Card Sorting Methoden für den benutzungsfreundlichen Fahrscheinautomaten

Marco Santi
Fraunhofer IAO
Nobelstraße 12
70569 Stuttgart
Marco.Santi@iao.fraunhofer.de
www.hci.iao.fraunhofer.de

Iris Niedermann
Dattelweg 22
70619 Stuttgart
Iris_Niedermann@web.de

Abstract

Dieser Artikel präsentiert das methodische Vorgehen und die Ergebnisse eines zweistufigen Card Sorting Ansatzes. Die vorgestellte Studie ist Teil eines größeren Projektes mit dem Ziel der Neukonzeption der Fahrkartenautomaten der DB. Eine in einem ersten Schritt durch ein offenes Card Sorting gefundene Fahrkarten-Kategorienstruktur wird mit einem geschlossenen Card Sorting validiert und gegen eine durch Experten entwickelte Struktur verglichen. Die Ergebnisse deuten auf die Verwendung von acht Kategorien auf der ersten Ebene hin, wobei einige Fahrkarten den Kategorien redundant zugeordnet werden sollten. Die Kategorienamen sollten so konkret wie möglich sein, da Entscheidungszeiten im Vergleich zu abstrakten Kategoriennamen kürzer waren.

Keywords

Card Sorting, Fahrscheinautomat, User Centred Design, Informationsstruktur, Redundanzen

1.0 Einleitung

Das Fraunhofer Institut für Arbeitswirtschaft und Organisation (IAO) wurde von der DB Personenverkehr GmbH mit der Neukonzeption der Fahrscheinautomaten beauftragt. Der vorliegende Artikel greift einen Aspekt aus diesem Projekt heraus.

1.1 Hintergrund

Der öffentliche Personennahverkehr in Deutschland ist so gestaltet, dass Fernverkehrsverbindungen vornehmlich durch die Deutsche Bahn (DB) und der lokale Verkehr durch 70 unabhängig operierende Verkehrsverbünde abgedeckt wird.

Da in den meisten Gebieten sowohl Fernverkehrszüge als auch die Züge der Verkehrsverbünde die gleiche Infrastruktur benutzen, gibt es 2335 Fahrscheinautomaten, die sowohl Fahrkarten der DB also auch der lokalen Verkehrsverbünde anbieten.

Fahrgäste entscheiden auf dem Startbildschirm bisher, ob sie (1) ein Ziel eingeben wollen, (2) einen DB-Fahrschein oder (3) ein Verbundfahrschein kaufen wollen.

Besonders Fahrgäste ohne Wissen über die lokale Tarifstruktur haben Probleme, sich für eine dieser Option zu entscheiden. In einem Nutzertest mit 42 Fahrgästen, bei dem pro Versuchsperson 10 Fahrscheine einer dieser drei Optionen zugeordnet werden sollten, wurde für 46 % der Fahrscheine von tarifunkundigen Fahrgästen und für 26 % der Fahrscheine von tarifkundigen Fahrgästen eine falsche Option gewählt. Darüber hinaus gaben 72 % der Befragten an, dass es ihnen bei einer Fahrt von A nach B unwichtig ist zu wissen, ob es sich bei dem Fahrschein um einen DB- oder VVS-Fahrschein handelt.

1.2 Neukonzeption

Diese Befunde haben zu einer der zentralen Konzeptprinzipien des Gesamtprojektes geführt: Fahrgäste sollen nicht entscheiden müssen, ob der gewünschte Fahrschein vom Anbieter A oder Anbieter B verkauft wird. Stattdessen wird angenommen, dass Fahrgäste zwei Motiven beim Fahrkartenkauf folgen.

- (1) „Ich möchte von A nach B fahren“
- (2) „Ich weiß genau, welchen Fahrschein ich kaufen möchte“.

Im ersten Fall erwarten die Fahrgäste, dass der Automat den richtigen Fahrschein ausgibt, nachdem notwendige Informationen eingegeben wurden. Im zweiten Fall ist das Fahrziel zweitrangig. Die Fahrgäste wissen genau, welchen Fahrschein sie kaufen möchten und kennen die tariflichen Bestimmungen.

Im vorliegenden Artikel wird vorgestellt, wie mit einem zweistufigen Card Sorting Vorgehen eine Kategorienstruktur für Fahrkarten entwickelt wurde, die es den tarifkundigen Fahrgästen ermöglicht, ihren Fahrschein schnell und sicher zu finden.

2.0 Methode

Card Sorting ist eine weit verbreitete Methode, mit der einzelne Elemente gruppiert werden können. „It is a great

reliable, inexpensive method for finding patterns in how users would expect to find content or functionality.“ (Maurer & Warfel 2004). Die meisten Autoren unterscheiden zwischen offenen und geschlossenen Cart Sorting Methoden (Bleiker 2005; Maurer & Warfel 2004; Moreville & Rosenfeld 2002). Während in einem offenen Cart Sorting:

“Participants are given cards showing content with no pre-established groupings. They are asked to sort cards into groups that they feel are appropriate and then describe each group. Open card sorting is useful as input to information structures in new or existing sites and products. “(Maurer & Warfel 2004).

In einem geschlossenen Card Sorting:

“[…] participants are given cards showing […] content with an established initial set of primary groups. Participants are asked to place cards into these pre-established primary groups. Closed card sorting is useful when adding new content to an existing structure, or for gaining additional feedback after an open card sort.“ (Maurer & Warfel 2004).

In Anlehnung an diese Beschreibung war die vorliegende Studie zweigeteilt. Im ersten Teil sollte eine Kategorienstruktur entwickelt werden. Deshalb wurde ein offenes Card Sorting durchgeführt. Im zweiten Teil wurde diese Struktur in einem geschlossenen Card Sorting validiert. Das geschlossene Card Sorting wurde modifiziert, so dass objektive Daten wie Zuordnungsfehler und Entscheidungszeiten zusätzliche Aussagen zur Güte der im ersten Teil entwickelten Struktur ermöglichten.

2.1 Stichprobe

Im ersten Teil nahmen 6 Männer und 4 Frauen mit einem mittleren Alter von 41 Jahren und einer Alterspanne von 20 bis 66 Jahren teil. Im zweiten Teil waren es 4 Männer und 6 Frauen mit einem Altersmittel von 37 Jahren und einer Altersspanne von 24 bis 57 Jahren.
Die Studie wurde innerhalb von einer Woche im Usability Labor des Competence Centers Human Computer Interaction des Fraunhofer IAO durchgeführt.

2.2 Vorauswahl der zu kategorisierenden Fahrkarten

Das Ziel der Studie war es, eine Kategorienstruktur für Fahrkarten zu entwickeln, die für alle Automaten in Deutschland gültig sein würde. Deshalb wurden im Vorfeld die Produkte verschiedener Verkehrsverbünde und der DB auf Ähnlichkeiten analysiert. Darauf aufbauend wurden 42 repräsentative Fahrscheine der DB und des Verkehrsverbundes Stuttgart (VVS) ausgewählt.

Jeder der 42 Fahrscheinnamen wurde auf eine Karte gedruckt. Da es für beide Teile der Studie sehr wichtig war, dass die Versuchspersonen sich mit den tariflichen Bestimmungen der Fahrscheine auskannten, wurden alle Fahrscheine im Vorfeld erläutert. Zusätzlich waren die tariflichen Bestimmungen auf der Rückseite einer jeden Karte abgedruckt.

2.3 Versuchsdurchführung

Der erste Teil der Gesamtstudie wurde durch ein offenes Card Sorting bestimmt. Zwei Versuchspersonen arbeiteten während zwei Stunden in einem Raum. Während der Kategorisierung wurden sie gebeten, nicht miteinander zu reden. Sie wurden instruiert, die Fahrscheine so zu kategorisieren, wie sie die Fahrscheine im Fahrscheinautomaten suchen würden. Die Karten wurden auf einem Tisch gruppiert und jede Gruppe durch einen Klebezettel gekennzeichnet. Die Benennung der Gruppen war möglich aber nicht notwendig. Die einzigen Beschränkungen waren, maximal 8 Gruppen auf jeder Kategorienebene zu bilden und die Kategorien „DB-Fahrscheine“ und „VVS-Fahrscheine“ auf der ersten Ebene zu vermeiden. Es war möglich, Fahrscheine redundant mehreren Kategorien zuzuordnen. Hierfür stand jede der 42 Fahrscheine dreifach zur Verfügung. Nach einer Stunde des Gruppierens der 42 Fahrscheine präsentierten beide Versuchspersonen ihre Lösungen. Der Grund für dieses Vorgehen lag darin, qualitative Informationen zur jeweiligen Lösung zu erhalten. Die Lösungen wurden schließlich fotografiert, um sie später analysieren zu können.

Im zweiten Teil wurden zwei Aspekte kombiniert: das geschlossene Card Sorting und die Testung der Kategorien durch die Erfassung von Entscheidungszeiten und Zuordnungsfehlern.
Es wurden zwei Kategorienstrukturen getestet. Die eine entstammte der Analyse des offenen Card Sorting des ersten Teils (Struktur CSE). Die zweite Struktur wurde im Vorfeld von Usability Experten und Fahrscheinexperten in einem internen Workshop erarbeitet (Struktur WSE). Diese Struktur zeichnete sich durch acht Kategorien auf der ersten Ebene und eine starke Verwendung von Redundanzen aus.
Beide Strukturen wurden nacheinander auf einem Touchscreen dargestellt (siehe Abb. 1). Der Versuchsleiter nannte den Namen des Fahrscheins. Die Aufgabe der Versuchsperson bestand darin, solange durch die Kategorienstruktur zu navigieren, bis der Fahrschein gefunden war. Es wurden die gleichen 42 Fahrscheine wie im ersten Teil verwendet. Die Navigationspfade wurden aufgezeichnet und konnten später bezüglich der Entscheidungszeiten und Zuordnungsfehler analysiert werden. Obwohl die aufgezeichneten Daten eine reiche Basis für ausführliche Auswertungen bot, lag der Fokus für die vorliegende

Studie auf den Kategorien der ersten Ebene.

Abb. 1. Versuchsperson navigiert mit einem Touchscreen durch eine im PC abgebildete Kategorienstruktur.

2.4 Clusteranalyse

Das offene Card Sorting aus dem ersten Teil resultierte in 10 unabhängigen Kategorienstrukturen. Eine Lösung musste von den statistischen Analysen ausgeschlossen werden. Die Versuchsperson war bei der Lösung von den Instruktionen abgewichen, so dass eine statische Analyse nicht möglich war.

Um die Kategorienstrukturen statistisch auszuwerten, wurden die verbleibenden neun Kategorienlösungen untersucht. Abb. 2 zeigt schematisch, wie dabei vorgegangen wurde. Der Abstand zwischen „Fahrschein A“ und „Fahrschein B“ beispielsweise wurde mit „0“ bewertet, da beide Fahrscheine in der gleichen Gruppe platziert sind. Um vom „Fahrschein B“ zum „Fahrschein D“ zu gelangen, müssen zwei „Klicks“ durchgeführt werden; einen, um den Inhalt der „Hauptkategorie 1“ und einen, um die Subkategorie 1.2“ zu erreichen.

Dieser Abstand wurde deshalb mit „2“ bewertet. Analog zu diesem Vorgehen beträgt der Abstand zwischen „Fahrschein B“ und „Fahrschein F“ „4“.
Die Abstände einer jeden Fahrkarten-Fahrkarten-Kombinationen bildeten eine Matrix, die als Grundlage für eine anschließende Clusteranalyse diente.

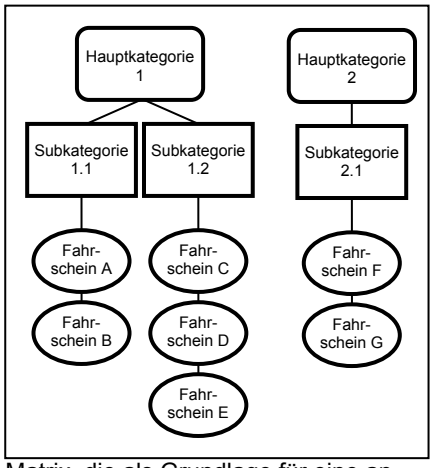

Abb. 2: Schematische Darstellung des Ergebnisses eines Card Sorting. Die Abstände zwischen den Fahrkarten waren die Basis für die clusteranalytischen Auswertungen.

Die statistischen Analysen wurden mit SPSS 14.0 für Windows, durchgeführt. Bei der Clusteranalyse wurde die Complete Linkage Methode angewendet (Bortz 1999).

3.0 Ergebnisse

3.1 Offenes Card Sorting

Obwohl die individuellen Kategorienstrukturen sehr verschieden voneinander ausfielen, war die durch die Clusteranalyse gewonnene Struktur klar und einfach zu interpretieren. Es konnten acht Kategorien auf der ersten Ebene identifiziert werden. Die Kategorien wurden von vier Usability Experten in einem internen Workshop benannt. Es ergaben sich fünf konkrete Kategoriennamen und drei abstrakte Kategoriennamen. Als konkret wurden solche Kategorien bezeichnet, bei denen der Kategorienname Teil des Fahrscheinnamens war. So beinhaltete beispielsweise die Kategorie „Gruppenkarten“ hauptsächlich Fahrscheine, die in ihrem Namen das Wortfragment „Gruppe“ trugen (bspw. „Gruppen-Tageskarte“). Im Gegensatz dazu gab es bei den abstrakten Kategoriennamen (bspw. „Freizeit und Veranstaltungen“) keine Wortfragmente, die sowohl im Kategorien- als auch im Fahrscheinnamen vorkamen.

Alle Versuchspersonen nahmen von der Möglichkeit Gebrauch, einzelne Fahrscheine mehreren Kategorien zuzuordnen. Insgesamt wurden 12 % der 42 Fahrscheine redundant zugeordnet. Die durch die Clusteranalyse resultierende Struktur wies bedingt durch den Analysealgorithmus keine Redundanzen mehr auf.

3.2 Geschlossenes Card Sorting

Ein direkter Vergleich der Entscheidungszeiten für eine der acht Kategorien gemessen in Sekunden zwischen den Strukturen CSE (M = 4.38, SD = 2.47) und WSE (M = 4.92, SD = 4.33), ergab keinen statisch signifikanten Unterschied, $t(41) = -1.53$, $p = .13$.

Auch die statistische Analyse der Zuordnungsfehler zeigte, dass sich die Strukturen CSE (51 Zuordnungsfehler) und WSE (61 Zuordnungsfehler) nicht signifikant voneinander unterschieden, $\chi^2(1, N = 112) = 0.83$, $p > .05$.

Die unterschiedliche Benennung der Kategoriennamen innerhalb der Struktur CSE machte eine Analyse der Entscheidungszeiten in Abhängigkeit von der Kategorienbezeichnung möglich. Der Vergleich zwischen konkreten und abstrakten Bezeichnungen resultierte in einem signifikanten Unterschied der Entscheidungszeiten, $t(9) = 2.78$, $p < .05$. Die Versuchspersonen benötigten bei abstrakten Kategorienbezeichnungen wie „Freizeit und Veranstaltungen“ oder „Service- und Zusatzkarten“ länger für eine Entscheidung (M = 4.64, SD = 1.60) als bei konkreten Bezeichnungen (M = 3.87, SD = 1.51) wie beispielsweise „Gruppenkarten“.

Eine weitere Frage bezog sich auf die Nutzung der in der Struktur WSE vorgesehenen Redundanzen. Insgesamt wurden 27 % der zu mehreren Kategorien zugeordneten Fahrscheine von mindestens einer Versuchsperson genutzt.

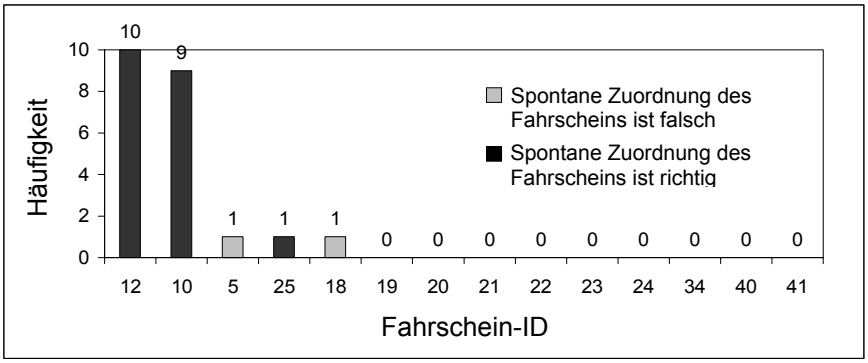

Abb. 3. Häufigkeiten der spontanen Suche der Fahrscheine (Fahrschein-ID) in der ersten Kategorienebene „Gruppenkarten Eine Häufigkeit von „0“ bedeutet, dass keiner der Versuchspersonen diesen redundanten Fahrschein in dieser Kategorie gesucht haben.

Abb. 3 zeigt, wie bei der Analyse vorgegangen wurde. Alle Argumente mit einer Häufigkeit von Null zeigen, dass diese Fahrscheine zwar Bestandteil dieser Kategorie waren, aber keine der Versuchspersonen diese Fahrscheine in dieser Kategorie gesucht haben. Diagramme, wie in Abb. 3 wurden ebenfalls genutzt um herauszufinden, welche Fahrschein-Kategorien-Kombination sehr bedeutsam zu sein schien (hohe Häufigkeiten spontaner Auswahl, schwarze Säulen) und welche Kategorienbezeichnung falsche Erwartungen weckte (hellgraue Säulen).

Derartige Analysen wurden für beide Kategorienstrukturen vorgenommen, um ein tieferes Verständnis der Bezeichnung der Kategorien und der Zuordnung der Fahrscheine zu erhalten.

4.0 Diskussion und weiteres Vorgehen

Mit Blick auf die Ergebnisse scheint die Verwendung von acht Kategorien auf der ersten Ebene angezeigt. Sowohl im offenen Card Sorting (CSE) als auch die von Experten im Vorfeld der Nutzertests entwickelten Kategorienlösungen (WSE) ergaben sich acht Kategorien auf der ersten Ebene.

Es hat sich gezeigt, dass konkrete Kategorienbezeichnungen zu kürzeren Entscheidungszeiten führen. Die Ergebnisse lassen sich so interpretieren, dass abstrakte Gruppenbezeichnungen einen kognitiven Mehraufwand erfordern. Dieser Mehraufwand äußert sich in längeren Entscheidungszeiten. Obwohl keine konkreten Untersuchungen zum subjektiven Empfinden gemacht wurden, lässt sich vermuten, dass konkrete Bezeichnungen den abstrakte Bezeichnungen subjektiv bevorzugt werden. Das subjektive Empfinden ist eine sehr zentrale Größe bei der Interaktion mit Fahrkartenautomaten. Deshalb sollte die Bezeichnung von Kategorien so konkret wie möglich erfolgen.

Die Ergebnisse der Studie unterstützen ebenfalls die Verwendung von Redundanzen. Die verbleibende Frage ist, in welchem Ausmaß. Im offenen Card Sorting haben die Versuchspersonen moderat von der Möglichkeit zur mehrfachen Zuordnung eines Fahrscheins zu verschiedenen Kategorien Gebrauch gemacht. Die Analyse der Struktur WSE, bei der Redundanzen stark zur Verfügung standen, hingegen zeigt, dass einige der Redundanzen überhaupt nicht genutzt wurden und deshalb eingeschränkt werden könnten. Für Walk-Up-And-Use Systeme wie Fahrscheinautomaten sollte jeder Versuch, den gewünschten Fahrschein auf Anhieb zu finden erfolgreich sein, ohne dass der Fahrgast in Kategorienstrukturen vor- und zurück navigieren muss. Diese Forderung und die Existenz verschiedener mentaler Modelle auf Seiten der Fahrgäste macht die Verwendung von Redundanzen notwendig. Der Nachteil liegt im Anstieg der Komplexität der Kategorienstruktur, was wiederum die Erlernbarkeit erschwert. Deshalb wird in den späteren Projektphasen die Verwendung von Redundanzen optimiert und die Erkenntnisse diese Studie angewendet werden, um eine weiter verbesserte Kategorienstruktur zu entwickeln, mit der alle Fahrkarten deutschlandweit gruppiert werden können. Ein iteratives Vorgehen unter kontinuierlicher Einbeziehung von Nutzern ist deshalb zu berücksichtigen.

Um die Verständlichkeit der Kategorieninhalte weiter zu verbessern, werden im nächsten Schritt Symbole für die Kategorien entwickelt und iterativ getestet werden.

Obwohl hier nur ein Ausschnitt der Ergebnisse dargestellt wurde und der Fokus auf der ersten Ebene der Kategorienstrukturen lag, lassen sich bezüglich des methodischen Vorgehens einige generelle Schlussfolgerungen ziehen. Die Kombination aus einem offenen Card Sorting zur Entwicklung einer nutzerzentrierten Informationsstruktur und einem geschlossenen Card Sorting zur anschließenden Validierung, erwies sich als sehr effektiv und effizient. Besonders die Kombination des geschlossenen Card Sortings mit der Aufzeichnung des Nutzerverhaltens bot große Vorteile bei der Auswertung. Spontane und intuitive

Verhaltensweisen, die zum gewünschten Erfolg führen, formen maßgeblich das Benutzungserleben. Die Möglichkeit, Entscheidungszeiten und Zuordnungsfehler objektiv zu erfassen, stellt eine deutliche Erweiterung der Aussagefähigkeit von geschlossenen Card Sorting dar.

Die gewonnen Erkenntnisse zur Gestaltung der Struktur zur Kategorisierung von Fahrkarten bildet eine gute Basis für die Weiterentwicklung der Struktur.

5.0 Einschränkungen

Trotz der Vorteile ergeben sich einige Einschränkungen insbesondere für das offene Card Sorting. Zunächst konnte nur ein Ausschnitt der verfügbaren Fahrscheine verwendet werden. Die Verwendung weiterer oder einfach nur anderer Fahrscheine hätte zu einer anderen Kategorienstruktur führen können. Es ist deshalb unbedingt erforderlich, dass bei offenen Card Sorting Methoden die zu gruppierenden Materialien eine sehr gute Repräsentation der Gesamtheit aller Elemente darstellt. Anderenfalls könnten die resultierenden Kategorienlösungen verfälscht sein. Leider sind der Anzahl der im offenen Card Sorting verwendbaren Elemente Grenzen gesetzt. Einige der Versuchspersonen äußerten, dass sie selbst nach einer Stunde intensiver Arbeit mehr Zeit gebraucht hätten, um eine für sie zufriedenstellende Lösung zu generieren. Neben der zeitlichen Einschränkung muss auch betont werden, dass die Komplexität der Strukturen mit der Anzahl der zu gruppierenden Elemente wächst.
Weiterhin waren die meisten der Versuchspersonen mit dem lokalen Tarifsystem und dem aktuellen Fahrscheinautomaten vertraut. Dadurch waren sie es gewohnt, zwischen der DB und dem VVS zu unterscheiden. Diese Unterscheidung schien für einige Versuchspersonen so prominent zu sein, dass sie sie deutliche Schwierigkeiten hatten, diese Kategorien entsprechend der Instruktion nicht zu bilden. Schließlich war es durch die Verwendung der Clusteranalyse nicht möglich, die von den Versuchspersonen verwendeten Redundanzen zu erhalten. Die resultierende Struktur war insofern nur eingeschränkt eine Aggregation der in die Analyse eingehenden Lösungen.

5.1 Literatur

Bleiker, A. (2005). Usability News >> Card Sorting - Viele Wege führen zu einer guten Informationsarchitektur. Letzter Zugriff: 03.11. 2006: http://www.usability.ch/News_D_Textonly/T_cards.htm

Bortz, J. (1999). Statistik für Sozialwissenschaftler. (5. Auflage). Berlin: Springer.

Maurer, D., & Warfel, T. (2004). Card sorting: A definitive guide. Letzter Zugriff 02.01.2007: http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide

Morville, P. & Rosenfeld, L. (2002). Information architecture for the world wide web. Sestabol, CA: O'Reilley and Associates, Inc.