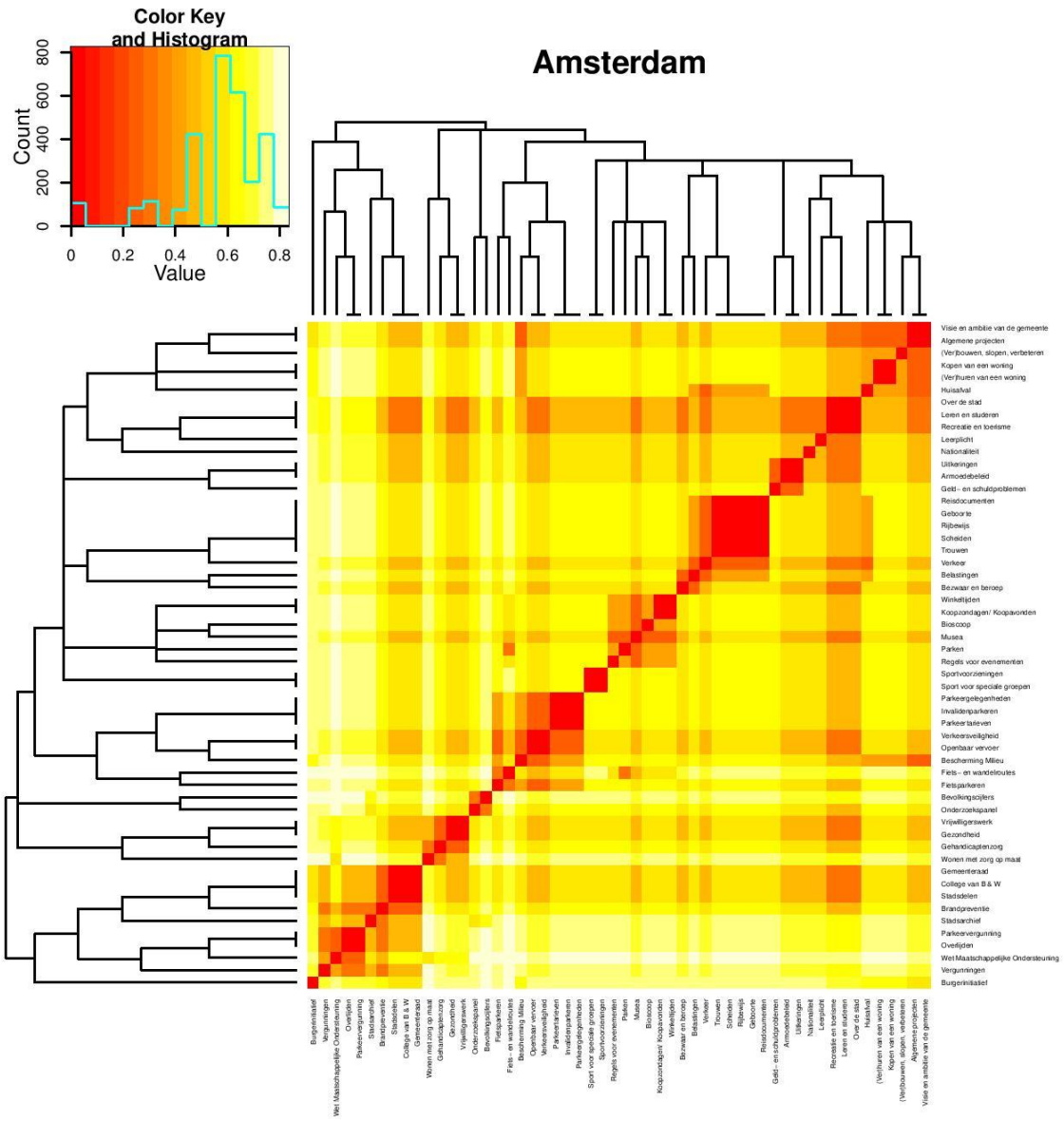

Nijmegen
Height
0.0
0.2
0.4
0.6
0.8
Wet Maatschappelijke Ondersteuning
Burgerinitiatief
Rekenkamer
Bevolkingscijfers
College van B & W
Verkiezing
Onderzoekspanel
Belastingen
Gemeenteraad
Brandpreventie
Algemene projecten
Trouwen
Nationaliteit
Begraafplaatsen
Scheiden
Rijbewijs
Reisdocumenten
Geboorte
Overlijden
Recreatie en toerisme
Bezwaar en beroep
Vergunningen
Parken
Koopzondagen/ Koopavonden
Winkeltijden
Fiets– en wandelroutes
Bioscoop
Musea
Wonen met zorg op maat
Geld– en schuldproblemen
Gehandicaptenzorg
Gezondheid
Invalidenparkeren
Parkeertarieven
Parkeervergunning
Openbaar vervoer
Fietsparkeren
Parkeergelegenheden
Regels voor evenementen
Huisafval
Bescherming Milieu
Water
Uitkeringen
Stadsarchief
Leerplicht
Sport voor speciale groepen
Sportvoorzieningen
Kopen van een woning
(Ver)bouwen, slopen, verbeteren
(Ver)huren van een woning
Verkeer
Leren en studeren
Stadsdelen
dm
hclust (*, "complete")

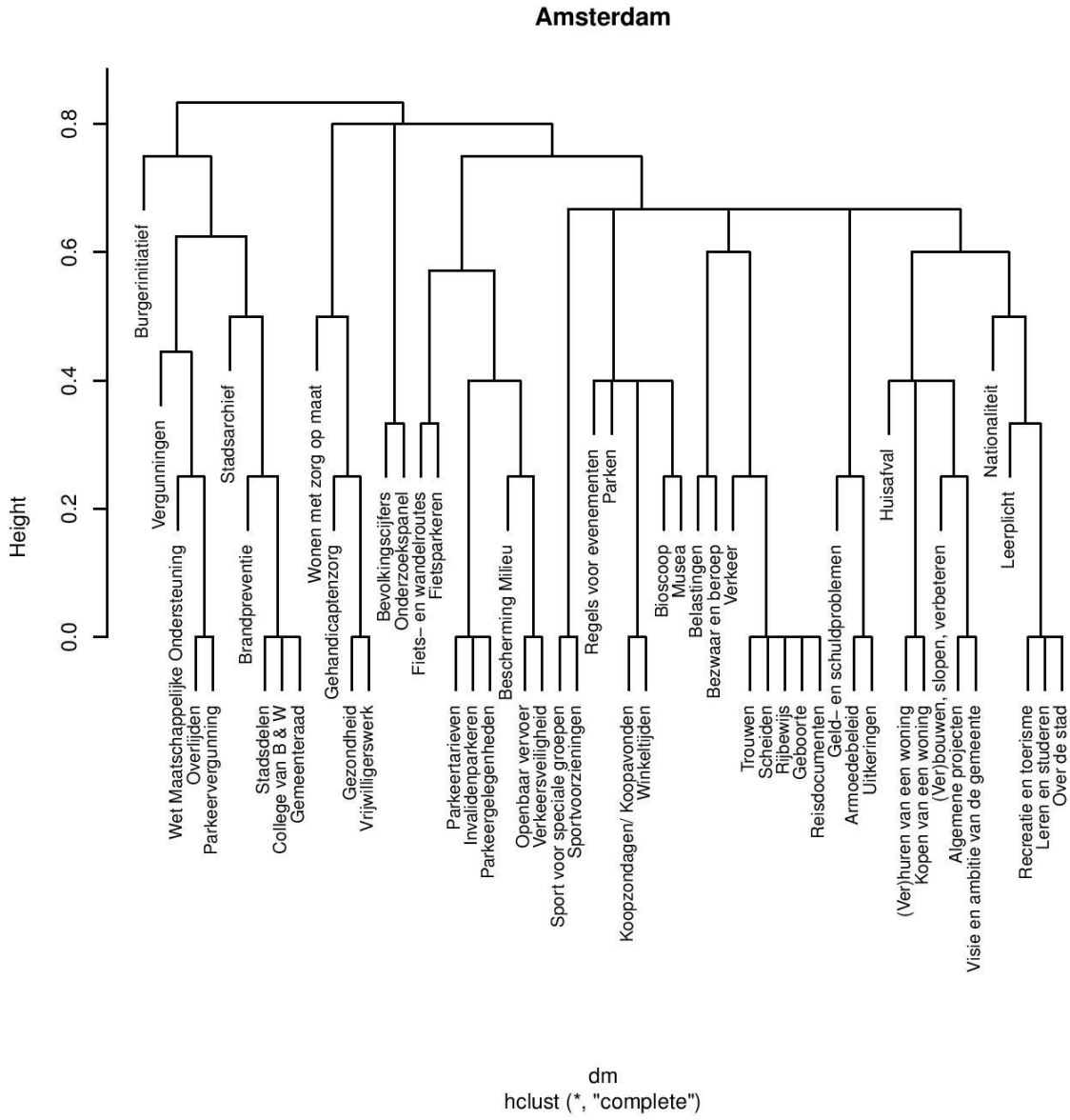

Color Key
and Histogram
Count
0
200
400
600
800
0
0.2
0.4
0.6
0.8
Value
Utrecht
Vergunningen
Belastingen
Bezwaar en beroep
Gemeenteraad
College van B & W
Onderzoekspanel
Reisdocumenten
Verkiezing
Burgerinitiatief
Rekenkamer
Vrijwilligerswerk
Weer
Overlijden
Nationaliteit
Scheiden
Trouwen
Geboorte
Sport voor speciale groepen
Sportvoorzieningen
Bevolkingscijfers
Recreatie en toerisme
Leren en studeren
Verkeer
Wet Maatschappelijke Ondersteuning
Gezondheid
Winkeltijden
Fietsparkeren
Openbaar vervoer
Water
Huisafval
Leerplicht
Musea
Bioscoop
Stadsarchief
Fiets- en wandelroutes
Koopzondagen/ Koopavonden
Uitkeringen
Geld- en schuldproblemen
Arbeidsbemiddeling
Gehandicaptenzorg
Verkeersveiligheid
Crisisbeheersing
Brandpreventie
Parkeertarieven
Parkeergelegenheden
Bereikbaarheid
Parkeervergunning
Invalidenparkeren
Algemene projecten
(Verhuren van een woning
Kopen van een woning
(Verbouwen, slopen, verbeteren
Bomen
Begraafplaatsen
Stadsdelen
Wonen met zorg op maat
Visie en ambitie van de gemeente
Parken
Bescherming Milieu


Utrecht
Height
0.0
0.2
0.4
0.6
0.8
Bescherming Milieu
Parken
Visie en ambitie van de gemeente
Wonen met zorg op maat
(Ver)bouwen, slopen, verbeteren
Kopen van een woning
(Ver)huren van een woning
Algemene projecten
Stadsdelen
Begraafplaatsen
Bomen
Invalidenparkeren
Parkeervergunning
Bereikbaarheid
Parkeergelegenheden
Parkeertarieven
Brandpreventie
Crisisbeheersing
Verkeersveiligheid
Gehandicaptenzorg
Arbeidsbemiddeling
Geld– en schuldproblemen
Uitkeringen
Geboorte
Trouwen
Scheiden
Nationaliteit
Overlijden
Vrijwilligerswerk
Weer
Rekenkamer
Burgerinitiatief
Verkiezing
Reisdocumenten
Onderzoekspanel
College van B & W
Gemeenteraad
Bezwaar en beroep
Belastingen
Vergunningen
Bevolkingscijfers
Sport voor speciale groepen
Sportvoorzieningen
Stadsarchief
Bioscoop
Musea
Fiets– en wandelroutes
Koopzondagen/ Koopavonden
Leerplicht
Huisafval
Water
Fietsparkeren
Openbaar vervoer
Winkeltijden
Gezondheid
Wet Maatschappelijke Ondersteuning
Verkeer
Leren en studeren
Recreatie en toerisme
dm
hclust (*, "complete")


Color Key and Histogram
Count
0
200
400
600
0
0.2
0.4
0.6
0.8
Value
Denhaag
Armoedebeleid
Arbeidsbemiddeling
Gelds - en schuldproblemen
Uitkeringen
Stadsdelen
Sportvoorzieningen
Sport voor speciale groepen
Kopen van een woning
(Ver)huren van een woning
(Ver)bouwen, slopen, verbeteren
Bevolkingscijfers
Onderzoekspanel
Gemeenteraad
College van B & W
Stadsarchief
Rekenkamer
Rijbewijs
Ovetlijden
Scheiden
Trouwen
Verkiezing
Geboorte
Regels voor evenementen
Nationaliteit
Bezwaar en beroep
Recreatie en toerisme
Reisdocumenten
Winkeltijden
Koopzondagen/ Koopavonden
Bioscoop
Musea
Gezondheid
Gehandicaptenzorg
Wet Maatschappelijke Ondersteuning
Vrijwilligerswerk
Verkeer
Leren en studeren
Leerplicht
Parken
Bomen
Huisafval
Water
Vergunningen
Belastingen
Parkeergelegenheden
Invalidenparkeren
Parkeervergunning
Openbaar vervoer
Fiets - en wandelroutes
Verkeersveiligheid
Bereikbaarheid
Fietsparkeren
Bescherming Milieu
Wonen met zorg op maat


Denhaag
Height
0.0
0.2
0.4
0.6
0.8
Wonen met zorg op maat
Fiets– en wandelroutes
Openbaar vervoer
Parkeervergunning
Invalidenparkeren
Parkeergelegenheden
Bescherming Milieu
Fietsparkeren
Bereikbaarheid
Verkeersveiligheid
Belastingen
Vergunningen
Bomen
Parken
Huisafval
Water
Vrijwilligerswerk
Wet Maatschappelijke Ondersteuning
Gehandicaptenzorg
Gezondheid
Leerplicht
Leren en studeren
Verkeer
Koopzondagen/ Koopavonden
Winkeltijden
Bioscoop
Musea
Reisdocumenten
Bezwaar en beroep
Recreatie en toerisme
Geboorte
Verkiezing
Trouwen
Scheiden
Overlijden
Rijbewijs
Nationaliteit
Regels voor evenementen
Rekenkamer
Stadsarchief
College van B & W
Gemeenteraad
Bevolkingscijfers
Onderzoekspanel
(Ver)bouwen, slopen, verbeteren
(Ver)huren van een woning
Kopen van een woning
Sport voor speciale groepen
Sportvoorzieningen
Stadsdelen
Uitkeringen
Geld– en schuldproblemen
Arbeidsbemiddeling
Armoedebeleid
dm
hclust (*, "complete")


Color Key
and Histogram
Count
0
200
400
600
800
0
0.2
0.4
0.6
0.8
Value
Amsterdam
Visie en ambitie van de gemeente
Algemene projecten
(Ver)bouwen, slopen, verbeteren
Kopen van een woning
(Ver)huren van een woning
Huisafval
Over de stad
Leren en studeren
Recreatie en toerisme
Leerplicht
Nationaliteit
Uitkeringen
Armoedebeleid
Geld- en schuldproblemen
Reisdocumenten
Geboorte
Rijbewijs
Scheiden
Trouwen
Verkeer
Belastingen
Bezwaar en beroep
Winkeltijden
Koopzondagen/ Koopavonden
Bioscoop
Musea
Parken
Regels voor evenementen
Sportvoorzieningen
Sport voor speciale groepen
Parkeergelegenheden
Invalidenparkeren
Parkeertarieven
Verkeersveiligheid
Openbaar vervoer
Bescherming Milieu
Fiets- en wandelroutes
Fietsparkeren
Bevolkingscijfers
Onderzoekspanel
Vrijwilligerswerk
Gezondheid
Gehandicaptenzorg
Wonen met zorg op maat
Gemeenteraad
College van B & W
Stadsdelen
Brandpreventie
Stadsarchief
Parkeervergunning
Overtijden
Wet Maatschappelijke Ondersteuning
Vergunningen
Burgerinitiatief



Appendix 5: AttrakDiff results

Evaluation Report for Products

"website 1" and "website 2"

Objectives:
How user-friendly and attractive is this product?

Contents of report

Method of investigation

AttrakDiff™ is an instrument for measuring the attractiveness of interactive products.

With the help of pairs of opposite adjectives, users (or potential users) can indicate their perception of the product. These adjective-pairs make a collation of the evaluation dimensions possible.

The following product dimensions are evaluated:

Pragmatic Quality (PQ):
- Describes the usability of a product and indicates how successfully users are in achieving their goals using the product.

- Hedonic quality - Stimulation (HQ-S):
  Mankind has an inherent need to develop and move forward. This dimension indicates to what extent the product can support those needs in terms of novel, interesting, and stimulating functions, contents, and interaction- and presentation-styles.

- Hedonic Quality - Identity (HQ-I):
  Indicates to what extent the product allows the user to identify with it.

- Attractiveness (ATT):
  Describes a global value of the product based on the quality perception.

Hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

Characteristics of investigation

Product title A:	website 1
Product title B:	website 2
Product industry:	Services
Duration of study A:	25.06.2012 - 23.09.2012
Duration of study B:	25.06.2012 - 23.09.2012
Project-type:	Comparison product A - product B, that means two different products are rated.
Variant:	The same test participants in both project parts.
Number of ratings for A:	20
Number of ratings for B:	20

Portfolio of results

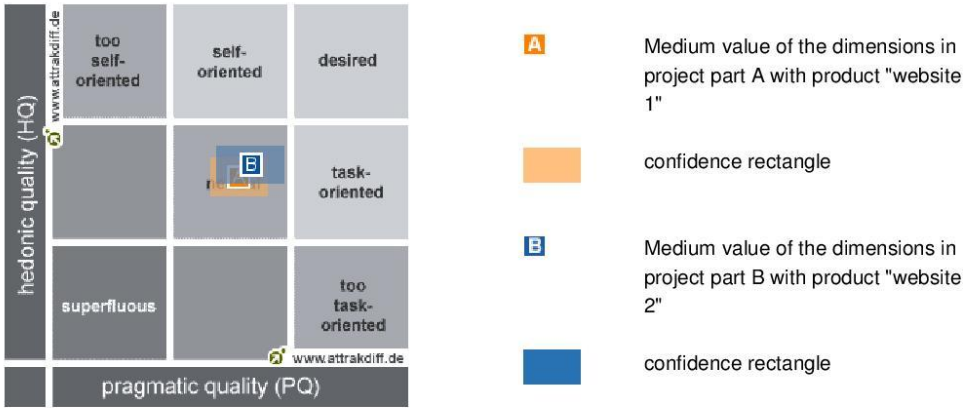


Diagram 1: Portfolio with average values of the dimensions PQ and HQ and the respective confidence rectangles of products "website 1" and "website 2"

In the portfolio-presentation the values of hedonic quality are represented on the vertical axis (bottom = low value). The horizontal axis represents the value of the pragmatic quality (i.e. left = a low value).

Depending on the dimensions values the product will lie in one or more "character-regions".

The bigger the confidence rectangle the less sure one can be to which region it belongs. A small confidence rectangle is an advantage because it means that the investigation results are more reliable and less coincidental.

The confidence rectangle shows, if the users are at one in their evaluation of the product.
The bigger the confidence rectangle, the more variable the evaluation ratings (more information is available in the appendix).

Interpretation for help

Project part A, product "website 1"

The products user interface was rated as "neutral".

Pragmatic quality is clearly the classification. The user is assisted by the product, however the value of pragmatic quality only reaches the average values.

Result: Consequently there is definite room for improvement.

In terms of hedonic quality the character classification applies positively. The user is stimulated by this product, however the hedonic value is only average.

Result: In terms of hedonic quality there is clearly room for improvement.

Attention! The values for HQ-I and HQ-s differ greatly. This differentiation is illustrated in the

diagram of mean values.

The confidence rectangle is small. The users are at one in their evaluation of the product.

Project part B, product "website 2"

The products user interface was rated as "neutral".

Pragmatic quality is clearly the classification. The user is assisted by the product, however the value of pragmatic quality only reaches the average values.

Result: Consequently there is room for improvement in terms of usability.

In terms of hedonic quality the character classification applies positively. The user is stimulated by this product, however the hedonic value is only average.

Result: Room for improvement exists in terms of hedonic quality.

Attention! The values for HQ-I and HQ-s differ greatly. This differentiation is illustrated in the diagram of mean values.

The confidence interval PQ is large. This could be attributed to limited sampling or to greatly differing product ratings.

Comparison of results of both project parts

Product website 2 performs better than product website 1. Pragmatic quality as well as hedonic quality is higher.

The difference in value between the pragmatic quality of product website 1 and product website 2 is statistically insignificant. It might therefore concern a chance fluctuation of judgement (see details in appendix).
The difference in value of the hedonic quality of product website 1 and product website 2 is statistically insignificant. It might therefore concern a chance fluctuation of judgement.

The confidence interval for pragmatic quality of product website 1 is smaller than for product website 2. In the rating of pragmatic quality of product website 1 the users are more at one in their evaluation. Thus the rating of product website 1 applies with greater certainty to the product.
The confidence interval for hedonic quality of product website 2 is smaller than for product website 1. In the rating of hedonic quality of product website 2 the users are more at one in their evaluation. Thus the rating of product website 2 applies with greater certainty to the product.

Diagram of average values

The average values of the AttrakDiff™ dimensions for the evaluated product are plotted on the diagram.

In this presentation hedonic quality distinguishes between the aspects of stimulation and identity. Furthermore the rating of attractiveness is presented.

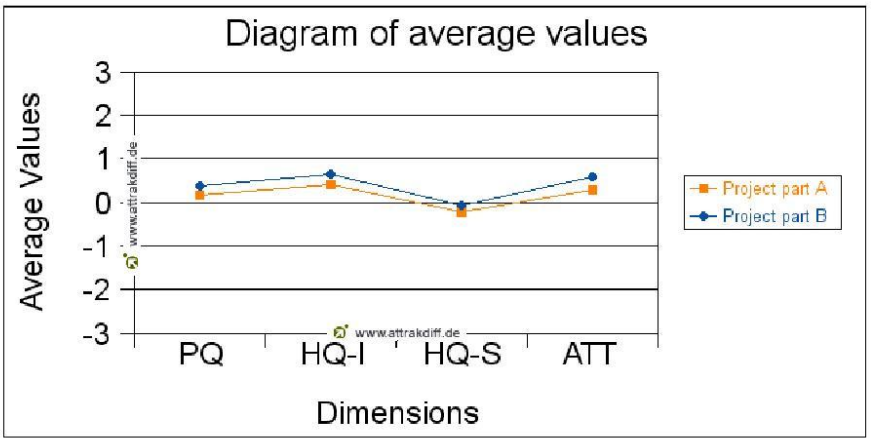


Diagram 2: Mean values of the four AttrakDiff™ dimensions for the products "website 1" (project part A) and "website 2" (project part B)

Interpretation for help

Project part A, product "website 1"

In terms of pragmatic quality the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to assist the user you must aim at improvement.

With regard to hedonic quality – identity, the product is located in the average region. It provides the user with identification and thus meets ordinary standards.

Result: Should you wish to bind the user more strongly to the product, you must aim at improvement.

With regard to hedonic quality – stimulation, the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to motivate, absorb and stimulate users, you must aim at improvement.

The product's attractiveness value is located in the average region.

Result: The overall impression of the product is moderately attractive.

Project part B, product "website 2"

In terms of pragmatic quality the product is located in the average region. It meets ordinary standards.

Result: Should you wish to provide the user with really great assistance you must strive to improve the quality even more so.

With regard to hedonic quality – identity, the product is located in the average region. It provides the user with identification and thus meets ordinary standards.

Result: Should you wish to bind the user more strongly to the product, you must aim at improvement.

With regard to hedonic quality – stimulation, the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to motivate, absorb and stimulate users, you must aim at improvement.

The product's attractiveness value is located in the average region.

Result: The overall impression of the product is moderately attractive.

Comparison of results of both project parts

In terms of pragmatic quality, product website 2 performs better than product website 1. This difference is however statistically insignificant.

In terms of the identity aspect of hedonic quality, product website 2 performs better than product website 1. This difference is however statistically insignificant.

In terms of the stimulation aspect of hedonic quality, product website 2 performs better than product website 1. This difference is however statistically insignificant.

In terms of attractiveness, product website 2 performs better than product website 1. This difference is however statistically insignificant.

Description of word-pairs

The mean values of the word pairs are presented here. Of particular interest are the extreme values. These show which characteristics are particularly critical or particularly well-resolved.

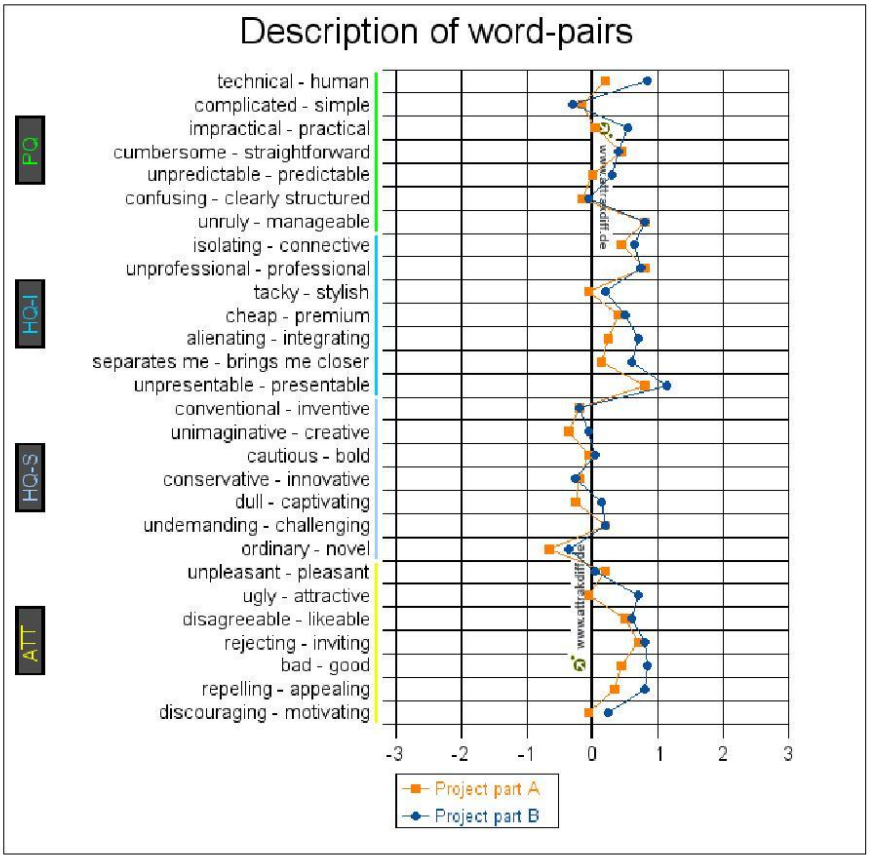


Diagram 3: Mean values of the AttrakDiff™ word pairs for products "website 1" (project part A) and "website 2" (project part B)

APPENDIX

Classification of test participants

Age

20 to 40:	20 test participants

Gender

Male:	6 test participants
Female:	14 test participants

School qualification

Higher Secondary Education:	1 test participants
University:	19 test participants

Profession

.:	1 test participants
.:	1 test participants
.:	1 test participants
.:	1 test participants
Student:	1 test participants
..:	1 test participants
..:	1 test participants
...:	1 test participants
jkj:	1 test participants
..:	1 test participants
Student:	1 test participants
Student:	1 test participants
..:	1 test participants
Student:	1 test participants
.:	1 test participants
.:	1 test participants
Student:	1 test participants
Bsc:	1 test participants
.:	1 test participants
cw:	1 test participants

Product experience
less than a month: 20 test participants

Confidence Intervals

The confidence intervals create a so-called confidence rectangle. As it is almost impossible to involve all users in the evaluation.

The project co-ordinator has to settle for a number of selected product users to evaluate the product. For this reason one can never be 100% sure that the outcome of the evaluation is representative of the collective users. It might be that the evaluation by the selected users differ from that of the collective users were it possible to ask them all.

The confidence interval outlines the area where the "true" values would lie were it possible to ask all the users.

The confidence rectangle suggests with what certainty the product equals the mean values of the characteristic dimensions.

Significance Tests

Significance tests make it possible to test whether the difference between 2 values can be attributed to the qualities of the product or whether the difference is the result of incidental fluctuations. E.g. If a product receives a higher pragmatic rating than another it does not necessarily mean that it is more pragmatic than the other.

Small, chance fluctuations of judgement can result in a higher value even when there is no systematic difference between the two products. In this case the difference measured is not very relevant.
T-Tests for independent random sampling to check whether there are in fact significant rating differences between the two products. The significance standard lies at 0,05.

This is interpreted as follows:
The difference in ratings is considered "significant" when one can assume with 95% certainty that there are no incidental fluctuations. A difference is considered "insignificant" when the probability of incidental fluctuation is greater than 5%.

Evaluation Report for Products

"website 1" and "website 2"

Objectives:
How user-friendly and attractive is this product?

Contents of report

Method of investigation

AttrakDiff™ is an instrument for measuring the attractiveness of interactive products.

With the help of pairs of opposite adjectives, users (or potential users) can indicate their perception of the product. These adjective-pairs make a collation of the evaluation dimensions possible.

The following product dimensions are evaluated:

Pragmatic Quality (PQ):
- Describes the usability of a product and indicates how successfully users are in achieving their goals using the product.

- Hedonic quality - Stimulation (HQ-S):
  Mankind has an inherent need to develop and move forward. This dimension indicates to what extent the product can support those needs in terms of novel, interesting, and stimulating functions, contents, and interaction- and presentation-styles.

- Hedonic Quality - Identity (HQ-I):
  Indicates to what extent the product allows the user to identify with it.

- Attractiveness (ATT):
  Describes a global value of the product based on the quality perception.

Hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

Characteristics of investigation

Product title A:	website 1
Product title B:	website 2
Product industry:	Services
Duration of study A:	04.07.2012 - 02.10.2012
Duration of study B:	04.07.2012 - 02.10.2012
Project-type:	Comparison product A - product B, that means two different products are rated.
Variant:	The same test participants in both project parts.
Number of ratings for A:	20
Number of ratings for B:	20

Portfolio of results

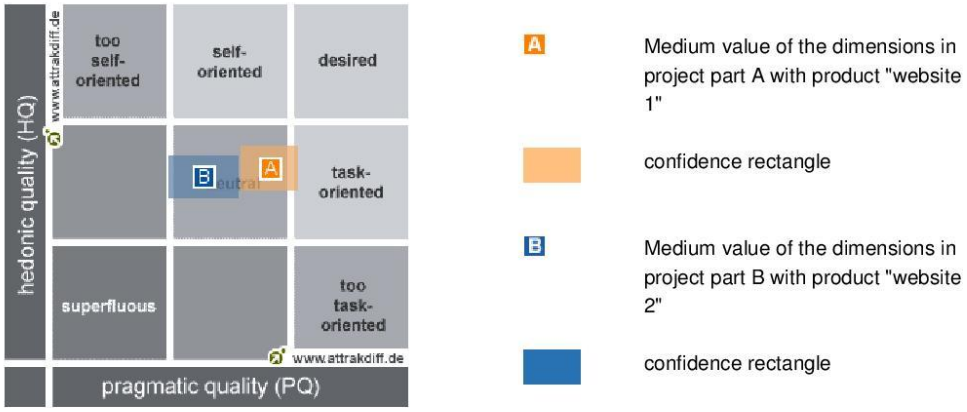


Diagram 1: Portfolio with average values of the dimensions PQ and HQ and the respective confidence rectangles of products "website 1" and "website 2"

In the portfolio-presentation the values of hedonic quality are represented on the vertical axis (bottom = low value). The horizontal axis represents the value of the pragmatic quality (i.e. left = a low value).

Depending on the dimensions values the product will lie in one or more "character-regions".

The bigger the confidence rectangle the less sure one can be to which region it belongs. A small confidence rectangle is an advantage because it means that the investigation results are more reliable and less coincidental.

The confidence rectangle shows, if the users are at one in their evaluation of the product.
The bigger the confidence rectangle, the more variable the evaluation ratings (more information is available in the appendix).

Interpretation for help

Project part A, product "website 1"

The products user interface was rated as "neutral".

The classification here is not clearly "pragmatic" because the confidence interval overlaps into the neighbouring character zone. The user is assisted by the product, however the value of pragmatic quality only reaches the average values.

Result: Consequently there is room for improvement in terms of usability.

In terms of hedonic quality the character classification applies positively. The user is stimulated by this product, however the hedonic value is only average.

Result: In terms of hedonic quality there is clearly room for improvement.

Attention! The values for HQ-I and HQ-s differ greatly. This differentiation is illustrated in the diagram of mean values.

The confidence rectangle is small. The users are at one in their evaluation of the product.

Project part B, product "website 2"

The products user interface was rated as "neutral".

The classification here is not clearly "pragmatic" because the confidence interval overlaps into the neighbouring character zone. The user is assisted by the product, however the value of pragmatic quality only reaches the average values.

Result: Consequently there is definite room for improvement.

In terms of hedonic quality the character classification applies positively. The user is stimulated by this product, however the hedonic value is only average.

Result: In terms of hedonic quality there is clearly room for improvement.

The confidence interval PQ is large. This could be attributed to limited sampling or to greatly differing product ratings.

Comparison of results of both project parts

Product website 1 performs better than product website 2. Pragmatic quality as well as hedonic quality is higher.

The difference in value of the pragmatic quality of product website 1 and product website 2 is statistically significant (see details in appendix).
The difference in value of the hedonic quality of product website 1 and product website 2 is statistically insignificant. It might therefore concern a chance fluctuation of judgement.

The confidence interval for pragmatic quality of product website 1 is smaller than for product website 2. In the rating of pragmatic quality of product website 1 the users are more at one in their evaluation. Thus the rating of product website 1 applies with greater certainty to the product.
The confidence interval for hedonic quality of product website 2 is smaller than for product website 1. In the rating of hedonic quality of product website 2 the users are more at one in their evaluation. Thus the rating of product website 2 applies with greater certainty to the product.

Diagram of average values

The average values of the AttrakDiff™ dimensions for the evaluated product are plotted on the diagram.

In this presentation hedonic quality distinguishes between the aspects of stimulation and identity. Furthermore the rating of attractiveness is presented.

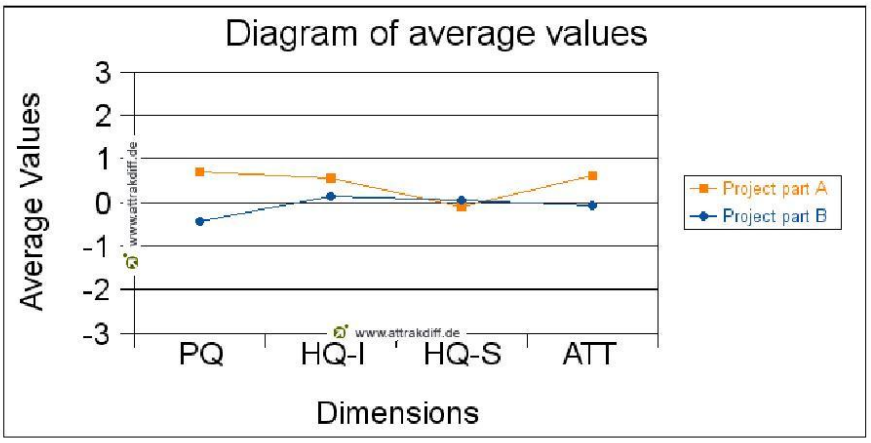


Diagram 2: Mean values of the four AttrakDiff™ dimensions for the products "website 1" (project part A) and "website 2" (project part B)

Interpretation for help

Project part A, product "website 1"

In terms of pragmatic quality the product is located in the average region. It meets ordinary standards.

Result: Should you wish to provide the user with really great assistance you must strive to improve the quality even more so.

With regard to hedonic quality – identity, the product is located in the average region. It provides the user with identification and thus meets ordinary standards.

Result: Should you wish to bind the user more strongly to the product, you must aim at improvement.

With regard to hedonic quality – stimulation, the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to motivate, absorb and stimulate users, you must aim at improvement.

The product's attractiveness value is located in the average region.

Result: The overall impression of the product is moderately attractive.

Project part B, product "website 2"

In terms of pragmatic quality the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to assist the user you must aim at improvement.

With regard to hedonic quality – identity, the product is located in the average region. It just meets ordinary standards.

Result: Should you wish to bind the user to the product, you must aim at improvement.

With regard to hedonic quality – stimulation, the product is located in the average region. It just about meets ordinary standards.

Result: Should you wish to motivate, absorb and stimulate users, you must aim at improvement.

The product's attractiveness value is located in the average region.

Result: The overall impression of the product is moderately attractive.

Comparison of results of both project parts

In terms of pragmatic quality, product website 2 performs less well than product website 1. This difference is statistically significant.

In terms of the identity aspect of hedonic quality, product website 2 performs less well than product website 1. This difference is however statistically insignificant.

In terms of the stimulation aspect of hedonic quality, product website 2 performs better than product website 1. This difference is however statistically insignificant.

In terms of attractiveness quality, product website 2 performs less well than product website 1. This difference is however statistically insignificant.

Description of word-pairs

The mean values of the word pairs are presented here. Of particular interest are the extreme values. These show which characteristics are particularly critical or particularly well-resolved.

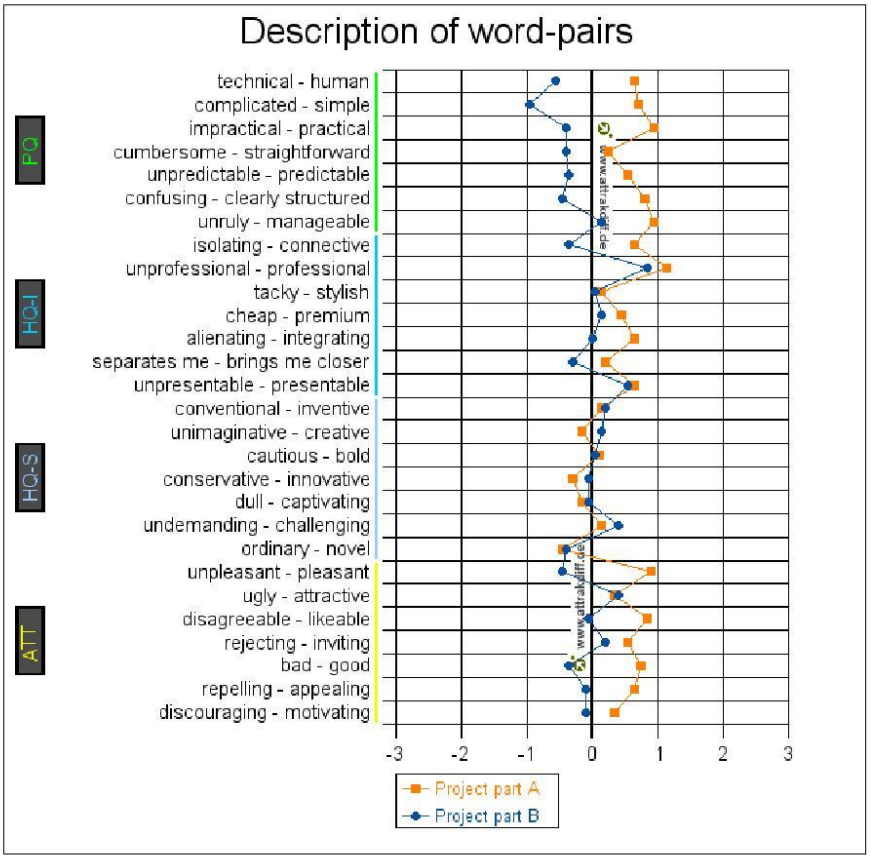


Diagram 3: Mean values of the AttrakDiff™ word pairs for products "website 1" (project part A) and "website 2" (project part B)

APPENDIX

Classification of test participants

Age

under 20:	1 test participants
20 to 40:	19 test participants

Gender

Male:	12 test participants
Female:	8 test participants

School qualification

University:	20 test participants

Profession

.:	1 test participants
.:	1 test participants
.:	1 test participants
.:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student Psychologie:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student Psychologie:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student:	1 test participants
Student Psychologie:	1 test participants
Student:	1 test participants
Student:	1 test participants

Product experience
less than a month: 19 test participants
less than a year: 1 test participants

Confidence Intervals

The confidence intervals create a so-called confidence rectangle. As it is almost impossible to involve all users in the evaluation.

The project co-ordinator has to settle for a number of selected product users to evaluate the product. For this reason one can never be 100% sure that the outcome of the evaluation is representative of the collective users. It might be that the evaluation by the selected users differ from that of the collective users were it possible to ask them all.

The confidence interval outlines the area where the "true" values would lie were it possible to ask all the users.

The confidence rectangle suggests with what certainty the product equals the mean values of the characteristic dimensions.

Significance Tests

Significance tests make it possible to test whether the difference between 2 values can be attributed to the qualities of the product or whether the difference is the result of incidental fluctuations. E.g. If a product receives a higher pragmatic rating than another it does not necessarily mean that it is more pragmatic than the other.

Small, chance fluctuations of judgement can result in a higher value even when there is no systematic difference between the two products. In this case the difference measured is not very relevant.
T-Tests for independent random sampling to check whether there are in fact significant rating differences between the two products. The significance standard lies at 0,05.

This is interpreted as follows:
The difference in ratings is considered "significant" when one can assume with 95% certainty that there are no incidental fluctuations. A difference is considered "insignificant" when the probability of incidental fluctuation is greater than 5%.