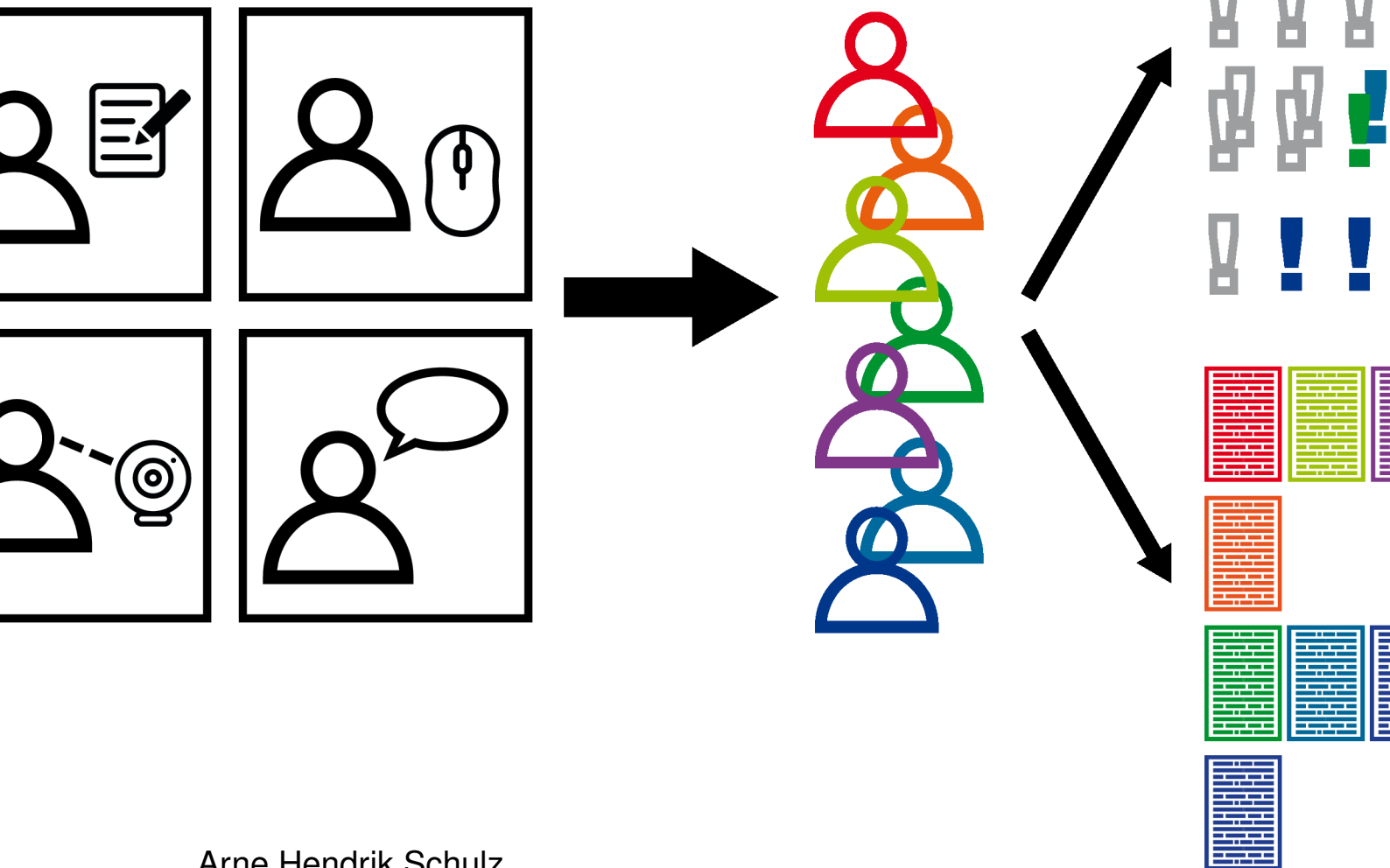


Usability in digitalen Kooperationsnetzwerken

Nutzertests und Logfile-Analyse
als kombinierte Methode



Dissertation von Arne Hendrik Schulz

zur Erlangung des Doktorgrades der Ingenieurwissenschaften (Dr.-Ing.), verliehen durch
den Fachbereich 3 (Mathematik und Informatik) der Universität Bremen.

Datum der Einreichung: 14.12.2015

Datum des Kolloquiums: 18.02.2016

1. Gutachter: Prof. Dr. Andreas Breiter (Universität Bremen)
2. Gutachterin: Prof. Dr. Ulrike Lucke (Universität Potsdam)

Für Kati

Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen, um mich bei einigen Personen zu bedanken, die mich während meiner Promotion begleitet und unterstützt haben. Der Entstehungsprozess einer Dissertation verläuft in der Regel nicht geradlinig, sondern ist mit vielen Hindernissen und Irrwegen gespickt. Erst zum Schluss wird ein Überblick erkennbar und lässt das Ergebnis in greifbare Nähe rücken. Es sei mir verziehen, wenn ich jemanden, der mich auf meinem Weg begleitet haben sollte, unerwähnt lasse.

Zunächst gilt mein Dank meinem Doktorvater Andreas Breiter für die großartige Betreuung während der Promotion. Er stand mir mit Ratschlägen und Anregungen zur Seite und hat sich stets Zeit für meine Fragen genommen. Gleichzeitig hat er mich beharrlich ermutigt, meinen Weg weiter zu verfolgen und mich nicht von Rückschlägen und Problemen verunsichern zu lassen. Ohne seine Hilfe und Betreuung wäre diese Dissertation heute nicht das, was sie ist.

Daneben bedanke ich mich herzlich bei Ulrike Lucke, die sich sofort bereit erklärt hat, meine Arbeit ebenfalls zu begutachten. Ihre Hinweise und ihr Feedback haben mir sehr weitergeholfen und eine noch deutlichere Fokussierung des Themas ermöglicht.

Während der Promotion haben mich auch meine Kolleginnen und Kollegen am „Institut für Informationsmanagement Bremen“ unterstützt und motiviert. Besonders bedanken möchte ich mich bei Anja, Dorothee, Jan, Jens und Stefan für ihre Zeit, Hilfestellungen und schier endlose Geduld bei der Beantwortung meiner (Nach-)Fragen. In diversen Diskussionen haben sie mir geholfen, einen kritischen Blick auf meine Dissertation zu werfen und diese inhaltlich weiter zu schärfen. Danken möchte ich auch Anja, Jan, Julia und Juliane für das inhaltliche Korrekturlesen meiner Arbeit.

Das Graduiertenkolleg „Advances in Digital Media“ hat mich herzlich als assoziierter Doktorand aufgenommen. Die beteiligten Hochschullehrerinnen und Hochschullehrer sowie Doktorandinnen und Doktoranden haben mir die Augen dafür geöffnet, was es heißt, zu promovieren. Gerade in der Anfangszeit meiner Promotion waren sie mir eine hilfreiche Stütze und kritischer Diskussionspartner, die mich bei der Formulierung und Eingrenzung meines Themas unterstützt haben. Hervorheben möchte ich an dieser Stelle Benjamin Walther-Franks und Rainer Malaka.

Entscheidend für die Empirie waren die Lehrerinnen und Lehrer, die sich bereit erklärt haben, an den Nutzertests teilzunehmen. Ohne ihre aufgebrachte Zeit wäre die Dissertation in dieser Form nicht möglich gewesen. Während der Datenauswertung konnte ich auf die Ressourcen des „Norddeutschen Verbunds für Hoch- und Höchstleistungsrechnen“ (HLRN) zurückgreifen, um umfangreiche Berechnungen durchzuführen. Wenngleich die Ergebnisse nicht in die Arbeit eingeflossen sind, möchte ich mich für die kostenlose Bereitstellung beziehungsweise Verlängerung meines Zugangs bedanken.

Weiterhin möchte ich mich bei allen anderen Personen bedanken, die mich während meiner Promotion unterstützt haben oder durch die eine oder andere Ablenkung dafür gesorgt haben, dass ich wieder Energie tanken konnte. Unter anderem sind hier Marie, Kai-Uwe und Verena, Juliane und Peter, Ceyhun, Stephanie und Benjamin, Christoph und Nelli zu nennen. Weitere Freunde und Familie, die ich hier nicht aufgeführt habe, dürfen sich gerne angesprochen fühlen.

Ein großer Dank gebührt meinen Eltern und meinen Schwiegereltern in spe. Durch ihren stetigen Zuspruch und ihr Vertrauen waren und sind sie eine große Stütze für mich, für diese Dissertation und darüber hinaus.

Zuletzt möchte ich mich bei Kati bedanken. Sie hat mich während der gesamten Zeit ohne eigene Rücksichtnahme unterstützt und dabei auf zahlreiche gemeinsame Stunden verzichtet, die ich am Schreibtisch verbracht habe. Gleichzeitig war sie immer ein umsichtiger und inspirierender Gesprächspartner mit dem ich viele Aspekte dieser Dissertation besprechen konnte. Durch ihr positives Denken und ihren Zuspruch hat sie mich stets aufs Neue motiviert. Ohne sie würde es die Dissertation nicht geben. Ich widme sie daher ihr!

Zusammenfassung

Bei der Entwicklung und Konzeption von neuen Anwendungen spielt die Usability (dt. Gebrauchstauglichkeit) eine wichtige Rolle. Anwendungen sollten in der Bedienung möglichst effizient, effektiv und einnehmend sein. Eine hohe Fehlertoleranz und eine einfache Erlernbarkeit sprechen ebenfalls für eine gute Usability. Die Messung der Usability erfolgt durch verschiedene, zumeist qualitative Methoden innerhalb der Produktentwicklung. Allen Methoden ist gemein, dass sowohl Planung, Durchführung als auch Auswertung zeitaufwendig sind. Bei der Einbeziehung von Endanwenderinnen und Endanwendern wird häufig auf Nutzertests zurückgegriffen. Durch den hohen Aufwand werden diese häufig mit zehn und weniger Testerinnen und Testern durchgeführt.

An diesem Punkt setzt diese Dissertation an. Sie untersucht, ob sich Nutzertests und Logfile-Analysen kombinieren lassen. Dazu werden die testweise in einem Lern-Management-System (LMS) durchgeführten Nutzertests im Hintergrund protokolliert und den gefundenen schwerwiegenden Nutzerproblemen zugeordnet. Anschließend erfolgte eine Kombination mit Logfiles aus einem Produktivsystem derselben Basis. Die „realen“ Logfiles protokollieren das Verhalten von 133 Nutzerinnen und Nutzern und erstrecken sich über einen Zeitraum von 300 Tagen. Vor der Kombination mittels der Kosinus-Ähnlichkeit wurden beide Logfiles in ein Format überführt, das einen Vergleich ermöglichte. Da es sich um ein neues Verfahren handelt, musste die endgültige Ähnlichkeit über deskriptive Statistiken und visuelle Inspektionen festgelegt werden. Die finale Kombination ermöglicht die Bestimmung der Schwere der gefundenen Usability-Probleme anhand „realer“ Nutzungsdaten. Die Methode zeigt ein genaueres Bild über das Auftreten der gefundenen Probleme außerhalb der Testsituation. Die Dissertation liefert einen Mehrwert in den Disziplinen des (Web) Data Mining, der Usability sowie der Human-Computer Interaction (HCI). Daneben werden Disziplinen wie die Softwareentwicklung, die quantitative und qualitative Sozialforschung sowie die Gebiete der rechnergestützten Gruppenarbeit (Computer-Supported Cooperative Work; CSCW) und Lern-Management-Systeme (LMS) tangiert.

Abstract

Usability is a key factor when developing new applications. The interaction between the users and the application should be efficient, effective and engaging. Furthermore, a good usability includes a high error tolerance and an good learnability. Different methods allow the measurement of usability throughout the development (process). All methods have in common that the different employed steps like planning, conducting and evaluating are rather time-consuming. When end-users are included as subjects, usability tests are employed. Due to the high time-effort, usually ten or less tests are conducted.

The thesis tries to solve this point by trying to combine usability tests and logfile analysis. The empirical work is two-folded. First, usability tests within a learning management system (LMS) are logged in the background. These logfiles are assigned to severe usability problems. Second, the paths of the severe usability problems are combined with logfile data from a “real-world” LMS that runs the same application. The “real-world” logfiles contain a period of about 300 days with 133 active users. Prior to the combination, both data sets converted into a similar format. Being a new procedure, the definite similarity value had to be specified by descriptive statistics and visual inspections. The final combination makes it possible to determine the severity of usability problems on the basis of “real-world” usage data. The proposed method offers a more precise overview of the occurrence of the found usability problems, independent of the test situation. This thesis provides additional value to the fields of (Web) Data Mining, Usability and Human-Computer Interaction (HCI). It also offers additional knowledge to the field of software development, quantitative and qualitative research as well as computer-supported cooperative work (CSCW) and learning management systems (LMS).

Inhaltsverzeichnis

1	Einleitung	1
2	Herleitung der Fragestellung	5
3	Vorgehensweise	9
4	Logfiles als Erfassungsinstrument	13
5	Wissensentdeckung in Logfiles	21
5.1	Wissensentdeckung in Datenbanken	22
5.2	Vorgehen in der Logfile-Analyse	23
5.2.1	Datenaufbereitung	24
5.2.2	Mustererkennung	32
5.2.3	Interpretation	40
6	Usability als Qualitätskriterium	41
6.1	Anforderungen an eine „gute“ Usability	43
6.2	Vorstellung der verschiedenen Usability-Methoden	47
6.2.1	Analytische Usability-Evaluationen	50
6.2.2	Empirische Usability-Evaluationen	58
6.2.3	Einordnung der Usability-Methoden	62
6.3	Nutzertests als zentrale Methode	64
6.3.1	Vorgehen	65
6.3.2	Techniken zur Aufdeckung von Problemen	78
6.4	Bisherige Ansätze zur Verbindung von Logfiles und Usability	81
7	Logfiles als quantitative Datenquelle	83
7.1	Digitale Kooperationssysteme	83
7.2	Evaluiertes System	87
7.3	Datenerfassung und Datenaufbereitung	91
7.4	Nutzung des Systems	93
8	Nutzertests als qualitative Datenquelle	101
8.1	Entwicklung der Tests/Erstellung des Testplans	101
8.1.1	Gründe und Ziele für den Nutzertest/Forschungsfragen	101
8.1.2	Merkmale der Testpersonen	103
8.1.3	Methodologie	104
8.1.4	Testumgebung und Testausrüstung	105
8.1.5	Aufgabenliste des Nutzertests	106
8.1.6	Rolle des Moderators	107
8.1.7	Testablauf	108
8.2	Auswertung der Tests	109
8.3	Bewertung der gefundenen Usability-Probleme	125

9	Kombination von Nutzertests und Logfiles	131
9.1	Umwandlung der Logfiles in eine Transaction Matrix	131
9.2	Vom Problempfad zum problematischen Pfadmerkmal	137
9.3	Kombination von Pfadmerkmalen	141
9.3.1	Übersicht über die Zuordnungen	142
9.3.2	Festlegung der Kosinus-Ähnlichkeit	145
9.3.3	Einordnung der Ergebnisse	153
10	Fazit	157
A	Heuristiken	163
A.1	Heuristiken von Nielsen	163
A.2	Heuristiken von Muller et al. (1998)	164
A.3	Heuristiken von Shneiderman (eight golden rules; 2010)	165
A.4	Heuristiken von Sarodnick und Brau (2011)	166
A.5	Heuristiken von Mehlenbacher et al. (2005)	168
A.6	Heuristiken von Reeves et al. (2002)	170
B	Tabellen	173
B.1	Zugriffe auf das Produktivsystem im Zeitverlauf	173
B.2	Zugriffe auf die Rubriken und Funktionen des Produktivsystems	180
C	Usability	197
C.1	Flyer für Lehrkräfte	198
C.2	Handout an die Testerinnen und Tester	199
C.3	Fragebogen für die Testerinnen und Tester	201
C.4	Zusätzliche Abbildungen	204
D	Kombination	207
	Literatur	211

Tabellenverzeichnis

4.1	Felder des Extended Logfile Format	16
4.2	HTTP 1.1 Statuscodes	17
5.1	Methoden zur Identifizierung der Nutzerinnen und Nutzer	26
5.2	Cluster Validation Indices	40
6.1	Vergleich Usability Dimensionen	45
6.2	Vorgehensweisen bei Nutzertests	66
6.3	Skalen zur Schwere von Usability-Problemen	77
7.1	Zugriffe auf die Rubriken	96
8.1	Alter der Testerinnen und Tester	109
8.2	Unterrichtete Fächer der Testerinnen und Tester	110
8.3	Gefundene Probleme je Aufgabe	126
8.4	Beschreibung schwerwiegende Probleme	129
9.1	Beispiel eines Logfiles	132
9.2	Darstellung als Rub-Funk-Matrix	134
9.3	Darstellung als Komb-Matrix	134
9.4	Bewertung der schwerwiegenden Usability-Probleme	140
9.5	Identifizierte generische problematische Pfadmerkmale in den Produktiv- Logfiles	142
9.6	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,7	147
9.7	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,75	148
9.8	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,8	148
9.9	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,85	149
A.1	Heuristiken nach Mehlenbacher et al. (2005)	168
B.1	Zugriffe auf InfoChange im Zeitverlauf	173
B.2	Zugriffe auf die Rubriken	180
B.3	Ähnlichkeit der Pfade	181
B.4	Bearbeitungsdauer der Aufgaben	195
D.1	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,6	207
D.2	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,65	208
D.3	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,9	208
D.4	Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,95	209

Abbildungsverzeichnis

2.1	TAM-Modell	6
2.2	Zentrale Fragestellung	7
3.1	Ursprüngliches Vorgehen	10
3.2	Übersicht Vorgehen	10
3.3	Vorgehen (Phase 1)	12
3.4	Vorgehen (Phase 2)	12
4.1	Beispielhafte Seitenstruktur	14
4.2	Datenquellen für Logfiles	19
5.1	KDD Ablauf	23
5.2	Ablauf Web Usage Mining	24
5.3	Path Completion Problem	31
5.4	Dendrogramm	38
6.1	System Accaptability	45
6.2	Einordnung der Usability-Evaluationen	51
7.1	Typologie von Schulinformationssystemen	84
7.2	Raum-Zeit-Matrix	85
7.3	3K-Modell	86
7.4	Einstiegsseite von InfoChange	88
7.5	Materialiensammlung von InfoChange	89
7.6	Terminübersicht von InfoChange	91
7.7	Besuche und Aktionen aller Nutzerinnen und Nutzer	94
7.8	Besuche und Aktionen ausgewählter Nutzerinnen und Nutzer	95
7.9	Besuche im Zeitverlauf	95
7.10	Treemap der Zugriffe auf Rubriken und Funktionen	97
7.11	Verlauf der Besuche	99
8.1	Benötigte Zeit pro Aufgabe	111
8.2	Antworten der SUS-Skala	112
8.3	Usability-Probleme in der Terminübersicht	113
8.4	Usability-Probleme bei der Zusage eines Termins	114
8.5	Usability-Probleme in der Eingabemaske für neue Termine	115
8.6	Usability-Probleme in der Materialübersicht	116
8.7	Usability-Probleme bei der Übersichtsseite der Personen	117

8.8	Usability-Probleme auf der Detailseite der Personen	117
8.9	Usability-Probleme in der Eingabemaske für neue Materialien	119
8.10	Usability-Probleme bei der Zuweisung von Schlagwörtern	120
8.11	Usability-Probleme bei der Zuweisung von Kategorien	121
8.12	Usability-Probleme bei der Kommentierung einer Ankündigung	122
8.13	Veränderungen im Suchfeld zwischen Startseite und Rubriken	123
8.14	Suche Stauffenberg über Ankündigungen	123
8.15	Suche Stauffenberg	123
8.16	Auflistung der Einträge mit Schlagwort	124
8.17	Suche nach Schlagwort	125
8.18	Suche nach Schlagwort und Rubrik	126
8.19	Schwerwiegende Probleme	127
9.1	Vergleich der Ähnlichkeiten der Transaction Matrix	136
9.2	Zugriffe der generischen problematischen Pfadmerkmale auf Rubriken und Funktionen	138
9.3	Ähnlichkeit der generischen problematischen Pfadmerkmale	139
9.4	Identifizierte generische problematische Pfadmerkmale in den Produktiv- Logfiles	143
9.5	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,7 .	147
9.6	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,75	150
9.7	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,8 .	151
9.8	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,85	152
C.1	Hover-Effekt bei Terminen	204
C.2	Hover-Effekt bei Symbolen	204
C.3	Eingblendeter Kalender	204
C.4	Übersichtsseite der Gruppen	205
C.5	Übersichtsseite der Ankündigungen	205
C.6	Fehlermeldung Ankündigungen bei fehlendem Titel	205
C.7	Suche Stauffenberg über Termine	206
C.8	Auflistung der Einträge mit Schlagwort in Rubrik	206
D.1	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,6 .	207
D.2	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,65	208
D.3	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,9 .	209
D.4	Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,95	209

1 Einleitung

In Deutschland nutzen aktuell fast achtzig Prozent der Bevölkerung das Internet und ein ebenso hoher Prozentsatz der Menschen hat einen Internetzugang zur Verfügung (van Eimeren und Frees, 2014). Parallel ist eine zunehmende Verbreitung digitaler Medien im privaten wie auch im beruflichen Umfeld festzustellen. Beide Phänomene führen zu einer zunehmend stärkeren Digitalisierung der Gesellschaft. Die Kommunikationswissenschaft beschreibt diesen Zusammenhang mit dem Begriff einer zunehmenden Mediatisierung (Hepp, 2011; Krotz, 2007). Durch die fortschreitende technische Entwicklung stehen den Endanwendern daneben unterschiedliche Geräte zur Verfügung. Insbesondere in den letzten Jahren wurden durch die Entwicklung und Einführung von Smartphones und Tablets zwei neue Geräteklassen geschaffen. Dies führt insbesondere in geschäftlichen Nutzungsszenarien zu einer Ausdifferenzierung der Arbeitskultur (Liang und Wei, 2004; Pitt et al., 2011; Yuan, Archer et al., 2010). Verschiedene Firmen, Behörden oder Verwaltungen setzen ein unterschiedliches Repertoire an Medien zur Kommunikation zwischen den einzelnen Mitarbeiterinnen und Mitarbeitern ein.

In dem DFG-Forschungsprojekt „Schulen als transnationale Netzwerkorganisationen“¹ wurden die Veränderungen der Kommunikation innerhalb des Kollegiums durch die zunehmende Verbreitung digitaler Medien untersucht (Welling et al., 2014). Dort konnte der zunehmende Medienwandel auch für Schulen beobachtet und belegt werden. Zudem wurde ersichtlich, dass digitale Kooperationsysteme primär dann von Lehrkräften angenommen und eingesetzt werden, wenn sie neben dem rationalen Effekt auch einfach und intuitiv zu bedienen sind.

In der Informatik wird eine gute Bedienbarkeit mittels der *Usability* (dt. Gebrauchstauglichkeit) erklärt.² Eine „gute“ Usability steht für eine effiziente, effektive, angenehme, fehlertolerante und intuitive Anwendung (vgl. u. a. ISO, 1998; Nielsen, 1993; Quesenbery, 2003). Die Überprüfung hinsichtlich der Gebrauchstauglichkeit erfolgt bisher vor allem mit qualitativen Methoden. Teilweise kommen quantitative Methoden wie Umfragen zum Einsatz. Grundsätzlich haben alle Methoden gemein, dass eine Überprüfung der Usability zeitintensiv und je nach verwendeter Methode auch kostenintensiv ist. Parallel kommt es durch die oben angesprochene Mediatisierung und Digitalisierung zu einer Zunahme an Anwendungen und Programmen. Durch diese Zunahme reichen die heute verfügbaren qualitativen und quantitativen Methoden aufgrund ihres hohen Aufwands nicht mehr aus, um alle Anwendungen hinsichtlich der Gebrauchstauglichkeit zu überprüfen. Aus diesen Erkenntnissen entstand die Idee zu dieser Dissertation. Es soll untersucht werden, ob sich

¹Das Projekt „Mediatisierte Organisationswelten in Schulen“ war Teil des DFG-Schwerpunktprogramms 1505 „Mediatisierte Welten“ (Förderkennzeichen: BR 2273/10-1).

²Der Begriff *Usability* wird in Kapitel 6.1 vertieft.

klassische qualitative Usability-Methoden mit der quantitativen Logfile-Analyse kombinieren lassen, um eine einfachere, schnellere und teilweise automatisierte Überprüfung hinsichtlich der Gebrauchstauglichkeit zu ermöglichen.

Die Fragestellung der Dissertation liefert einen Erkenntnisgewinn für die *Informatik*.³ Zunächst lässt sich die *Logfile-Analyse* dem Gebiet des *Data Minings* (Han, Kamber et al., 2012) beziehungsweise der *Wissensentdeckung in Datenbanken* (KDD; Fayyad et al., 1996) im Allgemeinen und des *Web Usage Mining* (Srivastava et al., 2000) im Besonderen zuordnen. Das Vorhaben, aus den Logfiles auf Nutzerprobleme zu schließen, bezieht das Gebiet der *Usability* (dt. Gebrauchstauglichkeit; Quesenbery, 2003; Sarodnick und Brau, 2011) mit ein. Das Gebiet ist ein wichtiger Aspekt der *Human-Computer Interaction* (HCI, dt. Mensch-Computer-Interaktion; Dix et al., 2004). Gleichzeitig spielt es in dem Gebiet der *Softwareentwicklung* eine wichtige Rolle (Balzert und Liggesmeyer, 2011). Neue Software sollte bereits während des Entwicklungsprozesses evaluiert werden, um mögliche Probleme und Fehlerquellen ausschließen zu können. Gleiches gilt für die Weiterentwicklung und Überprüfung bereits vorhandener Software.

Daneben werden weitere Fachgebiete und Disziplinen in dieser Dissertation tangiert. Die Durchführung von Nutzertests (engl. Usability Tests; Barnum, 2011; Rubin und Chisnell, 2008) basiert auf verschiedenen Testverfahren und experimentellen Designs, die der *qualitativen* und *quantitativen Sozialforschung* (Diekmann, 2010; Lamnek und Krell, 2010; Mayring, 2002; Schnell et al., 2008) entnommen wurden. Bei der Verknüpfung von Logfile-Analyse und Nutzertests handelt es sich um eine Methoden-Triangulation (Flick, 2008), die in der Usability-Forschung bisher geringere Verbreitung findet (Mackay, 1998). Zwar werden häufig verschiedene Methoden wie Fragebögen und Nutzertests parallel eingesetzt, allerdings werden die Ergebnisse weniger kombiniert (Wilson, 2006). Das verwendete und untersuchte Software-Produkt ist ein *Lern-Management-System* (LMS; Baumgartner et al., 2004; Schulmeister, 2009). Es kann der *rechnergestützten Gruppenarbeit* beziehungsweise dem *Computer Supported Cooperative Work* (CSCW; Gross und Koch, 2007; Grudin, 1994) zugeordnet werden. Die Funktionen zur Koordination können auch unter dem Begriff der *Groupware* (Ellis et al., 1991) subsumiert werden.

Die vorliegende Dissertation spaltet sich damit von der Fragestellung des DFG-Forschungsprojekts ab. Das Projekt ist primär der Kommunikationswissenschaft (Beck, 2015; Burkart, 2002) zuzuordnen und versucht einen Beitrag zur Mediatisierungsforschung (Hepp, 2011; Krotz, 2007) und zur Schulforschung (Altrichter, 2006; Kohn, 2011; van Weert und Munro, 2003) zu liefern. Die Analysen des Lern-Management-Systems liefern zumindest einen rudimentären Beitrag zu den oben genannten Gebieten. Sie wurden ausführlicher und prägnanter an anderen Stellen publiziert (Schulz und Breiter, 2012, 2013a,b; Schulz und Jeske, 2014; Welling et al., 2014). Die Besonderheit der Analysen ist, dass es im Schulkontext kaum Untersuchungen zur internen Nutzung von CSCW oder Groupware gibt. Im bisherigen Fokus der logfile-gestützten Forschung (Bienkowski et al., 2012) steht vor allem der Lehr- (engl. teaching analytics; Xu und Recker, 2012) beziehungsweise der Lernprozess (engl. learning analytics; Berland et al., 2014; Greller und Drachsler, 2012;

³Das DFG-Forschungsprojekt hatte einen kommunikationswissenschaftlichen Schwerpunkt.

Kiy und Lucke, 2014). Dies steht im Gegensatz zur traditionellen Wirtschaftsinformatik (Heinrich et al., 2011; Lassmann et al., 2006; Laudon und Laudon, 2015), die den Gebrauch dieser Anwendung in der Wirtschaft und Verwaltung deutlich intensiver erforscht und untersucht hat.

Der Aufbau der Arbeit gliedert sich grob in drei Teile. Im ersten Teil werden die Vorüberlegungen genauer ausgeführt. Dazu zählt die thematische und inhaltliche Herleitung der Fragestellung und die genaue Beschreibung der Vorgehensweise. Die beiden Kapitel hängen eng voneinander ab.

Die Einführung in die Grundlagen stellt den zweiten Teil dar. Dazu zählt die Einführung in die Logfile-Analyse, ihre Herleitung und Abgrenzung sowie das Vorgehen im Sinne des Data Minings beziehungsweise genauer des Web Usage Minings. Gleichzeitig muss die Logfile-Analyse innerhalb des Data Minings verortet werden. Neben der Logfile-Analyse wird der Begriff der Gebrauchstauglichkeit vertiefend behandelt. Vor der Beschreibung des optimalen Vorgehens und der Durchführung eines Nutzertests erfolgt die Bestimmung und Beschreibung des Begriffs „Usability“. Dieser Bereich tangiert teilweise psychologische Grundlagen sowie die qualitative und quantitative Sozialforschung. Der Nutzertest ist nur eine von vielen Möglichkeiten, auf die Usability einer Anwendung zu schließen. Eine Einordnung in die geläufigsten Methoden ist daher angeraten. Die Durchführung eines Nutzertests behandelt sowohl das Vorgehen und die Durchführung als auch eine Übersicht über die verfügbaren Techniken zur Aufdeckung von Problemen. Zudem erfolgt eine Aufzählung bereits vorhandener Ansätze, Usability automatisiert und auf Basis von Logfiles zu bestimmen. Der Ansatz dieser Dissertation stellt dabei einen eigenen Ansatz dar. Zuletzt erfolgt eine Einführung in Digitale Kooperationssysteme und die Verortung von Lern-Management-Systemen innerhalb des Spektrums von CSCW und Groupware.

Der dritte Teil beinhaltet die Ergebnisse und gliedert sich ebenfalls in drei Teile. Nach der theoretischen Einführung in die Themenkomplexe wird zunächst die untersuchte Anwendung genauer vorgestellt. Dies schließt die Beschreibung des Vorgehens der Logfile-Analyse mit ein. Abschließend wird die Nutzung der Anwendung durch die Lehrerinnen und Lehrer dargelegt und teilweise mit den Ergebnissen der Gruppendiskussionen (Welling et al., 2014) kontrastiert. Der Logfile-Analyse folgen die Nutzertests. Ähnlich wie im vorherigen Kapitel werden zunächst die Vorüberlegungen, die Vorbereitungen und die Durchführung der Nutzertests dargelegt. Die Beschreibung orientiert sich so nah wie möglich an den Vorgaben der Literatur. Abweichungen werden benannt und, wenn nötig, begründet. Die in der Auswertung gefundenen Usability-Probleme werden anschließend dargestellt und zur besseren Veranschaulichung mit Screenshots belegt. Danach werden die gefundenen Probleme hinsichtlich ihrer Schwere und Reichweite bewertet. Im letzten Teil der Empirie erfolgt die Kombination von den Logfile-Analysen und den Nutzertests. Zunächst wird das Vorgehen vorgestellt und theoretisch untermauert. Dies betrifft in diesem Teil der Arbeit vor allem die Aufbereitung der verschiedenen Datenquellen, um einen Vergleich zu ermöglichen. Die Ergebnisse der Kombination und die Entscheidung für ein finales Modell schließen den empirischen Teil ab. Im Fazit der Dissertation werden die Ergebnisse zusammengefasst,

diskutiert und kritisch beleuchtet. Dies schließt die Darlegung von Limitierungen mit ein.

Die Dissertation ist wie bereits erwähnt aus dem DFG-Projekt „Schulen als translokale Netzwerkorganisationen“ hervorgegangen. Verschiedene Kapitel kommen in ähnlicher Form in dem mittlerweile veröffentlichten Endbericht vor (vgl. Welling et al., 2014). Eine Ähnlichkeit wird zu Beginn der entsprechenden Kapitel vermerkt. Die Textstellen wurden auch in dem Endbericht ausschließlich von dem Autor dieser Dissertation verfasst.

2 Herleitung der Fragestellung

Computergestützte Arbeitsabläufe in Unternehmen oder Behörden finden in *Informationssystemen* statt (Laudon und Laudon, 2015). Dabei handelt es sich um sozio-technische Systeme, die sowohl technische, organisatorische als auch menschliche Bestandteile enthalten. In der Wirtschaftsinformatik wird von sogenannten MAT-Systemen gesprochen (Heinrich et al., 2011). Die Abkürzung steht für *Mensch-Aufgabe-Technik-System* und macht deutlich, dass es sich nicht um eine reine technische Lösung handelt. Grundsätzlich soll ein bestimmter (Arbeits-)Ablauf systematisiert, vereinfacht oder optimiert werden. Auf der technischen Ebene geschieht dies durch die Einführung einer *Anwendung*. Während die Nutzung im privaten Umfeld meist freiwillig erfolgt, muss die Einführung und Etablierung neuer Anwendungen in Unternehmen zwischen den einzelnen Individuen beziehungsweise Abteilungen ausgehandelt werden (Bergmann und Garrecht, 2008; Keen, 1981). Dies bedeutet allerdings nicht, dass die Mitarbeiterinnen und Mitarbeiter eine entsprechende Anwendung beziehungsweise die Modifizierung des Informationssystems akzeptieren und anstandslos nutzen. Ein bekanntes Modell zur Beschreibung der Veränderungen ist das *Technology-Acceptance-Model* (TAM, vgl. Abbildung 2.1; Davis, 1989). Dieses Modell postuliert eine kausale Abhängigkeit von verschiedenen Einstellungen, die zur tatsächlichen Nutzung einer neuen Technologie (in diesem Fall einer Anwendung) führen. Neben externen Einflüssen beeinflussen die wahrgenommene Nützlichkeit („Perceived Usefulness“) und die wahrgenommene Gebrauchstauglichkeit („Perceived Ease of Use“) die Einstellung zur Nutzung („Attitude toward Using“). Die tatsächliche Nutzung („Actual System Usage“) einer Anwendung hängt vor allem von der Intention zur Nutzung („Behavioral Intention to Use“) einer Mitarbeiterin bzw. eines Mitarbeiters ab, welche wiederum vor allem durch die Einstellung zur Nutzung sowie der wahrgenommenen Gebrauchstauglichkeit beeinflusst wird. Das Modell wurde mehrfach in (Meta-)Analysen evaluiert und als robust und valide bestätigt (Agarwal und Prasad, 1999; Holden und Karsh, 2010; King und He, 2006; Turner et al., 2010). Zudem wird die Relevanz für ITK-Anwendungen über Unternehmen hinaus festgestellt, unter anderem auch für den Bildungsbereich (vgl. u. a. Chung et al., 2012; De Smet et al., 2012). Es ist ersichtlich, dass die *Benutzbarkeit*, unter die auch die *Gebrauchstauglichkeit* (*Usability*) fällt, einen wichtigen Einfluss auf die Akzeptanz und den endgültigen Einsatz von Anwendungen hat. Eine hohe Benutzbarkeit ist daher bei der Auswahl und Entwicklung von Anwendungen ein entscheidendes Merkmal. In den vergangenen Jahren wurde das TAM-Modell zu TAM2 weiterentwickelt (Venkatesh und Davis, 2000). Das erweiterte Modell verfügt über mehrere zusätzliche Eingangsvariablen. Diese beeinflussen primär die wahrgenommene Nützlichkeit und haben keinen Einfluss auf die wahrgenommene Gebrauchstauglichkeit. Daher wird TAM2 an dieser Stelle nicht weiter betrachtet.

Berücksichtigt werden sollten allerdings die Erkenntnisse von Venkatesh (2000). Der Autor erweitert die wahrgenommene Nützlichkeit um zusätzliche Eingangsvariablen und kann belegen, dass zu Beginn einer Systemnutzung vor allem interne und externe Kontrolle, intrinsische Motivation und Emotionen einen Effekt aufweisen. Diese Faktoren sind allgemeingültig zu bewerten und beziehen sich nicht auf die verwendete Anwendung. Mit steigender Nutzungsdauer werden systemspezifische und regulierende Faktoren wirksam. Zum einen steigt die objektive Gebrauchstauglichkeit. Daneben kann längere Nutzung einer Anwendung einen positiven Effekt auf die intrinsische Motivation haben. Dieser ist allerdings systemspezifisch.

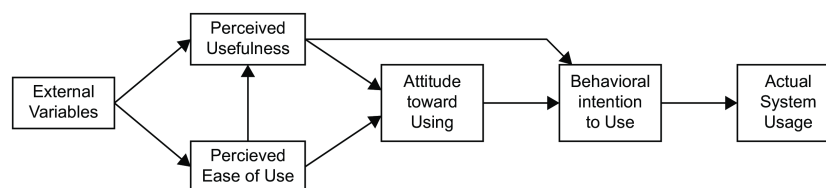


Abbildung 2.1: TAM-Modell nach Davis (1989)

Die *Gebrauchstauglichkeit* (Usability) einer Anwendung lässt sich durch verschiedene Methoden ermitteln. Häufig wird auf Verfahren wie *Inspektionsmethoden* („Usability Inspection Methods“) oder *Nutzertests* („Usability Testing“) zurückgegriffen. Sie bieten den Vorteil, dass sie umfangreich erprobt und etabliert sind. Bis auf wenige Ausnahmen (unter anderem *Fragebögen*) haben alle Methoden ihren Ursprung in der qualitativen Forschung. Dies stellt zunächst keinen Nachteil dar, sondern bietet durch die umfangreichen Interaktionsmöglichkeiten zwischen Forscherin beziehungsweise Forscher und Nutzerin beziehungsweise Nutzer den Vorteil, dass Probleme und Schwierigkeiten in der Bedienung der Anwendung detailliert und umfangreich erfasst werden können. Diese Vorteile gehen häufig mit einem höheren Zeitaufwand einher, was insbesondere auf die Durchführung und die Auswertung zutrifft. Zudem müssen weitere relevante Gütekriterien wie Reaktivität, Reliabilität oder die Untersuchungssituation mit berücksichtigt werden (Brüsemeister, 2008; Flick, 2007; Meyen et al., 2011; Schnell et al., 2008). Die Berechtigung etablierter Methoden wird in dieser Arbeit nicht hinterfragt oder kritisiert, sondern es soll versucht werden, die Nachteile durch eine Kombination mit Logfiles zu kompensieren oder zu vermindern. Die Kombination mehrerer Methoden fällt unter den Begriff der *Triangulation* und bietet einen tieferen Einblick und eine höhere Erklärungskraft als einzelne Ansätze (Flick, 2008).

Durch die zunehmende Verlagerung von clientgestützten zu serverbasierten Anwendungen bieten sich die Logfiles als zusätzliches Beobachtungsinstrument an. Das Verhalten der Nutzerinnen und Nutzer kann vergleichsweise einfach nachvollzogen werden, da die serverbasierten Anwendungen häufig bereits Möglichkeiten zur Erstellung von Logfiles bieten. Es wird davon ausgegangen, dass verschiedene Nutzerinnen und Nutzer sich unterschiedlich in Anwendungen bewegen. Diese Vermutungen konnten vom Autor mehrfach bestätigt werden (Breiter, Averbek et al., 2015; Schulz und Breiter, 2013b; Schulz und Jeske, 2014; Welling et al., 2014). Das kann mit dem aktuellen Ziel in der Anwendung zu-

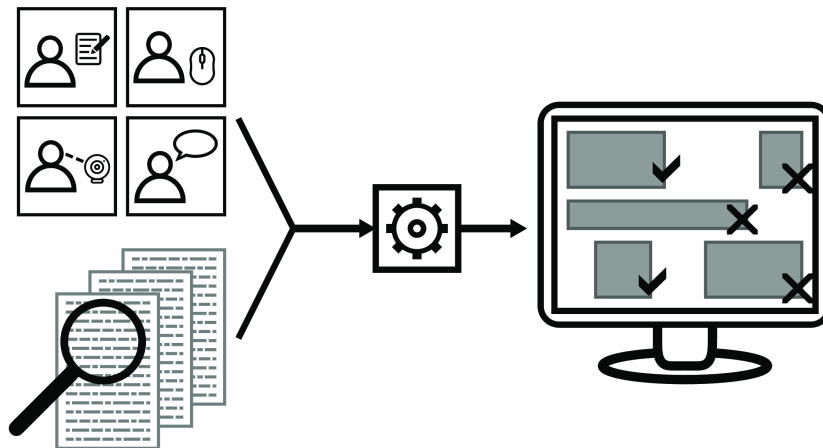


Abbildung 2.2: Zentrale Fragestellung

sammenhängen oder persönliche Vorlieben der Nutzerinnen und Nutzern widerspiegeln. Unterschiedliche Nutzerinnen und Nutzer sollten daher empfindlicher auf bestimmte Probleme hinsichtlich der Gebrauchstauglichkeit einer Anwendung reagieren. Beispielfhaft erläutert ist dies an dem folgenden Fall: Wenn sich in den Logfiles eine Nutzergruppe identifizieren lässt, die sich vorwiegend über Suchfunktionen in einer Anwendung bewegt, ist es denkbar, dass dieser Nutzertyp empfindlich auf eine problematische Bedienung der Suchfunktion reagiert, während dies unter Umständen für eine Nutzerin beziehungsweise einen Nutzer, die über Menüs und Menüleisten navigiert, nicht der Fall sein sollte. Zwei Vorteile machen die Erfassung des Verhaltens durch Logfiles besonders attraktiv: Zum einen kann das Verhalten aller Nutzerinnen und Nutzer im Hintergrund aufgezeichnet werden. Zum anderen erfolgt die Erhebung der Daten im Hintergrund und ist für die Nutzerinnen und Nutzer nicht ersichtlich, was die Ergebnisse der Feldforschung kaum verzerren dürfte (non-Reaktivität).

Die zentrale Fragestellung dieser Dissertation untersucht, ob sich Logfile-Analysen und Nutzertests zur besseren Messung der Usability kombinieren lassen (vgl. Abbildung 2.2).

Das Ziel soll sein, Usability-Probleme anhand von Logfiles zu erkennen beziehungsweise den Aufwand an Nutzertests zu verringern. Die *einmalige* Durchführung von Nutzertests soll Probleme und generische Pfade herausarbeiten. Werden diese Pfade mit weiteren „realen“ Logfiles zusammengebracht, lassen sich die Probleme in der Anwendung genauer nachvollziehen.

Die Kombination der beiden Methoden bietet aus Sicht des Autors fünf Vorteile. (1) Zentraler Punkt ist die Idee, Nutzertests durch Logfile-Analysen auf eine breitere Basis zu stellen. Bisher werden bei Nutzertests zwischen fünf bis zehn Tests durchgeführt (vgl. Kapitel 6.3.1). Von dieser Basis ausgehend wird entschieden, welche Probleme am schwerwiegendsten sind und die höchste Reichweite haben. Bei steigender Komplexität der Anwendungen und steigenden Nutzerzahlen erhöhen sich die Anforderungen an eine möglichst repräsentative Auswahl der Testpersonen. Eine Kombination mit Logfiles bietet die Möglichkeit, die vorhandenen Nutzertests mit einer größeren Datenbasis an „realen“ Logfiles zu verknüpfen. Dadurch kann darauf geschlossen werden, welche gefundenen

Probleme häufiger auftreten und welche Probleme in der Praxis irrelevant sind. (2) Gleichzeitig bedeutet der Einsatz von Nutzertests und Logfile-Analyse die Kombination von einer qualitativen Methode (Flick, 2007; Lamnek und Krell, 2010; Mayring, 2002) und einer quantitativen Methode (Diekmann, 2010; Schnell et al., 2008). Während die qualitative Methode des Nutzertests vor allem ein tieferes Verständnis der Nutzerprobleme liefern soll, wird durch die Logfile-Analyse als quantitative Methode eine Übertragung der Ergebnisse auf eine hohe Anzahl an Nutzerinnen und Nutzern möglich. (3) Die Kombination von Nutzertests und Logfiles bietet eine nachträgliche Einschätzung der erfolgreichen Problembhebung. Nach der Durchführung und Auswertung der Nutzertests folgt üblicherweise die Behebung der Probleme durch die Entwicklerinnen und Entwickler. Werden die Änderungen in einer neuen Version veröffentlicht, bieten Logfiles die Möglichkeit, das Auftreten der Probleme in der aktuellen Version erneut zu beobachten und zu erfassen. Sollten sich entsprechende Pfade nicht mehr finden lassen, kann dies als Indikator für eine erfolgreiche Problembhebung gewertet werden. (4) Im weiteren Verlauf der Produkt(weiter)entwicklung werden eventuell weitere Nutzertests durchgeführt und Logfiles erhoben. Solange sich der Aufbau der Anwendung nicht grundlegend ändert, kann eine Datenbank mit Pfaden für Probleme aufgebaut werden. Dies kann im Laufe der Zeit zu einer weiteren Verringerung der Nutzertests führen. (5) Zuletzt bietet sich die Nutzung von Logfiles an, weil eine bereits vorhandene und vorher ungenutzte Datenquelle erschlossen wird.

3 Vorgehensweise

Das ursprünglich geplante Vorgehen war, anhand der Produktiv-Logfiles Clusteranalysen (Everitt et al., 2011) durchzuführen. Aus jedem gebildeten Cluster sollten anschließend generische Pfade pro Cluster gebildet werden. Da die Cluster über die Eigenschaften der Pfade der einzelnen Besuche gebildet werden, sollten die generischen Pfade auf diesen Einzelpfaden basieren. In einem zweiten Schritt sollten Nutzertests durchgeführt werden. Nach der Auswertung und Analyse der Nutzertests sollten für jede Testerin beziehungsweise für jeden Tester spezifische Probleme identifiziert werden. Die dazugehörigen Pfade der Testerinnen und Tester sollten an die Pfade der gebildeten Cluster aus dem ersten Schritt angefügt werden. Dieses Matching sollte über die im Hintergrund protokollierten Logfiles erfolgen (vgl. Abbildung 3.1). Durch die Verwendung desselben Clusterverfahrens sollte die Gefahr einer Verzerrung minimiert werden.

Dieses Vorgehen hat einen zentralen Schwachpunkt. Die Zuordnung von Usability-Problemen zu vorher gebildeten Clustern funktioniert als abgeschlossenes Verfahren gut. Problematisch ist allerdings die Aufnahme weiterer Logfiles in das Verfahren (beispielsweise durch eine weitere Iterationsschleife). Es ist unklar, wie ein erneutes Matching funktionieren könnte. *Dieser Ansatz wird daher verworfen.*

Um das Verfahren nach einer ersten Iterationsschleife offen für neue Logfiles aus der Produktivanwendung zu gestalten, wird der Ablauf des Verfahrens geändert. An der ursprünglichen Idee und Fragestellung ändert sich nichts. Das neue Vorgehen ist in Abbildung 3.2 abgebildet. Analog zur ursprünglichen Idee, spaltet sich das Vorgehen in zwei Phasen auf. Die erste Phase besteht aus den Nutzertests (vgl. Kapitel 8) und der Generierung von problematischen Pfadmerkmalen (vgl. Kapitel 9.2). Danach folgt die zweite Phase. In dieser findet die Kombination der problematischen Pfadmerkmale mit Produktiv-Logfiles aus einer Anwendung mit Nutzerinnen und Nutzern statt (vgl. Kapitel 9.3). Die zweite Phase ist wiederholbar ausgelegt und kann mit weiteren Logfiles erneut durchgeführt werden. Dies trifft auf den angesprochenen Aufbau einer Datenbank und dem Nachverfolgen von Veränderungen in der Anwendung zu.

Der Großteil der *ersten Phase* (vgl. Abbildung 3.3) besteht aus Nutzertests. Die Planung, Durchführung und Auswertung der Tests erfolgt durch geläufige Usability-Methoden (Dix et al., 2004; Freiberg und Baumeister, 2008; Sarodnick und Brau, 2011). Grundsätzlich stehen alle Methoden wie lautes Denken, Mouse-Tracking, Eye-Tracking oder Fragebögen zur Auswahl. Dies trifft auch auf die theoretisch frei wählbaren Aufgaben zu. In diesem Fall werden diese aus den deskriptiven Analysen der Logfiles der Produktivanwendung und den Ergebnissen des DFG-Projekts (Welling et al., 2014) gewonnen. Die Durchführung der Nutzertests erfolgt in einer Referenzanwendung, welche über eine integrierte Protokollfunktion verfügt. Die Auswertung erfolgt zweigeteilt, bevor beide Teile verknüpft

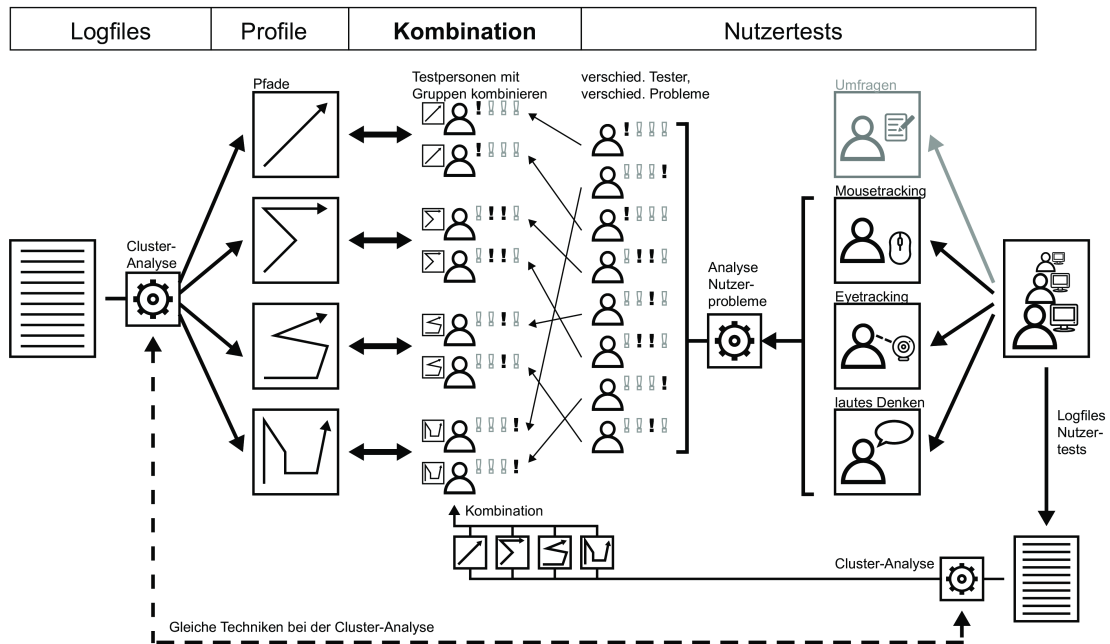


Abbildung 3.1: Ursprüngliches Vorgehen

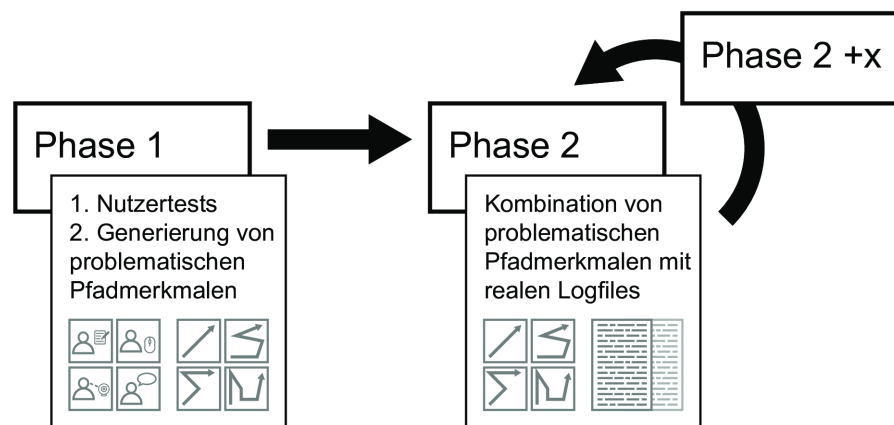


Abbildung 3.2: Übersicht Vorgehen

werden. Zunächst werden die Nutzertests ausgewertet und Usability-Probleme (UP_{p_i}) identifiziert. In dem Beispiel (vgl. Abbildung 3.3) finden die sieben Testpersonen vier verschiedene Probleme, wobei dies in unterschiedlicher Häufigkeit passiert. Das erste Problem bereitet drei Testpersonen Probleme, das zweite Problem wird von einer Testperson identifiziert. Das dritte Problem wird von zwei weiteren Testerinnen und Testern gefunden. Das letzte Problem wird von einer Testperson identifiziert. Diese „stolpert“ auch über das dritte Problem. Zu den herausgearbeiteten Problemen werden die entsprechenden Pfade aus den Logfiles der Testanwendung extrahiert. Testpersonen, die keine Probleme bei der Lösung der jeweiligen Aufgabe hatten, fallen aus der weiteren Analyse dieses Problems heraus, können aber bei weiteren Problemen noch eine Rolle spielen. Die Anzahl der verfügbaren Pfade kann sich auf eine Zahl zwischen eins und der Anzahl der Testerinnen und Tester belaufen. In dem gezeigten Beispiel wird das erste Problem von drei Testpersonen gefunden. Folglich können aus den Logfiles drei Fehlerpfade (FP_{p_i}) extrahiert werden. Für das erste Problem stehen damit drei individuelle Fehlerpfade zur Verfügung. Ihre Logfiles sind für die anderen drei Probleme nicht von Bedeutung. Das zweite Problem wird von einer Testperson identifiziert. Daher steht nur ein Pfad zur Verfügung. Das dritte Problem wird von drei Testerinnen und Testern gefunden. Die Besonderheit in diesem Fall ist, dass die dritte Testerin daneben das zweite Problem findet. Werden mehrere Probleme während derselben Aufgabe entdeckt, ist derselbe Pfad für beide Probleme heranzuziehen. Lassen sich die gefundenen Probleme verschiedenen Aufgaben zuordnen, so sind die jeweiligen Teilpfade der Aufgaben zu extrahieren. Das letzte Problem wird von einer Testperson identifiziert. Als Zwischenergebnis lassen sich jedem Usability-Problem (UP_{p_i}) einer oder mehrere Fehlerpfade (FP_{p_i}) zuordnen. Aus den Fehlerpfaden werden in einem letzten Schritt die *problematischen Pfadmerkmale* (pPM_{p_i}) extrahiert. Die Abstraktion von den Pfaden bietet eine höhere Verallgemeinerung der Probleme und dürfte bei Veränderungen der Anwendung zu einer längeren Anwendbarkeit der Ergebnisse führen. Dabei wird darauf geachtet, dass die Umwandlung von Pfaden zu Pfadmerkmalen zu einem möglichst geringen Informationsverlust führt (vgl. Kapitel 9.1).

Die problematischen Pfadmerkmale werden in der *zweiten Phase* (vgl. Abbildung 3.4) herangezogen und mit Logfiles aus einer Produktivanwendung verknüpft. Dazu werden *alle* Pfade der Produktiv-Logfiles in ein vergleichbares Format wie die problematischen Pfadmerkmale gebracht (vgl. Kapitel 9.1). Es wird ermittelt, in welchen Besuchen der Produktiv-Logfiles sich problematische Pfadmerkmale der Test-Logfiles finden lassen. Eine Übereinstimmung deutet darauf hin, dass die Nutzerin beziehungsweise der Nutzer während des Besuchs unter Umständen dasselbe Problem hatte wie die Testpersonen im Nutzertest. Wird das Beispiel aus der Abbildung 3.3 herangezogen, zeigt sich durch die Kombination der Test-Logfiles mit den Produktiv-Logfiles, dass die problematischen Pfadmerkmale des ersten Usability-Problems sechsmal, die Pfadmerkmale des zweiten Problems achtmal, die Pfadmerkmale des dritten Problems zehnmal und die Problempfade des vierten Problems zweimal identifiziert werden können. Anhand dieser Zahlen lässt sich die Schwere eines Usability Problems quantitativ einordnen.¹ Das zweite Pro-

¹Eine qualitative Einordnung der Reichweite kann nicht erfolgen, sondern ist nur über die Auswertungen der Nutzertests möglich.

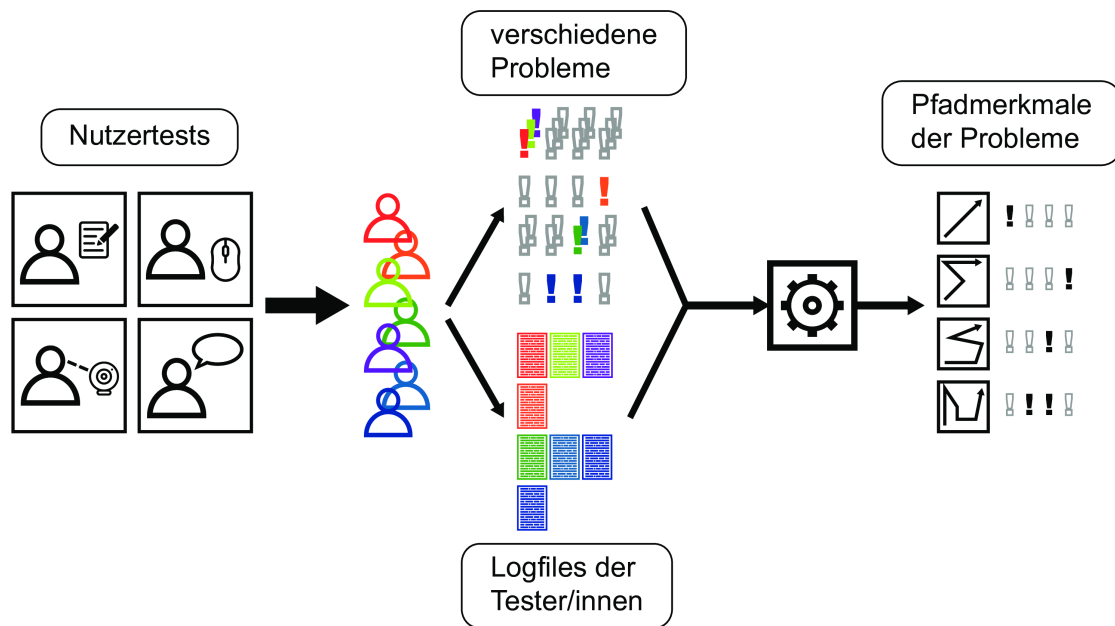


Abbildung 3.3: Vorgehen (Phase 1)

blem wird von einer Testperson identifiziert. Der problematische Fehlerpfad findet sich achtmal in den Produktiv-Logfiles. Dies deutet darauf hin, dass dieses Problem häufiger auftritt als der Nutzertest vermuten lässt. Die Ähnlichkeit zweier Pfadmerkmale wird über Distanzmaße (vgl. Gleichung 5.2.2) bestimmt. Die Methode ist so ausgelegt, dass sich die zweite Phase beliebig oft wiederholen lässt. Es lassen sich beispielsweise Produktiv-Logfiles nach der Behebung von Problemen analysieren. Dies würde einen Indikator dafür liefern, ob die entsprechenden Pfadmerkmale weiterhin auftreten oder behoben werden konnten.

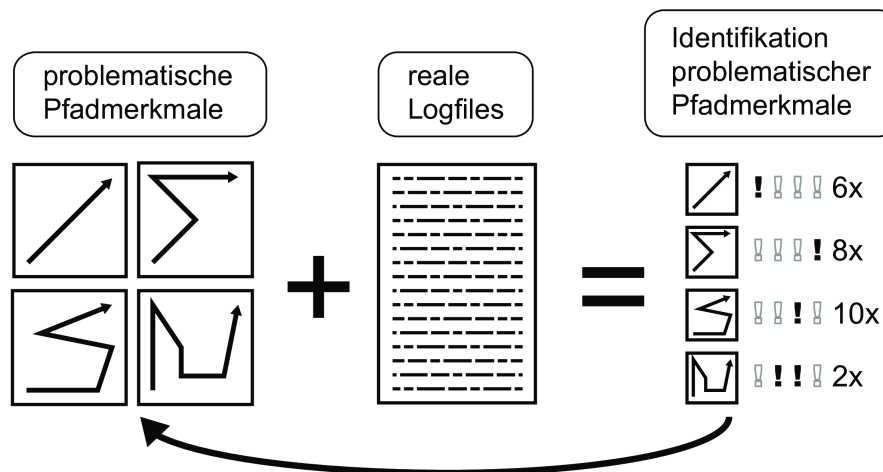


Abbildung 3.4: Vorgehen (Phase 2)

4 Logfiles als Erfassungsinstrument

Die Logfile-Analyse ist historisch gesehen ein technisches Verfahren. Das Verfahren wird mittlerweile in vielen Bereichen eingesetzt (Priemer, 2004). Der klassische Schwerpunkt sind die Programmentwicklung und die Systemadministration (Lutzky und Teichmann, 2002; Oliner et al., 2012; Suneetha und Krishnamoorthi, 2009). In der Programmentwicklung helfen Logfiles beim Auffinden von Fehlern („Bug-Tracking“, „Debugging“), im Bereich der Systemadministration werden mithilfe von Logfiles Probleme und Fehler in Systemen nachvollzogen und überprüft. Daneben gab es früh Interesse an Logfiles im Bereich des (E-)Commerce und des Marketings (Spiliopoulou, 1999). Erste Anstrengungen lassen sich bis Anfang der 1980er-Jahre zurückverfolgen (Tolle, 1983). Wichtig ist die Analyse von Logfiles für das *Customer Relationship Management* (CRM; Leußer et al., 2011), das *Enterprise Resource Planning* (ERP; Shad et al., 2012) oder als Grundlage für *Business Intelligence* (BI; Schildhauer, 2004). Logfile-Analysen werden zur Kundengewinnung, Kundenbetreuung und zum Erhalten und Ausbauen der Kundenbeziehung verwendet (Ansari et al., 2001; Bounsaythip und Rinta-Runsala, 2001; Eirinaki und Vazirgiannis, 2003; Facca und Lanzi, 2005). Mit dem Wissen aus Logfile-Analysen können Webseiten optimiert und angepasst werden (Tyagi et al., 2010). Dabei geht es darum, das Nutzungsverhalten von (vorrangig) Webseiten zu eruieren und auf Basis dieses Wissens die Struktur zu optimieren beziehungsweise angepasste Werbeinhalte einzublenden. Büchner und Mulvenna (1998) optimieren durch Logfiles das (E)Marketing. Daneben spielt die Logfile-Analyse im Bereich der Werbung eine bedeutende Rolle (Cooley et al., 1997b). Gleiches gilt für den Bereich der Empfehlungen (Recommendation Systems; Esslimani et al., 2008; Li und Zaïane, 2004; Mobasher, Dai, Luo, Sun et al., 2000). Dazu gehören automatische Anwendungen wie beispielsweise weitere Kaufvorschläge bei Amazon. Insbesondere auf diesem Gebiet wurde frühzeitig mit *Data Mining* Methoden auf Basis von Logfiles gearbeitet (Larose, 2006; Song und Shepperd, 2006). Ein weiteres Anwendungsfeld ist die Optimierung von Servern und Anwendungen durch intelligente Caching-Strategien. Auf Basis der Logfiles wird versucht, die Bewegungen von Nutzerinnen und Nutzern vorherzusagen (Nasraoui et al., 1999) beziehungsweise häufig frequentierte Bereiche der Seite oder Anwendung im Zwischenspeicher vorzuhalten (Bonchi et al., 2001; Chang und Chen, 2002; Lan et al., 2000; Nanopoulos et al., 2002; Wu und Chen, 2002). Die Bereiche sind nicht trennscharf. Intelligentes Caching, das dem Bereich der Systemadministration zugeordnet ist, spielt auch im Bereich des E-Commerce eine Rolle, wenn es um einen möglichst schnellen Aufbau der Webseiten geht. Ähnliches gilt für den Bereich der *Personalisierung* (Pierrakos et al., 2003). Die Inhalte von Webseiten werden dynamisch generiert und basieren auf bisher getätigten Besuchen der Nutzerin beziehungsweise des Nutzers (Adomavicius und Tuzhilin, 2001; Anderson, 2002; Mobasher, Dai, Luo und Nakagawa, 2002; van der Meer et al., 2000; Wong

et al., 2001). Das betrifft prinzipiell alle dynamisch generierten Webseiten (Mobasher, Cooley et al., 2000, S. 143).¹ Die Logfile-Analysen im geschäftlichen Umfeld werden häufig als *Web Analytics* (Hassler, 2010) bezeichnet. Unter dieser wird allerdings keine Logfile-Analyse im klassischen Sinne verstanden, sondern die Analyse der Webseitennutzung durch *Page Tagging* und spezielle Programme (beispielsweise Google Analytics).

Neben diesen vier groben Bereichen der Logfile-Analyse (Systemadministration, E-Commerce, Marketing, Caching) wird Web Usage Mining zunehmend in der Forschung angewandt (Priemer, 2004). Lee, Chen et al. (2009) untersuchen die Nutzung eines Lern-Management-Systems (LMS) durch Logfile-Analysen. Im Bereich des eLearnings hat sich mit dem *Educational Data Mining* in den letzten Jahren eine eigene Kategorie innerhalb des Data Minings herausgebildet (Baker und Yacef, 2009; Berland et al., 2014; Bienkowski et al., 2012; Romero und Ventura, 2007, 2010). Schulz und Breiter (2012, 2013a) analysieren die Nutzung von Schulinformationssystemen (SIS) auf der Basis von Logfiles.

Der Begriff *Logfile* bezieht sich nach dem heutigen Sprachgebrauch auf die von Webservern erstellten Logfiles und wird in dieser Arbeit entsprechend verwendet. Die nachfolgende Abbildung zeigt die Struktur einer Webseite. Die Kästen stehen für einzelne (Unter-)Seiten, die Pfeile für Hyperlinks zwischen den einzelnen Seiten.

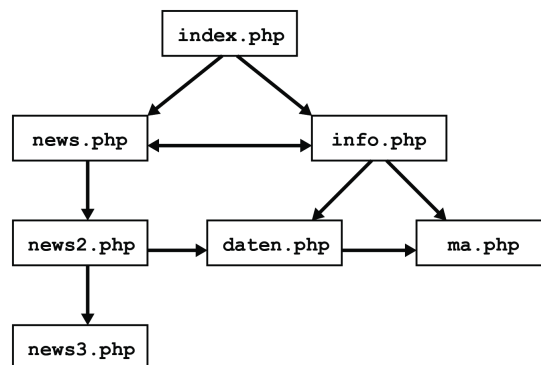


Abbildung 4.1: Beispielhafte Seitenstruktur

Für die Erfassung der Logfiles kommen in der Regel zwei Formate zum Einsatz: das *Common Logfile Format* (World Wide Web Consortium (W3C), 1995) und das *Extended Logfile Format* (Hallam-Baker und Behlendorf, 1996). Beide Formate werden vom *World Wide Web Consortium (W3C)*² definiert und aktualisiert. Wird ein Besuch mit der folgenden Reihenfolge an Aufrufen einzelner Unterseiten angenommen (*index.php* → *news.php* → *info.php*), erscheint der folgende Eintrag in der Logfile nach dem *Common Logfile Format*:

¹Darüber hinaus werden Logfiles auch in anderen Forschungsbereichen wie der Psychologie oder den Sozialwissenschaften angewandt. Allerdings ist die Verbreitung bisher vergleichsweise gering.

²<http://www.w3.org>

Listing 4.1: Common Logfile Format

```
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /index.php HTTP/1.1" 200 4521
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /nav.js HTTP/1.1" 200 8713
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /title.jpg HTTP/1.1" 200 34076
10.231.576.83 - - [12/Oct/2013:13:02:10 +0100] "GET /news.php HTTP/1.1" 200 6549
10.231.576.83 - - [12/Oct/2013:13:02:47 +0100] "GET /info.php HTTP/1.1" 200 2854
```

Der Aufbau des *Extended Logfile Format* ist ähnlich, die Einträge wurden um zusätzliche Informationen erweitert:

Listing 4.2: Extended Logfile Format

```
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /index.php HTTP/1.1" 200 4521
"- " "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /nav.js HTTP/1.1" 200 8713
"- " "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
10.231.576.83 - - [12/Oct/2013:13:01:56 +0100] "GET /title.jpg HTTP/1.1" 200 34076
"- " "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
10.231.576.83 - - [12/Oct/2013:13:02:10 +0100] "GET /news.php HTTP/1.1" 200 6549
"http://seite.de/index.php" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
10.231.576.83 - - [12/Oct/2013:13:02:47 +0100] "GET /info.php HTTP/1.1" 200 2854
"http://seite.de/news.php" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
```

Die beispielhaften Auszüge der beiden Logfile-Dateien enthalten folgende Informationen: Neben den drei besuchten Seiten (Zeile 1, 4, 5) werden automatisch eine Skriptdatei (`nav.js`) und ein Bild (`title.jpg`) geladen. Die Einstiegsseite lädt diese Dateien automatisch im Hintergrund.³ Im ersten Feld wird der *Remotehost* angegeben. Dabei handelt es sich entweder um den DNS-Namen oder die IP-Adresse des anfragenden Computers (hier `10.231.576.83`). Die IP-Adresse ist nur bedingt eindeutig, da in der Regel die IP-Adresse des Routers sichtbar ist. Ein DNS-Name beziehungsweise eine IP-Adresse erlauben es, das Land aus dem die Anfrage kommt, zu ermitteln. Die nächsten beiden Felder geben den Nutzernamen bei integrierten Authentifizierungsverfahren an (*Rfc931*, *Authuser*). Beide Verfahren werden selten verwendet und sind vernachlässigbar. Der überwiegende Teil der Authentifikationen findet heute nicht auf der Serverebene, sondern intern durch die Software statt. Als nächstes werden Datum, Uhrzeit und Zeitzone des Zugriffs gespeichert. Der folgende Eintrag wird als *Request* bezeichnet und gibt darüber Auskunft, welche Aktion durchgeführt wurde, speichert die aufgerufene Seite und das verwendete HTTP-Protokoll. Bei der Art der Aktion wird primär „GET“ für das Abrufen/Herunterladen von Dateien und „POST“ für das Senden von Dateien an den Server verwendet. Die entsprechende Seite wird im URI-Format angezeigt (ohne `http://www.domain.de`). Das geläufige HTTP-Protokoll ist HTTP/1.1 (Fielding et al., 1999). Ältere Versionen sind die Ausnahme. Die Anfrage wird mit einem Statuscode versehen (vgl. Tabelle 4.2 und Fielding et al. (1999)). Wichtig sind der Statuscode „200“ für eine erfolgreiche Anfrage/Übertragung, der Statuscode „404“ („Error 404“) für eine nicht existente Seite sowie der Statuscode „500“ für einen Serverfehler. Neben dem Statuscode wird die Größe der übertragenen Datei in Bytes angegeben (4521 Byte in der ersten Zeile). Bis zu diesem Punkt gleichen sich das Common Logfile Format und das Extended Logfile Format. Die verbleibenden beiden Einträge sind nur im Extended Logfile Format definiert. Der vorletzte

³Der genauere Umgang mit den protokollierten Inhalten wird in Kapitel 5.2.1 erklärt. An dieser Stelle wird zunächst auf die einzelnen Felder der Logfile-Formate eingegangen.

Eintrag ist der *Referer*. Dieser gibt an, von welcher Seite die Nutzerin beziehungsweise der Nutzer kommt. Die Nutzerin ruft in der Listing 4.2 die Seite `index.php` auf. Gleichzeitig werden ein JavaScript (`nav.js`) und ein Bild übertragen (`title.jpg`). In der vierten Zeile wechselt die Nutzerin dann auf die Seite `news.php`. In dem Referer ist die vorherige Seite (`index.php`) angegeben. Im Gegensatz zum *Request* wird beim Referer die komplette URL angegeben. Dies macht es möglich, externe verweisende Seiten zu identifizieren. Der Referer kann nur erfasst werden, wenn die Nutzerin beziehungsweise der Nutzer einem Link folgt. Die direkte Eingabe einer URL in die Adressleiste des Browsers wird nicht erfasst. In dem letzten Feld (*User-Agent*) wird der Browser inklusive der Version und dem verwendeten Betriebssystem protokolliert.

Tabelle 4.1: Felder des Extended Logfile Format (nach Hallam-Baker und Behlendorf, 1996)

Feld	Inhalt	Beispiel
Remotehost	DNS Name oder IP Adresse des Clients	10.231.576.83
Rfc931	Der entfernte Login des Nutzers (selten verwendet)	-
Authuser	Der Login, der zur Authentifizierung verwendet wurde (selten verwendet)	-
Date	Datum und Uhrzeit mit Angabe der Zeitzone (relativ zu GMT)	[12/Oct/2013:13:01:56 +0100]
Request	Die Anfrage des Clients, bestehend aus drei Teilen: Typ (GET/POST), angeforderte Seite (index.php), verwendetes Protokoll ("HTTP/1.1")	GET /index.php HTTP/1.1
Status	Statuscode des Webservers für die Anfrage (vgl. Tabelle 4.2)	200
Bytes	Größe der Antwort/ausgelieferten Datei/Seite in Bytes	4521
Referer	Herkunft des Nutzers, Angabe über die zuletzt aufgerufene Seite	http://seite.de/index.php
User-Agent	Angabe des verwendeten Browsers mit Angaben der Version, des Betriebssystems und weiterer Informationen	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)

Die Logfiles erlauben es, die Bewegungen einer Nutzerin beziehungsweise eines Nutzers zu verfolgen und zu analysieren. Besonders wichtig sind der DNS-Eintrag/IP-Adresse, die aufgerufene URI und der Referer. Mit diesen Informationen können unter optimalen Umständen die einzelnen Nutzerinnen und Nutzer voneinander separiert werden und die jeweiligen Bewegungsmuster erkannt werden. DNS-Eintrag/IP-Adresse und die Nutzerinformationen helfen dabei, echte Nutzer von Suchmaschinen zu unterscheiden und zu trennen.

Der Nachteil von Logfiles ist, dass nur das Verhalten der Nutzerinnen und Nutzer in den Logfiles erhoben wird. Weitere Merkmale der untersuchten Population oder Stichprobe wie Alter, Geschlecht oder Erfahrung im Umgang mit Computern können nicht aus den Logfiles entnommen werden. Sofern verfügbar, können die Nutzerprofile mit diesen Informationen verknüpft werden. Dieser Vorgang ist nicht unproblematisch: Zum einen muss aus technischer Sicht gewährleistet sein, dass sich die verschiedenen Datenquellen verknüpfen lassen. Dies ist nicht immer gegeben, da Logfiles von Webservern in der Regel kein Nutzerfeld aufweisen. Sofern die IP-Adressen oder sonstigen Merkmale dem Nutzer nicht zugeordnet werden können, ist die Zusammenführung mit anderen Daten nur mit

Tabelle 4.2: HTTP 1.1 Statuscodes (nach Fielding et al., 1999)

Code	Beschreibung
100	Continue (Weiter)
101	Switching Protocols (Umstellung der Protokolle)
200	OK (Erfolgreich)
201	Created (Erstellt)
202	Accepted (Akzeptiert)
203	Non-Authoritative Information (Nicht verbindliche Informationen)
204	No Content (Kein Content)
205	Reset Content (Content zurücksetzen)
206	Partial Content (Teil-Content)
300	Multiple Choices (Mehrfachauswahl)
301	Moved Permanently (Dauerhaft verschoben)
302	Found (Zeitweilig verschoben)
303	See Other (Anderen Speicherort aufrufen)
304	Not Modified (Nicht geändert)
305	Use Proxy (Proxy verwenden)
306	Unused (Nicht vergeben)
307	Temporary Redirect (Vorübergehende Weiterleitung)
400	Bad Request (Ungültige Anfrage)
401	Unauthorized (Nicht autorisiert)
402	Payment Required (Zahlung erforderlich) - nicht genutzt!
403	Forbidden (Verboten)
404	Not Found (Nicht gefunden)
405	Method Not Allowed (Methode nicht zulässig)
406	Not Acceptable (Nicht möglich)
407	Proxy Authentication Required (Proxy-Authentifizierung erforderlich)
408	Request Timeout (Zeitüberschreitung bei Anfrage)
409	Conflict (Konflikt)
410	Gone (Gelöscht)
411	Length Required (Längenangabe erforderlich)
412	Precondition Failed (Vorbedingung fehlgeschlagen)
413	Request Entity Too Large (Angeforderte Entität zu groß)
414	Request-URI Too Long (Angeforderte URI ist zu lang)
415	Unsupported Media Type (Nicht unterstützte Medientypen)
416	Requested Range Not Satisfiable (Ungültiger Anfragebereich)
417	Expectation Failed (Nicht wie erwartet)
500	Internal Server Error (Interner Serverfehler)
501	Not Implemented (Nicht implementiert)
502	Bad Gateway (Falsches Gateway)
503	Service Unavailable (Service nicht verfügbar)
504	Gateway Timeout (Gateway-Timeout)
505	HTTP Version Not Supported (HTTP-Version nicht unterstützt)

unverhältnismäßig hohem Aufwand möglich. Ist eine eindeutige Identifizierung gegeben, bedarf es vorsichtigem Umgang mit den Daten. Der Umgang ist zum einen gesetzlich geregelt⁴ und eingeschränkt, zum anderen kann die Zustimmung der Nutzerinnen und Nutzer zu nicht anonymisierten Daten unter Umständen gering sein.

Datenquellen

Die Protokollierung der Logfiles kann auf drei verschiedenen Ebenen erfolgen (vgl. Abbildung 4.2; Srivastava et al., 2000). Bei der *Server Level Collection* werden alle Aktionen auf dem Server protokolliert. Üblicherweise erfolgt die Protokollierung durch den Webserver. Diese Methode stellt die geringsten Anforderungen an die Protokollierung und ist einfach zu etablieren. Die Logfiles liegen in einem standardisierten Format (Common/Extended Logfile Format oder eigenes definiertes Format) vor. Es handelt sich um das am häufigsten angewandte Verfahren. Daneben kann auch die Anwendung selber Protokollfunktionen beinhalten. Unter Umständen kann dies zu mehr Informationen in den Logfiles führen (Baglioni et al., 2003). Der Vorteil dieser Methode ist, dass die Daten im Hintergrund und unbemerkt von der Nutzerin beziehungsweise dem Nutzer erfasst werden. Dem stehen entsprechende datenschutzrechtliche Bedenken gegenüber, weshalb zumindest in Deutschland viele Logfiles anonymisiert gespeichert werden (beispielsweise durch das Entfernen/Anonymisieren eines Teils der IP-Adresse). Wird anonymisierte Speicherung angewandt, liegt der Nachteil darin, dass die einzelnen Nutzer unter Umständen nicht mehr eindeutig voneinander zu trennen sind. Dies kann dazu führen, dass irrtümlich zwei oder mehr einzelne Sessions zusammengefasst werden. Diese Probleme lassen sich teilweise umgehen (vgl. Kapitel 5.2.1). Bei der *Proxy Level Collection* erfolgt die Protokollierung über einen Proxy-Server. Dabei handelt es sich um einen zwischengeschalteten Server, der die Anfragen des Clients annimmt und die ausgelieferten Webseiten des Servers zwischenspeichert. Internetprovider und Firmen/Institutionen setzen Proxy-Server zur Minderung des Internettraffics ein. Die Form der Protokollierung erfolgt in ähnlichen Formaten wie bei der Server Level Collection. Wie bei dieser erfolgt die Erhebung der Logfiles im Hintergrund. Die Proxy Level Collection stellt höhere Anforderungen an den Datenschutz, da der gesamte unverschlüsselte Internetverkehr der Nutzerinnen und Nutzer mitgeschnitten wird. Die letzte Möglichkeit zur Erfassung der Logfiles stellt die *Client Level Collection* dar. Dabei wird die Aktivität auf dem Rechner der Nutzerin beziehungsweise des Nutzers protokolliert. Dies kann unter anderem durch ein Zusatzprogramm, Skripte auf der Webseite (Shahabi und Kashani, 2002) oder einen modifizierten Browser (Catledge und Pitkow, 1995) erfolgen. Diese Art der Datensammlung erfordert die volle Kooperation der Nutzerin beziehungsweise des Nutzers, da diese neue Programme oder Skripte zulassen beziehungsweise installieren müssen. Der Vorteil liegt darin, dass die Probleme des Caching und der Sessions nicht mehr existieren: „Client-side collection has an advantage over server-side collection because it ameliorates both the caching and session identification problem.“ (Srivastava et al., 2000, S. 13) Zudem können nicht nur die reinen Aktionen („Klicks“) erfasst werden, sondern das generelle Verhalten der Nutzerinnen und Nutzer (Fenstermacher und Ginsburg,

⁴Siehe dazu § 11 des Bundesdatenschutzgesetzes (BDSG)

2002). Dem gegenüber stehen die größten Datenschutzbedenken aller drei genannten Verfahren.

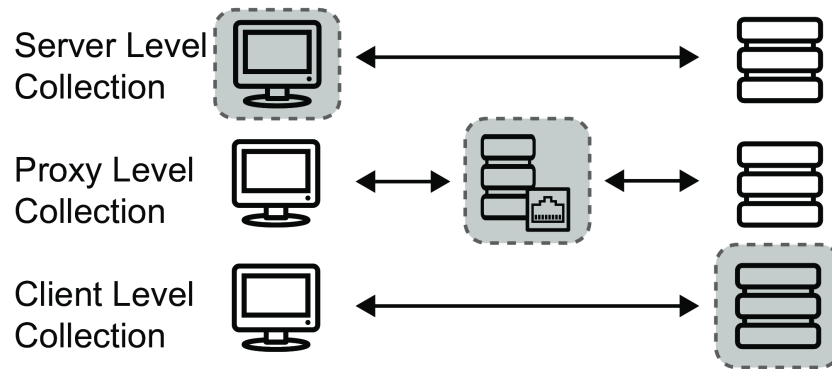


Abbildung 4.2: Datenquellen für Logfiles

5 Wissensentdeckung in Logfiles

Beim *Web Mining* handelt es sich um ein relativ junges Feld, was in erster Linie mit der Verbreitung des Internets in den letzten 20 Jahren zusammenhängt. Der erste Webserver wurde von Tim Berners-Lee 1991 vorgestellt, der erste Webbrowser (Mosaic) 1993. Das *Web (Data) Mining* (Liu, 2011; Markov und Larose, 2007; Theodoridis und Koutroumbas, 2009) geht aus dem klassischen Data Mining/KDD (vgl. Kapitel 5.1) hervor, unterscheidet sich allerdings teilweise davon. Die Methode ist speziell auf das Internet als Datenquelle zugeschnitten: „*Web mining* can be broadly defined as the discovery and analysis of useful information from the World Wide Web“ (Cooley et al., 1997b, 1, H.i.O.) Es unterteilt sich grob in drei verschiedene Untergebiete: *Web Content Mining*, *Web Structure Mining* und *Web Usage Mining* (Borges und Levene, 2000; Cooley et al., 1997b; Kosala und Blockeel, 2000; Liu, 2011; Madria et al., 1999).

Web Content Mining (vgl. Balabanovic und Shoham, 1995; Chakrabarti, 2000, 2002; Cohen et al., 2000) beschäftigt sich mit dem Inhalt von Webseiten und Webanwendungen. Ein Anwendungsgebiet ist die Klassifizierung und Gruppierung von Webseiten (Liu, 2011, S. 7). Hier kommen vor allem textanalytische Verfahren zum Einsatz (*Text Mining*; Gupta und Lehal, 2009; Heyer et al., 2006), aber auch die Darstellung der Inhalte spielt eine Rolle. Für Suchmaschinen ist der Inhalt einer Webseite von hoher Bedeutung (Cooley et al., 1997b).

Dies trifft auch auf das *Web Structure Mining* zu (vgl. Chakrabarti, Dom, Kumar et al., 1999; Chakrabarti, Dom und Indyk, 1998; Han und Elmasri, 2004; Henzinger, 2000; Hou und Zhang, 2003; Kleinberg et al., 1999; Kuo und Wong, 2000). Hier geht es darum, die Strukturen von Webseiten zu erfassen. Die Hyperlinks zwischen den einzelnen Seiten ermöglichen es, die Struktur der Gesamtseite (oder auch mehrerer zusammenhängender Seiten) zu analysieren. Die Forscher sind nicht zwangsläufig auf vorhandene Daten angewiesen, sondern können Verlinkungen und die sich daraus ergebenden Strukturen in vielen Fällen auch per *Crawler* (Chakrabarti, 2002) erfassen.

Dies trifft nicht auf das *Web Usage Mining* zu (Cooley et al., 1997b, 1999; Srivastava et al., 2000). Um das Nutzerverhalten zu analysieren, muss dieses während der Nutzung der Webseite aktiv mit protokolliert werden. Eine Erfassung im Nachhinein ist nicht oder nur mit unverhältnismäßig hohem Aufwand möglich, weshalb in der Regel auf aufgezeichnete Logfiles zurückgegriffen wird. Das generelle Erkenntnisinteresse von *Web Usage Mining* besteht darin, auffälliges und relevantes Verhalten der Nutzerinnen und Nutzer zu identifizieren (Facca und Lanzi, 2005). Das eigentliche Generieren der Nutzerprofile erfolgt über verschiedene Verfahren. Neben klassischen Methoden, die auch in anderen wissenschaftlichen Disziplinen Verbreitung finden, können auch speziellere Verfahren angewendet werden (vgl. für einen Überblick Han, Kamber et al., 2012; Larose, 2005; Theo-

doridis und Koutroubas, 2009). Unter die klassischen Verfahren fallen unter anderem *clusteranalytische Verfahren* (die Zuordnung von Nutzern zu Gruppen), *Bayes-Verfahren* (die Zuordnung von Nutzern zu Gruppen mittels Wahrscheinlichkeiten), *Support Vector Machines* (die Zuordnung von Nutzern zu Gruppen unter Verwendung sogenannter Kernel) oder *Neuronale Netze*. Spezielle Verfahren sind beispielsweise *Association Rules* (die Identifizierung von zusammenhängenden Daten; Agrawal, Imieliski et al., 1993) oder *Sequential Pattern* (die Identifizierung von zusammenhängenden Daten unter Berücksichtigung der Reihenfolge; Agrawal und Srikant, 1995).

Die Unterscheidung zwischen den drei Gebieten ist nicht klar trennbar, sondern es kommt häufig zu Überschneidungen. Dies ist meistens dann der Fall, wenn Inhalte von Webseiten für die Erstellung von Nutzerprofilen mit herangezogen werden. Li und Zaïane (2004) kombinieren beispielsweise alle drei Verfahren für ein *Recommendation System*.

5.1 Wissensentdeckung in Datenbanken

Die Datenbestände, mit denen in Logfile-Analysen gearbeitet wird, belaufen sich in der Mehrzahl der Fälle auf mehrere hundert Megabyte und reichen bis in den Gigabyte-Bereich hinein (Pabarskaite und Raudys, 2007). Um Datenbestände dieser Größe erfassen und auswerten zu können, wird auf *Knowledge Discovery in Databases* (KDD) (Fayyad et al., 1996) zurückgegriffen. Die vorgegebene Struktur der Logfiles entspricht einer tabellenähnlichen Form, die sich in eine Datenbank überführen lässt. Daher lässt sich die Logfile-Analyse ebenfalls in dem Bereich einordnen. Ziel dieses Verfahrens ist es, durch den Einsatz entsprechender Algorithmen und Verfahren Muster und Auffälligkeiten in den Daten zu erkennen. Der Begriff wurde erstmalig von Piatetsky-Shapiro (1990) geprägt. Damit sollte ausgedrückt werden, dass es sich um einen Prozess handelt, an dessen Ende ein Erkenntnisgewinn steht: „The phrase *knowledge discovery in databases* was coined at the first KDD workshop in 1989 (Piatetsky-Shapiro, 1990) to emphasize that knowledge is the end product of a data-driven discovery.“ (Fayyad et al., 1996, 39, H.i.O.) Der Begriff der KDD ist insbesondere in den Bereichen der *Artificial Intelligence* (Russell et al., 2010) und dem *Machine Learning* (Hastie et al., 2011) geläufig (Fayyad et al., 1996, S. 30). Geläufiger ist heute allerdings der Begriff des *Data Minings* (Han, Kamber et al., 2012; Kargupta, 2004; Larose, 2005, 2006; Markov und Larose, 2007; Nisbet et al., 2009; Witten und Frank, 2005). Nach dem Verständnis der Knowledge Discovery in Databases ist das Data Mining die eigentliche Datenanalyse: „The DM (Data Mining, Anm.) phase concerns, mainly, to the means by which the patterns are extracted and enumerated from data.“ (Azevedo und Santos, 2008, S. 182) In der Praxis steht der Begriff des Data Minings allerdings als Synonym für den Begriff des KDDs (Benoît, 2002).

Zur Analyse der Daten wurden in den vergangenen Jahren mehrere Prozessmodelle entwickelt (vgl. für einen Überblick Cios et al., 2007, 11ff.). Brachman und Anand (1996) definieren KDD als mehrstufigen Prozess (vgl. Abbildung 5.1), ausgehend von den Rohdaten (*Data*) und endend bei neu gewonnenem Wissen (*Knowledge*). Es handelt sich um ein akademisches Modell.

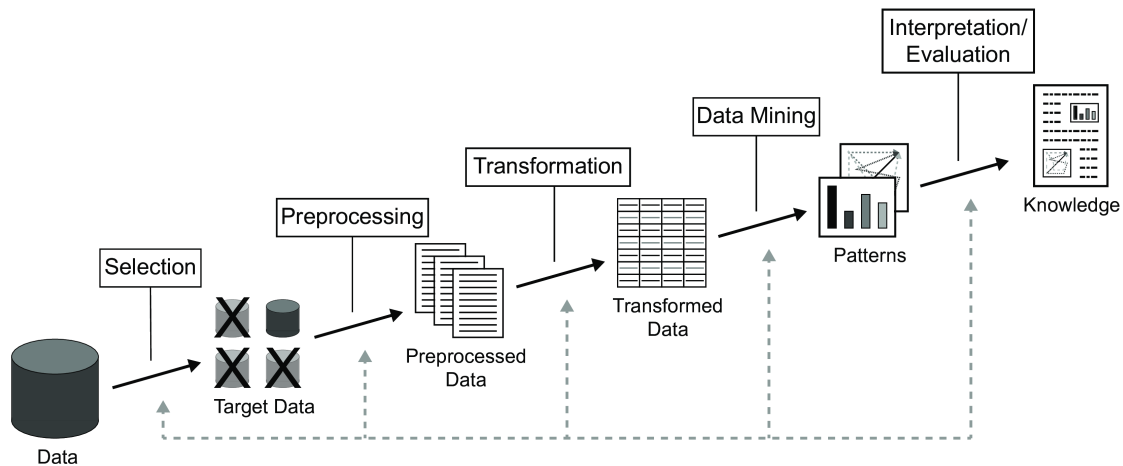


Abbildung 5.1: KDD Ablauf nach Brachman und Anand (1996)

Zunächst ist es wichtig, im Hinblick auf die Ziele beziehungsweise die Forschungsfragen aus allen verfügbaren Daten die relevanten Informationen zu extrahieren (*Selection*). Aufgrund der hohen Datenmenge und der Komplexität der Daten ist häufig domänen-spezifisches Wissen nötig. Die ausgewählten Daten werden in einem nächsten Schritt bereinigt (*Preprocessing*) und in das richtige (Daten-)Format überführt (*Transformation*). Die Wahl des richtigen Formats ist essentiell für die weitere Verwendung der Datenanalyse, das *Data Mining*. Ein falsches Format der Daten führt eventuell dazu, dass die Daten nicht ausgewertet werden können. Die Daten können visuell analysiert werden (Few, 2009; Lima, 2011; Yau, 2011) oder komplexere statistische Methoden angewendet werden (Cios et al., 2007; Han, Kamber et al., 2012). Darunter fallen Clusteranalysen (Berkhin, 2002; Everitt et al., 2011, vgl. auch Gleichung 5.2.2), Assoziationsanalysen (Agrawal, Imieliski et al., 1993, vgl. auch Kapitel 5.2.2), oder Klassifikationen (Gordon, 1999, vgl. auch Gleichung 5.2.2). In einem letzten Schritt werden die aus den Daten generierten Ergebnisse analysiert und interpretiert (*Interpretation*). Werden in dem Schritt Fehler oder Lücken in den Analysen entdeckt, kann beliebig viele Schritte zurückgegangen werden. So wird je nach Komplexität des Problems beziehungsweise der benötigten Daten an einem der vorherigen Schritte (*Selection*, *Preprocessing*, *Transformation*, *Data Mining*) erneut angesetzt. Gleiches gilt für den Fall, dass sich aus den Ergebnissen neue Fragestellungen ergeben.

5.2 Vorgehen in der Logfile-Analyse

Das *Web Usage Mining* setzt sich aus drei Schritten zusammen: *Datenaufbereitung* (engl. *Data Preprocessing*), *Mustererkennung* (engl. *Pattern Discovery*) und *Interpretation* (engl. *Pattern Analysis*) (vgl. Abbildung 5.2; Mobasher, 2008). Die vorhandenen Logfiles werden in der *Datenaufbereitung* bereinigt und aufbereitet. Am Ende dieses Vorgangs sind die Logfiles nach den Nutzerinnen beziehungsweise Nutzern und nach den einzelnen Besuchen (*Sessions* beziehungsweise *Visits*) sortiert. Tanasa und Trousse (2004) fassen das Erheben und Zusammenfügen von Logfiles unter dem Schritt *data fusion* zusammen, die die Datenaufbereitung beinhaltet. Diese dienen der Mustererkennung als Basis für die

Berechnung und Durchführung der Analyse. Abschließend werden die Ergebnisse in der Interpretation analysiert und interpretiert. Der aufgezeigte Weg sollte idealtypisch verstanden werden und kann je nach Anforderung und/oder Forschungsfrage der Situation entsprechend abgewandelt werden.

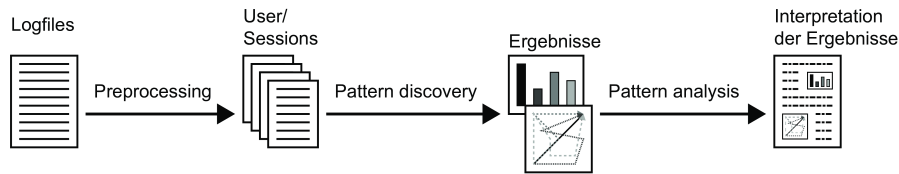


Abbildung 5.2: Ablauf Web Usage Mining

5.2.1 Datenaufbereitung

In dem ersten Schritt, *Datenaufbereitung* genannt, werden die Daten zunächst bereinigt. Anschließend werden sie so aufbereitet, dass sie in dem nächsten Schritt, der *Mustererkennung*, direkt verwendet werden können. Insgesamt kann dieser erste Schritt in vier Unterschritte eingeteilt werden: Darunter fallen *Datenbereinigung* (engl. Data Cleaning), *Identifizierung der Nutzer_innen* (engl. User Identification), *Identifizierung der Besuche* (engl. Session Identification) und *Vervollständigung der Pfade* (engl. Path Completion). Auf die Datenaufbereitung entfällt der größte Arbeitsaufwand. Etwa 80 Prozent des Gesamtaufwandes muss für die Aufbereitung der Daten eingeplant werden (Chitraa und Thanamani, 2011; Grossman und Frieder, 2004; Pabarskaite, 2002; Tyagi et al., 2010). Der tatsächliche Aufwand hängt in erster Linie von der Struktur der vorhandenen Rohdaten ab. Sind diese bereits automatisiert aufbereitet worden, kann dies den Aufwand deutlich reduzieren.

Datenbereinigung¹

Ziel der *Datenbereinigung* ist es, die Logfiles soweit zu bereinigen, dass unerwünschte Einträge nicht mehr vorhanden sind. Unerwünscht sind alle Einträge, die nicht direkt von der Nutzerin beziehungsweise dem Nutzer aufgerufen werden. Wird beispielsweise eine Seite mit einem Bild (gif, png, jpg/jpeg) aufgerufen, können je nach Konfiguration sowohl der Aufruf der Seite als auch der Aufruf des eingebundenen Bildes protokolliert werden. Der Aufruf des Bildes erfolgt automatisch im Hintergrund und würde weitere Analysen verfälschen. Bei den Bildern handelt es sich häufig um Buttons oder Grafiken in der Titelleiste beziehungsweise des Hintergrunds.² Gleiches gilt für eingebundene Skripte wie Javascript und Cascading Style Sheets (CSS). CSS sind Dateien, die Gestaltungsanweisungen beinhalten, die das Aussehen und die Reaktion der Seite beeinflussen. Javascript ermöglicht dynamische und interaktive Inhalte. Beide werden automatisch mit der Seite geladen und teilweise protokolliert. Daneben werden Zugriffe auf fehlerhafte und nicht vorhandene Seiten entfernt. Nicht vorhandene Seiten sind mit dem Error 404 („Seite kann nicht gefunden werden“) gekennzeichnet. Fehlerhafte Seiten weisen den

¹In diesem Kapitel entstammen Teile aus der Publikation des DFG-Projekts (Welling et al., 2014).

²Werden Bilddateien direkt aufgerufen und nicht in eine Seite eingebettet, muss dies berücksichtigt und diese Dateien nicht ausgeschlossen werden.

Error 500 („Interner Serverfehler“) auf.³ Erfolgreiche Seitenaufrufe haben dagegen einen Statuscode im 200er Bereich (vgl. Tabelle 4.2; Fu et al., 2001).

In einem letzten Schritt werden Zugriffe von automatisierten Systemen (sogenannte „Crawler“ oder „Bots“) entfernt. Dabei handelt es sich um Zugriffe, die von Suchmaschinen (Google, Yahoo, Bing) stammen. Es handelt sich nicht um menschliche Besucher und damit um keine inhaltlich verwertbaren Ergebnisse. Die einfachste Methode Bots zu identifizieren, ist die Angabe im *User-Agent*-Feld der Logfiles. Bei echten Nutzerinnen und Nutzern findet sich die Angabe des Browsers und des Betriebssystems. In der Regel geben sich Bots und Crawler zu erkennen und können identifiziert und entfernt werden. Daneben können alle Zugriffe auf die Datei `/robots.txt` ausgeschlossen werden. Die Datei ermöglicht es, das Verhalten von Bots zu definieren, unter anderem, welche Seiten und Bereiche indiziert beziehungsweise welche Seiten und Bereiche ausgeschlossen werden sollen (Peacock, 1998). Um die Bots und Crawler zu identifizieren, die sich nicht über das *User-Agent*-Feld oder den Aufruf der `/robots.txt` zu erkennen geben, schlagen Tan und Kumar (2000; 2002) eine Heuristik vor, um echte Besucherinnen und Besucher von Bots und Crawlern zu trennen. Das Hauptmerkmal liegt auf den Bewegungen auf der Seite, die sich von menschlichen Besuchern unterscheiden (mehrere Zugriffe pro Sekunde, keine logische Reihenfolge der Seitenbesuche). Zudem haben Bots und Crawler Probleme, dynamische Seiten mit Benutzereingaben (u. a. Captchas) zu besuchen (Yuan, MacGregor et al., 2002).

Die beispielhaften Logfile-Einträge in Listing 4.2 enthalten mehr Informationen (Spalten) als für spätere Analysen benötigt werden. Dazu gehört die Angabe des HTTP-Protokolls (hier HTTP/1.1) und die Größe der angeforderten Seite (4521 Bytes in der ersten Zeile). Diese und weitere nicht benötigte Zeilen (Statuscode, Methode) werden vor der weiteren Analyse entfernt.

Identifizierung der Nutzerinnen und Nutzer

Nach der Bereinigung der Logfiles erfolgt die Identifikation der einzelnen Nutzerinnen und Nutzer. Erste Ansätze, die teilweise heute noch gültig sind, wurden von Catledge und Pitkow (1995), Pirolli et al. (1996) und Pitkow (1997) entwickelt. Cooley (2000, vgl. Tabelle 5.1) listet sechs verschiedene Möglichkeiten auf. Die Aufzählung wird um die zweite Methode von Pitkow (1997) erweitert.

Am geläufigsten ist die Identifikation über die *IP-Adresse* und die *User-Agent-Angabe*. Kombiniert man beide Felder, ist eine Trennung zwischen den Nutzerinnen und Nutzern anhand dieses Merkmals möglich (Pirolli et al., 1996). Der größte Vorteil dieser Methode ist die einfache Implementierung. Es müssen keine weiteren Daten erhoben werden, da die Protokollierung der benötigten Angaben automatisiert erfolgt. Dies führt zu weniger Bedenken hinsichtlich des Datenschutzes. Das Vorgehen weist aber zwei Probleme auf. Zum einen kann es vorkommen, dass sich mehrere Nutzerinnen und Nutzer einen Internetzugang teilen. Sollten diese Nutzerinnen und Nutzer den gleichen Browser mit derselben Version und auch das gleiche Betriebssystem nutzen, ist eine Unterscheidung zwischen ihnen nicht mehr möglich. Dies tritt vor allem in standardisierten Umgebungen

³Ist die Forscherin beziehungsweise der Forscher genauer an der Analyse von Fehlern (insbesondere 400er und 500er Statuscodes) interessiert, sollten diese nicht ausgeschlossen werden.

Tabelle 5.1: Methoden zur Identifizierung der Nutzerinnen und Nutzer nach Cooley (2000), übersetzt und erweitert durch Pitkow (1997)

Methoden	Beschreibung	Datenschutz	Vorteile	Nachteile
IP-Adresse & User-Agent-Angabe	Annahme, dass jede einzigartige IP-Agent-Kombination einen Nutzer darstellt	gering	Daten vorhanden; einfach durchzuführen	Einzigartigkeit der IP-Agent-Kombination nicht gewährleistet; Probleme bei Anonymisierung
IP-Adresse & User-Agent-Angabe mit Seitenstruktur	Annahme, dass jede einzigartige IP-Agent-Kombination einen Nutzer darstellt; Bewegungen werden zusätzlich mit der Seitentypologie verglichen und verifiziert	gering	Daten vorhanden; relativ einfach durchzuführen	Einzigartigkeit der IP-Agent-Kombination nicht gewährleistet; Probleme bei Anonymisierung; eventuelle Probleme lokalen Caching
Eingefügte Session-ID	Bei dynamischen Seiten wird eine Session-ID mit in die URL integriert	gering/mittel	Immer vorhanden; unabhängig von der IP-Adresse	Session-ID wird bei jedem Besuch neu generiert; komplett dynamische Seite erforderlich; Anpassungen bei vorhandenen Produkten nötig
Registrierung	Nutzer müssen sich bei der Seite registrieren und bei jedem Besuch anmelden	mittel	Einzelne Benutzer und Sessions können genau verfolgt werden	Registrierung nicht immer erwünscht
Cookie	Gespeichertes Cookie auf dem Client ermöglicht eine Identifizierung	mittel/hoch	Kann theoretisch auch Besuche anderer Seiten verfolgen	Kann deaktiviert werden; wird allgemein negativ gesehen
Software	Programm protokolliert und sendet Informationen zurück	hoch	Zuverlässige und genaue Daten für eine einzelne Seite	Kann abgelehnt werden; wird allgemein negativ gesehen
Modifizierter Browser/Browser Plug-In	Browser oder Plug-In zeichnen Surfverhalten auf	sehr hoch	Zuverlässige und genaue Daten über das gesamte Surfverhalten	Software muss von Nutzerinnen und Nutzern installiert werden; wird allgemein negativ gesehen

auf. Zudem verfügt die Mehrheit aktueller Browser über eine automatische Updateroutine. Dies erhöht die Wahrscheinlichkeit, dass alle Nutzerinnen und Nutzer eines Zugangs dieselbe Version eines Browsers nutzen. Daneben kommt es vor, dass Logfiles aus Gründen des Datenschutzes anonymisiert werden. Dies führt ebenfalls zu einer erschwerten Identifizierung der einzelnen Nutzerinnen und Nutzer. In der Regel werden Oktetts der IP-Adresse anonymisiert: Anstatt 10.231.576.83, 10.231.576.90 und 10.231.576.112 werden 10.231.576.0, 10.231.576.0 und 10.231.576.0 gespeichert. Sollten diese drei Nutzerinnen und Nutzer dieselben Browser verwenden, würden sie fälschlicherweise als eine Nutzerin identifiziert.

Eine Erweiterung der Methode stellt die Verwendung der Seitentypologie dar. Die Struktur der Seite, sofern bekannt, kann zur Nutzererkennung hinzugezogen werden. Wenn eine Nutzerin beziehungsweise ein Nutzer eine Seite aufruft und es zwischen dieser und der letzten aufgerufenen Seite keinen Hyperlink gibt, wird davon ausgegangen, dass es sich um eine neue Nutzerin handelt (Chitraa und Thanamani, 2011). Die Problematik dieses Ansatzes liegt darin, dass einzelne Aktionen serverseitig nicht mitgeloggt werden (beispielsweise die Nutzung des „Zurück“-Buttons).

Die dritte vorgestellte Methode integriert eine kryptische Session-ID mit in die URL, sprich die URL `webseite.php` wird um die Session-ID erweitert wie beispielsweise `webseite.php?session=223as128`. Da heute vor allem Skriptsprachen (beispielsweise PHP, CGI) für die Darstellung von Webseiten verwendet werden, ist die Implementierung aus technischer Sicht einfach. Die Session-ID wird randomisiert serverseitig erzeugt und die integrierte URL in den Logfiles gespeichert. Diese Methode ist wie die Verwendung der IP-Browser-Kombination von keinen weiteren Voraussetzungen seitens der Nutzerinnen und Nutzer abhängig. Die generierte Session-ID ist in der URL-Leiste des Browsers sichtbar und kann bei den Nutzerinnen und Nutzern der Webseite für Irritationen sorgen beziehungsweise zu Bedenken hinsichtlich des Datenschutzes führen. Die Methode lässt sich nur auf komplett dynamischen Seiten einsetzen und führt nach einem Verlassen der Webseite zu einer neuen Session-ID, unabhängig davon, ob der Besuch bereits beendet oder das Fenster versehentlich geschlossen wurde.

Die bisher genannten Ansätze kommen ohne Aktionen der Nutzerin beziehungsweise des Nutzers aus. Der vierte Ansatz besteht darin, eine *Registrierung* der Nutzerinnen und Nutzer zu verlangen. Dadurch lässt sich jede Nutzerin beziehungsweise jeder Nutzer individuell erfassen. Fraglich ist allerdings, ob alle Nutzerinnen und Nutzer gewillt sind, sich zu registrieren und bei jedem Besuch erneut einzuloggen. Loggt sich die Nutzerin beziehungsweise der Nutzer bei einem neuen Besuch nicht ein, kann sie nicht erfasst werden.

Daneben ist die Verwendung von Cookies möglich. Ein *Cookie* wird dauerhaft auf dem Client der Nutzerin beziehungsweise des Nutzers gespeichert, was eine Verfolgung über einen längeren Zeitraum ohne Registrierung ermöglicht. Problematisch ist, dass Cookies generell über ein negatives Image verfügen und deaktiviert werden können. Die aktuelle europäische Datenschutzrichtlinie 2009/136/EG (EU, 2009) sieht vor, dass Nutzerinnen und Nutzer über die Verwendung von Cookies informiert werden müssen. Parallel dazu gibt es eine „Do-not-track“-Initiative, die in der Zwischenzeit u. a. durch das

W3C aufgenommen wurde (Singer und Fielding, 2014). Diese sieht vor, dass Nutzerinnen und Nutzer über den Browser angeben können, ob sie verfolgt („getrackt“) werden möchten oder nicht. Eine Verneinung geht in der Regel mit einem Verzicht von Cookies einher.

Die letzten beiden Erweiterungen setzen die Installation von Software auf dem Client-PC voraus. Auf der einen Seite wird ein weiteres Programm installiert, das die Nutzung des Webbrowsers protokolliert. Alternativ ist es möglich, einen modifizierten Browser oder ein Browser-Add-On zu installieren. Das Browser-Add-On stellt die einfachere Möglichkeit der Installation dar. Der Vorteil dieser Methode besteht in den umfangreichen Protokollmöglichkeiten, die über die Webseite hinausgehen. Es ist theoretisch möglich zu erfassen, ob der Browser im Vordergrund des Betriebssystems läuft oder über „Mouse-tracking“ die Bewegungen und „Klicks“ der Maus nachzuvollziehen (Arroyo et al., 2006; Atterer und Schmidt, 2007; Atterer, Wnuk et al., 2006; Fujioka et al., 2007). Die Installation von Programmen oder Add-Ons setzt ein hohes Vertrauen der Nutzerinnen und Nutzer voraus, da die Programme theoretisch sensible Daten erfassen beziehungsweise die gesamte Computernutzung protokollieren können. Gleichzeitig sind die Anforderungen des Datenschutzes zu beachten. Diese beiden Möglichkeiten werden daher primär in kontrollierten Umgebungen (Laborsituation) eingesetzt.

Identifizierung der Besuche

Nach der Identifizierung und Trennung der Nutzerinnen und Nutzer werden in dem dritten Schritt die einzelnen Besuche („Sessions“) der Nutzerin oder des Nutzers getrennt (Spiliopoulou, Mobasher et al., 2003). Definiert wird ein Besuch wie folgt: „A user session can be defined as a set of pages visited by the same user within the duration of one particular visit to a website.“ (Chitraa und Thanamani, 2011, S. 24) Dies ist die kleinste Einheit, in die Logfiles beim Web Usage Mining aufgeteilt werden (Cooley et al., 1997a, S. 4). Der Grund für das Aufsplitten in Besuche ist, dass eine Nutzerin beziehungsweise ein Nutzer eine Seite über einen längeren Zeitraum hinweg wahrscheinlich mehrmals besucht. „Die Intention der Besuche kann sich voneinander unterscheiden.“ (Welling et al., 2014, S. 83)

Für die Trennung der Sessions wird in der Literatur auf *drei Heuristiken* zurückgegriffen.⁴ Die ersten beiden Heuristiken orientieren sich an der Verweildauer der Nutzerin beziehungsweise des Nutzers auf der Seite, die letzte Heuristik orientiert sich dagegen an den Pfaden der Nutzerin beziehungsweise des Nutzers auf den einzelnen Unterseiten. Berendt et al. (2002) und Spiliopoulou, Mobasher et al. (2003) sprechen von zwei *time-orientated heuristics* und einer *referer-based heuristic*. Die Heuristiken werden als *h1*, *h2* und *h-ref* bezeichnet (Liu, 2011, S. 536).

h1 orientiert sich an der Gesamtzeit, die eine Nutzerin beziehungsweise ein Nutzer auf der Seite verbringt. Die zeitliche Schwelle wird mit θ definiert, die zwischen dem aktuellen Zeitstempel t und dem anfänglichen Wert t_0 gemessen wird. Diese Heuristik geht auf

⁴In den letzten Jahren wurden weitere Heuristiken entwickelt und vorgestellt, die aber auf mindestens einem der drei ursprünglichen Heuristiken basieren und gegen Ende des Absatzes genauer vorgestellt werden (Bayir et al., 2008; Chen, LaPaugh et al., 2002; Dell et al., 2008; Huiying und Wei, 2004; Zhou et al., 2006).

Catledge und Pitkow (1995) zurück, die einen Mittelwert von 25,5 Minuten pro Besuch ermittelt haben. Der ermittelte Wert gilt bis heute (Facca und Lanzi, 2005), wird häufig der Einfachheit halber mit 30 Minuten ersetzt (Cooley et al., 1999, S. 13). Konkret wenden Borges und Levene (2000), Catledge und Pitkow (1995), Domenech und Lorenzo (2007), Etminani, Akbarzadeh-T. et al. (2009), Etminani, Rezaeian Delui et al. (2009), Nasraoui et al. (1999), Paliouras et al. (2000), Pei et al. (2000), Pitkow (1997), Wu, Yu et al. (1998) und Zhou et al. (2006) dieses Verfahren an.

h_2 zieht nicht die gesamte Länge des Besuchs heran, sondern die verstrichene Zeit zwischen zwei Aktionen. Die Schwelle δ definiert den Wert, der maximal zwischen zwei Zeitpunkten (t_2 und t_1) erreicht werden darf. Wird der Wert überschritten, kann davon ausgegangen werden, dass es sich um einen neuen Besuch handelt. Die Methode findet eine breitere Verwendung (Banerjee und Ghosh, 2000; Cooley et al., 1999; Spiliopoulou und Faulstich, 1999). Als Schwellenwert gelten in der Regel 10 Minuten (Berendt et al., 2002; Spiliopoulou, Mobasher et al., 2003). Die Heuristik lässt sich gut bei Anwendungen mit einem benötigten Login einsetzen. Häufig werden Nutzerinnen und Nutzer nach einer definierten Zeitspanne automatisch ausgeloggt. Weiteren Aktionen in der Anwendung geht ein erneutes Einloggen voraus.

$h\text{-ref}$ geht einen anderen Weg und definiert einen Besuch S danach, ob der Referer der aktuellen Seite mit der vorherigen Seite übereinstimmt. Ist dies der Fall, kann davon ausgegangen werden, dass es sich um den gleichen Besuch handelt. Der Vorteil dieses Verfahrens liegt in der Variabilität, da die Dauer eines Besuchs (h_2) beziehungsweise die Dauer zwischen zwei Aktionen (h_1) nicht fest definiert werden muss. Die Methode wird ungenau, sobald eine Nutzerin beziehungsweise ein Nutzer mehrere parallele Sessions zum Server aufbaut (beispielsweise in mehreren Fenstern oder Tabs) oder mehrere Nutzerinnen und Nutzer denselben Internetzugang und Browser benutzen. In diesem Fall kann es vorkommen, dass die Referer nicht mehr übereinstimmen und die Heuristik mehr Besuche definiert als tatsächlich vorhanden sind. Gleiches trifft zu, wenn die Logfiles unvollständig sind (beispielsweise durch die Nutzung des „Zurück“-Buttons). Dies kann dazu führen, dass die Referer nicht korrekt sind.

Ein Vergleich der drei Heuristiken durch Spiliopoulou, Mobasher et al. (2003) zeigt, dass die *time-orientated* Heuristiken ein besseres Ergebnis als die *referer-based* Heuristik liefern, wobei die zweite *time-orientated* Heuristik (h_2) leicht bessere Ergebnisse als die h_1 -Heuristik zeigt. Die Schwelle θ für h_1 liegt bei 30 Minuten, die Schwelle δ für h_2 liegt bei 10 Minuten. Die Ergebnisse zeigen zudem, dass die Qualität davon abhängt, mit welchen Methoden die *Identifikation der Nutzer_innen* durchgeführt wurde. Das Setzen eines Cookies führt zu einer genaueren Rekonstruktion der Besuche als die IP-User-Agent-Kombination. Daneben vergleichen Berendt et al. (2002) die drei Heuristiken miteinander. Die *time-orientated* Heuristiken (h_1 und h_2) liefern auf Seiten mit Frames gute Ergebnisse⁵, während die $h\text{-ref}$ Methode bei kurzen Besuchen zu guten Ergebnissen führt. Diese werden

⁵Die Ergebnisse sind heute weniger relevant, da Frames kaum noch zum Einsatz kommen und durch dynamische Seitengenerierung verdrängt wurden. In dem kommenden HTML-Standard (Version 5; Hickson et al., 2014) werden sie aus Gründen der Usability und Barrierefreiheit nicht mehr vorhanden sein (Pieters, 2014).

auch durch die Heuristiken $h1$ und $h2$ erreicht.

Neben den drei etablierten Heuristiken gibt es weitere Ansätze, die Identifikation und Trennung von Besuchen zu verbessern. Alle neueren Ansätze basieren zumindest teilweise auf den oben vorgestellten Ansätzen. Banerjee und Ghosh (2000) erweitern die *referrer-based* Heuristik mit einem Graphen der Webseitenstruktur und überprüfen anhand von diesem, ob die entsprechenden Pfade real sind oder nicht. Im Vergleich zu der $h2$ -Heuristik ist die Methode langsamer und ineffektiver. Bayir et al. (2008) kombinieren alle drei Heuristiken miteinander. Zunächst erfolgt eine Einteilung in Besuche durch die maximale Aufenthaltsdauer ($h1$) und/oder der maximalen Aufenthaltsdauer pro Seite ($h2$). In einem zweiten Schritt werden anhand der Seitentypologie Besuche nach nicht möglichen Pfaden, die durch die Referer erkannt werden, getrennt (*h-ref*). Chen, LaPaugh et al. (2002) erweitern die $h2$ Heuristik um einen *Adaptive Approach*, der das bisherige Surfverhalten der Nutzerinnen und Nutzer mit einbezieht, um Besuche dynamisch zu trennen. Chitraa und Thanamani (2011) entwickeln einen Algorithmus, der den Referer als auch die verbrachte Zeit auf den Webseiten als Datenquelle heranzieht. Ein neuer Besuch wird dann gesetzt, wenn der Referer nicht übereinstimmt oder die Zeit zwischen zwei Seiten deutlich über dem Durchschnitt der anderen Nutzerinnen und Nutzer beziehungsweise Besuchen liegt. Dell et al. (2008) verwenden *Integer Programming*, um Besuche zu identifizieren. Der Vorteil liegt darin, dass alle Besuche parallel identifiziert und weitere Daten der Webseiten und der Seitentypologie hinzugezogen werden können. Huiying und Wei (2004) setzen einen Algorithmus ein, der die User-ID, die IP, den User-Agent, die URL, den Referer und den Zeitstempel in einem Kubus speichert. Der Algorithmus dient dazu, neben den Besuchen die Nutzerinnen und Nutzer zu identifizieren. Zhang und Ghorbani (2004) erweitern die *time-orientated* Heuristiken um die *Maximal Frequent Sequence (MFS)*.

Vervollständigung der Pfade⁶

Als letzter Schritt erfolgt die Vervollständigung unvollständiger Pfade (Path Completion). Browser speichern die letzten aufgerufenen Seiten in dem lokalen *Cache* ab. Wenn die Nutzerin beziehungsweise der Nutzer den „Zurück“-Button anklickt, wird die vorherige Seite nicht erneut vom entfernten Server angefordert, sondern aus dem lokalen Cache geladen. Dies hat zur Folge, dass kein neuer Eintrag in den Logfiles erscheint und ein „Loch“ (Markov und Larose, 2007, S. 170) entsteht, wie in Abbildung 5.3 ersichtlich wird. Es handelt sich um die gleiche Seitenstruktur wie in Abbildung 4.1.

Dargestellt ist eine Webseite, wobei die Kästen eine Seite und die grauen Pfeile Links zwischen den Seiten darstellen. Die Nutzerin beziehungsweise der Nutzer bewegt sich wie folgt durch die Seite (dargestellt durch die gestrichelten Pfeile):

index.php $\xrightarrow{\text{Link}}$ news.php $\xrightarrow{\text{Link}}$ news2.php $\xrightarrow{\text{Link}}$ news3.php $\xrightarrow{\text{Zurück-Button}}$ news2.php
 $\xrightarrow{\text{Zurück-Button}}$ news.php $\xrightarrow{\text{Link}}$ info.php $\xrightarrow{\text{Link}}$ daten.php

⁶In diesem Kapitel entstammen Teile aus der Publikation des DFG-Projekts (Welling et al., 2014).

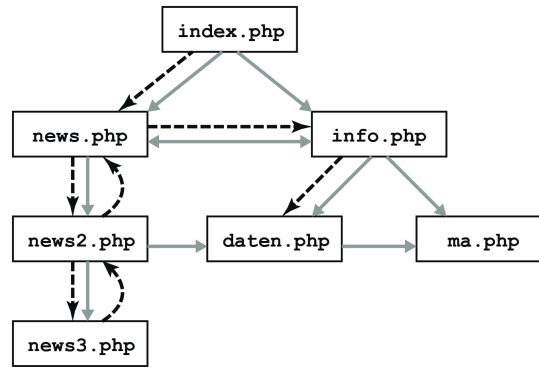


Abbildung 5.3: Path Completion Problem (nach Liu, 2011, 539; graue Pfeile stehen für vorhandene Links, gestrichelte Pfeile für die Bewegungen der Nutzerin)

In dem entsprechenden Logfile fehlen die Bewegungen durch die Nutzung des „Zurück“-Buttons:

	<u>URL</u>	<u>Referer</u>
(1)	index.php	-
(2)	news.php	index.php
(3)	news2.php	news.php
(4)	news3.php	news2.php
(5)	info.php	news3.php
(6)	daten.php	info.php

Das Logfile ist nicht nur unvollständig, sondern auch inkonsistent. Die ersten vier Aktionen auf der Seite konnten richtig und vollständig erfasst werden. Zwischen dem letzten und vorletzten Eintrag (Zeile 4 und 5) entsteht ein „Loch“. Die beiden letzten Aktionen wurden nicht erfasst, weshalb der Eintrag der Zeile 5 der Aufruf der Seite info.php mit dem Referer news3.php ist. Alle Aktionen zwischen diesen Seiten fehlen.

Der geläufigste Weg die fehlerhaften Pfade zu vervollständigen, basiert auf einem einfachen Algorithmus (Cooley, 2000; Cooley et al., 1999; Liu, 2011; Markov und Larose, 2007).⁷ Dieser geht davon aus, dass die Nutzerinnen und Nutzer bei unvollständigen Pfaden stets den „Zurück“-Button verwenden. Der Algorithmus geht schrittweise in dem Logfile zurück bis er auf der Seite landet, die mit dem Referer der vorherigen Seite übereinstimmt. Die zurückgegangenen Schritte werden nachträglich eingefügt. Das vervollständigte Logfile sieht in diesem Fall wie folgt aus:

⁷Daneben existieren noch weitere Algorithmen, die die entsprechenden Pfade vervollständigen (vgl. Li, Feng et al., 2008).

		<u>URL</u>	<u>Referer</u>	
(1)	(1)	index.php	-	
(2)	(2)	news.php	index.php	
(3)	(3)	news2.php	news.php	
(4)	(4)	news3.php	news2.php	
(-)	(5)	news2.php	news3.php	← eingefügt
(-)	(6)	news.php	news2.php	← eingefügt
(5)	(7)	info.php	news.php	
(6)	(8)	daten.php	info.php	

Die fehlenden Einträge zu vervollständigen ist wichtig, da sowohl Catledge und Pitkow (1995) als auch Tauscher und Greenberg (1997) zeigen konnten, dass etwa 30 bis 40 Prozent aller Aktionen die Nutzung des „Zurück“-Buttons ist. Ein Auslassen dieses Schrittes hätte eine unter Umständen starke Verfälschung der Daten zur Folge.

5.2.2 Mustererkennung

Nachdem die Logfiles in der *Datenaufbereitung* bereinigt und vervollständigt wurden, werden im zweiten Schritt, der *Mustererkennung*, die Daten nach Auffälligkeiten und Mustern untersucht. In der Literatur wird vor allem zwischen den folgenden Analysemethoden unterschieden: *Statistische Analysen*, *Assoziationsanalyse*, *Sequential Patterns*, *Clustering* und *Klassifikation* (Baglioni et al., 2003; Cooley et al., 1999; Facca und Lanzi, 2005; Mobasher, Cooley et al., 2000; Srivastava et al., 2000). Alle Verfahren sind nicht für das Web Usage Mining entwickelt, sondern von anderen Forschungsfeldern (KDD, Statistik, KI, Maschinelles Lernen) adaptiert worden (Etminani, Akbarzadeh-T. et al., 2009; Etminani, Rezaeian Delui et al., 2009).

Statistische Analysen

Bei den *statistischen Analysen* (engl. Statistical Analysis) handelt es sich vorrangig um deskriptive Methoden. Diese Analysen geben einen ersten Überblick über die Struktur der vorhandenen Daten (bspw. Nutzung der Webseite). In der Regel handelt es sich um aggregierte Daten, welche aussagen, wie häufig sich jemand in einem bestimmten Bereich aufgehalten hat oder wie viele Personen beispielsweise die Suchfunktion verwendet haben (Lazar et al., 2010). Sie überlappt sich zu einem großen Teil mit den *Web Analytics* (Hassler, 2010). Die deskriptive Analyse der Logfiles kann entweder händisch erfolgen oder automatisiert durch Programme oder Anbieter. Geläufige Programme sind *Webalizer*⁸, *AWStats*⁹ oder *piwik*¹⁰. Bei den (teilweise kommerziellen) Angeboten von Anbietern ist *Google Analytics*¹¹ ein prominentes Beispiel. Je nach Vorgehensmodell des KDD (vgl. Kapitel 5.1) fallen diese Analysen bereits während der Datenaufbereitung an. Han, Kamber et al. (2012) sprechen auch von explorativer Datenanalyse („Exploratory Data Analysis“). Die deskriptive Analyse der Daten ist daneben ein Teil wissenschaftlicher Forschung.

⁸<http://www.webalizer.org> [Aufgerufen am 21.07.2015]

⁹<http://awstats.sourceforge.net> [Aufgerufen am 21.07.2015]

¹⁰<http://piwik.org> [Aufgerufen am 21.07.2015]

¹¹<http://www.google.com/analytics> [Aufgerufen am 21.07.2015]

Teilweise lassen sich auch andere Verfahren unter dem Begriff subsumieren. Dazu zählen zum einen ausgewählte multivariate Analysemethoden (Backhaus et al., 2011). Die Trennung ist aber klar, da streng genommen auch strukturentdeckende Verfahren wie die Clusteranalyse unter die multivariaten Analysemethoden fallen, in der Literatur des Data Minings aber als eigene Methode geführt werden. Es lassen sich vermehrt strukturprüfende Verfahren wie Regressionsanalysen aufnehmen. Die Anwendung ist denen der deskriptiven Analysen ähnlicher, da bereits Annahmen über die Struktur der Daten bestehen und nicht entdeckt werden sollen. Zum anderen fällt die Analyse (*sozialer*) Netzwerke (Hogan, 2008; Jansen, 2006; Stegbauer, 2010; Stegbauer und Hennig, 2012; van der Aalst et al., 2005; Wasserman und Faust, 1994) unter die statistischen Analysen. Bei der Analyse sozialer Netzwerke geht es darum, wie die Beziehungen von Individuen untereinander aussehen. Das wohl prominenteste Merkmal des Verfahrens ist die Darstellung als Netzwerkgrafik. Die Anwendung überlappt sich häufig mit dem *Web Content Mining* wie beispielsweise bei der Analyse von Wikipedia (Stegbauer, 2009).

Statistische Analysen sind für die meisten Analysen von hoher Relevanz. Dies trifft insbesondere auf die deskriptiven Analysen zu, da diese Ergebnisse häufig zur groben Beschreibung der Daten herangezogen werden.

Assoziationsanalyse

Beim Einsatz der *Assoziationsanalyse* (engl. Association Rules) wird versucht, Beziehungen zwischen Variablen zu identifizieren und daraus Regeln zu extrahieren. Es handelt sich um eines der populärsten KDD/Data Mining Verfahren (Hipp et al., 2000, S. 58). Bekannt ist das Verfahren durch die Warenkorbanalyse (Masseglia et al., 2000). Diese versucht herauszufinden, ob sich durch den Kauf von bestimmten Artikeln mit einer Wahrscheinlichkeit voraussagen lässt, dass weitere Artikel gekauft werden. Das Ergebnis der Warenkorbanalyse wird häufig in diesen oder vergleichbaren Sätzen ausgegeben: „Kunden, die Brot kaufen, kaufen mit einer Wahrscheinlichkeit von 60% auch Butter:“

$$\text{Brot} \Rightarrow \text{Butter} [\text{Support} = 10\%, \text{Confidence} = 60\%] \quad (5.1)$$

Die Gleichung 5.1 verfügt über die Kenngrößen *Support* und *Confidence*.

$$\text{Support}, s(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (5.2)$$

Der Support (vgl. Gleichung 5.2) gibt an, wie häufig die Bestandteile der Regel in allen Transaktionen vorkommen. In diesem Fall kommen sowohl Brot als auch Butter in zehn Prozent aller Warenkörbe vor. Dieses Merkmal ist nicht exklusiv. Es ist möglich, dass in einigen Warenkörben neben Brot und Butter auch Milch (o. ä.) vorhanden ist.

$$\text{Confidence}, c(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (5.3)$$

Daher wird neben dem Support auf die Confidence geachtet. Diese gibt an, in wie viel Prozent aller Transaktionen, in denen Brot vorkommt, Butter enthalten ist. Dazu wird der Support beider Items durch den Support für Butter geteilt. Bei der Berechnung der Assoziationsanalyse werden ein *minimaler Support* als auch eine *minimale Confidence* definiert,

um die Anzahl der extrahierten Regeln möglichst klein zu halten. Die Wahl der Werte bestimmt maßgeblich die Anzahl der extrahierten Rules.

Neben der Warenkorbanalyse wird das Verfahren in weiteren Forschungsgebieten wie der Bioinformatik, der medizinischen Forschung, der Textanalyse, Versicherungsanalysen als auch im Web Mining angewendet (Höppner, 2005; Tan, Steinbach et al., 2010). Das Verfahren wurde erstmals von Agrawal, Imieliski et al. (1993) vorgestellt. Häufig verwendet wird der von Agrawal und Srikant (1994) vorgestellte *Apriori-Algorithmus*. In letzter Zeit hat der *FP-Growth-Algorithmus* (Han, Kamber et al., 2012) zunehmende Verbreitung gefunden, da er eine bessere Performance aufweist (Kumar und Rukmani, 2010).

Assoziationsanalysen wurden bis etwa 2005 vergleichsweise häufig bei der Analyse von Lern-Management-Systemen (LMS) eingesetzt (Romero und Ventura, 2007). Ein Grund ist die einfache Anwendbarkeit des Verfahrens (Merceron und Yacef, 2007). Allerdings ist die detailliertere Verwendung des Verfahrens schwieriger. García et al. (2007) benennen die Einstellung der Parameter der Algorithmen, die Entdeckung von zu vielen Beziehungen sowie die Entdeckung von schwachen Beziehungen als problematisch. Dies mag einer der Gründe sein, warum die Assoziationsanalyse weniger eingesetzt wird und zunehmend vor allem Cluster-Analysen zum Einsatz kommen (Mohamad und Tasir, 2013).

Klassifikation

Während die Assoziationsanalyse versucht, ähnliche Handlungsmuster aller Fälle eines Datensatzes zu analysieren, wird mittels Klassifikationsverfahren versucht, neue Daten in vorhandene Kategorien oder andere Einteilungen einzuordnen. Prominente Beispiele für die Anwendung von Klassifikationsverfahren sind die Schufa-Auskunft zur Einschätzung der Bonität oder die Einordnung von E-Mails hinsichtlich der Wahrscheinlichkeit, dass es sich um Spam handelt oder nicht. In beiden Anwendungen wird anhand der Eigenschaften eines neuen Falles (bei E-Mails unter anderem ausgehender Mailserver und Inhalte der E-Mail) geprüft, in welche Kategorie dieser eingeordnet wird (Spam beziehungsweise kein Spam). Die Zuordnung neuer Fälle wird durch das vorherige *Trainieren* ermöglicht. Bereits klassifizierte Fälle (manuell oder maschinell) werden in einem ersten Schritt analysiert. Auf der Basis werden neue Fälle klassifiziert, indem ihre Eigenschaften mit denen der Trainingsdaten verglichen werden. Die Methode wird als maschinelles Lernen bezeichnet (Alpaydin, 2014) und ist dem überwachten Lernen zuzuordnen (Mohri et al., 2012).

Es existieren eine ganze Reihe an unterschiedlichen Klassifikationsmethoden (vgl. für einen Überblick Alpaydin, 2014; Cios et al., 2007; Han, Kamber et al., 2012; Larose, 2005; Mohri et al., 2012; Tan, Steinbach et al., 2010; Witten und Frank, 2005). Häufig genannte Verfahren sind unter anderem *Entscheidungsbäume* (engl. Decision Trees; Rokach und Maimon, 2008), *(Naïve) Bayes Verfahren* (Duda und Hart, 1973), *K-nearest neighbor Verfahren* (Altman, 1992; Cover und Hart, 1967), *Support Vector Machines* (SVM; Cortes und Vapnik, 1995; Vapnik, 1998, 2000) und *Neuronale Netze* (engl. Neuronal Networks). Je nach verwendeter Methode kann die Klassifizierung endgültig oder anhand einer Wahrscheinlichkeit bestimmt werden (Cichosz, 2015).

Das Vorgehen ist bei allen Verfahren gleich. In einem ersten Schritt wird mittels des Trainingsdatensatzes ein Modell (oder auch Kernel) berechnet. In einem zweiten Schritt

werden anhand des Modells neue Fälle und Daten klassifiziert. Das Augenmerk liegt besonders auf der Güte und Qualität des Modells. Diese kann mittels verschiedener Gütekriterien wie der F-Score (auch F-Measure genannt) nach Özgür et al. (2005), van Rijsbergen (1979) und Yang und Liu (1999) erfolgen.

Die Wahl der Methode ist durch die Struktur der vorhandenen Daten und persönlichen Vorlieben der Forscherinnen und Forscher bestimmt. Um die Modelle möglichst einfach zu halten, werden die vorhandenen Eigenschaften der Daten vor der Berechnung des Trainingsmodells häufig verkleinert. Dies wird als *Dimensionality Reduction* (Alpaydin, 2014) oder *Feature Selection* (Cios et al., 2007) bezeichnet. Damit soll zudem eine Überspezifizierung der Modelle verhindert werden. Hinzu kommt, dass insbesondere neuere Verfahren (Neuronale Netze und Support Vector Machines) komplizierter und unübersichtlicher hinsichtlich der zu spezifizierenden Parameter sind.

Klassifikationsverfahren stellen eine zentrale Säule der Mustererkennung dar (Gordon, 1999; Han, Kamber et al., 2012), auch im Bereich des Web Usage Mining (Cooley, 2000; Liu, 2011). Besonders im Bereich der Personalization spielt die Einordnung neuer Fälle auf Basis bereits bekannter Daten eine wichtige Rolle (Anderson, 2002; Eirinaki und Vazirgiannis, 2003; Pierrakos et al., 2003; Pretschner und Gauch, 1999). Die vorhandene Datenlage verhindern allerdings Klassifikationsverfahren, da klassifizierte Daten erst nach dem abschließenden Schritt der Kombination vorliegen.

Cluster-Analyse

Die letzte große Gruppe in der Mustererkennung sind die *Cluster-Analysen*. Mittels der Cluster-Analyse wird versucht, die Nutzerinnen und Nutzer zu Gruppen (*Cluster* genannt) zusammenzufassen. Die den unterschiedlichen Clustern zugeordneten Fälle sollten innerhalb eines Clusters möglichst homogen und im Vergleich zu anderen Clustern möglichst heterogen sein (Berkhin, 2002, S. 1). Das Ziel des Clusters ist vor allem deskriptiver Natur: „Clustering is the process of partitioning a set of items (or grouping individual items) into a set of categories.“ (Vayssières und Plant, 1998, S. 14)

Erste Ansätze aus dem Web Usage Mining Besuche zu gruppieren, stammen von Xie und Phoha (2001). Seitdem ist Clustering eines der am häufigsten eingesetzten Verfahren, sowohl im Web Usage Mining (vgl. für einen Überblick Facca und Lanzi, 2005; Ivancsy und Kovacs, 2006; Liu, 2011; Park et al., 2008) als auch im Data Mining generell (Berkhin, 2002; Tan, Steinbach et al., 2010).

Im Bereich des Clustering gibt es eine hohe Anzahl an unterschiedlichen Verfahren. Dies ist dadurch bedingt, dass der Begriff „Cluster“ nicht eindeutig definiert ist (Estivill-Castro und Yang, 2000). Nach Berkhin (2002) und Ivancsy und Kovacs (2006) kann grundsätzlich zwischen den folgenden Verfahren unterschieden werden: *Hierarchische Verfahren*, *Partitionierende Verfahren*, *Dichte-basierte Verfahren*, *Raster-basierte Verfahren*, *Graphen-basierte Verfahren* und *unscharfe (fuzzy) Verfahren*. Diese Unterteilung hat sich in weiten Teilen der Literatur durchgesetzt (Berkhin, 2002; Everitt et al., 2011; Halkidi et al., 2001; Han, Kamber et al., 2012; Jain et al., 1999; Kaufman und Rousseeuw, 2005; Tan, Steinbach et al., 2010). Neuere Verfahren sind zum Teil durch moderne (und schnellere) Computertechnik populär geworden, da die Berechnungen deutlich aufwendiger sind (vgl. u. a. Hocking et al., 2011).

Zentral war bisher die Unterscheidung zwischen *Hierarchischen* und *Partinierenden Verfahren*. Sie sind auch in anderen Anwendungsgebieten wie der Soziologie oder der Psychologie weit verbreitet. Bekannte partitionierende Verfahren sind der *K-Means Algorithmus* (MacQueen, 1967) oder *Partitioning around Means* (Kaufman und Rousseeuw, 2005, PAM). Bei diesen Verfahren wird eine vorher festgelegte Partitionierung (Anzahl der Cluster) solange durch die Zuordnung der Fälle verändert, bis ein durch den Algorithmus festgelegtes Optimum erreicht ist. Beide Verfahren berechnen Cluster mit *exklusiven* Zuordnungen. Ein Fall wird fest einem Cluster zugeordnet. Das Gegenteil stellen *überlappende* Verfahren dar. Bei diesen Verfahren ist es möglich, dass ein Fall zwei oder mehr Clustern zugewiesen wird. Einen Mittelweg bieten *unscharfe (fuzzy)* Verfahren. Alle Fälle erhalten eine Wahrscheinlichkeit der Zugehörigkeit zu jedem Cluster zwischen 0 und 1. Diese Verfahren werden mittlerweile als eigene Gruppe (*Fuzzy Algorithms*) angesehen. Zum einen handelt es sich meist um die Erweiterung bereits bestehender Verfahren wie das *Fuzzy C-Means Verfahren* (Bezdek et al., 1984). Zum andern spielt der mögliche Ausschluss von Fällen eine Rolle. In der Regel werden alle Fälle in die Analyse mit einbezogen (*complete*). Dies schließt beispielsweise Ausreißer mit ein, die im Zweifel vorher identifiziert und gelöscht werden müssen. Im Gegensatz dazu können *partielle* Verfahren selbstständig Fälle ausschließen. Dadurch können Ausreißer automatisch eliminiert werden. Diese Möglichkeit besteht vor allem bei *Dichte-basierten Verfahren* und *Raster-basierten Verfahren*, da bei diesen Verfahren die Parameter zur Größe der Cluster im Merkmalsraum (häufig als *Vektor Space* definiert) angegeben werden. Es ist möglich, dass Fälle, die von keinem Cluster eingeschlossen werden, von der weiteren Analyse ausgeschlossen werden. Dabei muss es sich aber nicht zwangsläufig um Ausreißer handeln. Dichte-basierte Verfahren bieten den Vorteil, dass sie Datenstrukturen einer beliebigen Form (eng. *Arbitrary Shape*) identifizieren können. Ein beliebtes Beispiel in der Literatur sind Daten in einer S-Kurve. Die Algorithmen suchen zunächst zentrale Fälle. In einem weiteren Schritt fügen sie diesen anhand eines vorher definierten Radius weitere Punkte in der Umgebung hinzu. Der Nachteil der Verfahren ist, dass Parameter wie der maximale Radius vorher festgelegt werden müssen. Bekannte Verfahren sind *DBSCAN* (Ester et al., 1996) oder *OPTICS* (Ankerst et al., 1999).

Unter der *Hierarchischen Clusteranalyse* werden distanzbasierte Clusterverfahren zusammengefasst. Die Verfahren fassen Fälle zu Gruppen zusammen, die hinsichtlich ihrer Merkmale eine geringe Distanz zueinander aufweisen (sich möglichst ähnlich sind). Die Ähnlichkeit der Fälle wird mittels verschiedener Distanzmaße berechnet. Die Distanzen werden anschließend über sogenannte Fusionsalgorithmen schrittweise zusammengefasst. Es wird zwischen dem *agglomerativen* und dem *divisiven* Verfahren unterschieden (Kaufman und Rousseeuw, 2005). Bei dem *agglomerativen Verfahren* handelt es sich um einen „Bottom-up approach“ (Ivancsy und Kovacs, 2006, S. 239), bei dem zunächst n Cluster existieren, die zunehmend zu übergeordneten Clustern fusioniert werden. Das *divisive Verfahren* verfolgt einen „Top-down approach“ (Ivancsy und Kovacs, 2006, S. 239), bei dem zunächst ein Cluster vorhanden ist und zunehmend in kleinere Subcluster aufgeteilt wird. Das agglomerative Verfahren ist in der Praxis geläufiger. Es handelt sich um eines der am häufigsten eingesetzten Clusterverfahren überhaupt (Tan, Steinbach et al., 2010, S. 515). Im Folgenden wird daher nur die *agglomerative hierarchische Clusteranalyse* vertieft.

Die Berechnung der Cluster findet in zwei Schritten statt. Zunächst wird die Unterschiedlichkeit zwischen allen Fällen mittels eines Distanzmaßes berechnet. In einem zweiten Schritt werden die Fälle zu Clustern über einen Fusionierungsalgorithmus gruppiert.¹² Bei der Berechnung der Distanzen werden diese für alle Fallkombinationen berechnet, wodurch eine Distanzmatrix mit der Größe $(n \times n)$ entsteht (vgl. Tan, Steinbach et al. (2010, 65ff) und Rokach (2010, 270ff) für einen Überblick).¹³ Zur Berechnung der Distanzen wird auf die Eigenschaften und Merkmale der einzelnen Fälle zurückgegriffen. Bei der Verwendung von Merkmalen mit verschiedenen Ausprägungen empfiehlt es sich, die Daten vorher zu standardisieren (*z-Transformation*). Dies macht die Variablen vergleichbar. Ansonsten droht die Gefahr, dass Variablen mit einer großen Spannweite und höheren Werten stärker berücksichtigt werden. Bekannte Distanzmaße sind beispielsweise der *Jaccard Metrik*, die *Euklidische Distanz* oder die *Manhattan-Block Distanz* (vgl. für einen Überblick Deza und Deza, 2009). Die *Jaccard Metrik* (vgl. Gleichung 5.4) stammt aus der Mengenlehre. Mit dieser wird die Unähnlichkeit zweier Mengen berechnet (Jaccard, 1912). Je unähnlicher die Merkmale, desto geringer ist die Ähnlichkeit. Der Wertebereich liegt im Wertebereich von $W = [0, 1]$. Bei den invertierten Werten wird von dem Jaccard Koeffizienten gesprochen.

$$d_{Jacc}(x, y) = 1 - J(x, y) = \frac{|x \cup y| - |x \cap y|}{|x \cup y|} \quad (5.4)$$

Im Vergleich zu der Jaccard Metrik berechnen sowohl die *Manhattan-Block Distanz* als auch die *Euklidische Distanz* die Entfernung zwischen zwei Punkten. Bei der *Manhattan-Block Distanz* (vgl. Gleichung 5.5) werden die absoluten Differenzen zwischen den Koordinaten x und y aufsummiert. Die *Euklidische Distanz* (vgl. Gleichung 5.6) berechnet die direkte Distanz zwischen den beiden Koordinaten x und y .

$$d_{Man}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (5.5)$$

$$d_{Eu}(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5.6)$$

Im Bereich des *Information Retrieval* (Grossman und Frieder, 2004) ist die Verwendung von anderen Maßen zur Bestimmung der Ähnlichkeit von beispielsweise Textdokumenten und Textfragmenten weiter verbreitet. Häufig wird auf Winkelmaße wie die *Kosinus-Ähnlichkeit* zurückgegriffen (vgl. Gleichung 5.7; Liu, 2011, 215ff). Diese besagt, dass bei einer vollständigen Gleichheit der als Vektoren präsentierten Besuche ein Winkel von 1 vorliegt (entspricht 0° Grad), bei einer vollständigen Verschiedenheit ein Kosinus von -1 berechnet wird (entspricht 180° Grad). Bei der Analyse von Besuchen macht es Sinn, die (Un-)Ähnlichkeit anhand eines Winkels zu bestimmen (Tan, Steinbach et al., 2010, 74ff). Das Maß wurde bereits in dem Bereich des Web Usage Mining angewendet (Shahabi,

¹²Bei dem diversiven Verfahren spricht man von einem Partitionierungsalgorithmus.

¹³Da die Matrix gespiegelt ist, werden effektiv nur $n \times n/2$ Distanzen berechnet.

Zarkesh et al., 1997; Xiao und Zhang, 2001; Xiao, Zhang et al., 2001).

$$d_{\text{Cos}}(x, y) = 1 - \cos(x, y) = 1 - \frac{x \cdot y}{\|x\| \|y\|} = 1 - \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{k=i}^n y_i^2}} \quad (5.7)$$

Die Wahl des optimalen Distanzmaßes ist nicht trivial. Zum einen lässt sich die Ähnlichkeit von Besuchen unterschiedlich definieren. Eine möglichst hohe Unterschiedlichkeit zwischen zwei Besuchen kann durch das Aufrufen unterschiedlicher Seiten verursacht werden. Dann sollte entweder auf die Jaccard Metrik oder die Kosinus-Ähnlichkeit zurückgegriffen werden. Ist dagegen die Distanz von zwei Besuchen von gesteigertem Interesse, so kann auf metrische Maße wie die Euklidische Distanz oder die Manhattan-Block Distanz zurückgegriffen werden. Bei der Manhattan-Block Distanz bleibt zu beachten, dass diese vor allem bei Gleichabständigkeit einzusetzen ist: „The use of the Manhattan distance is advised in those situations where for example a difference of 1 in the first variable, and of 3 in the second variable is the same as a difference of 2 in the first variable and of 2 in the second.“ (Kaufman und Rousseeuw, 2005, S. 12)

Nach der Berechnung der Distanzmatrix werden mittels eines Fusionierungsalgorithmus die Fälle zu Clustern zusammengezogen. Bei einer agglomerativen hierarchischen Clusteranalyse werden in einem ersten Schritt n Cluster (entsprechend der Fallzahl) gebildet. Schrittweise werden die einzelnen Fälle zu Clustern *fusioniert* bis final ein Cluster übrig bleibt. Geläufig ist, den Prozess der Fusionierung visuell durch ein Dendrogramm (vgl. Abbildung 5.4) darzustellen. Anhand des Dendrogramms wird erstens ersichtlich, welche Fälle beziehungsweise Sub-Cluster schrittweise zusammengeführt werden und zweitens kann nachvollzogen werden, wann Fälle und Sub-Cluster fusioniert wurden. Je höher (y-Achse) die Zusammenführung stattfindet, desto später werden im Fusionierungsprozess die Fälle beziehungsweise Sub-Cluster zusammengeführt.

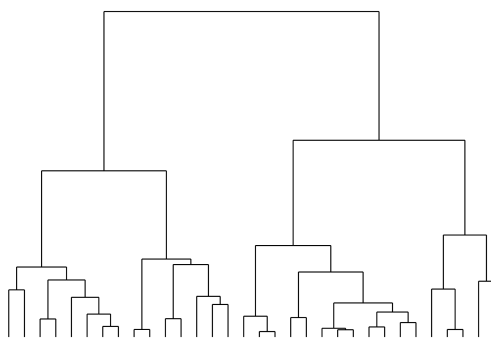


Abbildung 5.4: Dendrogramm

Die Zahl der relevanten Fusionierungsalgorithmen ist gering. In der Literatur werden primär sieben verschiedene Verfahren genannt (vgl. u.a. Backhaus et al., 2011; Everitt et al., 2011): Single Linkage (Sneath, 1957), Complete Linkage (Sørensen, 1948), Average Linkage (Sokal und Michener, 1958), Weighted Average Linkage (McQuitty, 1966), Centroid Linkage (Sokal und Michener, 1958), Median Linkage (Gower, 1967) und Ward's Linkage (Ward,

1963). Bei dem Single Linkage Verfahren werden Fälle und Cluster anhand ihres geringsten Abstandes zusammengefasst. Das Complete Linkage Verfahren zieht dagegen die größte Distanz zwischen den Fällen beziehungsweise Clustern heran. Sowohl das Average Linkage als auch das Weighted Average Linkage ziehen die Vergleiche aller Fälle der zu vergleichenden Sub-Cluster heran. Das Verfahren gewichtet die durchschnittlichen Distanzen zusätzlich mit den Clustergrößen, was insbesondere bei größeren Unterschieden Sinn macht. Daneben berechnet die Centroid Linkage zunächst die Zentren der Sub-Cluster und vergleicht dann die Distanzen. Die Median Linkage basiert auf der Centroid Linkage, gewichtet die Zentren vorher, um Größenunterschiede der Sub-Cluster herauszurechnen. Das letzte häufig angewendete Verfahren ist die Ward's Linkage. Hier wird keine Distanz zwischen Fällen oder Sub-Clustern berechnet, sondern diese werden so fusioniert, dass die Zunahme der Fehlerquadratsumme (sum-of-squares) möglichst gering bleibt. Das Verfahren neigt dazu, möglichst homogene Cluster zu bilden.

Eine besondere Herausforderung ist die Festlegung der Clusteranzahl. Im Gegensatz zu anderen Clustermethoden wird bei der hierarchischen Clusteranalyse zu Beginn der Analyse keine endgültige Anzahl der Cluster festgelegt. Die Zahl der Cluster kann auf theoretischen Überlegungen basieren. Daneben macht es Sinn, auf *Cluster Validation Techniques* oder *Cluster Validation Indices* (Charrad et al., 2014; Cichosz, 2015; Desgraupes, 2014; Halkidi et al., 2001; Hennig, 2014; Liu, Li, Xiong, Gao und Wu, 2010; Liu, Li, Xiong, Gao, Wu und Wu, 2013; Milligan und Cooper, 1985) zurückzugreifen. Diese Algorithmen versuchen anhand von Maßen und Kennzahlen, eine optimale Clusteranzahl mathematisch zu ermitteln. Alle Indices nehmen an, dass die Ergebnisse der Clusteranalyse möglichst kompakte Cluster identifizieren, die gleichzeitig eine hohe Distanz zu den anderen Clustern aufweisen. Sie lassen sich in zwei Gruppen unterteilen: *internal cluster validation indices* und *external cluster validation indices* (Rendón et al., 2011). Der Unterschied zwischen den beiden Gruppen ist, dass externe Verfahren zusätzliche Informationen über die Daten benötigen, um die Clusterlösung zu validieren. Interne Verfahren benötigen lediglich die verwendeten Rohdaten der Clusteranalyse. Wichtig ist, dass die objektive Qualität nicht zwangsläufig zu einer hohen subjektiven Qualität der Ergebnisse führen muss. Es existiert eine ganze Reihe an verschiedenen internen Clusterkriterien. Vergleichsstudien (Liu, Li, Xiong, Gao und Wu, 2010; Liu, Li, Xiong, Gao, Wu und Wu, 2013; Milligan und Cooper, 1985; Rendón et al., 2011) lassen es sinnvoll erscheinen, auf mehrere Verfahren zurückzugreifen und die Ergebnisse miteinander zu vergleichen. Häufig genannt werden die folgenden fünf Cluster Validation Indices: *Average Silhouette Width* (Kaufman und Rousseeuw, 2005; Rousseeuw, 1987), *C-Index* (Hubert und Schultz, 1976), *Calinski-Harabasz Index* (Caliński und Harabasz, 1974), *Dunn Index* (Dunn, 1974) und *Davies-Bouldin Index* (Davies und Bouldin, 1979).

Mit steigenden Datenbeständen wurde in den vergangenen Jahren die Berechnung der Distanzen zunehmend aufwendiger. Daher gibt es vermehrt hierarchische Clusterverfahren, die sich von der herkömmlichen Vorgehensweise unterscheiden. Ein prominentes Beispiel ist *BIRCH* (Balanced Iterative Reducing and Clustering using Hierarchies; Zhang, Ramakrishnan et al., 1996). Der Algorithmus benötigt weniger Speicher, da die Struktur (CF-Tree) iterativ aufgebaut wird und neue Fälle nachträglich zugeordnet werden kön-

Tabelle 5.2: Cluster Validation Indices

Index	Optimierung	Wertebereich	Optimaler Verlauf
Average Silhouette Width	max	$W = [-1, 1]$	
C-Index	min	nein	Abnehmend
Calinski-Harabasz Index	max	nein	Anstieg, Peak, Abfall
Dunn Index	max	$W = [0, unendlich]$	
Davies-Bouldin Index	min	nein	

nen. Gleichzeitig kann die Zuordnung der Cluster nachträglich verändert werden. Die klassische hierarchische Clusteranalyse hat daneben Probleme, Strukturen in beliebigen Merkmalsräumen (*Arbitrary Shapes*) zu entdecken. Die Algorithmen *CURE* (Guha et al., 1998) und *CHAMELEON* (Karypis et al., 1999) versuchen, diese Limitierung zu umgehen. *CURE* wurde zudem ebenfalls für die Analyse großer Datensätze entwickelt. Eine Einschränkung ist, dass nur mit numerischen Attributen gearbeitet werden sollte. Dieselben Entwickler haben daher *ROCK* (Guha et al., 1999) entwickelt, welcher auch kategoriale Daten verarbeiten kann.

Für die empirischen Arbeiten sind primär die Distanzmaße von Interesse. In dem ursprünglich angedachten Vorgehen sollten die Pfadmerkmale aus den Produktiv-Logfiles per Cluster-Analyse zu mehreren Gruppen zusammengefasst werden. Das Vorgehen wurde dahingehend abgewandelt, dass die Pfadmerkmale der Testerinnen und Tester mit denen der Produktiv-Logfiles kombiniert werden sollen. Eine Gruppierung der Fälle ist nicht mehr Bestandteil des empirischen Vorgehens. Die Bestimmung der Ähnlichkeiten mittels Distanzmaßen spielt bei der Zuordnung allerdings nach wie vor eine zentrale Rolle.

5.2.3 Interpretation

Der letzte Schritt des Web Usage Minings ist die *Interpretation* (engl. pattern analysis). In diesem werden die während der Mustererkennung berechneten Ergebnisse evaluiert und bewertet (Etminani, Akbarzadeh-T. et al., 2009; Etminani, Rezaeian Delui et al., 2009): „In this final phase the objective is to convert discovered rules, patterns and statistics into knowledge or insight involving the Website being analyzed.“ (Scime, 2005, S. 34) Der Schritt findet sich auch in dem KDD-Modell nach Brachman und Anand (1996, vgl. auch Kapitel 5.1). Das Vorgehen ähnelt sich daher. Es gilt zunächst, die interessanten von den uninteressanten Ergebnissen zu trennen (Srivastava et al., 2000, S. 17). Dies ist häufig der Fall, wenn mehrere Clusteranalysen durchgeführt oder Klassifikationsmethoden angewendet wurden. Dann sollte das Modell mit der größten Erklärungskraft ausgewählt werden (siehe dazu auch Cooley, 2000). Gleiches gilt für deskriptive oder andere Analysen. Anschließend sollten die Ergebnisse verständlich aufbereitet werden, entweder durch Berichtslegungen oder Virtualisierungen der Ergebnisse (Pierrakos et al., 2003, S. 321). Die Ergebnisse lassen sich darüber hinaus auch maschinell weiterverwenden, beispielsweise für Recommendation Systems (Mobasher, 2008).

6 Usability als Qualitätskriterium

Bei der Entwicklung und der Nutzung von Anwendungen spielt die *Usability* eine entscheidende Rolle. Übersetzt ins Deutsche lässt sich der Begriff wohl am ehesten mit *Gebrauchstauglichkeit* beschreiben: „Dieser Begriff fokussiert dabei nicht nur eine komfortable Benutzung, sondern auch die Forderung nach einer geeigneten Unterstützung des Nutzers bei der Erreichung seiner Ziele in dem jeweiligen Einsatzfeld. Usability ist keine eigenständige Disziplin wie die Ergonomie, sondern eine Qualität eines technischen Systems.“ (Sarodnick und Brau, 2011, S. 19) Sowohl in der deutschen als auch in der englischsprachigen Literatur/Forschung wird häufig nur der Begriff Usability verwendet.

Die Usability bezieht sich stets auf eine Benutzerschnittstelle. Aufbau und Eigenschaften sind in der ISO 9241, Teil 110¹ definiert (ISO, 2006). Die *Grundsätze zur Dialoggestaltung* beschreiben sieben Anforderungen: *Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Erwartungskonformität, Lernförderlichkeit, Steuerbarkeit, Fehlertoleranz* und *Individualisierbarkeit*. Die ersten drei Grundsätze spiegeln einen logischen, angemessenen und nachvollziehbaren Aufbau des Dialogs wider. Gleichzeitig sollten neue Nutzerinnen und Nutzer den Dialog einfach erlernen können. Die Steuerbarkeit besagt, dass Richtung und Geschwindigkeit einstellbar sein sollen. Werden von der Nutzerin beziehungsweise dem Nutzer Fehler bei der Eingabe gemacht, soll die Aufgabe mit maximal kleinen Änderungen erfüllbar bleiben. Der Dialog sollte zudem an die individuellen Bedürfnisse anpassbar sein. Im Jahr 2002 wurde zudem die ISO 14915 *Software-Ergonomie für Multimedia-Benutzungsschnittstellen* veröffentlicht (ISO, 2003), um der zunehmenden Interaktivität und Verbreitung von Browser-Anwendungen Rechnung zu tragen. In dem ersten Teil, *Gestaltungsgrundsätze und Rahmenbedingungen*, sind vier weitere Anforderungen zu Benutzerschnittstellen aufgeführt, die die ISO 9241, Teil 110 ergänzen: *Eignung für das Kommunikationsziel, Eignung für Wahrnehmung und Verständnis, Eignung für die Exploration* und *Eignung für die Benutzungsmotivation*. Die ersten beiden Grundsätze beziehen sich auf die angemessene Präsentation und das einfache Verständnis der Informationen. Der dritte Grundsatz definiert, dass Nutzerinnen und Nutzer die Informationen auch ohne Vorkenntnisse und Erfahrungen finden können müssen. Die Benutzer motivation zielt auf die Anregung der Nutzerinnen und Nutzer ab. Die ersten drei Grundsätze haben durch Präsentation der Benutzerschnittstelle und Führung durch diese die Möglichkeit, eine möglichst hohe Motivation zu erzeugen. Das grundlegende Ziel bei beiden Standards ist stets die möglichst einfache Bedienung.

Die einfache Bedienbarkeit des Produkts spielt neben den inhaltlichen und formalen Kriterien, insbesondere die Erfüllung der Anforderungen an die Anwendung, eine wichtige Rolle. Bereits 1984 wies Eason darauf hin, dass eine schlechte Bedienbarkeit das Potenzial

¹Der Teil ersetzt den Teil 10 der ISO 9241.

eines technischen Produkts deutlich mindern kann.

Die Grundlagen der Usability beziehen sich in ihrer Entwicklung vor allem auf die Kognitionspsychologie (Wentura und Frings, 2013). Durch verschiedene Modelle wird versucht, den Ablauf der Informationsverarbeitung im Gehirn nachzustellen (vgl. u. a. Harvey et al., 2011; John, 2003; Perry, 2003; van Rijn et al., 2011). Durch Programme dargestellte Informationen müssen durch die Nutzerinnen und Nutzer aufgenommen und verarbeitet werden. Die *Wahrnehmung* der Informationen kann grundsätzlich über verschiedene Sinnesorgane erfolgen (visuell, auditiv, haptisch, olfaktorisch, gustatorisch). Im Bereich der Computer sind primär visuelle und teilweise auditive Reize von Bedeutung (Dix et al., 2004; Sarodnick und Brau, 2011), zunehmend auch haptische Reize durch beispielsweise „Wearable Computing“ (Mann, 1997; Starner, 1996). Die Wahrnehmung ist begrenzt, sodass nur ein Teil der dargestellten Informationen gleichzeitig fokussiert werden kann. Visuelle Reize können nur von einem kleinen Bereich des menschlichen Auges wahrgenommen werden (*foveales Sehen*; Sarodnick und Brau, 2011, S. 53). Dieser Bereich ist fokussiert und im Zentrum der *Aufmerksamkeit*. Weitere visuelle Reize können über das *periphere Sehen* aufgenommen werden. Die visuelle Aufmerksamkeit kann mit *raumbasierten* oder *objektbasierten* Modellen erklärt werden (Malaka et al., 2009). Ein prominenter Vertreter der raumbasierten Modelle ist Michael Posner (Posner, Petersen et al., 1988; Posner, Snyder et al., 1980). Die Aufmerksamkeit stellen sich die Autoren als Scheinwerfer vor, welcher auf Teilbereiche gerichtet werden kann. Je kleiner der Ausschnitt, desto genauer ist die Fokussierung und Aufmerksamkeit. Gleichzeitig kann nur noch ein kleiner Teil der Informationen wahrgenommen werden. Behrmann et al. (1998) konnten dagegen zeigen, dass eher Objekte als Räume fokussiert und wahrgenommen werden. Die *Wahrnehmung von Objekten* kann mit der Gestalttheorie und den daraus abgeleiteten *Gestaltgesetzen* erklärt werden (Metzger et al., 1986). Nach der Wahrnehmung der Objekte erfolgt mental die *Identifikation der Objekte*. Dies kann über einen *Schablonenvergleich* erfolgen. Hauptkritikpunkt an dem Verfahren ist, dass bei steigender Anzahl der Objekte die Zahl der Schablonen steigt und zu einer höheren kognitiven Belastung führt (Funke und Anderson, 2013). Präferiert wird deshalb der *Merkmalsvergleich*. Dieser zeichnet sich durch seine höhere Flexibilität und einfachere Struktur aus. Die Erkennung von Merkmalen bildet zudem die Verknüpfung zum Gedächtnis. In der Theorie wird zwischen drei verschiedenen Speichern unterschieden (*Multi-Speicher-Modell*; Dix et al., 2004; Sarodnick und Brau, 2011): das *sensorische Gedächtnis*, das *Kurzzeitgedächtnis* und das *Langzeitgedächtnis*. Der sensorische Speicher nimmt die Informationen der Sinnesorgane auf. Er verfügt über eine hohe Kapazität bei kurzer Speicherdauer. Anschließend gelangen die Informationen in das Kurzzeitgedächtnis. Dort erfolgt die bewusste Verarbeitung der über die Sinnesorgane gewonnenen Informationen. Wie der sensorische Speicher verfügt das Kurzzeitgedächtnis über eine kurze Speicherdauer. Dix et al. (2004, S. 29) sprechen von 200 Millisekunden. Die Größe ist ebenfalls auf etwa sieben Einheiten beschränkt. Zuletzt gelangen die Informationen in das Langzeitgedächtnis. Dieses verfügt über eine hohe Kapazität und eine lange Speicherdauer. Gleichzeitig ist der Zugriff auf die Informationen langwieriger. Der interne Ablauf des Gedächtnisses geschieht wie folgt (Sarodnick und Brau, 2011, S. 61): Die Informationen werden in dem sensorischen Gedächtnis kategorisiert

und gefiltert. Sie werden mit den vorhandenen Informationen des Langzeitgedächtnisses zur Mustererkennung (Merkmalsvergleich) verglichen. Abschließend werden die Informationen in das Kurzzeitgedächtnis überführt, wo sie bewusst verarbeitet werden können. Diese Überlegungen zur Informationsverarbeitung durch die Nutzerin beziehungsweise den Nutzer bilden die Grundlage für die bisherige Entwicklung der Usability.

In diesem Kapitel wird zunächst der Begriff „Usability“ bestimmt und mit verschiedenen Definitionen verglichen. Anschließend wird ein Überblick über vorhandene Methoden zur Messung von Usability gegeben. Ein besonderes Augenmerk liegt auf den Nutzertests, da diese Methode im empirischen Teil dieser Dissertation angewandt wird.

6.1 Anforderungen an eine „gute“ Usability

Die genaue Definition des Begriffs *Usability* gestaltet sich im Hinblick auf die Literatur als schwierig. Nielsen (1994e, S. 3) definiert Usability als „(...) a fairly broad concept that basically refers to how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use.“

Geläufig ist die ISO 9241, Teil 11 der Normenreihe zur *Ergonomie der Mensch-Maschine-Interaktion*² (ISO, 1998) mit dem Titel *Anforderungen an die Gebrauchstauglichkeit – Leitsätze*. Auf diese verweisen bis heute diverse Autoren (u. a. Barnum, 2011; Dix et al., 2004; Hegner, 2003; Stone et al., 2005; Stowasser, 2006). Die ISO-Norm definiert *Usability* wie folgt: „Das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“ (ISO, 1998, S. 4) Wie Barnum (2011, S. 11) und Sarodnick und Brau (2011, S. 38) herausstellen, bezieht sich die Definition auf einen *bestimmte* Benutzerin beziehungsweise einen *bestimmten* Benutzer in einem *bestimmten* Umfeld, der *bestimmte* Ziele erreichen will. Sowohl Nutzerkreis, Nutzungskontext als auch Nutzungsziele sind nicht beliebig, sondern festgelegt. Weiterhin werden drei Anforderungen an die Anwendungen definiert.³ Sie sollen über eine hohe *Effektivität* und *Effizienz* beim Erreichen der Ziele verfügen sowie eine hohe *Zufriedenstellung* bei den Nutzerinnen und Nutzern hervorrufen. Die Effektivität ist „die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen.“ (ISO, 1998, S. 4) Unter der Effizienz wird das Verhältnis zwischen eingesetztem Aufwand zur Genauigkeit und Vollständigkeit oder auch „Kosten-Nutzen-Verhältnis“ (Sarodnick und Brau, 2011, S. 37) verstanden. Diese beiden ersten Dimensionen sind einfach zu erfassen und zu messen. Barnum (2011, S. 11) spricht von dem „value-added part of usability“. Die Zufriedenstellung ist ein subjektives Maß, das die „Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produkts“ (ISO, 1998, S. 4) beinhaltet. Daher ist sie schwieriger zu operationalisieren und durch die subjektive Wahrnehmung der Nutzerinnen und Nutzer möglicherweise verfälscht. Gleichzeitig besteht eine Korrelation zwischen der Effektivität und der Effizienz auf der einen Seite und der Zufriedenstellung auf der anderen Seite. Anwendungen, die bei den

²Bis 2006 hieß die Norm *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten*. Der aktuelle englische Titel lautet *Ergonomics of Human System Interaction*.

³Analog zu den Ausführungen von Nielsen (1993) werden diese ebenfalls Dimensionen genannt.

Nutzerinnen und Nutzern eine hohe Zufriedenstellung erfahren, werden von diesen häufig auch effektiver und effizienter bedient. Wie bereits für Nielsen (1993), stellt die Zufriedenstellung für Barnum (2011, S. 12) die wichtigste Dimension dar.

In den vergangenen 25 Jahren hat sich innerhalb der Usability-Forschung ein grundlegender Konsens gebildet, was unter dem Begriff der Usability verstanden wird. Jakob Nielsen (1993) war einer der ersten, der die vorherigen ersten Definitionen von „user friendliness“ (Nielsen, 1993, 24f) systematisch aufgearbeitet und definiert hat (vgl. Abbildung 6.1). Nach seiner Definition ist zunächst entscheidend, ob die Anwendung von den Nutzerinnen und Nutzern akzeptiert wird (*System acceptability*).⁴ Diese Akzeptanz untergliedert sich in eine soziale und eine praktische Komponente. Die soziale Akzeptanz (*Social acceptability*) hinterfragt, ob die Nutzerinnen und Nutzer moralische beziehungsweise ethische Gründe gegen die Nutzung der Anwendung haben. Die praktische Akzeptanz (*Practical acceptability*) trifft eine Aussage darüber, ob die Nutzerinnen und Nutzer die Anwendung hinsichtlich technischer Aspekte für bedienbar halten.⁵ Die praktische Akzeptanz untergliedert sich in die verschiedenen Dimensionen Kosten (*Cost*), Kompatibilität zu bestehenden Anwendungen (*Compatibility*), Reliabilität (*Reliability*), vorhandene Unterstützung (*Support*) und Brauchbarkeit (*Usefulness*): „*Usefulness* is the issue of whether the system can be used to achieve some desired goal.“ (Nielsen, 1993, 24, H.i.O.) *Usability* ist neben der Funktionalität (*Utility*) eine Unterkategorie von Brauchbarkeit (*Usefulness*). Während die Funktionalität daran gemessen wird, ob die Nutzerin oder der Nutzer mit der Anwendung das anvisierte Ziel erreichen kann, ist „usability (.) the question of how well users can use that functionality.“ (Nielsen, 1993, S. 25) Die beiden Kategorien finden sich auch in dem TAM-Modell (vgl. Abbildung 2.1; Davis, 1989). Die Funktionalität („*Utility*“) lässt sich in dem Modell durch die wahrgenommene Nützlichkeit („*Perceived Usefulness*“) und die Usability mit der wahrgenommenen Gebrauchstauglichkeit („*Perceived Ease of Use*“) verknüpfen. Das theoretische Modell lässt sich daher gut in die empirische Analyse übertragen.

Usability hat nach Nielsen (1993, 26ff.) fünf Dimensionen: *Learnability*, *Efficiency*, *Memorability*, *Few Errors* und *Satisfaction*. Unter *Learnability* versteht Nielsen einen einfachen Zugang zu der Anwendung, der Nutzerinnen und Nutzern einen schnellen Einstieg ermöglicht. Die Lernkurve sollte insbesondere bei unerfahrenen Nutzerinnen und Nutzern nicht zu steil sein. Die *Learnability* kann als die elementarste Dimension angesehen werden. Sobald eine Nutzerin beziehungsweise ein Nutzer ein grundlegendes Wissen über die Benutzung der Anwendung erlernt hat, spielt die Effizienz (engl. *Efficiency*) eine wichtige Rolle. Sie stellt sicher, dass die Nutzerinnen und Nutzer mit einer hohen Produktivität arbeiten können. Die Anwendung sollte so aufgebaut sein, dass Nutzerinnen und Nutzer sich auch nach längerer Nichtnutzung schnell wieder in der Anwendung zurecht finden und das vorherige Produktivitätslevel erreichen können (*Memorability*).

⁴Obwohl Nielsen von „Systemen“ spricht, wird in dieser Arbeit durch die vorherige Einordnung in die Terminologie der Wirtschaftsinformatik (vgl. Kapitel 2) weiterhin der Begriff „Anwendung“ verwendet.

⁵Eine hohe praktische Akzeptanz geht nicht zwangsläufig mit einer sozialen Akzeptanz der Anwendung einher. Die Einführung einer digitalen Zeiterfassung kann beispielsweise mit einer hohen technischen Akzeptanz erfolgen. Die Mitarbeiterinnen und Mitarbeiter des Unternehmens müssen die Anwendung dagegen nicht für sozial akzeptabel halten.

Die drei angesprochenen Punkte betreffen primär den Aufbau und das Design der Anwendung. Bei der Programmierung sollte auf eine geringe Fehlerquote geachtet werden (Few Errors). Wenn möglich, sollten das System fehlerfrei programmiert sein. Gleichzeitig sind die Nutzerinnen und Nutzer durch die Anwendungsgestaltung davon abzuhalten werden, Fehler zu machen. Wenn doch, sollte es möglich sein, den vorherigen Zustand wiederherzustellen (beispielsweise durch *Undo*-Funktionen). Die fünfte Dimension betrifft die Zufriedenstellung (engl. Satisfaction) der Nutzerinnen und Nutzer. Um eine möglichst hohe Zufriedenstellung zu erreichen, sollte das System angenehm und flüssig zu bedienen sein.

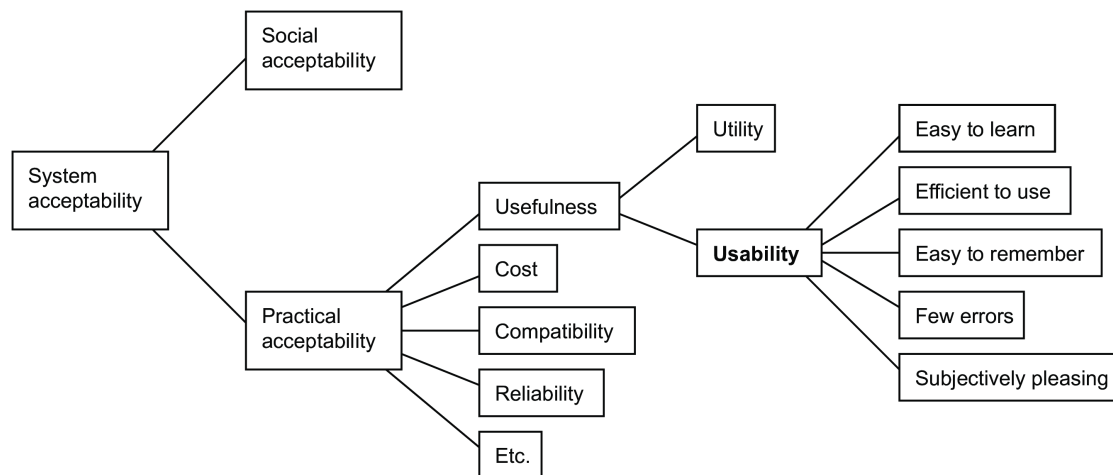


Abbildung 6.1: System Acceptability nach Nielsen (1993, S. 25)

Tabelle 6.1: Vergleich Usability Dimensionen

Nielsen (1993)	ISO (1998)	Lecerof und Paternò (1998)
Learnability	Effektivität ¹	Relevance
Efficiency	Effizienz ²	Efficiency
Memorability	Zufriedenstellung ³	Attitude
Few errors		Learnability
Satisfaction		Safety
Preece et al. (2002)	Quesenberg (2003)	Rubin und Chisnell (2008)
Effectiveness	Effective	Usefulness
Efficiency	Efficient	Efficiency
Safety	Engaging ⁴	Effectiveness
Utility	Error tolerant	Learnability
Learnability	Easy to learn	Satisfaction
Memorability		

¹ Effectiveness in der internationalen Version.

² Efficiency in der internationalen Version.

³ Satisfaction in der internationalen Version.

⁴ Abgeleitet von Satisfaction.

In den letzten Jahren wurde neben den Definitionen von Nielsen und der ISO (1998) vor allem auf Whitney Quesenberg (2003) verwiesen. Sie geht zunächst von der Definition der ISO 9241-11 aus, kritisiert aber die starke Fokussierung auf definierte Aufgaben und Arbeitsumfelder. Dies trifft vor allem auf Anwendungen zu, die der Nutzerin be-

ziehungsweise dem Nutzer Freiheiten gewährleisten und keine eindeutigen Aufgaben vorgeben. Zudem würde der „Spaß“-Aspekt nicht abgedeckt, trotz seiner Wichtigkeit im Freizeitbereich: „This has been especially true for designers of products without a strong task orientation, such as large information systems. In addition, designers of consumer products criticize usability as not accommodating the concept of ‚fun‘ as a goal, even though it can be important in the marketplace.“ (Quesenbery, 2003, S. 82) Sie erweitert die Dimensionen der ISO 9241-11 um die Fehlertoleranz (engl. *Error tolerant*) und die Lernbarkeit (engl. *Easy to learn*). Gleichzeitig benennt sie die Dimension der Zufriedenstellung (engl. *Satisfaction*) in *Engaging* um. Damit soll der Aspekt der Emotionalität und die zunehmend dynamischere Interaktion gestärkt werden.⁶ Die Dimensionen der Fehlertoleranz und der Lernbarkeit finden sich bereits in der Definition von Nielsen (1993). Die fünf Dimensionen sind wie folgt definiert (Quesenbery, 2003, 83ff.):

- *Effective*: The completeness and accuracy with which users achieve their goals.
- *Efficient*: The speed (with accuracy) with which users can complete their tasks.
- *Engaging*: The degree to which the tone and style of the interface makes the product pleasant or satisfying to use.
- *Error tolerant*: How well the design prevents errors or helps with recovery from those that do occur.
- *Easy to learn*: How well the product supports both initial orientation and deepening understanding of its capabilities.

Daneben existieren eine ganze Reihe weiterer Definitionen (vgl. für einen Überblick Madan und Dubey, 2012). Tabelle 6.1 stellt die Dimensionen von sechs ausgewählten Autoren gegenüber. Es wird sichtbar, dass es einen allgemeinen Konsens über einige Dimensionen gibt. Die Effizienz (engl. *Efficiency*) wird in allen sechs Quellen aufgeführt, die Lernbarkeit (engl. *Learnability*) in jeder Definition bis auf die ISO 9241-11 genannt. Die Zufriedenstellung (engl. *Satisfaction*) wird fünfmal aufgeführt. Quesenbery (2003) erweitert die Definition der Dimension etwas, Lecerof und Paternò (1998) verstehen unter *Attitude* die subjektiven Einstellungen und Gefühle der Nutzerinnen und Nutzer gegenüber dem System. Diese Definition ist zu einem hohen Teil deckungsgleich mit der Zufriedenstellung. Die Effektivität kann, je nach Auslegung bei vier bis sechs Autoren gefunden werden. Bei Nielsen (1993), ISO (1998), Preece et al. (2002) und Quesenbery (2003) wird sie explizit genannt. Lecerof und Paternò (1998) verstehen unter *Relevance of the system*, „(...) how well it serves the users' needs.“ Darunter fallen implizit die verfolgten Ziele. Rubin und Chisnell (2008) definieren *Usefulness* als die Möglichkeit der Nutzerinnen und Nutzer, mit der Anwendung ein festgelegtes Ziel zu erreichen und die allgemeine Einstellung der Nutzerinnen und Nutzer gegenüber dem Produkt. Zumindest der erste Teil der Dimension deckt sich mit den anderen Definitionen der Zufriedenstellung. Die Fehlertoleranz wird von vier Autorinnen und Autoren genannt (Lecerof und Paternò (1998) sowie Preece et al. (2002) beschreiben dies mit dem Begriff *Safety*). Kaum relevant sind die Einprägsamkeit (engl. *Memorability*) der Anwendung sowie die *Utility*. Letztere wird von Nielsen (1993) als separates Konstrukt gesehen (vgl. Abbildung 6.1). Die Einprägsamkeit

⁶ „Satisfying‘ becomes ‚engaging‘ not merely to preserve the alliteration, but to raise the emotional level and create a sense of a dynamic interaction.“ (Quesenbery, 2003, S. 83)

überlappt sich teilweise mit der Lernbarkeit der Anwendung.

Im weiteren Verlauf dieser Arbeit wird auf die Dimensionen von Quesenbery (2003) zurückgegriffen. Zum einen stützen sie sich explizit auf die Definition nach der ISO 9241-11, zum anderen fügt Quesenbery ein dynamisches Element in die Dimensionen mit ein. Einige Autorinnen und Autoren betonen die Wichtigkeit von bestimmten Dimensionen, legen allerdings keine explizite Rangfolge vor. Quesenbery (2003) dagegen äußert, dass die Dimensionen von Usability durchaus unterschiedlich gewertet werden. Je nach Anwendung erscheinen bestimmte Dimensionen wichtiger als andere. Gleichzeitig dürfen die anderen Dimensionen auf keinen Fall als vernachlässigt werden: „When one dimension is dominant, it is easy to ignore others, especially when they introduce competing requirements.“ (Quesenbery, 2003, S. 89) Die dynamische Anpassbarkeit ermöglicht die Anpassung an verschiedene Anwendungen und Lebenssituationen wie Arbeit, Freizeit etc.

6.2 Vorstellung der verschiedenen Usability-Methoden

Zur Messung und Bewertung der Usability wurden in den vergangenen Jahrzehnten verschiedene Methoden entwickelt. Grob lassen sich diese in zwei Gruppen einteilen: die *analytischen Methoden* und die *empirischen Methoden* (Sarodnick und Brau, 2011; Te'eni et al., 2006). Analytische Methoden werden von Expertinnen und Experten durchgeführt, die sich während der Evaluation in die Rolle der Nutzerinnen und Nutzer hineindenken. In der Literatur wird auch von *Expert Analysis* (Dix et al., 2004), *Expert Evaluation* (Freiberg und Baumeister, 2008) oder *Expert Reviews* (Shneiderman und Plaisant, 2010) gesprochen. Zu den bekanntesten Methoden in diesem Bereich zählen die *heuristische Evaluation* und das *Cognitive Walkthrough* (vgl. Kapitel 6.2.1). Im Gegensatz dazu werden die empirischen Methoden mit tatsächlichen Nutzerinnen und Nutzern der Anwendung durchgeführt. Teilweise wird von *User Participation* (Dix et al., 2004), *User-based Evaluation* (Freiberg und Baumeister, 2008) oder *User-observation* (Stone et al., 2005) gesprochen. In diesem Bereich sind die Methoden des *Nutzertests* und der *Fragebögen* hervorzuheben (vgl. Kapitel 6.2.2).

Die Auswahl der geeigneten Methode wird maßgeblich durch den Evaluationszeitpunkt definiert. Ursprünglich wurden Usability-Evaluationen gegen Ende der Produktentwicklung durchgeführt, um ein abschließendes Feedback über das (fast) fertige Produkt zu erhalten. Dies wird als *summative Evaluation*⁷ (Barnum, 2011; Hegner, 2003; Sarodnick und Brau, 2011; Stowasser, 2006) bezeichnet. Dumas und Fox (2008, S. 1135) sprechen von *High-fidelity Prototypes*. Vorteilhaft an dem späten Zeitpunkt der Evaluation ist die Möglichkeit, alle Eigenschaften der Anwendung testen und bewerten zu können. Darunter fallen komplexere Bedienungsabläufe, die sich über mehrere Teile der Anwendung erstrecken, welche in früheren Entwicklungsphasen nur teilweise zur Verfügung stehen. Grundlegende Entscheidungen zu Funktionalität und Design sind zu diesem späten Zeitpunkt der Entwicklung bereits abgeschlossen. Nachteilig ist, dass tiefgreifende Änderungen der Anwendung, die durch tiefgehende Probleme wie grundsätzliche Design-Entscheidungen

⁷Die Begriffe gehen auf die FS-Klassifikation nach Scriven (1967) zurück.

verursacht werden, nur mit einem hohen Aufwand lösbar sind.

Um gravierende Usability-Probleme in früheren Stadien der Entwicklung aufzudecken, wurden Methoden zur *formativen Evaluation* (Barnum, 2011; Hegner, 2003; Sarodnick und Brau, 2011; Stowasser, 2006) entwickelt, die an sogenannten *Low-fidelity Prototypes* (Dumas und Fox, 2008, S. 1135) getestet werden. Zu diesem Zeitpunkt innerhalb der Produktentwicklung werden grundlegende Entscheidungen hinsichtlich der Funktionalität und des Designs der Anwendung getroffen. Wichtig ist, dass umfangreiche und kritische Probleme erkannt werden (Lazar et al., 2010, S. 260) und konkrete Verbesserungsmöglichkeiten oder Ansatzpunkte aufgezeigt werden (Sarodnick und Brau, 2011, S. 120). Das Simulieren komplexer Arbeitsabläufe ist von nachrangiger Bedeutung. Daher muss zu dem Zeitpunkt kein fertiges und benutzbares System vorhanden sein. Je nach Fragestellung kann mit Prototypen, simulierte Reaktionen (engl. *Wizard of Oz*; Kelley, 1984) oder Zeichnungen gearbeitet werden (engl. *Paper Prototyping*; Snyder, 2003). Die Methoden können auch in späteren Schritten des Entwicklungsprozesses eingesetzt werden, da die Vorbereitung und Durchführung durch Expertinnen und Experten erfolgt, die eine kürzere Einführungszeit als echte Nutzerinnen und Nutzer der Anwendung benötigen. Die formativen Methoden werden auch eingesetzt, um kleinere Änderungen oder Teilentwicklungen zeitnah durch Expertinnen und Experten testen zu lassen.

Der Einsatz einer formativen oder summativen Evaluation hängt mit den Evaluationszielen zusammen. Nach Gediga und Hamborg (2002) lassen sich die Evaluationsziele in drei Fragestellungen aufteilen: „Which is better?“, „How good?“ und „Why bad?“. Bei „Which is better?“ geht es um den direkten Vergleich von zwei oder mehr Anwendungen. Mit der Frage nach „How good?“ wird versucht, bestimmte Eigenschaften der Anwendung genauer zu bewerten. Dies findet üblicherweise gegen Ende des Entwicklungsprozesses statt. Beide Evaluationsziele sind eher der summativen Evaluation zuzuordnen. Das Ziel bei „Why bad?“ ist dagegen mit der formativen Evaluation zu verbinden. Die Fragestellung zielt auf Schwachstellen und Probleme ab, die im weiteren Entwicklungsprozess behoben werden sollen. Gould und Lewis (1985) rieten Mitte der 1980er-Jahre dazu, in frühen Phasen der Produktentwicklung auf Usability Evaluationen zurückzugreifen und diese kontinuierlich weiterzuführen. Verschiedene Autoren (Barnum, 2011; Dix et al., 2004; Krug, 2006; Rubin und Chisnell, 2008; Sarodnick und Brau, 2011; Stone et al., 2005) folgen dieser Empfehlung. In frühen Phasen der Entwicklung bietet sich laut den Autoren an, auf formative Evaluationen zurückzugreifen. Dies habe einen positiveren Effekt auf die Usability der späteren Anwendung als gegen Ende auf eine summative Evaluation zurückzugreifen.

Der dritte wichtige Punkt bei der Planung einer Evaluation ist die Art der gewonnenen Daten. Es wird zwischen *qualitativen* und *quantitativen* Daten unterschieden (Cooper et al., 2007; Dix et al., 2004; Hegner, 2003; Sarodnick und Brau, 2011). *Quantitative Daten* lassen sich durch den Einsatz von Methoden wie Fragebögen oder Logfile-Analysen gewinnen. Summative Nutzertests fallen je nach Testaufbau ebenfalls in diese Kategorie, insbesondere dann, wenn Maßzahlen wie die benötigte Zeit für bestimmte Operationen (engl. *Task Duration*) oder die Anzahl der gemachten Fehler (engl. *Error Rate*) gemessen werden. Die gewonnenen Daten lassen sich mittels statistischer Methoden auswerten. Dix et al. (2004)

und Hegner (2003) sprechen in diesem Zusammenhang von *objektiven Methoden*, Sarodnick und Brau (2011) von *objektiven Daten*. Werden Nutzerinnen und Nutzer beziehungsweise Expertinnen und Experten nach Einschätzungen und Meinungen gefragt, werden diese als *qualitative Daten* bezeichnet. Sie liefern nach Hegner (2003, S. 15) „weiche Daten“. Darunter fallen Beurteilungen, ob die Benutzung angenehm, bequem, klar und einsichtig ist oder konkrete Probleme bei der Verwendung der Anwendung bestehen. Je freier und weniger durch konkrete Anweisungen eingeschränkt die Nutzerinnen und Nutzer beziehungsweise die Expertinnen und Experten die Anwendung verwenden können, desto subjektiver sind ihre Einschätzungen.

Für die Durchführung von analytischen und empirischen Usability-Evaluationen gelten verschiedene Gütekriterien (Diekmann, 2010; Flick, 2007; Lamnek und Krell, 2010; Mayring, 2002; Schnell et al., 2008). Die Einhaltung der Gütekriterien ist aufwendig, garantiert aber eine zuverlässige Messung, Interpretation und daraus folgend valide und reliable Ergebnisse. Für qualitative und quantitative Daten existieren verschiedene Gütekriterien, die jeweils auf die anfallenden Daten zugeschnitten sind. Bei *quantitativen Daten* liegen Kriterien für Objektivität, Reliabilität und Validität vor. Unter einer hohen *Objektivität* wird verstanden, dass ein Erhebungsinstrument unabhängig von der Expertin beziehungsweise dem Experten stets dieselben Ergebnisse erzielt. Die *Reliabilität* gibt die Zuverlässigkeit der Messmethode an. Eine hohe Reliabilität ist dann gegeben, wenn eine wiederholte Messung des Untersuchungsgegenstands bei gleichbleibenden Rahmenbedingungen dasselbe Ergebnis liefert. Dazu können die Untersuchungen gleichzeitig durchgeführt werden („Paralleltest-Methode“), zu einem späteren Zeitpunkt wiederholt werden („Test-Retest-Methode“) oder doppelte Indikatoren für die Messinstrumente geteilt werden („Split-Half-Methode“). Das letzte Gütekriterium ist die *Validität*. Darunter wird die Zuverlässigkeit der Messung verstanden, dass „das Messinstrument tatsächlich das misst, was es messen sollte.“ (Schnell et al., 2008, S. 154) Die Validität gilt für den Inhalt („Inhaltsvalidität“), die Erfassung und Überprüfbarkeit der gewünschten und tatsächlichen Inhalte („Konstruktvalidität“) sowie die Korrelation mit anderen empirischen Ergebnissen, die statistisch nachweisbar sind („Kriteriumsvalidität“). Eine reliable Untersuchung muss nicht zwangsläufig auch valide sein. Die Gütekriterien für *qualitative Erhebungen* sind dagegen weniger strikt, da der Gegenstand weniger allgemein zu definieren ist. Mayring (2002, 144ff.) nennt sechs verschiedene Gütekriterien, die auch von Flick (2007) sowie Lamnek und Krell (2010) übernommen wurden. Das erste Kriterium ist die *Verfahrensdokumentation*. Durch die Einzigartigkeit der jeweiligen Forschung (jeder Untersuchungsgegenstand ist anders) muss das Vorgehen detailliert dokumentiert werden. Eine umfangreiche Dokumentation ist ebenfalls bei dem zweiten Gütekriterium, der *argumentativen Interpretationsabsicherung*, wichtig. Da die Interpretation häufig anhand eines Einzelfalls erfolgt, müssen der Verlauf und die aus der Interpretation gezogenen Schlüsse intersubjektiv nachvollziehbar sein. Die *Regelgeleitetheit* bezieht sich auf den Ablauf des Forschungsprozesses. Qualitative Forschung sollte so systematisch wie möglich erfolgen. Ansonsten drohen wie bei der quantitativen Forschung Verzerrungen der Ergebnisse. Ohne Regeln und Vorgaben wird „qualitative Forschung wertlos bleiben.“ (Mayring, 2002, S. 146) Forscherinnen und Forscher müssen eine *Nähe zum Gegenstand* einhalten.

Darunter fällt die Einbeziehung der natürlichen Umwelt des Untersuchungsgegenstands. Ein Verstoß kann eine eingeschränkte Wahrnehmung des untersuchten Phänomens zur Folge haben. Daher kann eine Laborsituation unter Umständen irritierend wirken und sollte möglichst realitätsnah aufgebaut sein. Um die Interpretationen und Ergebnisse zu überprüfen, können diese einer *kommunikativen Validierung* unterzogen werden. Dazu werden diese den Teilnehmerinnen und Teilnehmern erneut vorgelegt und mit diesen diskutiert. Dieses ist in der qualitativen Forschung möglich, weil den „Beforschten mehr Kompetenz zugebilligt [wird, Anm.] als üblich.“ (Mayring, 2002, S. 147) Das letzte Gütekriterium ist die *Triangulation*. Dazu werden mehrere Interpretationsansätze oder Methoden miteinander kombiniert, um „Phänomene umfassender, abgesicherter und gründlicher zu erfassen.“ (Lamnek und Krell, 2010, S. 132) Die Triangulation wird auch in der quantitativen Forschung eingesetzt. Geläufig ist ebenfalls, quantitative und qualitative Methoden miteinander zu kombinieren, da beide Traditionen Vor- und Nachteile haben. Mittels der Triangulation versucht man, die Vorteile beider Methoden zu maximieren und gleichzeitig die Nachteile zu minimieren.

Für eine weitere Unterscheidung der verschiedenen Evaluationsmethoden wird sich vor allem auf die bisherigen Unterscheidungen von Dix et al. (2004), Freiberg und Baumeister (2008), Te'eni et al. (2006) und Stone et al. (2005) gestützt. Zentral ist die Unterscheidung zwischen empirischen und analytischen Methoden. Wie Stowasser (2006, S. 121) kritisiert, ist der Aufbau einer eindeutigen Klassifizierung der Methoden nur bedingt möglich. Dies liegt daran, dass es bisher keine eindeutige Terminologie zur Beschreibung der Methoden gibt. Dies wird dadurch deutlich, dass die zentrale Unterscheidung zwischen empirischen und analytischen Methoden innerhalb der Literatur nicht eindeutig ist, sowohl was die Bezeichnung als auch die Zuordnung der Methoden betrifft. Gleichzeitig sind die Methoden zu einem großen Teil nicht standardisiert, beziehungsweise es ist keine allgemein gültige Operationalisierung vorhanden. Die Folgen davon sind, dass die Ergebnisse von zwei verschiedenen Methoden nur bedingt vergleichbar sind. Auf der anderen Seite hat dies unter Umständen zur Folge, dass zwei Untersuchungen mittels derselben Methode schwer vergleichbar sind. Letztlich ist keine allgemeine Vorgehensweise zur Ermittlung der anzuwendenden Methode vorhanden. Diese beruhen in der Regel auf pragmatischen Ansätzen. Die folgende Einordnung und kurze Beschreibung der vorhandenen Methoden ist weder als vollständig noch als allgemein gültig zu betrachten. Ausführlicher wird die Methode der *Nutzertests* (vgl. Kapitel 6.3) beschrieben, da diese auch in dieser Dissertation angewendet wird.

6.2.1 Analytische Usability-Evaluationen

Die analytischen Usability-Evaluationen sind durch die Durchführung von Expertinnen und Experten gekennzeichnet. Zudem finden sie eher in frühen Phasen des Entwicklungsprozesses statt, was sie den formativen Evaluationen zuordnet.⁸ Innerhalb der Literatur zu den analytischen Usability-Evaluationen lassen sich vor allem zwei größere Gruppen

⁸Wie im vorherigen Kapitel beschrieben, ist weder die allgemeine Zuordnung von Methoden noch die thematische Zusammengehörigkeit der Methoden untereinander klar definiert.

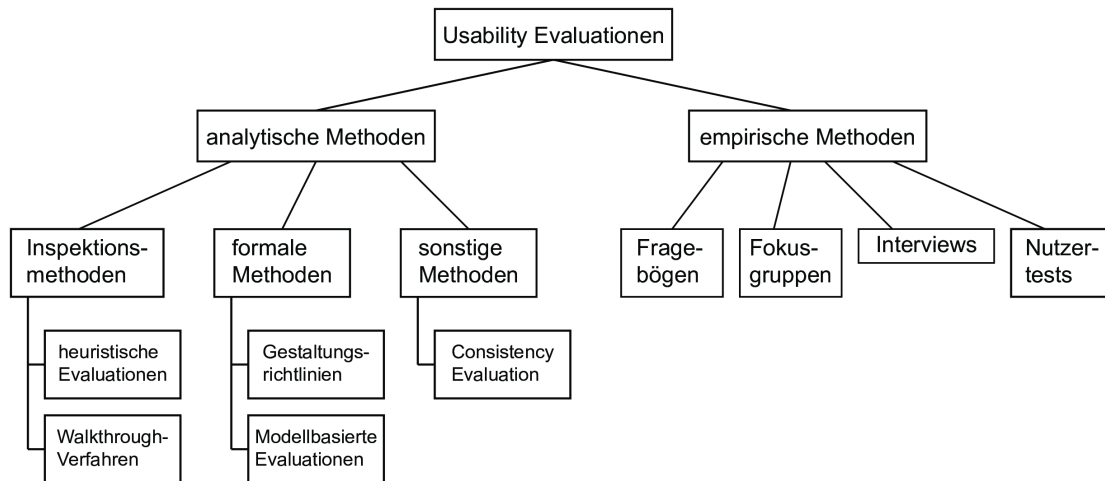


Abbildung 6.2: Einordnung der Usability-Evaluationen

identifizieren: *Inspektionsmethoden*⁹ und *formale Methoden*. Daneben existieren noch weitere Methoden, die sich entweder nicht eindeutig einer der beiden Gruppen zuordnen lassen oder komplett eigenständig sind. Innerhalb der Inspektionsmethoden (Cockton et al., 2008; Hollingsed und Novick, 2007; Virzi, 1997) wird zwischen der *heuristischen Evaluation* und den *Walkthrough-Verfahren* unterschieden. Die beiden Verfahren gehören zu den am häufigsten in der Literatur genannten analytischen Usability-Methoden. Ein weiteres Feld sind die *formalen Methoden*. Diese Methoden verwenden zum einen *Gestaltungsrichtlinien* und *Guidelines* zur Ermittlung der Usability. Daneben zählen die *modellbasierten Evaluationen* zu den formalen Methoden. Innerhalb dieser Gruppe ist besonders das *GOMS-Modell* (Kieras, 2008) geläufig. Problematisch innerhalb dieser Gruppe erscheint die Verwendung der Begriffe *Guidelines* und *Leitfäden*, die je nach Autorin beziehungsweise Autor dieselben oder verschiedene Methoden beschreiben. Dem letzten Feld werden die *sonstigen Methoden* zugeordnet. Darunter fällt beispielsweise die *Consistency Evaluation*. Die aufgezeigte Kategorisierung ist nicht trennscharf. Viele der hier aufgeführten Methoden sind sich im Aufbau, der Durchführung oder ihrer Zielsetzung mitunter ähnlich.

Heuristische Evaluation

Das Verfahren der Heuristischen Evaluation wurde in der ersten Hälfte der 1990er-Jahre entwickelt. Federführend in der Entwicklung war Jakob Nielsen (1992; 1994; 1994; 1990), dessen entwickelte Grundlagen der Methode bis heute Gültigkeit haben. Die Methode sieht vor, dass *Usability-Experten* eine Anwendung anhand von festgelegten *Heuristiken*, sogenannten „rules of thumb“ (Cockton et al., 2008, S. 1175), nach Problemen hinsichtlich der Usability untersuchen und bewerten. Die Heuristiken können als Oberkategorien für eine Vielzahl von Usability-Problemen verstanden werden, wie beispielsweise die Einhaltung von Konsistenzen. Die ersten Heuristiken stammen von Nielsen (1994d) und wurden durch eine Faktorenanalyse von 249 Usability-Problemen erweitert (Nielsen,

⁹Diese sind auch unter den Begriffen *Usability Inspection Methods* (Hollingsed und Novick, 2007; Virzi, 1997) und *Inspection-based Evaluations* (Cockton et al., 2008) bekannt.

1994a, vgl. auch Kapitel A.1). Muller, Matheson et al. (1998) und Muller und McClard (1995) erweiterten die zehn Heuristiken von Nielsen, da diese den Nutzungskontext nicht berücksichtigen würden und zu produktorientiert seien (vgl. Kapitel A.2). Aktuelle Heuristiken finden sich bei Shneiderman und Plaisant (2010) und Sarodnick und Brau (2011). Erstere werden von den Autoren als die *eight golden rules* bezeichnet (vgl. Kapitel A.3). Die zwölf Heuristiken von Sarodnick und Brau basieren auf der ISO 9241, Teil 110, eigenen Erfahrungen als auch Literaturrecherchen (vgl. Kapitel A.4).

Die Expertinnen und Experten nehmen die Heuristiken als Grundlage und versuchen sich in die Sichtweise der Nutzerinnen und Nutzer hineinzudenken und entsprechend zu handeln. Dies stellt hohe Anforderungen an die Expertinnen und Experten, da sie neben ihrem Wissen über Usability auch Wissen über den Nutzungskontext, den Aufbau und das Design der Anwendung als auch über die Zielgruppe mit in die Evaluation aufnehmen müssen (Cockton et al., 2008, S. 1173). Zudem ist es relevant, ob eine Expertin beziehungsweise ein Experte bereits über Hintergrundwissen über die zu untersuchende Anwendung oder die Domäne besitzt. Die Ergebnisse durch ungeübte Evaluatorinnen und Evaluatoren unterscheiden sich deutlich von denen geübter Expertinnen und Experten (Shneiderman und Plaisant, 2010, S. 142).

Die Heuristische Evaluation wurde in der Vergangenheit kaum verändert. Es gibt zwei Weiterentwicklungen: Muller, Matheson et al. (1998) haben die *Participatory Heuristic Evaluation* entwickelt, Sarodnick und Brau (2011) beschreiben ihre Methode als *Kooperative Heuristische Evaluation*. Beide Methoden beziehen die Nutzerinnen und Nutzer der zu evaluierenden Anwendung mit in die Evaluation ein. In der *Participatory Heuristic Evaluation* sind sie gleichrangige Evaluatorinnen und Evaluatoren neben den Usability-Expertinnen und Usability-Experten. Muller, Matheson et al. (1998, S. 15) sprechen von „work-domain expert inspectors“. Als Vorteil wird gesehen, dass zwischen der Gruppe der Usability Expertinnen und Expertinnen und der Gruppe der Nutzerinnen und Nutzern unterschieden wird. Zudem müssen sich die Usability Expertinnen und Experten nicht mehr in die Rolle der Nutzerinnen und Nutzer hineindenken. Die Evaluationen sollten mit Heuristiken der Autorinnen und Autoren durchgeführt werden (Muller und McClard, 1995). Der Ablauf des Verfahrens bleibt gleich. In der *Kooperativen Heuristischen Evaluation* versuchen Sarodnick und Brau (2011) die Trennung zwischen Heuristischer Evaluation und Nutzertest teilweise aufzuheben. Hierzu werden Teams aus jeweils einer Usability-Expertin beziehungsweise Usability-Experten und einer Nutzerin beziehungsweise einem Nutzer gebildet. Die Usability Expertin beziehungsweise der Usability Experte führt einen Nutzertest nach vorher ausgearbeiteten Szenarien durch. Die Nutzerin beziehungsweise der Nutzer kommentiert das Vorgehen der Expertin beziehungsweise des Experten und steht für Rückfragen zur Verfügung. Dies soll der Expertin beziehungsweise dem Experten helfen, die Nutzersicht besser einzunehmen.

Cognitive Walkthrough

Das Cognitive Walkthrough-Verfahren ist ein weiterer bekannter Vertreter der Inspektionsmethoden. Das Verfahren wurde Anfang der 1990er-Jahre von Peter Polson, Clayton Lewis und Cathleen Wharton entwickelt (Lewis, Polson et al., 1990; Polson, Lewis et al.,

1992; Wharton, Bradford et al., 1992) und ist auf die Zielorientierung der Nutzerinnen und Nutzer ausgelegt. Die Vorgehensweise der ersten Version sah neben einer ganzen Reihe an Vorbereitungs-materialien einen hohen Aufwand für die Expertinnen und Experten durch eine Vielzahl von Fragen für jeden durchgeführten Schritt vor. Neben der Häufigkeit der Fragen wurde die stete Gleichheit der Fragen kritisiert (Hollingsed und Novick, 2007, S. 250). Auf Basis dieser Kritik veröffentlichten die Autorinnen und Autoren eine zweite Version der Methode (Lewis und Wharton, 1997; Wharton, Rieman et al., 1994). Beide Versionen basieren auf einer Theorie des explorierenden Lernens (Hollingsed und Novick, 2007). Konkret bezieht sich die Methode auf das CE+ Model (Polson und Lewis, 1990). Es geht davon aus, dass die Nutzerinnen und Nutzer zielorientiert (*goal-driven*) innerhalb einer Anwendung arbeiten. Daran anschließend ist davon auszugehen, dass diese in einer ihnen unbekannt oder wenig geläufigen Anwendung die Schritte gehen, die am besten zu ihren Zielen passen (Cockton et al., 2008, 1175f). Um die Kritikpunkte der ersten Version zu beheben, ist die zweite Version leicht von der Zielorientierung der Nutzerinnen und Nutzer losgelöst. Dies reduziert die Vorbereitung der Expertinnen und Experten. Gleichzeitig müssen sie in der aktuellen Version nur noch vier Fragen zu jedem Schritt beantworten.

Ähnlich der Heuristischen Evaluation versetzen sich die Expertinnen und Experten in die Rolle der Nutzerinnen und Nutzer hinein. Sie müssen sich an vorher festgelegte Handlungsabfolgen halten. Gleichzeitig ist die Durchführung wesentlich stringenter, was in erster Linie an definierten Aufgaben liegt. Dies stellt höhere kognitive Anforderungen an die Expertinnen und Experten sowie ein hohes Verständnis der Nutzerinnen und Nutzer. Gleichzeitig muss bei der Aufgabenstellung ein fundiertes Wissen über die Ziele und Aufgaben der Nutzerinnen und Nutzer in der Anwendung vorhanden sein (Freiberg und Baumeister, 2008, S. 11). Das Verfahren ist dennoch für Einsteigerinnen und Einsteiger geeignet (John und Packer, 1995). Das Verfahren kann in allen Phasen der Produktentwicklung eingesetzt werden (Hollingsed und Novick, 2007, S. 250).

Die Durchführung eines Cognitive Walkthroughs setzt sich aus einer Vorbereitungs- und einer Analysephase zusammen. Zunächst werden von den Evaluatoreninnen und Evaluatoren ein oder mehrere Nutzerprofile definiert. Anschließend werden die Aufgaben festgelegt, anhand derer die Anwendung evaluiert werden soll. Für jede Aufgabe wird eine idealtypische Musterlösung entwickelt. In einem letzten Schritt wird die Nutzerschnittstelle definiert. Dies ist bei nicht funktionsfähigen Anwendungen (Prototypen, Mock-Ups etc.) wichtig, damit die Evaluatoreninnen und Evaluatoren wissen, welche Interaktionsmöglichkeiten ihnen zur Verfügung stehen. Anschließend erfolgt die Analysephase. Hier beantworten die Expertinnen und Experten zu jeder entsprechenden Aufgabe die folgenden Fragen (Wharton, Rieman et al., 1994, S. 9):

1. Will the user try to achieve the right effect?
2. Will the user notice that the correct action is available?
3. Will the user associate the correct action with the effect that the user is trying to achieve?
4. If the correct action is performed, will the user see that progress is being made towards solution of the task?

Mittels der Methode lassen sich potentielle Probleme der Anwendung schrittweise analysieren. Gleichzeitig wird ersichtlich, ob die zukünftigen Nutzerinnen und Nutzer aufgrund ihres Vorwissens vor unlösbare Aufgaben gestellt werden.

Neben dem klassischen Cognitive Walkthrough wurden in den letzten Jahren weitere Varianten entwickelt. Bias (1994) hat das *Pluralistic Walkthrough* vorgestellt. Der größte Unterschied liegt darin, dass die Evaluatorinnen und Evaluatoren in Teams arbeiten, die sich aus Usability-Expertinnen und Usability-Experten, Entwicklerinnen und Entwicklern sowie aus potentiellen Anwenderinnen und Anwendern zusammensetzen. Die Gruppe wird von einer Administratorin beziehungsweise einem Administrator geleitet. Diese entwickelt die Aufgabenbeschreibungen und führt die Testpersonen anschließend über Screenshots und Aufgabenbeschreibung durch den Test. Der Vorteil der Methode liegt darin, dass durch die Anwesenheit der Entwicklerinnen und Entwickler der Aufwand der Entwicklung verringert wird. Das gleichzeitige Testen von Entwicklerinnen und Entwicklern sowie Nutzerinnen und Nutzern kann die Anzahl der gefundenen Probleme erhöhen (Cockton et al., 2008, S. 1177; Hollingsed und Novick, 2007, S. 251). Untersuchungen haben eine Steigerung von etwa 30 Prozent an gefundenen Problemen festgestellt (Zhang, Basili et al., 1998). Negativ ist zu bewerten, dass die Festlegung der Aufgaben nur noch von einer Person vorgenommen wird. Als Vereinfachung des Cognitive Walkthrough kann das *Streamlined Cognitive Walkthrough* (Spencer, 2000) angesehen werden. Um die Belastung der Expertinnen und Experten zu verringern, wurde die Anzahl der Fragen von vier auf zwei reduziert (Spencer, 2000, S. 355):

1. Will the user know what to do at this step?
2. If the user does the right thing, will they know that they did the right thing, and are making progress towards their goal?

Eine andere Methode ist das *Heuristische Walkthrough* (Sears, 1997). Hierbei handelt es sich um eine Kombination von Cognitive Walkthrough und Heuristischer Evaluation. Zunächst wird den Evaluatorinnen und Evaluatoren eine Aufgabenliste vorgelegt, die geläufigere als auch weniger geläufigere Aufgaben enthält. Sie sind mit einer Prioritätenliste gekennzeichnet. Den Evaluatorinnen und Evaluatoren ist freigestellt, welche dieser Aufgaben und in welcher Reihenfolge sie diese bearbeiten. Dabei sollen sie ihre Schritte, analog zum Cognitive Walkthrough, durch das Beantworten von vier Leitfragen begleiten. Diese Fragen sind im Vergleich zu Wharton, Rieman et al. (1994) leicht abgewandelt. Der erste Teil der Evaluation dient dazu, dass sich die Expertinnen und Experten einen ersten Überblick über die Anwendung verschaffen können. Anschließend erfolgt die „free-form evaluation“ (Spencer, 2000, S. 221). Dabei können sich die Evaluatorinnen und Evaluatoren frei in der Anwendung bewegen und sollen sowohl den Aufgaben folgen als auch die Heuristiken von Nielsen (1994d) heranziehen. Die Methode findet nach Sears (1997) mehr unmittelbare Probleme als das Cognitive Walkthrough und weniger fälschliche Probleme als die Heuristische Evaluation (Cockton et al., 2008, S. 1176).

Gestaltungsrichtlinien/Guidelinebasierte Evaluationen

Neben den modellbasierten Methoden stellen die *Gestaltungsrichtlinien* beziehungsweise *guidelinebasierenden Evaluationen* die erste große Gruppe der *formalen Usability-Methoden*

dar. Historisch sind sie die ältesten Usability-Evaluationen (Sarodnick und Brau, 2011, 122f). Wie bei den Inspektionsmethoden werden die Usability-Untersuchungen in dieser Gruppe ebenfalls von Usability-Expertinnen und Usability-Experten durchgeführt und ausgewertet. Die direkte Interaktion mit den Nutzerinnen und Nutzern der zu untersuchenden Anwendung ist nicht vorgesehen. Vanderdonckt (1999, S. 82) definiert Guidelines wie folgt: „We here define a guideline by a design and/or evaluation principle to be observed in order to get and/or guarantee the usability of a UI for a given interactive task to be carried out by a given user population in a given context.“ Dem Begriff können fünf Subgruppen zugeordnet werden (Sarodnick und Brau, 2011, S. 124; Vanderdonckt, 1999, 87f):

- *Gestaltungsrichtlinien/Design Rules*: Die erste Gruppe beinhaltet Richtlinien und Anweisungen zur Gestaltung der entsprechenden Anwendung. Diese kommen häufig von Designerinnen beziehungsweise Designern oder Entwicklerinnen beziehungsweise Entwicklern und sind selbsterklärend.
- *Sammlungen von Richtlinien/Guideline Sets*: Es handelt sich um zusammengefasste Anforderungen zu User Interfaces. Die Anforderungen sind thematisch gebündelt und für eine ganze Reihe an Schnittstellen geeignet. Jede Anforderung ist genau beschrieben und häufig mit Beispielen erläutert. Diese Zusammenfassungen werden als Sets oder Guides beschrieben und sind eng an eine Schnittstelle gebunden wie die „Research-based web design & usability guidelines“ (Department of Health and Human Services und General Services Administration, 2006) oder „Guidelines for designing user interface software“ (Smith und Mosier, 1986).
- *Standards*: Standards sind definierte Anforderungen und Grundlagen, die häufig von supra-nationalen Organisationen herausgegeben werden. Ein prominentes Beispiel ist die ISO 9241 (ISO, 1998, 2006, 2010). Ein Vorteil der Standards, im Vergleich zu den Gestaltungsrichtlinien, ist die Generalisierbarkeit. Sie sind nicht auf ein bestimmtes Interface zugeschnitten.
- *Styleguides/Style Guides*: Diese sind im Vergleich zu den Standards sehr konkret und genau gefasst. Es handelt sich um konkrete Regeln und Anweisungen, wie eine Anwendung zu gestalten ist. Ziel ist die möglichst hohe Verallgemeinerung. Wie der Name andeutet, zielen sie vor allem auf gestalterische Aspekte ab.
- *Ergonomische Algorithmen/Ergonomic Algorithms*: Diese versuchen, die Anforderungen zu systematisieren und legen fest, wie unter bestimmten Bedingungen eine Ausgestaltung auszusehen hat. Es handelt sich häufiger um einen Teil der Software, der bestimmte Anzeigoptionen steuert. Daher stammt die Verwendung der Beschreibung als „Algorithmus“.

Die Trennung zwischen den einzelnen Subgruppen ist schwierig, da die Bezeichnung innerhalb der Scientific Community nicht klar ist. Cockton et al. (2008) unterscheiden zwischen *Principles*, *Guidelines* und *Style* und *Rules*. Freiberg und Baumeister (2008) vermischen die Punkte der *Richtlinien-Sammlungen* und *Standards*. Zudem existieren weitere Begriffe für die Methoden wie *Styleguides* oder *ergonomische Kriterien* (Ergonomic Criteria; Bastien und Scapin, 1995; Scapin und Bastien, 1997). Die Guidelines gehen neben der Usability teilweise auch auf die *Barrierefreiheit* (Accessibility; Lazar et al., 2010) ein, da

beide Gruppen miteinander verwandt sind.

Die Durchführung ähnelt dem Vorgehen in anderen analytischen Methoden. Nach dem Einprägen der Guidelines sowie dem Kennenlernen der Anwendung folgt die Anwendung der Guidelines. Das Vorgehen ist eher iterativ und zufällig (Sarodnick und Brau, 2011, S. 126). Es ist unüblich, die Guidelines Punkt für Punkt durchzugehen. Um dem entgegenzuwirken, sollten die Guidelines nach einer eher freien Phase zusätzlich strukturiert durchgegangen werden. Guidelines lassen sich sowohl für formative als auch summative Evaluationen einsetzen.

Modellbasierte Evaluationen

Bei *modellbasierten Evaluationen* wird davon ausgegangen, dass sich die Aufgaben in Anwendungen in kleinere Unteraufgaben und Zwischenschritte unterteilen lassen. Die Unteraufgaben ergeben zusammengefasst ein Modell, welches sich separat evaluieren und detailliert messen lässt. Die Messungen sind stark auf quantitative Kennzahlen ausgelegt (Ausführungszeit, logische Abfolge der Schritte). Ein Vorteil der Methode ist, dass sich Prototypen modellbasiert überprüfen lassen, weshalb modellbasierte Evaluationen vor allem in frühen Entwicklungsphasen einer Software angewendet werden. Die Methode ist einfacher und schneller durchzuführen als beispielsweise Nutzertests (vgl. Kapitel 6.3), da das Verfahren sowohl von Nutzerinnen und Nutzern als auch Expertinnen und Experten durchgeführt werden kann (Kieras, 2008). Es existieren verschiedene Verfahren, wobei das *GOMS-Modell* (Card et al., 1983) die bekannteste Variante darstellt.¹⁰ GOMS steht als Akronym für *Goals, Operators, Methods* und *Selection Rules*, welche wie folgt definiert sind (John und Kieras, 1996a,b):

- *Goals*: Stellen die Ziele dar, die eine Nutzerin beziehungsweise ein Nutzer in der Anwendung erreichen soll.
- *Operators*: Dies sind die „elementaren Handlungen“ (Sarodnick und Brau, 2011, S. 132), die die Nutzerinnen und Nutzer in der Anwendung ausführen. Dabei kann es sich um die aktive Interaktion mit der Anwendung (Klicken mit der Maus, Drücken einer Taste), um das Aufnehmen von Informationen (Lesen des Bildschirm Inhalts) oder die Verarbeitung von Informationen handeln.
- *Methods*: Darunter sind die vorher definierten Sequenzen von einzelnen Handlungen (*Operators*) zusammengefasst. Die Sequenzen führen zum erfolgreichen Erreichen der Ziele.
- *Selection Rules*: Diese beschreiben, wann eine bestimmte Sequenz von Handlungen (*Methods*) nötig ist, um ein bestimmtes Ziel zu erreichen.

Anhand des festgelegten Modells können Expertinnen und Experten die Anwendung evaluieren. Zunächst wird diese detailliert beschrieben und prototypische Nutzerinnen und Nutzer definiert. Anschließend wird der optimale Weg zur Erfüllung der Aufgabe definiert. Nach der Durchführung der Tests durch die Expertinnen und Experten werden die Ergebnisse zusammengefasst, mit den vorher festgelegten Sequenzen und Handlungen verglichen und entsprechende Hinweise zur Weiterentwicklung oder Überarbeitung der Anwendung gemacht (Freiberg und Baumeister, 2008, S. 10). Durch die strikte Vorgabe

¹⁰Dix et al. (2004), Kieras (2008) und Paterno (2009) stellen weitere Verfahren vor.

von Zielen und Handlungen eignet sich das GOMS-Modell besonders für Aufgaben, die sich genau beschreiben und eng fassen lassen (Strybel, 2011). Darunter fallen auch wiederholbare und lernbare Aufgaben wie beispielsweise das Buchen eines Tickets. Weniger geeignet ist das Verfahren, wenn die Aufgaben kreativ, offen oder problemlösend ausgelegt sind. Entsprechende Aufgaben wie das Abgeben einer Bewertung oder Verfassen eines Textes lassen sich nicht unproblematisch zeitlich messen. Die genaue Definition des entsprechenden Ziels ist ebenfalls schwierig.

Eine besondere Spielart der modellbasierten Evaluation stellen die *Szenarien* und *Use Cases* dar (Carroll, 2000, 2003). Bei beiden handelt es sich um die Beschreibung der Interaktion zwischen Nutzerin beziehungsweise Nutzer und der Anwendung. Beide Verfahren kommen vor allem in der Programmentwicklung zum Einsatz und sollen Designern und Programmierern einen Eindruck darüber vermitteln wie das endgültige Produkt aussehen soll. Szenarien sind „stories“ (Goodman et al., 2012, S. 501), die die Durchführung einer Aufgabe in einem realen Anwendungsfall zusammenfassen. Neben der Interaktion zwischen Mensch und Maschine werden auch die Umstände der Nutzung und Eigenschaften der Nutzerinnen und Nutzer beschrieben. Das Use Case ist dagegen die genaue Beschreibung des Interaktionsablaufs der Aufgabe (Constantine und Lockwood, 1999). Es ist abstrakter gehalten und konzentriert sich genauer auf die Handlungen wie beispielsweise bei Stone et al. (2005, 67ff) beschrieben wird. Streng genommen handelt es sich dabei um keine Methode zur Usability-Evaluation. Allerdings spielt die Usability bei der Beschreibung der Szenarien und Use Cases eine wichtige Rolle (Carroll, 2000; Stone et al., 2005). Szenarien werden auch bei den Nutzertests verwendet, um die Aufgaben der Testpersonen zu entwickeln (vgl. Kapitel 6.3.1).

Consistency Inspection

Bei der Consistency Inspection wird darauf geschaut, ob das Design der Anwendung stets gleich dargestellt wird. Die Konsistenz ist nach Nielsen (1993, S. 132) ein zentraler Usability-Baustein: „Consistency is one of the most basic usability principles.“ (Nielsen, 1993, S. 132) Shneiderman und Plaisant (2010, 74f) definieren Konsistenz wie folgt: „Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; consistent color, layout, capitalization, fonts, and others should be employed throughout. Exceptions should be comprehensible and limited in number.“ Eine Verletzung der Konsistenz kann bei Nutzerinnen und Nutzern zu Irritation und zu Problemen in der Bedienung der Anwendung führen. Dabei wird zwischen interner und externer Konsistenz unterschieden (Chen, Huang et al., 2007), wobei die interne Konsistenz bei der singulären Betrachtung einer Anwendung von höchster Bedeutung ist. Eine Fehlermeldung sollte beispielsweise unter allen Bedingungen gleich aussehen. Unter der externen Konsistenz wird eine gleichbleibende *User Experience* zwischen verschiedenen Anwendungen (vom gleichen Hersteller oder innerhalb eines Cooperative Designs) verstanden. Methodisch wird die Consistency Inspection durch den Vergleich verschiedener Screenshots realisiert. Diese werden von einer Expertin beziehungsweise einem Experten verglichen und auf Verletzungen der Konsistenz geprüft. Optional ist die direkte Interaktion mit der Anwendung möglich. Die

prominente Stellung der Konsistenz wird durch die Erwähnung in weiteren Guidelines und Heuristiken untermauert. In Shneiderman und Plainsants (2010) *acht goldenen Regeln* für das Interface Design und in den 10 Heuristiken nach Nielsen wird jeweils auf die Konsistenz verwiesen (siehe Anhänge A.1 und A.3).

6.2.2 Empirische Usability-Evaluationen

Die empirischen Usability-Evaluationen sind die zweite große Gruppe neben den analytischen Usability-Evaluationen. Der maßgebliche Unterschied zu den analytischen Evaluationen ist, dass bei den empirischen Methoden immer Nutzerinnen und Nutzer mit in die Tests einbezogen werden. Die Evaluationen werden nicht von Expertinnen und Experten durchgeführt, sondern die Eindrücke und Erfahrungen der Nutzerinnen und Nutzer stehen im Zentrum des Interesses. Konzeption, Durchführung und Auswertung der Evaluationen erfolgen durch Expertinnen und Experten. Grundsätzlich können drei verschiedene Methoden herausgearbeitet werden, die sich den empirischen Usability-Methoden zuordnen lassen: *Fragebögen*, *Interviews* und *Nutzertests* (vgl. Kapitel 6.3). Die beiden erstgenannten Methoden werden teilweise unter dem Begriff der „Query Techniques“ (Dix et al., 2004, 348ff. Freiberg und Baumeister, 2008) zusammengefasst. Fragebögen scheinen eine höhere Relevanz innerhalb der Literatur zu besitzen als Interviews. Dies liegt möglicherweise daran, dass es sich um die einzige Methode handelt, mittels derer viele Nutzerinnen und Nutzer bei geringem Aufwand und in einem kleinen Zeitfenster evaluiert werden können. Am relevantesten für die Gruppe der empirischen Usability-Evaluationen sind die Nutzertests. Dabei testen Nutzerinnen und Nutzer eine Anwendung und bewerten die Nutzbarkeit beziehungsweise decken Probleme bei der Bedienung auf.

Die Konzeption von empirischen Usability-Evaluationen ist nicht trivial, was in erster Linie an der Einbeziehung der Nutzerinnen und Nutzer liegt. Während die Expertinnen und Experten der analytischen Usability-Evaluationen häufig bereits über Erfahrungen bei der Durchführung der Evaluationen verfügen, ist dies bei normalen Nutzerinnen und Nutzern nicht der Fall. Daneben ist die Vorbereitung, Durchführung und Auswertung der Evaluationen häufig aufwendiger. Dies liegt neben der Unerfahrenheit der Testpersonen daran, dass die genaue Ausgestaltung der empirischen Usability-Evaluationen über mehr Optionen verfügt und daher variabler ist. Die Einbeziehung „realer“ Nutzerinnen und Nutzer sowie die hohe Optionsvielfalt stellen zugleich die größten Vorteile der empirischen Usability-Evaluationen dar. Die hohe Anzahl an Optionen bezieht sich nicht nur auf die verschiedenen Methoden (insbesondere Interviews, Fragebögen und Nutzertests), sondern auch auf die hohe Anzahl an verschiedenen Techniken der einzelnen Methoden. Eine besonders hohe Anzahl an verschiedenen Techniken ist bei den Nutzertests zu finden (u. a. „lautes Denken“, Mouse-Tracking, Eye-Tracking, Performance Measures). Das empirische Vorgehen von Nutzertests kann ebenfalls angepasst werden. Eine weitere Besonderheit der empirischen Usability-Evaluationen stellt die Möglichkeit dar, experimentelle Untersuchungsdesigns zu verwenden (Diekmann, 2010; Schnell et al., 2008). Ein „Experimentelles Design“ (Diekmann, 2010, 296f) liegt vor, wenn mindestens zwei Testgruppen gebildet werden, von denen jeweils eine Gruppe die Kontrollgruppe und die

andere Gruppe die Versuchsgruppe darstellt. Die Zuordnung muss randomisiert erfolgen. Den „Reiz“ oder „experimentellen Stimulus“ (Diekmann, 2010, S. 296) erfährt nur die Versuchsgruppe. Experimentelle Designs helfen daher vor allem bei der Kontrolle von Störfaktoren (zwischenzeitliches Geschehen, Reifungsprozesse der Probanden, Messfehler, Hilfsmittel, Verzerrte Auswahlen und Ausfälle; Schnell et al., 2008, 217f). Im Fall von Usability könnten beispielsweise verschiedenen Versuchsgruppen unterschiedliche Designs einer Anwendung in einem formativen Test präsentiert werden. Denkbar ist ebenfalls, bei einer (Weiter-)Entwicklung einer bestehenden Anwendung, eine Kontrollgruppe zu bilden, die im Rahmen eines summativen Tests die ursprüngliche Version untersucht, während die Versuchsgruppe die Neuentwicklung untersucht. Es existieren verschiedene Versionen des experimentellen Designs. Bekannt sind vor allem das Vorher-Nachher-Design, die einmalige Messung und das Solomon-Design. In der Umfrageforschung ist das Ex-Post-Facto-Design weiter verbreitet. Dort werden Versuchs- und Kontrollgruppe erst nach dem Test gebildet, was streng genommen gegen die Randomisierung verstößt. Die Methode wird daher auch als „quasi-experimentelles Design“ bezeichnet (Diekmann, 2010; Schnell et al., 2008).

Fokusgruppen

Die Methode der *Fokusgruppen* stellt die erste hier beschriebene Methode dar, in der die Nutzerinnen und Nutzer der Anwendungen im Mittelpunkt stehen. Das Vorgehen bei der Methode besteht darin, sechs bis neun Nutzerinnen und Nutzer in einer etwa zwei-stündigen, moderierten Gruppendiskussion zusammenzubringen (Lamnek und Krell, 2010; Mayring, 2002). Das Thema der Diskussion ist die Bewertung neuer Konzepte oder der bestehenden Anwendungen. Durch das direkte Einbeziehen von Anwenderinnen und Anwendern wird diese Methode teilweise den empirischen Usability-Evaluationen zugeordnet. Die Methode kommt üblicherweise in einem frühen Entwicklungsstadium der Software zum Einsatz (Freiberg und Baumeister, 2008, 3f). Die Rückmeldungen durch die Nutzerinnen und Nutzer sind in der Regel keine konkreten Usability-Probleme („(...) focus groups are a rather poor method for evaluating interface usability.“ (Nielsen, 1997, S. 94)), sondern eher Anregungen und Erwartungen an die Anwendung. Bei der Zusammenstellung der Gruppe sollte auf eine hohe Homogenität der Nutzerinnen und Nutzer geachtet werden (Hegner, 2003, 29f). Daneben sollte die Moderatorin beziehungsweise der Moderator im Umgang mit Gruppendiskussionen geschult sein. Das größte Problem stellt die soziale Erwünschtheit während der Diskussion dar. Die Methode protokolliert ausschließlich die Anregungen, Wünsche und Erwartungen der Nutzerinnen und Nutzer und geht weniger auf die tatsächliche Nutzung der Anwendung ein. Nielsen (1997) empfiehlt daher, Fokusgruppen immer mit einer weiteren Methode zu kombinieren. Die Zuordnung zu den empirischen Usability-Evaluationen ist in der Literatur nicht konsistent. In der Mehrzahl der Fälle wird sie den empirischen Methoden zugeordnet (u. a. bei Freiberg und Baumeister, 2008; Stone et al., 2005; Stowasser, 2006), teilweise wird sie allerdings auch zu den analytischen Usability-Evaluationen gezählt (unter anderem bei Hegner, 2003). Durch die Einbeziehung der (potenziellen) Nutzerinnen und Nutzern wird sie auch hier den empirischen Usability-Methoden zugeordnet.

Interviews

Bei einem Interview kommt es zu einer Diskussion zwischen der Nutzerin oder dem Nutzer auf der einen Seite und einer Interviewerin beziehungsweise einem Interviewer auf der anderen Seite. In Interviews wird zwischen drei Standardisierungsgraden unterschieden: *offen*, *halbstandardisiert* und *standardisiert* (Bortz und Döring, 2006; Gläser und Laudel, 2010). Komplett offen gehaltene Interviews haben eine einleitende Forschungsfrage. Die weitere Entwicklung des Interviews hängt von der sich entwickelnden Dynamik zwischen der Interviewerin beziehungsweise dem Interviewer auf der einen Seite und der Nutzerin beziehungsweise dem Nutzer auf der anderen Seite ab. Je offener das Interview gehalten ist, desto gefragter sind die kommunikativen Fähigkeiten der Interviewerin beziehungsweise des Interviewers. Gleichzeitig müssen diese über ein umfangreiches Wissen in Bezug auf die Fragestellung und die evaluierte Anwendung verfügen, um flexibel auf die Nutzerin beziehungsweise den Nutzer reagieren zu können. Je stärker das Interview standardisiert wird, desto mehr ist die Interviewerin beziehungsweise der Interviewer angehalten, sich an den vorher formulierten Leitfragen zu orientieren. Ein voll standardisiertes Interview verhält sich wie ein interviewgeführter Fragebogen. Eine höhere Standardisierung hat zur Folge, dass die Interviewerin beziehungsweise der Interviewer weniger Wissen über die Anwendung haben muss. Gleichzeitig geht eine höhere Standardisierung mit einer geringeren Flexibilität einher. Da ein gewisses Maß an Flexibilität häufig erwünscht ist, werden halbstandardisierte Interviews durchgeführt, die sich an vorher festgelegten Leitfragen orientieren (Freiberg und Baumeister, 2008). Diese Art bietet die größte Flexibilität und ermöglicht es der Interviewerin beziehungsweise dem Interviewer, dynamisch auf die Nutzerin beziehungsweise den Nutzer einzugehen und an bestimmten Stellen genauer nachzufragen (Dix et al., 2004, S. 348). Die Orientierung an den Leitfragen stellt sicher, dass alle relevanten Aspekte der Gebrauchstauglichkeit der Anwendung angesprochen werden. Durch den Einsatz der qualitativen Inhaltsanalyse (Gläser und Laudel, 2010; Mayring, 2010; Mayring und Gläser-Zikuda, 2008) können auch weniger standardisierte Interviews systematisch ausgewertet und verglichen werden.

Fragebögen

Neben einem Interview sind Fragebögen die zweite Möglichkeit, empirische Usability-Evaluationen durchzuführen. Es handelt sich um eine der ältesten Methoden, Anwendungen zu evaluieren (Baecker et al., 1995). Dies hängt damit zusammen, dass es sich bei der Umfrageforschung um eine insbesondere durch die Sozialwissenschaften hoch entwickelte Technik mit einer hohen Verbreitung handelt (De Leeuw et al., 2009; Dillman, 2000; Groves, Fowler et al., 2009; Schnell et al., 2008). Fragebögen bieten im Vergleich zu Interviews eine verminderte Flexibilität, da die Fragen vorher formuliert werden (Dix et al., 2004, 258f). Die Erstellung und Gestaltung eines Fragebogens ist nicht trivial und sollte die drei Gütekriterien *Objektivität*, *Reliabilität* und *Validität* (Bortz und Döring, 2006; Diekmann, 2010) erfüllen. Fragebögen, die diese Kriterien verletzen, können zu inkonsistenten Ergebnissen führen. Bei der Entwicklung des Fragebogens stehen der Expertin beziehungsweise dem Experten verschiedene Antworten zur Auswahl. Darunter fallen Multiple Choice Fragen, offene Fragen oder Ratingskalen. Die genaue Ausgestaltung hängt vom Standardisierungs-

grad des Fragebogens ab (Sarodnick und Brau, 2011, 181ff.). Gleichzeitig bieten sie die Möglichkeit, schnell und mit einem überschaubaren Aufwand, Meinungen und Eindrücke der Nutzerinnen und Nutzer zu sammeln (Ozok, 2008). Im Vergleich zu Nutzertests ist der Aufwand geringer. Dies trifft insbesondere auf webbasierte Umfragen zu. Diese erlauben eine einfachere Nutzeradministration und schnellere Auswertung. Zudem ist es durch Umfragen möglich, mit einem überschaubaren Zeitaufwand viele Nutzerinnen und Nutzer zu befragen. Die Rekrutierung von Testerinnen und Testern sollte gut überwacht werden. Diese werden in der Regel per Zufallsstichprobe gezogen. Neben der Sicherstellung der Zufälligkeit („Sampling Error“; Levy, 2009), muss vorher sichergestellt werden, dass die Population der Testpersonen richtig definiert wurde („Coverage Error“; Lohr, 2009). Bei Nicht-Einhaltung einer oder beider Fehler, ist eine Repräsentativität der Stichprobe unter Umständen nicht mehr gegeben. Dies hat zur Folge, dass eine Übertragung der Ergebnisse der Stichprobe auf die Population (Grundgesamtheit) nicht mehr oder nur mit hohem Aufwand möglich ist. Zuletzt ist bei der Rekrutierung darauf zu achten, dass die Anzahl der Verweigerer („Nonresponse“; Groves, Dillman et al., 2002) möglichst gering ist. Sollte eine bestimmte Teilgruppe der Stichprobe die Teilnahme verweigern, kann dies ebenfalls zu einer systematischen Verfälschung der Stichprobe führen. Gleiches gilt, wenn sich die Verweigerung auf bestimmte Fragen bezieht („Item-nonresponse“). Die hohe Standardisierung der Methode ermöglicht eine quantitative Auswertung der Daten mit statistischen Methoden (Backhaus et al., 2011; Ozok, 2008) und den Vergleich zwischen einzelnen Nutzerinnen und Nutzern beziehungsweise bestimmten Nutzergruppen.

In den letzten Jahren wurden einige Fragebögen zur Bestimmung der Gebrauchstauglichkeit entwickelt (Perlman, 2015). Darunter fallen der *System Usability Scale* (SUS; Brooke, 2013), die *IsoMetrics* (Gediga, Hamborg und Düntsch, 1999), der *Questionnaire for User Interaction Satisfaction* (QUIS; Shneiderman und Plaisant, 2010, 169ff) oder *AttrakDiff* (Hassenzahl et al., 2003). Der *SUS-Fragebogen* beinhaltet zehn Fragen zu der wahrgenommenen Gebrauchstauglichkeit der Anwendung. Er wird als „quick and dirty“ (Brooke, 1996) bezeichnet. Dennoch wird dieser als valide und reliabel beschrieben (Tullis und Stetson, 2004). Bei den *IsoMetrics* handelt es sich um eine Operationalisierung der ISO 9241, Teil 110¹¹ (ISO, 2006). Es existieren zwei Versionen: die *IsoMetrics^S* für summative Evaluationen und die *IsoMetrics^L* für formative Evaluationen. Die Länge des Fragebogens variiert zwischen den beiden Versionen, wobei der summative Fragebogen kürzer ist. Der *QUIS-Fragebogen* stellt die subjektive Zufriedenstellung mit der Anwendung in den Vordergrund. Es werden weitere Punkte wie Gestaltung, Erlernbarkeit und Effizienz abgefragt. Die überarbeitete Version von Chin et al. (1988) wird als reliabel und valide beschrieben. Ähnlich ist der *AttrakDiff-Fragebogen* einzustufen. Dieser stellt noch deutlicher als der *QUIS-Fragebogen* die „joy of use“ (Hatscher, 2001) in den Vordergrund. Dies geschieht aus der Kritik an vorhandenen Fragebögen (wie den *IsoMetrics*), die vor allem die Dimensionen Effizienz und Effektivität in den Vordergrund stellen und die Zufriedenstellung weniger betrachten.

Neben den Fragen zur Gebrauchstauglichkeit der Anwendung sollten soziodemografi-

¹¹Ursprünglich handelte es sich um die ISO 9241, Teil 10 (Gediga, Hamborg und Düntsch, 1999). Diese wurde im Jahr 2006 durch die ISO 9241, Teil 110 ersetzt.

sche Merkmale (Alter, Geschlecht, Bildungsstand) mit erfasst werden. Je nachdem, wo die Evaluation stattfindet, lässt sich diese Liste auf Spezifika der Organisation (Abteilung, Position, Beruf etc.) zuschneiden. Dies ermöglicht feinere Gruppenvergleiche.

6.2.3 Einordnung der Usability-Methoden

Die bisher vorgestellten Usability-Evaluationen können in analytische und empirische Methoden eingeteilt werden. Der größte Unterschied ist, dass die analytischen Methoden von Expertinnen und Experten durchgeführt werden, während bei den empirischen Methoden die Nutzerinnen beziehungsweise die Nutzer als zentrale Quelle des Erkenntnisgewinns fungieren.

Analytische Methoden haben den Nachteil, dass sie keine direkte Interaktion mit der Endnutzerin bzw. dem Endnutzer bieten (Holzinger, 2005). Bilden die Expertinnen und Experten die Interaktionen der Nutzerinnen und Nutzer nicht adäquat ab, ist eine Verzerrung der Ergebnisse die Folge. Die eigene Meinung und Vorstellungen der Expertinnen und Experten können die Ergebnisse ebenfalls negativ beeinflussen (Jaspers, 2009, S. 342). Der größte Nachteil ist allerdings auch der größte Vorteil der analytischen Methoden: Die Durchführung erfolgt durch Expertinnen und Experten. Diese besitzen häufig bereits ein fundiertes Wissen über das Themenfeld Usability und das Testen von Anwendungen. Tatsächliche oder potenzielle Nutzerinnen und Nutzer werden nicht einbezogen und müssen folglich nicht rekrutiert beziehungsweise begleitet werden (Holzinger, 2005). Dadurch können die Methoden zeitnah und ohne einen hohen vorherigen Aufwand eingesetzt werden. Insbesondere die heuristische Evaluation wird als effizient und kostengünstig beschrieben (Jeffries und Desurvire, 1992; Nielsen, 1994c). Zudem werden durch die heuristische Evaluation eine hohe Anzahl an leichten und schweren Usability-Problemen gefunden (Bailey et al., 1992; Sears, 1997). Die Anzahl ist im Vergleich zu Nutzertests (vgl. Kapitel 6.3) und zum Cognitive Walkthrough größer (Jeffries, Miller et al., 1991). Die verwendeten Heuristiken sind allgemein gehalten und können dadurch die Analyse behindern (Jaspers, 2009, S. 342). Zudem sind die Expertinnen und Experten nur bedingt mit den späteren Nutzerinnen und Nutzern vergleichbar (Lin et al., 1997). Das Cognitive Walkthrough bietet durch die strukturierte Vorgehensweise den Vorteil, dass die gefundenen Probleme detailliert analysiert und beschrieben werden (Jaspers, 2009, S. 344). Die Anzahl der gefundenen Probleme ist höher je genauer die Aufgabenbeschreibungen sind (Sears und Hess, 1999), bleibt in der Summe aber häufig hinter der heuristischen Evaluation zurück. Die Methode findet vor allem schwere Probleme (Sears, 1997). Insbesondere das Cognitive Walkthrough eignet sich besonders zur Evaluation in einer frühen Entwicklungsphase (Hegner, 2003).

GOMS als ein Vertreter der formalen Methoden liefert ebenfalls in einer frühen Entwicklungsphase einen genaueren Überblick über die Usability und Performance der Anwendung (John und Kieras, 1996b). Dabei muss kein funktionsfähiger Prototyp vorhanden sein. Problematisch kann die detaillierte Ausgestaltung des Modells sein (Freiberg und Baumeister, 2008). Dies betrifft insbesondere die frühe Entwicklungsphase, wenn zwar wesentliche Bestandteile der Anwendung festgelegt wurden, aber viele Funktionen

noch nicht oder rudimentär implementiert sind. Gleichzeitig haben die Modelle Probleme mit interpersonellen Unterschieden, insbesondere wenn diese nicht in dem Modell vorgesehen wurden (Gediga und Hamborg, 2002). Daneben wurden die Gestaltungsrichtlinien als formale Methode vorgestellt. Diese haben den Vorteil der einfachen Anwendbarkeit. So ist es möglich, eine Evaluation von nur einer Expertin bzw. Experten durchführen zu lassen (Stowasser, 2006). Praktisch wird die zu evaluierende Anwendung mit den vorhandenen Richtlinien abgeglichen. Da diese keine Aufgaben oder Ziele enthalten, ist der Mehrwert vor allem oberflächlicher Natur und betrifft das Layout oder die interne Konsistenz (Freiberg und Baumeister, 2008). Inhaltlich ähneln die Richtlinien der Consistency Inspection. Daher treffen für diese dieselben Vor- und Nachteile zu. Der größte Nachteil der analytischen Methoden ist der fehlende Bezug zu den Nutzerinnen und Nutzern. Nur durch die Einbeziehung dieser ist eine Kombination von Usability-Evaluation und Logfile-Analyse möglich.

Empirische Methoden beziehen dagegen immer tatsächliche oder potenzielle Endnutzerinnen und Endnutzer mit in die Analyse ein, was für die spätere Kombination nötig ist. Dies ermöglicht die tatsächliche Analyse des Nutzerverhaltens und wird nicht durch Expertinnen und Experten simuliert beziehungsweise modelliert (Holzinger, 2005). Der Nachteil ist der vergleichsweise hohe Aufwand der Rekrutierung der Testpersonen und der Auswertung. Die Fokusgruppen und die Interviews ermöglichen einen direkten Austausch mit den Nutzerinnen und Nutzern. Bei einer weniger starren Ausrichtung der Methoden ist durch Nachfragen der Moderatorin beziehungsweise des Moderators ein tieferes Verständnis der Usability-Probleme möglich. Zudem kann in den Fokusgruppen eine Interaktion zwischen den Teilnehmerinnen und Teilnehmern erfolgen. Durch die Diskussionen können neben den Ursachen weitere Punkte bezüglich der Usability herausgearbeitet werden. Alle Aussagen und Angaben in den Interviews und Fokusgruppen sind allerdings rekapitulierend, und die direkte Interaktion mit der Anwendung kann nicht beobachtet werden (Goodman et al., 2012). Rekapituliertes Verhalten ist darüber hinaus unter Umständen stark fehlerbehaftet (Schwarz, 2007). Umfragen und Fragebögen weisen dasselbe Problem der Rekapitulierung von vorhandenem Wissen auf (Schwarz und Oyserman, 2001). Der Vorteil im Vergleich zu Interviews und Fokusgruppen ist, dass mit vergleichsweise geringem Aufwand eine hohe Zahl an Nutzerinnen und Nutzern erreicht werden kann. Umfragen und Fragebögen stellen neben den Logfile-Analysen die quantitativste Methode dar. Der Inhalt und die Fragen müssen im Vorfeld erstellt werden und erfordern daher bereits mehr oder weniger umfassendes Vorwissen (Goodman et al., 2012). Neben Fragen zur Usability können weitere nutzerspezifische Fragen gestellt werden, wie beispielsweise die Computeraffinität oder allgemeine Fragen zur Nutzung von bestimmter Software. Zur Messung der Usability stehen verschiedene bereits definierte Fragebögen zur Auswahl. Diese sind allerdings allgemein gehalten und wenig spezifisch. Für die weitere Auswertung sind unter Umständen auch demografische Merkmale der Nutzerinnen und Nutzer von Interesse. Zuletzt benötigen Umfragen und Fragebögen noch eine umfangreichere Kenntnis der empirischen Sozialforschung (Diekmann, 2010; Groves, Fowler et al., 2009; Schnell et al., 2008) zur Erstellung und Formulierung der Fragebögen und statistische Kenntnisse zur Auswertung (Backhaus et al., 2011; Bortz und Döring, 2006; Ozok, 2008).

Die bisher vorgestellten empirischen Usability-Evaluationen decken die Anforderungen dieser Dissertation nicht ab. Für eine erfolgreiche Kombination der Logfiles aus dem Testsystem mit den „realen“ Logfiles aus einem Referenzsystem und damit zur erfolgreichen Bearbeitung der Forschungsfrage müssen die Testerinnen und Tester während der Nutzung beobachtet werden. Dies schließt die analytischen Methoden von vornherein aus, da dort keine Nutzerinnen und Nutzer herangezogen werden, sondern Expertinnen und Experten. Die bisher vorgestellten drei empirischen Usability-Evaluationen beziehen zwar die Nutzerinnen und Nutzer mit in die Analyse ein, lassen aber ebenfalls keine direkte Beobachtung dieser während der Interaktion mit der Anwendung zu. Daher werden diese auch verworfen. Die gewählte Alternative sind *Nutzertests*. Bei dieser empirischen Usability-Evaluation sollen die Testerinnen und Tester vorgegebene Aufgaben lösen. Währenddessen wird ihr Verhalten beobachtet und aufgezeichnet. In Verbindung mit der Protokollierung auf Logfile-Ebene ermöglicht dies die angestrebte Kombination der Methoden. Die Methode der Nutzertests wird im nächsten Kapitel ausführlicher vorgestellt.

6.3 Nutzertests als zentrale Methode

Nutzertests stellen die dritte Methode der empirischen Usability-Evaluationen dar. Es handelt sich zusammen mit den Gestaltungsrichtlinien (vgl. Kapitel 6.2.1) um das älteste Verfahren. Die ersten Ansätze lassen sich bis Anfang der 1980er-Jahre zurückverfolgen (Roberts und Moran, 1983). Dies geschah zeitlich mit der zunehmenden Verbreitung von Computern und Anwendungen, die in der Folge für eine breitere Bevölkerungsschicht interessant und relevant wurden (Dumas und Fox, 2008, S. 1130). Die Methode ist wahrscheinlich diejenige mit der höchsten Verbreitung innerhalb der Usability-Forschung (Dumas und Fox, 2008, S. 1130). Dies drückt sich auch dadurch aus, dass der englische Begriff *Usability Testing* häufig als Synonym für Usability-Evaluationen verwendet wird (Lazar et al., 2010, S. 260). Ein Nutzertest definiert sich dadurch, dass eine Gruppe von Testpersonen repräsentative Aufgaben in einer Anwendung durchführt (Dumas und Fox, 2008; Dumas und Redish, 1999; Lazar et al., 2010; Rubin und Chisnell, 2008; Shneiderman und Plaisant, 2010; Stone et al., 2005). Die Methode kann für formative als auch für summative Tests eingesetzt werden. Der Einsatz in summativen Tests ist geläufiger, da zu diesem Zeitpunkt der Entwicklung mindestens ein funktionierender Prototyp zur Verfügung steht, an dem die entsprechenden Aufgaben durchgeführt werden können (Dix et al., 2004, S. 327). Durch den Einbezug von Nutzerinnen und Nutzern bietet die Methode in der Regel den höchsten positiven Effekt auf die Weiterentwicklung der Anwendung und auf die Entdeckung von Problemen (Rosenbaum et al., 2000). Die Einbeziehung wird als größter Vorteil der Methode gesehen (Freiberg und Baumeister, 2008, S. 3), da dies die Möglichkeit bietet, Nutzerinnen und Nutzer während der Interaktion mit der Anwendung zu beobachten und Probleme „live“ zu entdecken. Gleichzeitig stellt die Methode hohe Anforderungen an die Entwicklung und Durchführung von Testaufgaben und die Rekrutierung von Nutzerinnen und Nutzern (Dumas und Fox, 2008, 1131f). Die Testerinnen und Tester folgen dabei den gestellten Aufgaben. Alle Teile der Anwendung,

die nicht in den Aufgaben vorkommen, werden nicht oder kaum evaluiert. Gleichzeitig sollten die Nutzerinnen und Nutzer möglichst repräsentativ ausgewählt werden, da es sonst zu einer Verfälschung der Ergebnisse kommen kann. Lazar et al. (2010, S. 256) schlagen vor, Nutzertests mit weiteren Evaluationsmethoden, insbesondere analytischen, zu kombinieren. Hier bietet sich an, auf *Cognitive Walkthroughs* oder *Heuristische Evaluationen* zurückzugreifen (Cockton et al., 2008; Nielsen, 1994d; Virzi, 1997). Diese sollten *vor* den Nutzertests durchgeführt werden.

Nutzertests entdecken vor allem das Synonym für ein Usability-Problem, ohne häufig den dahinterliegenden Grund sowie die Lösungsmöglichkeiten des Problems zu benennen (Doubleday et al., 1997). Dies muss entweder durch Interpretation oder die Evaluation durch Experten erfolgen. Da es in dieser Dissertation aber vor allem um das Entdecken der Probleme geht, kann dieser Nachteil vernachlässigt werden. Nutzertests liefern dabei eine große Menge an Daten aus verschiedenen Quellen (vgl. Kapitel 6.3.2; Holzinger, 2005). Zwar werden weniger Probleme als bei heuristischen Evaluationen gefunden. Jeffries, Miller et al. (1991) sprechen von etwa einem Drittel. Allerdings sind die gefundenen Probleme häufiger schwerwiegender als bei der heuristischen Evaluation. Verschiedene Studien haben zudem gezeigt, dass Nutzertests mehr schwerwiegende Probleme als das Cognitive Walkthrough-Verfahren entdecken (Karat et al., 1992; Lewis, Polson et al., 1990).

6.3.1 Vorgehen

Der Ablauf eines Nutzertests kann als allgemein gesichert gelten. Trotz leichter Unterschiede im idealen Vorgehen (Barnum, 2011; Dumas und Redish, 1999; Lazar, 2006; Rubin und Chisnell, 2008, vgl. auch Tabelle 6.2;) ist grundsätzlich eine hohe Ähnlichkeit zwischen den einzelnen Autoren erkennbar. Die Aufzählungen sind in den letzten Jahren eher gestrafft worden.¹² Inhaltlich wurden die einzelnen Unterpunkte eher ausgebaut als gekürzt. Dies bedeutet, dass der Gesamtumfang gleich geblieben ist. Bei Dumas und Redish (1999) ist der Bereich des Internets aufgrund des Alters der Veröffentlichung größtenteils ausgeklammert oder durch neuere Entwicklungen (Skriptsprachen etc.) irrelevant geworden. Die grundsätzlichen Aussagen der Autoren sind nach wie vor relevant. Dumas und Loring (2008) und Barnum (2011) beziehen die Definition von Nutzer- als auch Testzielen explizit mit in die Aufzählung ein. Bei Rubin und Chisnell (2008) findet sich diese implizit mit im ersten Punkt. Aufgrund der hohen Verbreitung¹³ wird das Vorgehen von Rubin und Chisnell (2008) hier übernommen. Das Debriefing (Punkt 6) wird bei den Ausführungen mit in die Durchführung der Nutzertests (Punkt 5) integriert.

(1) Testplan aufstellen¹⁴

In der ersten Phase wird ein Testplan entwickelt. Dieser beinhaltet alle Informationen, die zur Durchführung der Nutzertests notwendig sind und dient als zentrales Dokument für das weitere Vorgehen. Ein Testplan sollte die folgenden Informationen enthalten: *Gründe und Ziele für den Nutzertest, Forschungsfragen, Merkmale der Testpersonen, Methodolo-*

¹²Bei Dumas und Redish (1999) wurden einzelne Punkte per Hand zusammengezogen.

¹³Mit über 2.500 Zitationen bei Google Scholar.

¹⁴Engl.: Develop the test plan (vgl. Tabelle 6.2)

Tabelle 6.2: Vorgehensweisen bei Nutzertests

(Rubin und Chisnell, 2008)	(Lazar, 2006)
<ol style="list-style-type: none"> 1. Develop the test plan 2. Set up the test environment 3. Find and select test participants 4. Prepare test materials 5. Conduct the test sessions 6. Debrief the participants 7. Analyze data and observations 8. Report findings and recommendations 	<ol style="list-style-type: none"> 1. Select representative users 2. Select the setting 3. Decide what tasks users should perform 4. Decide what type of data to collect 5. Before the test session 6. During the test session 7. Debriefing the session
(Dumas und Redish, 1999)	(Barnum, 2011)
<ol style="list-style-type: none"> 1. Defining goals and concerns 2. Deciding who should be the participants 3. Recruiting participants 4. Selecting and organizing tasks to test 5. Deciding how to measure usability 6. Preparing test materials and environment 7. Conduct a pilot test 8. Conducting the test 9. Tabulating and analyzing data 10. Recommending changes and communicate the results 	<ol style="list-style-type: none"> 1. Understanding users and their goals 2. Planning for usability testing 3. Preparing for usability testing 4. Conducting a usability test 5. Analyzing the findings 6. Reporting the findings

gie, Aufgabenliste, Testumgebung und Testausrüstung, Rolle des Moderators, Datensammlung, Berichterlegung und Präsentation (Rubin und Chisnell, 2008, S. 67). Barnum (2011, 145ff.) erweitert die Aufzählung um die Punkte *Zusammenfassung, Incentives für Testpersonen* und *Szenarien*.

Zunächst sollten die *Gründe und Ziele des Nutzertests* festgelegt werden. Durch die Ausführungen wird deutlich, warum Nutzertests in der Anwendung sinnvoll und nötig sind. Optimal ist eine Einbindung in die laufende Produktentwicklung oder in ein wissenschaftliches Interesse. Auf jeden Fall sind Nutzertests des „Testens wegen“ zu vermeiden. Aus den Gründen und Zielen lassen sich in einem nächsten Schritt die *Forschungsfragen* ableiten. Nach Rubin und Chisnell (2008, S. 69) handelt es sich um den wichtigsten Abschnitt des gesamten Testplans. Die Forschungsfragen sollten so stark wie möglich zugespitzt und allgemeine Fragen (zum Beispiel „Ist das Produkt nutzbar?“) vermieden werden.¹⁵ Es bietet sich an, zunächst analytische Usability-Evaluationen durchzuführen, mittels derer Expertinnen und Experten Probleme und Unstimmigkeiten in der Anwendung offenlegen. Daraus können sowohl Forschungsfragen als auch Ziele des Nutzertests abgeleitet werden (Lazar et al., 2010, 256ff.).

Anschließend sollten die *Testerinnen und Tester genauer beschrieben werden*. Eine genaue Beschreibung der Testerinnen und Tester erleichtert die Rekrutierung. Die Teilnahmequote kann durch ein kleines Dankeschön (*Incentive*) gesteigert werden und muss im Testplan erwähnt werden. Wichtig ist, dass die zu rekrutierenden Personen zur aktuellen oder potenziellen Zielgruppe der Anwendung passen (Dumas und Fox, 2008, S. 1131). Zur Definition der Zielgruppe bietet sich ein zweistufiges Verfahren an. In einem ersten Schritt werden *Merkmale der Testerinnen und Tester* beschrieben. Dazu zählen neben demo-

¹⁵Rubin und Chisnell (2008, 71f) bieten eine Übersicht inhaltlich sinnvoller Forschungsfragen und Negativ-Beispielen an.

grafischen Merkmalen (Alter, Geschlecht, Bildungsniveau, Ethnie etc.), Kenntnisse und Kompetenzen im Bereich der zu testenden Anwendung und der Beruf beziehungsweise die berufliche Stellung (Barnum, 2011, 117f; Dumas und Redish, 1999, 121f; Rubin und Chisnell, 2008, 72f). Aus diesen Merkmalen sollten in einem zweiten Schritt *Profile* (Dumas und Redish, 1999; Rubin und Chisnell, 2008) oder *Persona* (Barnum, 2011; Pruitt und Aldin, 2006) entwickelt werden. Ein Profil definiert eine prototypische Testerin beziehungsweise einen prototypischen Tester anhand der definierten Merkmale wie Altersspanne, Bildungsniveau, Arbeitsverhältnis etc. Je nach gewünschter Anzahl der Testpersonen können zwei oder mehr unterschiedliche Profile definiert werden. Dadurch können verschiedene Nutzergruppen erfasst werden (Barnum, 2011, S. 116). Diese sollten abschließend beschrieben und nach ihren Merkmalen charakterisiert werden (Dumas und Redish, 1999, 125ff).¹⁶ Anschließend sollte die *Methodologie* beschrieben werden. Je nach Anzahl der Testerinnen und Tester kann ein experimentelles Design mit zwei oder mehr Gruppen durchgeführt werden (Dix et al., 2004, 329ff). Die Gruppen können verschiedene Versionen der Anwendung evaluieren. Alternativ kann die Reihenfolge der Aufgaben variiert werden oder die Gruppen führen unterschiedliche Aufgaben durch. Wichtig ist ebenfalls die in Kapitel 6.2.2 beschriebenen Gütekriterien einzuhalten. Neben dem Design des Nutzertests sollten die angewandten Techniken (vgl. Kapitel 6.3.2) aufgeführt werden. Die Anwendung einer Technik ist zwingend erforderlich, da das unstrukturierte Beobachten der Testerinnen und Tester nicht zielführend ist (Dix et al., 2004, S. 343). Die Kombination von quantitativen und qualitativen Techniken sollte erwähnt und begründet werden. Den Testerinnen und Testern kann nach dem Nutzertest ein Fragebogen (vgl. Kapitel 6.2.2) vorgelegt werden. Dies sollte erwähnt und der Fragebogen in den Anhang des Testplans angefügt werden.

Neben der Beschreibung, „wie“ die Nutzertests erhoben werden sollen, muss beschrieben werden, „was“ genau erhoben werden soll. Dies geschieht über die *Aufgabenliste*. Da sowohl die Anzahl der Testerinnen und Tester als auch die zur Verfügung stehende Zeit limitiert sind, können nicht alle Aspekte einer Anwendung getestet werden. In der Aufgabenliste sind alle Aufgaben und die Ziele aufgeführt, die die Testerinnen und Tester erreichen sollen (Barnum, 2011, 99f). Die Auswahl der Aufgaben erfolgt aus mehreren Gründen. Zunächst können diese aus einer Forschungsfrage oder einem Ziel abgeleitet werden. Daneben definieren Dumas und Redish (1999, S. 160) und Dumas und Fox (2008, 1131f) sechs weitere Gründe, warum eine Aufgabe in einem Nutzertest ausgewählt werden kann¹⁷: (1) Wichtige Aufgaben, die häufig durchgeführt werden oder kritisch für die weitere Nutzung der Anwendung sind (z. B. Einloggen). (2) Aufgaben, die Teile einer Anwendung betreffen, in denen häufiger Probleme in der Gebrauchstauglichkeit auftreten oder wahrscheinlich sind. Daneben können (3) Aufgaben auch das Design der Anwendung betreffen, wie die Nutzung von tiefen Menüs oder Toolbars. (4) Sofern eine Funktion zentral für den (ökonomischen) Erfolg einer Anwendung ist, kann dies ein Grund für die Aufnahme in einen Nutzertest sein. (5) Gleiches trifft für neue Funktionen innerhalb einer Anwendung zu. Zuletzt können (6) Aufgaben wiederholt durchgeführt werden, um

¹⁶Beispiele für Nutzerprofile finden sich in Barnum (2011, 94ff.), Dumas und Redish (1999, 129ff.) und Rubin und Chisnell (2008, 122ff.).

¹⁷Die letzten drei Gründe sind nach Ansicht der Autoren sekundär.

die Verbesserung der Gebrauchstauglichkeit nach einer Veränderung der Anwendung aufgrund eines vorherigen Nutzertests zu untersuchen. Neben der Aufzählung lassen sich Aufgaben aus eigenen Überlegungen und Vorerfahrungen ableiten. Rubin und Chisnell (2008, 85ff.) greifen die Punkte der *Häufigkeit* (engl. Frequency), *Wichtigkeit* (engl. Critically) und *problematischen Bereichen* (engl. Vulnerability) auf und fügen den Punkt der *Bereitschaft* (engl. Readiness) im Sinne einer abgeschlossenen Entwicklung hinzu. Die Beschreibung der Aufgabenliste enthält daneben die nötigen Materialien oder Voreinstellungen, um die Aufgabe durchzuführen (Rubin und Chisnell, 2008, 79ff.). Die Auflistung kann in Hardware und Software unterteilt werden (Dumas und Redish, 1999, 165ff.). Zudem sollte eine Beschreibung des erfolgreichen Abschlusses der Aufgabe vorhanden sein. Es ist möglich, zusätzliche Benchmark-Informationen anzugeben wie die benötigte Zeit, die Anzahl der Versuche oder die benötigten „Mausklicks“. Dies macht nur bei summativen Nutzertests Sinn und kann bei bestimmten Techniken nicht zielführend sein (z. B. lautes Denken; Rubin und Chisnell, 2008, S. 80). Den Testerinnen und Testern kann direkt die Aufgabenliste vorgelegt werden. Geläufiger ist die Arbeit mit *Szenarien* oder *Use Cases*, in denen die Aufgaben integriert sind und den Testpersonen einen verständlichen Kontext liefern sollen (vgl. Unterpunkt (4)).

Der Beschreibung der Aufgaben schließt sich die *Beschreibung der Testumgebung und der Testausrüstung* an. Die Beschreibung der Testumgebung beinhaltet den Ort, an dem die Nutzertests stattfinden. In der Vergangenheit war das Testen in Laboren oder unter Laborbedingungen geläufig (Dumas und Fox, 2008, S. 1136). Dies diente dazu, Störfaktoren und andere Einflüsse zu kontrollieren. Heute finden Nutzertests zunehmend in der für die Testerinnen und Tester gewohnten Umgebung statt, um diese in ihrer gewohnten Umgebung beobachten zu können (Stoessel, 2002). Parallel zu der Verbreitung von Laptops ist die Verfügbarkeit von Internetanschlüssen und deren Geschwindigkeit gestiegen (Statistisches Bundesamt, 2015). Dies ermöglicht *entfernte Nutzertests* (engl. *Remote Usability Tests*), bei denen die Testleiter räumlich von den Testpersonen getrennt sind. Neben der Beschreibung des Ortes sollte die Ausrüstung aufgeführt werden. Hier sollte sowohl Hardware (stationärer PC, Laptop, Tablet etc.) als auch die verwendete Software erwähnt werden. Es ist sinnvoll, zwischen Software für die Testpersonen und für die Aufzeichnung des Nutzertests (Recording-Software etc.) zu unterscheiden.

Der Nutzertest wird von einer *Moderatorin* oder einem *Moderator* begleitet und durchgeführt. Es kann vorkommen, dass Moderatorin und Testleiterin dieselbe Person sind. Diese steht den Testerinnen und Testern bei Nachfragen oder als Hilfestellung zur Verfügung. Gleichzeitig ist diese für die *Einleitung* (engl. Briefing) und Ausführungen nach dem Nutzertest verantwortlich (engl. *Debriefing*). In dem Testplan sollten die Aufgaben der Moderatorinnen und Moderatoren vermerkt werden. Gleichzeitig sollten Anweisungen für Verhaltensweisen im Problemfall erwähnt werden, beispielsweise ab wann Hilfestellungen für die Testerin beziehungsweise den Tester gegeben werden dürfen. Es empfiehlt sich, neue Moderatorinnen beziehungsweise Moderatoren vor den ersten Nutzertests zu schulen (Dumas und Loring, 2008). Wird erwartet, dass die Moderatorin beziehungsweise der Moderator aktiv Testdaten sammelt (handschriftliche Notizen, Zeitmessungen etc.), sollte dies detailliert beschrieben werden. In dem Abschnitt der *Datensammlung* sollen

quantitative und qualitative Daten aufgeführt werden. Die quantitativen Daten beziehen sich auf direkt messbares Verhalten der Testerinnen und Tester wie Anzahl der Klicks, benötigte Zeit oder Anzahl der Aufrufe bestimmter Bereiche oder Seite der Anwendung. Diese Daten werden als „Performance Data“ (Barnum, 2011, S. 137) oder „Performance Measures“ (Dumas und Redish, 1999, 184ff) bezeichnet und können automatisiert im Hintergrund erfasst werden (beispielsweise durch Logfiles). Sofern Fragebögen an die Testpersonen ausgeteilt werden, zählen diese zu den quantitativen Daten. Sie werden unter dem Begriff „Preference Data“ zusammengefasst (Barnum, 2011, 137f). Die Zuordnung von Fragebögen zu quantitativen Daten ist umstritten. Dumas und Redish (1999, 187f) ordnen sie den qualitativen Daten zu. Qualitative Daten sind häufig Meinungen der Testerinnen und Tester wie sie durch „lautes Denken“ erhoben werden. Dazu zählen Beobachtungen der Testpersonen, entweder durch die Moderatorin beziehungsweise den Moderator oder automatisiert im Hintergrund (z. B. durch Screencapturing). Weitere qualitative Techniken sind „Mouse-Tracking“ (Arroyo et al., 2006; Atterer und Schmidt, 2007; Atterer, Wnuk et al., 2006; Fujioka et al., 2007) oder „Eye-Tracking“ (Schmidts, 2007). Sie werden auch unter dem Begriff „Subjective measures“ (Dumas und Redish, 1999, 187f) zusammengefasst.¹⁸

Zuletzt sollte in dem Testplan erwähnt werden, wie die erzielten Ergebnisse der Nutzertests ausgewertet werden. In der Regel erfolgt dies über einen Bericht. In dem Testplan sollte grob aufgeführt werden, wie der Bericht aufgebaut und welche inhaltlichen Teile beziehungsweise Ergebnisse er enthalten wird. Erfolgt der Nutzertest innerhalb eines Entwicklungsteams, kann es ausreichend sein, die Ergebnisse weniger formal (beispielsweise im Rahmen einer Präsentation) festzuhalten und weiterzugeben.

(2) Testumgebung vorbereiten¹⁹

Nach dem Erstellen des Testplanes sollte die Testumgebung eingerichtet und vorbereitet werden. Wo und wie die Nutzertests durchgeführt werden sollen, ist zu diesem Zeitpunkt bereits durch den Testplan festgelegt. Als wichtigstes Merkmal ist zu nennen, an welchem Ort getestet werden soll. Durchgesetzt haben sich vor allem drei Orte beziehungsweise Formen der Testumgebung: *Nutzertests in einem Labor*, *Nutzertests vor Ort* mittels eines mobilen Labors und *entfernte Nutzertests* (per Telefon oder Internet; Lazar et al., 2010, S. 264). Alle drei Methoden haben Vor- und Nachteile. Wichtig ist, dass die Orte den Forschungsfragen beziehungsweise den Gründen und Zielen des Nutzertests folgen (Lazar et al., 2010, S. 264).

Nutzertests in einem Labor sind die klassische Form des Nutzertests. Bei einem Labor handelt es sich um zwei Räume, die über einen Einwegspiegel miteinander verbunden sind. In einem Raum findet der Nutzertest statt. In diesem befindet sich eine Büroeinrichtung mit Computer, an welchem die Testerinnen und Tester sitzen und den Nutzertest durchführen. Rubin und Chisnell (2008, S. 95) empfehlen, den Raum so realistisch wie möglich zu gestalten. Neben dem Arbeitsplatz kann weiteres Equipment wie Kameras und Mikrofone aufgebaut werden. In dem zweiten Raum befinden sich die Testleiterinnen

¹⁸Rubin und Chisnell (2008, 88ff.) bieten einen Überblick über verschiedene qualitative und quantitative Daten.

¹⁹Engl.: Set up the test environment (vgl. Tabelle 6.2)

und Testleiter beziehungsweise Moderatorinnen und Moderatoren und beobachten die Testerin beziehungsweise den Tester. In dem Raum können weitere Monitore aufgestellt werden, die parallel die Mimik der Testperson anzeigen. Über Lautsprecher kann die Testleiterin beziehungsweise der Testleiter oder die Moderatorin beziehungsweise der Moderator mit der Testerin beziehungsweise dem Tester kommunizieren. Je nach Variation des Laboraufbaus kann sich die Moderatorin beziehungsweise der Moderator im Testraum befinden. Vorteilhaft an der Laborumgebung ist, dass sich äußere Einflüsse in der experimentellen Situation kontrollieren lassen (Schnell et al., 2008; Stoessel, 2002). Störende Einflüsse an anderen Orten wie am Arbeitsplatz oder Zuhause lassen sich eliminieren. Der Testleiter hat zudem einen guten Überblick und die Kontrolle über die gesamte Testsituation. Der Auf- und Abbau des Equipments entfällt. Als Nachteil steht dem gegenüber, dass die für die Testperson ungewohnte Situation das Ergebnis verzerren kann (Dumas und Fox, 2008, S. 1136). Dies kann unter Umständen den positiven Effekt der kontrollierten Situation aufheben. Früher wurden die hohen Kosten für die Einrichtung eines Labors als Nachteil gesehen. Dies gilt heute nicht mehr (Barnum, 2011, 34f). Einfache Labore lassen sich bereits für unter 10.000\$ einrichten.

Der Preisverfall für Laptops in Verbindung mit guter Audio- und Videoleistung von beispielsweise Webcams und leistungsstärkeren Computern machen die Verwendung eines *mobilen Labors* (oder *Onsite Testing*) attraktiv. Der kleine Formfaktor bedingt, dass weder Transport noch Auf- und Abbau des Equipments ein Problem darstellen. Der größte Vorteil dieser Methode ist, dass die Nutzertests in der gewohnten Umgebung der Testerrinnen und Tester stattfinden. Zudem müssen diese nicht im Labor erscheinen, sondern werden vor Ort aufgesucht. Für die Testleiter hat dies den Vorteil, dass weitere Eindrücke über die Arbeits- beziehungsweise Wohnumgebung der Testpersonen gesammelt werden können (Dumas und Fox, 2008, S. 1137). Mobile Labore sind kostengünstiger als fest installierte Labore und die Wahrscheinlichkeit, dass Testpersonen rekrutiert werden können, ist höher (Rubin und Chisnell, 2008, S. 100). Der größte Nachteil ist die fehlende Kontrolle der Testsituation hinsichtlich äußerer Einflüsse. Es ist nicht sichergestellt, dass die Nutzertests an einem ruhigen Ort oder zum optimalen Zeitpunkt stattfinden, da die Testleiter die Örtlichkeiten in der Regel erst am Testtag betreten. Besteht das Testteam aus mehreren Personen, ist nicht gesichert, dass alle Personen vor Ort genügend Platz finden oder sich im Hintergrund halten können. Gleichzeitig steigt der organisatorische Aufwand, alle Nutzertests zu koordinieren, insbesondere wenn die Orte der Nutzertests geografisch auseinander liegen. Nutzertests vor Ort benötigen zudem mehr Zeit als Tests unter Laborbedingungen (Dumas und Fox, 2008, 1137f; Rubin und Chisnell, 2008, 100f).

Die dritte Möglichkeit ist das *entfernte Testen* (engl. Remote Testing). Dabei handelt es sich um ein neues Testverfahren. Die Testleiterin beziehungsweise der Testleiter und die Testperson sind räumlich (und optional zeitlich) voneinander getrennt (Paterno, 2009). Die Kommunikation erfolgt hauptsächlich über das Internet. Durch die zunehmende Verbreitung von Internetanschlüssen und Breitbandinternet kann diese Methode an mehr Standorten durchgeführt werden. Diese bieten neben der Möglichkeit, Testerinnen und Tester unabhängig von ihrem Standort zu rekrutieren, die Vorteile, dass mit einer höheren Teilnahmebereitschaft gerechnet werden kann und der Möglichkeit, die Nutzertests

an einem gewohnten Ort (Büro oder Zuhause) durchzuführen (Dumas und Fox, 2008, S. 1137). Für die Testleiterin beziehungsweise den Testleiter ergibt entferntes Testen eine Kostenersparnis, da kein Equipment benötigt wird und organisatorische Tätigkeiten zu einem großen Teil entfallen (Freiberg und Baumeister, 2008, S. 5). Die Testpersonen führen die Nutzertests mittels ihrer eigenen Hardware durch. Technische Probleme sind nicht ausgeschlossen und können nicht von der Testleiterin oder dem Testleiter gelöst werden. Gleiches trifft auf weitere technische Probleme wie Firewalls oder Softwarefehler zu. Durch die räumliche Distanz kann keine Kontrolle durch die Testleitung erfolgen (Barnum, 2011, S. 45). Stone et al. (2005, 475f) empfehlen daher den Einsatz spezieller Tools zum Screencapturing oder Protokollieren der Testerin beziehungsweise des Testers. Es wird zwischen *synchronem* (oder *moderiertem*) und *asynchronem* (oder *unmoderiertem*) entferntem Testen unterschieden. Bei einer *synchronen Testsituation* sind die Testleiterin beziehungsweise der Testleiter und die Testperson direkt miteinander verbunden (Dumas und Fox, 2008, S. 1137). Optional kann die Kommunikation durch eine Webcam unterstützt werden. In einer *asynchronen Testsituation* führt die Testerin beziehungsweise der Tester die vorgegebenen Aufgaben ohne weitere Unterstützung aus. Der Vorteil der asynchronen Testsituation ist, dass auf Seiten der Testleiterin beziehungsweise des Testleiters keine Ressourcen gebunden werden. Zudem können mehrere Testpersonen gleichzeitig einen Nutzertest durchführen. Tullis, Fleischman et al. (2002) konnten mittels dieser Methode in einem kurzen Zeitraum 88 Nutzertests durchführen. Gleichzeitig kann auf Crowdsourcing Anwendungen wie „Amazon Mechanical Turk“²⁰ zurückgegriffen werden (Kittur et al., 2008). Die mögliche Hilfestellung durch die Testleiterin beziehungsweise den Testleiter ist eingeschränkt (Barnum, 2011, S. 45). Daher werden vor allem synchrone entfernte Tests durchgeführt.

Die drei vorgestellten Testumgebungen sind variabel. Insbesondere der Aufbau eines Usability-Labors kann auf unterschiedliche Arten erfolgen (vgl. u. a. Rubin und Chisnell, 2008). Entfernte Nutzertests sind am einfachsten zu beschreiben, da Vor- und Nachteile am deutlichsten erkennbar sind. Gleichzeitig wird die Testumgebung von einigen Autoren (u. a. Freiberg und Baumeister, 2008) eher als Technik denn als Umgebung verstanden. Bei allen drei Testumgebungen hat sich bisher noch kein allgemein gültiger Quasi-Standard in der Literatur herausgebildet.

(3) Testerinnen und Tester rekrutieren²¹

In dem nächsten Schritt werden die Testerinnen und Tester rekrutiert. Dieser Schritt überschneidet sich eventuell mit der Einrichtung der Testumgebung. Je nachdem wo die Tests durchgeführt werden sollen, kommt die Testleiterin beziehungsweise der Testleiter vorher mit den Testpersonen in Kontakt, insbesondere bei mobilen Nutzertests. Die Teilnehmerinnen und Teilnehmer können über verschiedene Wege gewonnen werden (Barnum, 2011, 158ff. Dumas und Redish, 1999, 141ff. Rubin und Chisnell, 2008, 132ff.). Die einfachste und kostengünstigste Methode ist die Rekrutierung von eigenen Mitarbeiterinnen und Mitarbeitern. Diese sollten nicht in die Entwicklung und Vermarktung der Anwendung

²⁰<https://www.mturk.com/mturk/welcome> [Aufgerufen am 02.06.2015]

²¹Engl.: Find and select test participants (vgl. Tabelle 6.2)

involviert sein. Zudem kann die Loyalität zum Arbeitgeber zu Antworten im Sinne der sozialen Erwünschtheit führen (Schnell et al., 2008, 355f). Die Mitarbeiterinnen und Mitarbeiter des Unternehmens sind darüber hinaus häufig nicht die Zielgruppe der Anwendung. Daher ist diese Quelle an Testpersonen eher zu vermeiden.²² Bei der Rekrutierung der Testerinnen und Tester kann auf den Testplan zurückgegriffen werden. Die in diesem definierten Profile bilden die Grundlage für die Auswahl der Testpersonen. Da nicht alle potenziellen Testerinnen und Tester in das ausgewählte Profil passen, empfiehlt sich ein „Screening“ (Rubin und Chisnell, 2008, S. 126). Das geeignete Mittel zur Vorauswahl stellt ein *Screening Fragebogen* dar (Barnum, 2011, 124ff. Rubin und Chisnell, 2008, 126ff.). In diesem Fragebogen werden die relevanten Merkmale des definierten Profils abgefragt (z. B. Alter, Geschlecht, Einkommen, Vorwissen über die Anwendung). Es bietet sich an, nach weiteren Informationen wie der Vorerfahrung im Bereich der Gebrauchstauglichkeit oder Nutzertests zu fragen. Dies erlaubt eine genauere Einschätzung der Eignung der Testerin oder des Testers. Je größer der zeitliche Vorlauf des Nutzertests ist, desto gründlicher kann die Vorauswahl der Testpersonen erfolgen. Separate Anschreiben oder Nachfassaktionen können die Teilnahme positiv beeinflussen (Dillman, 2000). Sollen die Testpersonen außerhalb des eigenen Unternehmens gewonnen werden, kann die Rekrutierung durch eine Agentur erfolgen. Diese übernimmt das Screening, was eine weitere Zeitersparnis bedeutet. Spezialisierte Agenturen verfügen über einen Pool an Testpersonen, die sich zeitnah aktivieren lassen. Allerdings ist diese Methode kostenintensiv. Eine weitere Quelle ist die Nutzung von beruflichen und privaten Netzwerken („Kennst du jemanden?“). Diese Methode hat den Vorteil, dass wenig zusätzliche Kosten entstehen. Bei beruflichen Netzwerken kann unter Umständen auf Kundenlisten zurückgegriffen werden. Bei Geschäftspartnern und Kunden lässt sich die bisherige Vorerfahrung relativ gut einschätzen. Bei privaten Kontakten (u. a. auch über Vereine etc.) kann ebenfalls die Eignung im Vorhinein besser abgeschätzt werden. Eine weitere Möglichkeit ist, eine Selbstselektion der Testerinnen und Tester durchzuführen. Dies kann über Werbung oder einen Banner auf der eigenen Webseite erfolgen. Durch Crowdsourcing (Franke, 2011; Hammon und Hippner, 2012) können weitere Testpersonen aktiviert werden. Sofern absehbar ist, dass in der Zukunft weitere Nutzertests durchgeführt werden, sollte über den Aufbau einer eigenen Personendatenbank nachgedacht werden (Dumas und Redish, 1999, S. 141). Die Verwendung von Selbstselektion sollte streng überwacht werden, da eventuell nur bestimmte Teile der Zielgruppe bereit sind, teilzunehmen („Coverage“-Problem Couper, 2000; Faulbaum et al., 2012). Die Teilnahmebereitschaft kann durch ein kleines Dankeschön (engl. *Incentive*) erhöht werden (Dumas und Redish, 1999, 147f; Rubin und Chisnell, 2008, S. 150). In der Umfrageforschung (Dillman, 2000; Groves, Fowler et al., 2009) konnte gezeigt werden, dass Incentives die Teilnahmebereitschaft signifikant steigern können. Zuletzt sollte die genaue Anzahl an benötigten Testpersonen festgelegt werden. Diese hängt von dem Versuchsaufbau ab. Sollen Tests mit mehreren Testgruppen oder vergleichende Nutzertests mit verschiedenen Interfaces durchgeführt werden, ist mit mehr Testpersonen zu planen.

Die genaue Anzahl der benötigten Testpersonen wird in der Literatur bis zum heutigen

²²Sie lässt sich allerdings gut für Pretests und formative Tests an Mockups oder Prototypen verwenden.

Zeitpunkt kontrovers diskutiert (Barnum et al., 2003; Freiberg und Baumeister, 2008, S. 6). Anfang der 1990er-Jahre kamen mehrere Studien zu dem Ergebnis, dass etwa fünf Testpersonen ausreichen um etwa 80 Prozent aller Usability-Probleme zu finden (Nielsen, 1994b; Nielsen und Landauer, 1993; Virzi, 1990, 1992). Virzi (1992) verweist darauf, dass schwere Probleme eher mit weniger Testerinnen und Testern (in diesem Fall fünf) gefunden werden. Diese Aussage konnte mehrfach bestätigt werden (u. a. Law und Vanderheiden, 2000). Tullis und Albert (2008) schlagen vor, pro Profil mindestens fünf Testpersonen zu rekrutieren. Diese Annahme wird von anderen Forscherinnen und Forschern zum Teil stark kritisiert und als „HCI Weisheit“ bezeichnet (Lazar et al., 2010, S. 263). Sowohl Faulkner (2003), Lewis (1994), Lewis (2001), Lindgaard und Chattratichart (2007) und Spool und Schroeder (2001) konnten zeigen, dass fünf Testpersonen in einem Nutzertest weniger als 80 Prozent der Probleme finden. Fast alle Forscherinnen und Forscher arbeiten mit Formeln (Faulkner, 2003; Lewis, 2006; Nielsen und Landauer, 1993; Virzi, 1992), die unter Berücksichtigung der Testpersonen und der Häufigkeit des Auftretens beziehungsweise des Entdeckens von Problemen die Anzahl der gefundenen Probleme berechnen. Insgesamt sind diese Formeln höchstens als Ansatzpunkt sinnvoll, da die endgültige Effektivität und Effizienz der Nutzertests stark von der Durchführung der Nutzertests und der Testpersonen abhängt (Lewis, 2006). Die genaue benötigte Zahl ist nach wie vor unklar (Faulkner, 2003). Die Bestimmung der Zahl hängt auch vom Umfang des Nutzertests und der Komplexität der zu untersuchenden Anwendung ab (Nielsen und Landauer, 1993).

(4) Nutzertests vorbereiten²³

Vor der Durchführung der Nutzertests, müssen diese vorbereitet werden. Dies dient der Testleiterin beziehungsweise dem Testleiter zur Orientierung während der Tests. Zum anderen werden in diesem Schritt alle benötigten Materialien für die Testpersonen vorbereitet. Grob lässt sich der Vorgang in das *Orientierungsskript*, *Fragebögen*, *rechtliche Aspekte*, *Datenerfassung* und *Szenarien* unterteilen (Rubin und Chisnell, 2008, 153f). Das *Orientierungsskript* wird für die Moderatorin beziehungsweise den Moderator entworfen. Darin sollte auf die folgenden Punkte eingegangen werden (Barnum, 2011, 167ff.): Begrüßung der Testpersonen, Begründungen und Hintergründe des Nutzertests, Datenschutz- und andere Erklärungen, Erklärung des Testablaufs und der Testsituation, Beschreibung der verwendeten Methoden und die Erwähnung der Möglichkeit, jederzeit Fragen zu stellen. Darüber hinaus kann das Skript Ausführungen zu den Erwartungen an die Testerin beziehungsweise den Tester enthalten (Rubin und Chisnell, 2008, 160f). Längere Sprechpassagen der Moderatorin beziehungsweise des Moderators sollten vorformuliert werden (Rubin und Chisnell, 2008, S. 158). Optional kann eine Checkliste für die Moderatorinnen und Moderatoren erstellt werden (vgl. Barnum, 2011, S. 165). Sofern den Testpersonen Fragebögen vorgelegt werden, müssen diese vorbereitet und in ein einheitliches Layout gebracht werden. Es wird zwischen drei verschiedenen Fragebögen unterschieden: *Pre-test* Fragebogen, *Post-task* Fragebogen, *Post-test* Fragebogen (Barnum, 2011, 173ff. Dumas und Redish, 1999, 209ff). Der *Pre-test* Fragebogen kann ein Screening durchführen, sofern

²³Engl.: Prepare test materials (vgl. Tabelle 6.2)

nicht bereits passiert und zum anderen weitere Hintergrundinformationen abfragen, die zur Interpretation der Moderatorin beziehungsweise des Moderators während des Tests beitragen.²⁴ Der *Post-task Fragebogen* wird der Testperson direkt nach Beendigung einer Aufgabe während des Nutzertests vorgelegt. Dies macht Sinn, wenn direkt nach einer Aufgabe das Feedback einer Testerin beziehungsweise eines Testers eingeholt werden soll. Zudem können mittels Post-task Fragebögen die Veränderung der Meinungen und Wahrnehmungen während des Nutzertests erfasst werden. Diese Fragebögen sollten kurz gehalten sein (max. eine Seite mit vier bis fünf Fragen; Dumas und Redish, 1999, S. 211). Nach Abschluss des Nutzertests können die Testerinnen und Tester aufgefordert werden, einen *Post-test Fragebogen* auszufüllen. Dieser soll den Gesamteindruck der Anwendung wiedergeben. Es kann auf existierende Fragebögen wie die System Usability Scale (SUS, vgl. Kapitel 6.2.2) zurückgegriffen werden. Darüber hinaus sollten persönliche Informationen wie demografische Merkmale etc. abgefragt werden. Vor Beginn der Nutzertests sind rechtliche Aspekte wie der Datenschutz oder die Einwilligung zur Datenerfassung des Nutzertests zu beachten. Je nach Wichtigkeit kann eine Verschwiegenheitserklärung vorbereitet werden. Bei der Zustimmung zur Datenerfassung sollte explizit aufgeführt werden, welche Verfahren angewendet werden. Darunter fallen qualitative und quantitative Techniken (vgl. Kapitel 6.3.2). Werden Daten automatisiert im Hintergrund erfasst (durch „Data Loggers“; Rubin und Chisnell, 2008, 168f), müssen teilweise zusätzliche Programme installiert, eingerichtet und konfiguriert werden. Gleiches gilt für Audio- und Videoaufnahmen. Zuletzt sollten für die Durchführung des Tests Szenarien aus den Profilen und Aufgaben entwickelt werden (Barnum, 2011, 99ff; Dumas und Redish, 1999, 171ff; Rubin und Chisnell, 2008, 182ff). Die Testerinnen und Tester sollten statt der vorher entwickelten Aufgabenliste ein Szenario zu einer Aufgabe vorgelegt bekommen. Das Szenario setzt die Aufgabe in einen für die Testpersonen verständlichen Kontext und versucht gleichzeitig die Testsituation in den Hintergrund zu rücken, indem das Szenario eine für die Testerin oder den Tester alltägliche Situation beschreibt. Das Szenario sollte kurz und für alle Testpersonen verständlich formuliert werden. Gleichzeitig sollte die Formulierung genügend Informationen enthalten, damit die Testerinnen und Tester die Aufgabe erfolgreich ausführen können und das Ziel der Aufgabe in das Zentrum des Szenarios rückt. Zudem sollte eine direkte Verbindung zu den Aufgaben und Forschungsfragen herstellbar sein (Carroll, 2003).

(5) Nutzertests durchführen & (6) Debriefing²⁵

Nach den Vorbereitungen wird der Nutzertest selber durchgeführt. Es sollte über einen Pre-Test (oder auch „pilot test“ (Rubin und Chisnell, 2008, S. 215)) nachgedacht werden, der etwa eine Woche (Rubin und Chisnell, 2008, S. 214) bis zwei Tage (Dumas und Redish, 1999, S. 264) vor den Nutzertests erfolgt. Dadurch werden neben dem Vorgehen auch die eingesetzten Techniken, die formulierten Szenarien, eventuelle Fragebögen und die veranschlagte Zeit überprüft. Gleichzeitig können vorhandene Checklisten getestet werden. Für die Pre-Tests sind Testpersonen aus der Zielgruppe nicht zwingend erforderlich.

²⁴Sind diese Informationen nicht für den Test wichtig, können sie auch in einem Post-test Fragebogen abgefragt werden.

²⁵Engl.: Conduct the test sessions & Debrief the participants (vgl. Tabelle 6.2)

Der Fokus liegt auf dem Ablauf des Tests und nicht auf dem Inhalt. Die wichtigste Rolle während des Nutzertests nimmt neben der Testperson die Moderatorin beziehungsweise der Moderator ein. Eine „gute“ Moderatorin beziehungsweise ein „guter“ Moderator hat einen direkten Einfluss auf das Ergebnis der Nutzertests, da dies die Person ist, die die Testerinnen und Tester den gesamten Nutzertest begleitet, beginnend bei der Begrüßung und endend bei dem Debriefing. Wichtig ist, dass die Moderatorinnen und Moderatoren einen Leitfaden für ihr weiteres Vorgehen erhalten. Darauf kann verzichtet werden, wenn Testleitung und Moderation ein und dieselbe Person sind. In dem Leitfaden sollten Verhaltensweisen für den Umgang mit den Testerinnen und Testern festgehalten werden. Zudem ist die Rolle der Moderatorin beziehungsweise des Moderators in problematischen Situationen zu definieren, insbesondere ab wann die Moderatorin beziehungsweise der Moderator bei Problemen der Testpersonen eingreift. Ein sofortiges Eingreifen ist in der Regel nicht erwünscht, da diese Situationen zu einem tieferen Verständnis der darunterliegenden Probleme führen (Rubin und Chisnell, 2008, 209f). Hilfestellungen sollten nur in Ausnahmefällen gegeben werden, um Einflüsse auf das Testergebnis zu vermeiden (Lazar et al., 2010, S. 273): „As soon as you assist, you are affecting the test results in a major way.“ (Rubin und Chisnell, 2008, S. 211) Eingriffe sollten nur erfolgen, wenn die Testerin beziehungsweise der Tester nicht weiterkommt oder sich in der Anwendung verirrt hat. Dabei sollten ausschließlich kleine Hilfestellungen und Tipps gegeben werden. Sollte die Testerin beziehungsweise der Tester Anzeichen zeigen, dass sie sich unwohl fühlt beziehungsweise zunehmend frustriert ist, sollte von Seiten der Moderation eingegriffen werden. Dies geschieht, um einen Abbruch durch die Testperson zu vermeiden. Handelt es sich um einen formativen Nutzertest mit einer unvollständigen Anwendung, ist die Moderatorin beziehungsweise der Moderator unter Umständen zum Eingreifen gezwungen, wenn Teile der Anwendung noch nicht voll funktionsfähig sind. Außerdem sollte auf jeden Fall ein Eingreifen erfolgen, wenn ein Fehler („Bug“) durch die Testerin oder den Tester aufgedeckt wird.²⁶ Es bietet sich an, Checklisten für die Moderatorinnen und Moderatoren und die Testleitung zu erstellen.²⁷ Nach Beendigung des Nutzertests erfolgt ein Debriefing mit den Testerinnen und Testern (Rubin und Chisnell, 2008, 229ff.). Dadurch erhält die Testleiterin beziehungsweise der Testleiter die Möglichkeit, weitere Informationen von den Testpersonen zu erhalten. Dies ist bei größeren Problemen während des Nutzertests oder bei Abbrüchen wichtig. Es kann genauer eruiert werden, warum die Testpersonen Probleme hatten. Gleichzeitig kann den Testerinnen und Testern die Möglichkeit eingeräumt werden, Fragen zu der Anwendung zu stellen. Auf der anderen Seite kann die Moderatorin beziehungsweise der Moderator weitere Informationen zu den Nutzertests wie Implikationen, Vorstellungen und Forschungsfragen geben. Diese Informationen führen vor dem Nutzertest eventuell zu einer Beeinflussung der Testpersonen. Das Debriefing findet nach dem Ausfüllen des Post-test Fragebogens statt und sollte von der Moderatorin beziehungsweise dem Moderator protokolliert werden. Es ist auch

²⁶Dumas und Loring (2008) beschreiben die Rolle der Moderatorin beziehungsweise des Moderators ausführlich.

²⁷Beispiele finden sich bei Barnum (2011, 163ff.), Dumas und Redish (1999, 250ff.) und Rubin und Chisnell (2008, 214ff.).

möglich, die nachfolgende Diskussion per Audio oder Video festzuhalten.

(7) Auswertung²⁸

Nach der Erhebung folgt die Auswertung der erhobenen Daten. Quantitative und qualitative Daten werden getrennt betrachtet. Während der Analysephase werden die Probleme genauer analysiert und benannt. Abschließend müssen die gefundenen Probleme nach ihrer Schwere („Severity Ratings“ (Barnum, 2011, 261ff.)) und Reichweite („Scope Ratings“) bewertet werden (Barnum, 2011, 261f.). Die Auswertung von quantitativen Daten ist weniger zeitintensiv. Dies liegt daran, dass die Daten bereits in digitaler Form vorliegen (bei Data Logging oder Online-Fragebögen) oder leicht in ein digitales Format überführt werden können. Die Auswertungen von „Performance Data“ (die Anzahl der Klicks, die benötigte Zeit oder die Häufigkeit von Fehlern) findet häufig über Randauszählungen oder Mittelwerte statt (Tullis und Albert, 2008). Diese werden teilweise als „Task Accuracy“ und „Task Timings“ bezeichnet (Rubin und Chisnell, 2008, 249ff.). Bei der Auswertung von Fragebögen kann ähnlich verfahren werden. Es empfiehlt sich das Heranziehen der einschlägigen Literatur der empirischen Sozialforschung (De Leeuw et al., 2009; Diekmann, 2010; Groves, Fowler et al., 2009; Schnell et al., 2008). Es kann neben Randauszählungen mit der Inferenzstatistik (Bortz und Döring, 2006) gearbeitet werden. Nutzertests werden häufig von einer geringen Anzahl an Testpersonen durchgeführt (weniger als 15-20), was besondere Anforderungen an die Qualität der Daten stellt. Ausreißer sollten genau betrachtet werden, da sie zum einen die Daten verfälschen können, zum anderen besonders interessant sind (Dumas und Redish, 1999, 313f). Qualitative Daten müssen anders aufgearbeitet werden, bevor sie ausgewertet werden können. Protokolle von Moderatorinnen beziehungsweise Moderatoren, Audiomitschnitte des lauten Denkens und Videobeziehungsweise Screencapturingaufnahmen müssen analysiert werden. Barnum (2011, 242f.) unterscheidet zwischen „Top-down“ und „Bottom-up“ Methoden. Beide Methoden ähneln der Inhaltsanalyse (Gläser und Laudel, 2010; Mayring, 2010) in der qualitativen Sozialforschung (Bohnsack, 2003; Flick, 2007), allerdings sind diese vereinfacht. Bei der Top-down-Methode werden zunächst Kategorien gebildet und entsprechende Text- beziehungsweise Videostellen diesen Kategorien zugeordnet. Der Vorteil dieser Methode ist, dass sie weniger zeitintensiv ist. Es muss fundiertes Vorwissen über die Anwendung und über vermeintliche Nutzerprobleme bestehen. Zudem sind fälschliche Zuordnungen nicht ausgeschlossen. Die Bottom-up-Methode folgt den umgekehrten Weg. Zunächst werden relevante Stellen identifiziert und aus diesen wiederum Kategorien gebildet. Der Vorteil dieser Methode ist die Freiheit bei der Analyse der Daten. So können unvorhergesehene Ereignisse und Elemente mit in die weitere Analyse aufgenommen werden. Nachteilig ist der höhere Zeitaufwand. Die Zusammenfassung der Einzelbeobachtungen ist zudem nicht trivial.

Sind die Probleme aus den Rohdaten herausgearbeitet, müssen sie nach ihrer Schwere und Reichweite bewertet werden. Bei der Reichweite wird zwischen lokalen und globalen Problemen unterschieden (Barnum, 2011, 261ff. Dumas und Redish, 1999, 322ff.). Lokale Probleme treten nur auf einer bestimmten Seite oder Einblendung auf. Ähnliche oder eng

²⁸Engl.: Analyze data and observations (vgl. Tabelle 6.2)

verwandte Seiten beziehungsweise Funktionen sind nicht betroffen. Ein Beispiel ist ein fälschlicher Verweis auf einer Hilfeseite, der auf allen anderen Hilfeseiten funktioniert. Globale Probleme sind häufiger in einer Anwendung zu finden. Darunter fallen nach Barnum (2011, S. 262) drei Kategorien: *Navigation/Informationsdarstellung*, *Konsistenz* und *Feedback*. Eine schlechte Navigation oder ungewöhnliche Informationsdarstellung führt dazu, dass Nutzerinnen und Nutzer Schaltflächen und Informationen an anderer Stelle vermuten als sie dargestellt werden. Wichtig ist daneben eine durchgehende Konsistenz des Inhalts. Damit ist gemeint, dass Seiten mit derselben Funktion über ein gleiches Aussehen verfügen sollten. Gleiches gilt für stilistische Merkmale wie beispielsweise Überschriften oder andere Schriftbilder. Fehlendes Feedback ist ebenfalls ein größeres Problem, beispielsweise wenn die Nutzerinnen und Nutzer keine Rückmeldung über eine erfolgreiche oder fehlerhafte Eingabe erhalten. Globale Probleme sind schwerwiegender und sollten primär behandelt werden. Neben der Reichweite ist die Schwere der Probleme entscheidend. Hier wird häufig auf eine vierstufige Skala zurückgegriffen (vgl. Tabelle 6.3). Die Einstufung eines Problems in eine Kategorie kann nach der Beurteilung der Testleitung, nach der Hartnäckigkeit (engl. „Persistence“; Barnum, 2011, S. 264) oder der Häufigkeit des Auftretens eines Problems erfolgen (Barnum, 2011, 263f). Rubin und Chisnell (2008, 261f) schlagen vor, die Gesamtschwere eines Problems aus der Schwere und der Häufigkeit des Auftretens zu bestimmen.

Tabelle 6.3: Skalen zur Schwere von Usability-Problemen

Barnum (2011, S. 264)	Dumas und Redish (1999, 324f)
1. catastrophe	Level 1: problems prevent completion of a task
2. major problem	Level 2: problems create significant delay and frustration
3. minor problem	Level 3: problems have a minor effect on usability
4. cosmetic problem	Level 4: problems are more subtle
<hr/>	
Rubin und Chisnell (2008, S. 262)	
4. Unusable	
3. Severe	
2. Moderate	
1. Irritant	

Bei der Verwendung mehrerer Methoden bietet es sich an, alle Daten zu nutzen und eine Methodentriangulation durchzuführen. Beispielsweise kann ein Fragebogen dazu genutzt werden, Testerinnen und Tester einzuschätzen und dieser Basis die Schwere der gefundenen Probleme jeder Testerin beziehungsweise jeden Testers zu gewichten.

(8) Berichtlegung und Empfehlungen²⁹

Abschließend sollten die Ergebnisse des Nutzertests in einem schriftlichen Bericht zusammengetragen werden. Der Aufbau des Berichts orientiert sich am Testplan. Viele der dort aufgeführten Inhalte können ausformuliert wieder aufgenommen werden. Der Aufbau sollte die folgenden Punkte beinhalten: *Zusammenfassung*, *Methodologie*, *Ergebnisse*, *Funde und Empfehlungen* und *Anhänge* (Rubin und Chisnell, 2008, 273ff). Teilweise werden Forschungsfragen und Ziele des Tests als separate Punkte geführt (Barnum, 2011, 283f).

²⁹Engl.: Report findings and recommendations (vgl. Tabelle 6.2)

Daneben gibt es Empfehlungen den Verlauf und die Vorbereitung der Tests genauer zu beschreiben (Dumas und Redish, 1999, S. 349). Rubin und Chisnell (2008, S. 274) führen diese Teile mit in der *Zusammenfassung* auf. Einige Teile wie die Aufgabenliste oder die Nutzerprofile sollten erwähnt und in den Anhang eingefügt werden. Zudem sollten verwendete Fragebögen und Einverständniserklärungen (Datenschutz etc.) im Anhang angefügt werden. Der Teil der Ergebnisse bietet zunächst einen groben Überblick. Hier sollten quantitative Auswertungen von Data Loggern und Fragebögen aufgeführt werden. In den *Funden und Empfehlungen* sollten detaillierte Beschreibungen der gefundenen Probleme erfolgen. Es bietet sich an, die Probleme mittels Screenshots, Tabellen oder Abbildungen zu visualisieren (Barnum, 2011, 290ff.). Je nach Datenlage kann das Kapitel in generelle und spezifische Probleme unterteilt werden. Den generellen Problemen sollten dabei globale Probleme zugeordnet werden, den spezifischen Problemen wiederum lokale Probleme (Rubin und Chisnell, 2008, S. 276). Zudem sollten diese zusammen mit den Empfehlungen aufgeführt werden. Wichtig ist, den Bericht an die jeweilige Zielgruppe der Ergebnisse und Empfehlungen anzupassen (Dumas und Redish, 1999, 343ff.). Entwickler haben andere Erwartungen an die Inhalte als die Geschäftsleitung. Neben einem Bericht kann es sinnvoll sein, die Ergebnisse in einer Präsentation aufzubereiten.

6.3.2 Techniken zur Aufdeckung von Problemen

Im vorherigen Kapitel wurden bereits vereinzelt mögliche Techniken zur Aufdeckung von Usability-Problemen erwähnt, darunter lautes Denken oder Performance Measurement. Es kann grob zwischen quantitativen und qualitativen Techniken unterschieden werden. Den qualitativen Techniken sind vor allem Verfahren wie das laute Denken (engl. *Think-Aloud*), Eye-Tracking, Mouse-Tracking, Audio- und Videomitschnitte sowie Aufzeichnungen der Moderatorinnen und Moderatoren beziehungsweise Testerinnen und Tester zuzuordnen. Unter quantitative Techniken fallen Kennzahlen wie benötigte Zeit oder Anzahl der Klicks (engl. *Performance Data*), Logfiles und Fragebögen. Die Verwendung des Begriffs „Techniken“ ist in der Literatur nicht eindeutig, da dieser Kategorie sowohl Methoden zur Datenerhebung (lautes Denken) als auch die Methoden selber (Nutzertests, Fragebögen) zugeordnet werden (u. a. bei Dix et al., 2004; Freiberg und Baumeister, 2008; Stone et al., 2005; Stowasser, 2006). Nutzertests und Fragebögen stellen für den Autor allerdings eine Methode dar, während unter Techniken Verfahren zur Erfassung und Messung der Probleme während der Nutzertests subsumiert werden. Die Kombination von beispielsweise Fragebögen, einer Heuristischen Evaluation und eines Nutzertests stellt eine *Methoden-Triangulation* (Flick, 2008) und keine Triangulation von Techniken dar. Eine *Techniken-Triangulation* wäre der Einsatz von lautem Denken, Screencapturing und Logfiles-Analyse.

Lautes Denken („Think-Aloud“)

Die Technik des „lauten Denkens“ (engl. „Think-Aloud“ oder „Thinking Aloud“) ist die ge­läufigste Technik (Dumas und Redish, 1999; Rubin und Chisnell, 2008; Sarodnick und Brau, 2011). Die Testerinnen und Tester werden während des Nutzertests dazu aufgefordert, ihr Vorgehen, ihre Eindrücke, auftretende Probleme und Frustrationen zu artikulieren (Lazar

et al., 2010, S. 270). Die Technik kann in formativen als auch summativen Nutzertests eingesetzt werden. Laut Lazar et al. (2010, S. 270) ist der Einsatz in formativen Nutzertests häufiger. Die Technik stammt ursprünglich aus der Kognitionspsychologie (Simon und Ericsson, 1993). Dort wird sie eingesetzt, um das Kurzzeitgedächtnis der Probanden zu studieren (Dumas und Fox, 2008, S. 1139). Die Übernahme der Methode in die Usability-Forschung geschah relativ unreflektiert. Mehrere Studien (Boren und Ramey, 2000; Dumas, 2001) konnten zeigen, dass dieses „blinde Kopieren“ der Technik dazu geführt hat, dass sich Einsatz und Anwendung der Technik in der Usability-Forschung und der Kognitionspsychologie deutlich voneinander unterscheiden. Von einer vergleichbaren Anwendung kann nicht gesprochen werden (Dumas und Fox, 2008, 1139f).

Die Wahrscheinlichkeit des Auffindens von Problemen wird höher, wenn lautes Denken eingesetzt wird (Dumas und Redish, 1999; Ebling und John, 2000; Virzi et al., 1993). Zudem kann die Technik bewirken, dass die Testpersonen ihr eigenes Handeln deutlicher und expliziter wahrnehmen (Rubin und Chisnell, 2008, S. 204). Gleichzeitig führt die Methode zu einer kognitiven Doppelbelastung der Testerinnen und Tester, da sie sowohl die Aufgabe ausführen als auch ihre Gedankengänge artikulieren müssen (Sarodnick und Brau, 2011, S. 171). Die Methode kann für ungeübte Testpersonen ungewohnt sein, weshalb eine vorherige Einführung angeraten wird (Barnum, 2011, 205ff). Lautes Denken führt häufig zu einer längeren Bearbeitung der Aufgaben (Dumas und Loring, 2008; Kraemer und Ummelen, 2004; van den Haak et al., 2003), insbesondere im mittleren Lebensalter (Olmsted-Hawala et al., 2010). Eine mögliche Folge ist, dass die gleichzeitige Erhebung von Performance Data zu verzerrten Daten führen kann. Teilweise wird dazu geraten, auf ein retrospektives lautes Denken („Retrospective Think Aloud“; Guan et al., 2006) zurückzugreifen, wobei die Verbalisierung nach der Aufgabe stattfindet. Allerdings sind die Ergebnisse bei dieser Technik weniger detailliert und unvollständiger (Bowers und Snyder, 1990). Zudem werden weniger Probleme gefunden (van den Haak et al., 2003).

Physiologische Messmethoden

Während bei der Technik des lauten Denkens die Testerinnen und Tester aufgefordert sind, ihre Eindrücke zu beschreiben, wird bei physiologischen Messtechniken (engl. Physiological Monitoring) die körperliche Resonanz der Testpersonen gemessen.

Weiter verbreitet ist das Eye-Tracking. Dabei werden mittels einer speziellen Brille oder einer am Bildschirm angebrachten Kamera die Bewegung und Fokussierung der Augen auf dem Bildschirm erfasst (Bojko und Stephenson, 2005; Duchowsky, 2007; Goldberg et al., 2002; Russell, 2005; Schmidts, 2007). In der Regel wird parallel der Bildschirminhalt aufgenommen. Beide Quellen werden kombiniert, sodass ein Video entsteht, in welchem die Interaktion mit dem Bildschirm per Maus und die Fokussierung der Augen der Testpersonen ersichtlich wird. Der Vorteil ist, dass die genauen Quellen für eine eventuelle Irritation beziehungsweise für ein Problem erfasst werden können. Der größte Nachteil ist, dass sich Eye-Tracking nicht mit allen Methoden kombinieren lässt. Bojko (2005) konnte zeigen, dass Testpersonen Webseiten anders betrachten, wenn sie neben dem Eye-Tracking die Technik des lauten Denkens anwenden. Die daraus eventuell resultierende Verzerrung der Daten sollte vermieden werden.

Bei den physiologischen Messtechniken wird die emotionale Reaktion der Testerinnen und Tester gemessen (Dix et al., 2004; Lazar et al., 2010; Ward und Marsden, 2003). Die körperliche Reaktion während des Testens kann Aufschlüsse über Probleme bei der Usability geben. Gemessen wird mit Geräten, die auch in der klinischen Forschung verwendet werden (Stern, 2001). Dabei wird unter anderem die elektrodermale Aktivität (Hautleitfähigkeit oder Galvanic Skin Response; Mandryk und Inkpen, 2004; Scheirer et al., 2002), der Blutfluss beziehungsweise Blutdruck (Mandryk und Inkpen, 2004; Scheirer et al., 2002), die Atmung (Mandryk und Inkpen, 2004), die Muskelanspannung (Mandryk und Inkpen, 2004) oder die Hirnaktivität (Lee und Tan, 2006) gemessen (vgl. auch für einen Überblick Dix et al., 2004; Lazar et al., 2010). Die Techniken bieten den Vorteil, dass sie objektiv sind. Gleichzeitig gelten die bereits erwähnten Gütekriterien. Darunter fallen vor allem Störeinflüsse, die die Testerinnen und Tester beispielsweise erschrecken oder zusätzlichem Stress aussetzen. Gleichzeitig kann neben der ungewohnten Situation eines Nutzertests die Verwendung von medizinischen Testgeräten weitere Irritationen bei den Testpersonen auslösen. Die negativen Effekte können wie bei dem Eye-Tracking zu einer Verzerrung der Daten führen, was zu vermeiden ist. Zudem lassen sich die Testgeräte unter Umständen schwierig transportieren, was sich insbesondere beim Onsite-Testing negativ bemerkbar macht.

Performance Measures

Während die bisher genannten Techniken qualitativer Natur waren, ist das Performance Measurement eine quantitative Technik. Es geht darum, bestimmte Kennzahlen zu erfassen beziehungsweise zu messen. Diese werden auch als *Usability Metrics* bezeichnet (Freiberg und Baumeister, 2008, S. 7). Die verschiedenen Kennzahlen reichen von der Messung der Zeit pro Aufgabe, der Anzahl der Klicks, über die „Task Accuracy“ (Anzahl der erfolgreichen Testpersonen) und die „Error Rate“ (Verhältnis von richtigen Aktionen zu fehlerhaften Aktionen) bis hin zu komplexeren Skalen wie die „Single Usability Metric“ (Sauro und Kindlund, 2005).³⁰ Die Daten werden entweder automatisch im Hintergrund durch „Data Logger“ (Barnum, 2011) erfasst, durch die Moderatorin beziehungsweise den Moderator protokolliert oder nach den Tests während der Analyse der Mitschnitte (Audio oder Video) gemessen. Durch die Technik lassen sich die Tests deskriptiv abbilden. Allerdings sagen sie wenig über die tatsächlichen Probleme aus, können aber hilfreich bei deren Identifikation sein. Bei der Nutzung von Data Loggern entfällt die Aufbereitung der Daten. Die Auswertung geht in der Regel schneller, da die Daten bereits in einem maschinenlesbaren Format vorliegen. Allerdings sollte bei einer Kombination mit anderen Techniken darauf geachtet werden, ob diese die Daten verfälschen können (siehe lautes Denken und Eye-Tracking im vorherigen Absatz).

³⁰Kennzahlen finden sich bei Nielsen (1993, 193f), Dumas und Redish (1999, S. 185), Rubin und Chisnell (2008, 166f, 249f), Sarodnick und Brau (2011, 172f), Bevan und Macleod (1994, S. 138). Freiberg und Baumeister (2008) stellen eine Übersicht der gebräuchlichsten Usability Metrics zusammen.

6.4 Bisherige Ansätze zur Verbindung von Logfiles und Usability

Logfiles werden bereits zur Usability-Evaluation eingesetzt. Im Vergleich zu etablierten Methoden wie der Heuristischen Evaluation oder des Nutzertests, die ihren Höhepunkt in der Forschung in den 1990er-Jahren hatten (Dumas und Fox, 2008; Hollingsed und Novick, 2007), steht die Verwendung von Logfiles noch am Anfang. Mitte der 1990er-Jahre wurde damit begonnen, Logfiles zur Analyse der Usability heranzuziehen. Einer der ersten Ansätze für eine automatisierte Analyse stammt von Smith, Smith et al. (1993). Die Autorinnen und Autoren protokollieren Nutzertests im Hintergrund. Zum einen können die Interaktionen dadurch erneut wiedergegeben werden. Gleichzeitig können die Interaktionen durch die Verwendung von *cognitive grammar* (Langacker, 1986) modelliert und sichtbar gemacht werden. Chang und Dillon (1997) protokollieren durch eine selbst entwickelte Testumgebung die Nutzung der Tastatur und der Maus sowie den Bildschirminhalt. Auf Basis der Daten werden Performance Measures berechnet und die Zahl von Fehlern gezählt (unter anderem Dialogfehler). Einen ähnlichen Ansatz verfolgen Svensson et al. (1998), die neben den Logfiles allerdings nur auf die Mausinteraktionen der Teste-rinnen und Tester zugreifen. Darunter fällt auch die Nutzung des „Zurück“-Buttons. Die Daten können anschließend visualisiert und gruppiert (durch Cluster-Analysen) werden. Die Methode identifiziert Probleme der Navigation, des Designs und des subjektiven Feedbacks, hat aber Schwierigkeiten bei weichen Usability-Problemen von Webseiten (Burger und Jackson, 1997). Lecerof und Paternò (1998) protokollieren die Interaktion der Testerinnen und Tester während der Nutzertests (vor allem Mouse-Tracking). Anschließend werden die Interaktionen mit einem vorher definierten „Task Model“ verglichen und automatisiert Probleme abgeleitet. Zudem kann die entwickelte Software Performance Measures ausgeben. Die Anstrengungen sind in diesem Teilgebiet in den vergangenen Jahren aktiv geblieben (Brinck und Hofer, 2002). Au et al. (2008) entwickeln eine Testumgebung, die speziell auf Mobilgeräte zugeschnitten ist. Diese nimmt neben dem Bildschirm alle Eingaben der Testerinnen und Tester auf. Das System ist in der Lage, Probleme in der Navigation zu erkennen. Gleichzeitig kann es die Ergebnisse aufbereiten und darstellen. Fujioka et al. (2007) halten die Mausnutzung (Mouse-Tracking) von Testerinnen und Testern mit vorher festgelegten optimalen Pfaden fest. Sie können zwei hauptsächliche Probleme bestimmen: überflüssige und fehlende Aktionen. Diese Methode zielt vor allem auf die Effektivität von Usability. Harty (2011) nutzt Selenium³¹, eine quelloffene Testumgebung für Webentwicklungen, um Nutzerinteraktionen zu automatisieren. Der Autor findet mit dem Tool vor allem Probleme bei der Tastaturnutzung und im Layout. Die Probleme der Tastaturnutzung tangieren unter anderem die Barrierefreiheit (Lazar et al., 2010). Chi et al. (2003) und Vanderdonck et al. (2004) stützen sich weniger auf das Web Usage Mining, sondern primär auf das Web Content Mining und das Web Structure Mining (vgl. Kapitel 5), um die Usability einer Webseite zu bestimmen. Chi et al. (2003) nutzen den Inhalt und die Struktur von Webseiten, um die „navigability“ (Chi

³¹<http://docs.seleniumhq.org/> [Aufgerufen am 17.07.2015]

et al., 2003, S. 508) zu bestimmen. Vanderdonckt et al. (2004) versuchen, die Evaluation durch Guidelines zu automatisieren. Sie greifen ebenfalls auf die Struktur und den Inhalt zurück. Hervorzuheben sind zudem die Forschungen von *Melody Ivory*. In ihrer Dissertation hat sie versucht, die Usability-Evaluation von Webseiten ausschließlich über Automatismen, darunter auch Logfiles, zu realisieren (Ivory, 2001). Gleichzeitig hat sie mit *Marti Hearst* den bis heute wohl umfangreichsten Überblicksartikel zu automatischen Usability-Evaluationen verfasst (Ivory und Hearst, 2001).

Bei den aufgezählten Artikeln wird vor allem die Protokollierung und die Analyse von Problemen vorgenommen. Die Analyse beschränkt sich häufig auf Maße wie die Anzahl der Verlinkungen oder das „Klick-Verhalten“. Alle bekannten Vorgehensweisen setzen zudem statische Webseiten voraus. Vor dem Hintergrund, dass sich in den letzten 10 Jahren dynamisch programmierte Webseiten durchgesetzt haben, erscheint dieser Ansatz nicht mehr zeitgerecht. Es gibt zwar Ansätze, zusätzliche Informationen dynamischer Seiten zu protokollieren (Atterer und Schmidt, 2007), deutliche Erweiterungen konnten in den vergangenen Jahren aber nicht festgestellt werden. Von daher scheinen viele dieser Methoden auf den ersten Blick nicht ausreichend angepasst zu sein. Alle Anwendungen haben einen weiteren Nachteil: Es wird versucht, allein auf Basis von Logfile-Analysen auf die Usability zu schließen. Die dahinter stehende Annahme, dass sich auf den Logfiles alleine die Probleme identifizieren lassen, kann bisher aber nicht schlüssig bewiesen werden. Daher wird in dieser Dissertation versucht, die Logfile-Analysen mit Nutzertests zu kombinieren.

Ein weiteres Feld sind die *semi-automatisierten* Verfahren. Das *Cognitive Walkthrough for the Web* (Blackmon, Kitajima et al., 2005; Blackmon, Polson et al., 2002; Karahoca und Karahoca, 2009) kombiniert automatische und manuelle Verfahren miteinander. Vor dem eigentlichen Walkthrough wird die Seite mittels einer *Latent Semantic Analysis* eruiert und auf Basis des Ergebnisses das Walkthrough gestaltet. Daneben gibt es ebenfalls Bestrebungen, Logfile-Daten nicht durch reine automatische Verfahren auszuwerten, sondern die Daten vorher zu aggregieren, zu sortieren oder zu filtern. Dieser Datensatz wird anschließend per Hand ausgewertet, entweder qualitativ (Kort und de Poot, 2005) oder quantitativ (Nakamichi et al., 2006).

7 Logfiles als quantitative Datenquelle

Im ersten Teil der empirischen Analyse werden Logfile-Analysen von einem Lern-Management-System (LMS) durchgeführt. Dazu erfolgt zunächst eine Einordnung des verwendeten Systems innerhalb des vorhandenen Spektrums sowie eine Übersicht über die zentralen Funktionen der Anwendung. Die Logfile-Analysen beziehen sich auf einen Zeitraum von etwa 300 Tagen, in denen 133 Nutzerinnen und Nutzer in dem System aktiv waren. Die Ergebnisse fließen in die Entwicklung der Aufgaben der Nutzertests mit ein.

7.1 Digitale Kooperationssysteme¹

Durch die zunehmende Verbreitung von digitalen Medien und der damit einhergehenden zunehmenden Mediatisierung dringen digitale Medien in immer mehr Bereiche ein. Das Aufkommen, die Etablierung und die Nutzung von digitalen Medien in schulischen Organisationskontexten außerhalb der pädagogischen Tätigkeiten sind im Vergleich zu anderen Organisationswelten (z. B. in Wirtschaft und Verwaltung) kaum erforscht. In der US-amerikanisch geprägten Forschung über *Information Systems* (vergleichbar zur deutschen Wirtschaftsinformatik) geht man davon aus, dass die Aneignung von Informations- und Kommunikationssystemen nicht nur die internen Kommunikationsprozesse in Unternehmen verändert, sondern auch zu neuen Formen der inter-organisatorischen Kooperation führt (Miles und Snow, 2003; Sydow, 2010). Eine Wechselwirkung zwischen organisatorischer Entwicklung und der Aneignung und Verbreitung von digitalen Medien zum „Wissensmanagement“ (Alavi und Leidner, 2001) kann daher als gegeben vorausgesetzt werden. Im Zuge des Medienwandels verändern sich die Zugänge zu explizitem Wissen und dessen Bereitstellung durch Informations- und Kommunikationssysteme. In Schulen ist dabei grundsätzlich zu differenzieren zwischen Softwaresystemen, die direkt auf die Unterstützung des Lern- und Lehrprozesses ausgerichtet sind, und Systemen, die der Verwaltung und Organisation oder strategischen (pädagogischen) Managemententscheidungen dienen (vgl. Abbildung 7.1). Gleichzeitig stehen den Lehrkräften weitere Medientechnologien zur Verfügung, um die Kommunikation innerhalb des Kollegiums neu zu definieren und zu strukturieren.

Schulinformationssysteme sind laut Visscher: „[An, Anm.] information system based on one or more computers, consisting of a data bank and one or more computer applications which altogether enable the computer-supported storage, manipulation, retrieval, and distribution of data to support school management.“ (Visscher, 2001, S. 4) Dabei wird nur der Aspekt der Verwaltungsunterstützung herausgestellt. *Verwaltungsinformationssysteme*

¹In diesem Kapitel entstammen Teile aus der Publikation des DFG-Projekts (Welling et al., 2014).

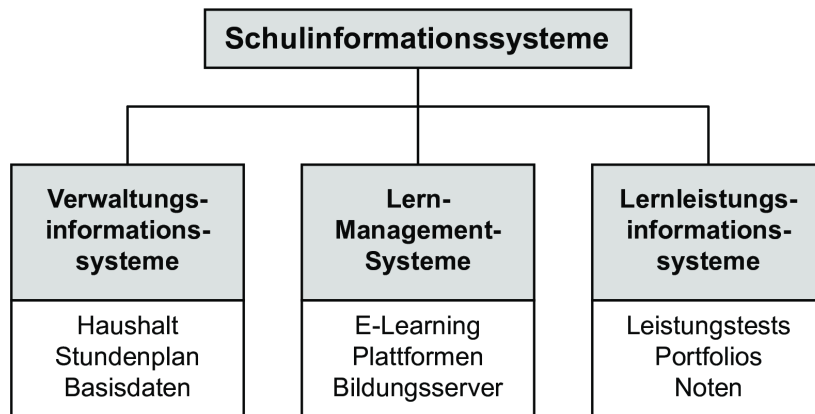


Abbildung 7.1: Typologie von Schulinformationssystemen nach Breiter, Lange et al. (2008)

bilden den gesamten Bereich der Basisdaten ab - von Adressen über Stundenpläne, Raumbelegung bis hin zur Buchhaltung und Finanzplanung. Davon bislang völlig unabhängig sind Lernleistungsinformationssysteme, die Daten über Schülerleistungen verfügbar machen, seien sie aus standardisierten Vergleichstests, aus internen Vergleichsarbeiten oder im Rahmen von Schüler-Portfolios. *Lern-Management-Systeme* (LMS) fokussieren die direkte Unterstützung des Lern- und Lehrprozesses. Dabei werden entweder Lernprozesse individualisiert gesteuert oder Materialien und Kommunikationswerkzeuge zur Verfügung gestellt. Oftmals werden derartige Anwendungen unter E-Learning-Plattformen subsumiert (Kerres, 2001; Schulmeister, 2009). In den vergangenen Jahren haben sich webbasierte Anwendungen durchgesetzt, die unabhängig von Betriebssystemen und installierter Software verwendet werden können und zunehmend mobile Endgeräte (Smartphones, Tablets) unterstützen. Daneben bieten LMS nicht nur Unterstützung für Lernende und Lehrende, sondern sie werden zunehmend Teil des schulischen Wissensmanagements (Breiter, 2002; Friehs, 2003; Maisch, 2006). Im Gegensatz zu Anwendungen aus dem Bereich der Verwaltungsinformationssysteme, die vor allem für Verwaltungsaufgaben gedacht und konzipiert wurden, zielen LMS besonders auf die Kommunikation zwischen Lehrendem und Lernendem ab. Gleichzeitig ermöglichen sie die medienvermittelte Kommunikation zwischen Lehrerinnen und Lehrern, Schülerinnen und Schülern sowie Eltern. Dazu werden neben der direkten Unterstützung des Lehrbetriebs Möglichkeiten des Datei- und Informationsaustausches sowie der Terminverwaltung integriert.

Beide Typen beinhalten bestimmte Funktionen, die eine Unterstützung von Gruppen ermöglichen („Groupware-Systeme“). Ellis et al. definieren Groupware als „computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment.“ (Ellis et al., 1991, S. 40). Unter Groupware fallen nicht nur entsprechend betitelte Anwendungen, die häufig Kooperationen in kleineren Arbeitsteams hervorheben, sondern eine ganze Reihe von verschiedenen Kommunikationsmöglichkeiten wie Videokonferenzen, E-Mail, Wikis, Wissensportale oder Abstimmungswerkzeuge. LMS und Groupware-Systeme hängen mit dem Begriff des *Computer Supported Cooperative Work* (CSCW) zusammen (Gross und Koch, 2007; Grudin, 1994). Carstensen und Schmidt (1999) fassen Groupware-Anwendungen und LMS unter

dem gleichen Begriff zusammen (Wilson, 1991). Bei einer genaueren Einordnung der Systeme und Anwendungen wird im Allgemeinen auf die Klassifikation nach Raum und Zeit nach Johansen (1991) und auf das 3-K-Modell nach Teufel et al. (1995) zurückgegriffen.

	Gleiche Zeit	Verschiedene Zeit
Gleicher Ort	Gruppenmoderationssystem Brainstormingunterstützung Abstimmungswerkzeuge	Schwarzes Brett Gruppenarbeitsraum LMS
Verschiedener Ort	Videokonferenzen Application Sharing Virtuelle Sitzungsräume	E-Mail Nachrichtensysteme Wissensmanagementsysteme Gruppenportale LMS

Abbildung 7.2: Raum-Zeit-Matrix nach Johansen (1991)

Johansen (1991) unterscheidet zwischen gleicher Zeit und verschiedener Zeit sowie gleichem Ort und verschiedenem Ort. Innerhalb dieser Matrix werden verschiedene Systeme klassifiziert. LMS lassen sich primär den Feldern „Gleicher Ort/Verschiedene Zeit“ und „Verschiedener Ort/Verschiedene Zeit“ zuordnen. In Abbildung 7.2 sind jeweils bereits entsprechende Anwendungen aufgeführt, die teilweise auch in den verwendeten LMS zu finden sind. So verfügen sie über Ankündigungen, die mit einem schwarzen Brett vergleichbar sind. Daneben sollen sie vor allem das schulische Wissensmanagement unterstützen und die Kommunikation erleichtern und vereinfachen. Teufel et al. (1995) unterscheiden die Groupware-Systeme im Hinblick auf die drei Elemente Kommunikation, Kooperation und Koordination und inwieweit die Anwendungen diese unterstützen. Grafisch tragen sie dieses in einem Dreieck ab (vgl. Abbildung 7.3).

LMS bilden in der Regel alle drei Elemente, sind aber stärker auf die Unterstützung der Koordination und der Kommunikation konzentriert. Die Kommunikation findet über Nachrichtensysteme, Ankündigungen, Dateiaustausch mit Kommentarfunktion oder auch Diskussionsforen statt. Die Aufgabe der Koordination bezieht sich fast ausschließlich auf Lehrprozesse (Dateiaustausch von Aufgaben und Materialien, Terminverwaltung von Ereignissen und Räumen) und Verwaltungsaufgaben (Lehrer- und Schülerverwaltung, Noten- und Zeugnisverwaltung, sofern integriert). Gleiches trifft auch auf Videokonferenzen und andere virtuelle Kommunikationsmedien zu (Verschiedener Ort/Gleiche Zeit).

Evaluation von Groupware- und Lern-Management-Systemen

Sowohl Groupware- als auch Lern-Management-Systeme wurden in der Vergangenheit evaluiert. Im Bereich der „klassischen“ Groupware erfolgt in der Regel eine qualitative

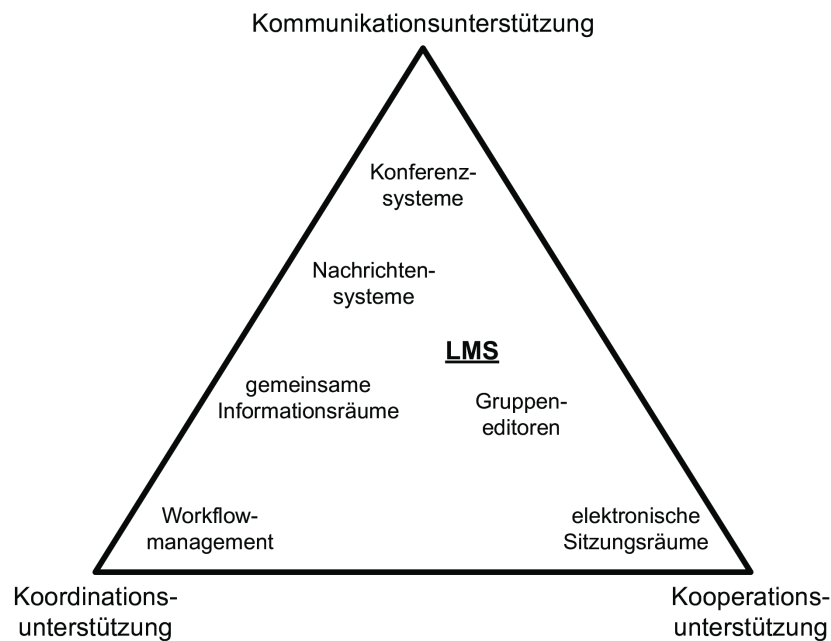


Abbildung 7.3: 3K-Modell nach Teufel et al. (1995)

Evaluation (Pinelle und Gutwin, 2000). Die Evaluatorinnen und Evaluatoren setzen methodologisch häufig Beobachtungen ein. Diese finden eher in laborähnlichen Situationen als im Feld statt. Dabei kommen auch angepasste Methoden wie das *Groupware Walkthrough* (Pinelle und Gutwin, 2002) zum Einsatz.

Im Gegensatz dazu merken Blecken et al. (2010, 8f) an, dass es bisher wenige Studien zur Usability von LMS gibt. Evaluationen werden vor allem mit Nutzertests, heuristischen Evaluationen und Fragebögen durchgeführt. In diesen Bereichen gibt es Ansätze, die Methoden auf den Bereich der E-Learning-Plattformen auszurichten. Mehlenbacher et al. (2005) stellen Heuristiken für heuristische Evaluationen vor, die auf E-Learning-Systeme zugeschnitten sind. Dabei erweitern sie die bekannten Heuristiken von Nielsen (1994d) um spezielle Anforderungen von LMS wie beispielsweise Instruktionen durch Lehrende. Zudem spielt das Wissen der Lernenden eine wichtige Rolle. Reeves et al. (2002) modifizieren die Heuristiken von Nielsen (1994d) und passen sie ebenfalls auf E-Learning-Systeme an (vgl. Kapitel A.6). Die Entwicklung von angepassten Fragebögen ist ein weiteres Feld (Zaharias und Poylymenakou, 2009). Daneben werden teilweise neue Verfahren entwickelt (Nguyen, 2012). Umfangreiche Entwicklungen gehen auf Maria F. Costabile et al. zurück (Ardito, Costabile, De Marsico, Lanzilotti, Levaldi, Plantamura et al., 2004; Ardito, Costabile, De Marsico, Lanzilotti, Levaldi, Roselli et al., 2006; Costabile, De Marsico, Maria et al., 2005; Costabile und Matera, 2001; Matera et al., 2002). Sie beziehen Usability auf vier Dimensionen (Presentation, Hypermediality, Application proactivity, User activity) und entwickeln darauf basierend Prinzipien, Kriterien und Guidelines für eine hohe Effektivität und Effizienz des Systems. Die Entwicklung angepasster Kriterien erscheint sinnvoll, da sich nicht alle Aspekte von E-Learning-Plattformen klassischen Heuristiken wie Nielsen (1994d) abbilden lassen (Rentróia-Bonito et al., 2011). Dies trifft insbesondere auf die kognitiven Prozesse zu (Martin et al., 2008).

Nutzertests stellen die dritte große Gruppe der Usability Evaluationen von LMS dar. Besonders häufig werden Usability Evaluationen von Moodle² durchgeführt (vgl. für einen Überblick Senol et al., 2014). Daneben wird vor allem Blackboard³ evaluiert. Die Fokussierung auf die beiden Systeme ist unter anderem durch die Verbreitung im Hochschulsektor bedingt. Evaluationen werden deutlich häufiger dort als im schulischen Bereich durchgeführt.

Die Ergebnisse lassen vor allem folgende Schlüsse zu. Mit Blick auf die heterogenen Nutzerinnen und Nutzer dieser Systeme (vgl. u. a. Rentróia-Bonito et al., 2011; Shehu et al., 2009), sind die Anforderungen an die Usability, insbesondere der Zufriedenstellung und Effektivität, hoch (Blecken et al., 2010). Die grundlegende Benutzung des Systems ist den Testerinnen und Testern in der Regel geläufig. Problematisch ist, dass die Nutzung webbasierter Systeme den Testpersonen Probleme bereitet (vgl. u. a. Baumgartner et al., 2004; Melton, 2006). Pretorius und van Biljon (2010) können bestätigen, dass erfahrenere Internetnutzer weniger Probleme bei der Bedienung von LMS haben. Die Nutzung von LMS besteht von Seiten der Lernenden primär aus dem Herunterladen von Dateien (Blecken et al., 2010, S. 8). Dies stellt bei mehreren Studien ein Problem für die Testerinnen und Tester dar (Blecken et al., 2010; Melton, 2006). Lehrende nutzen dagegen mehrere Aspekte der Interaktion und Kommunikation. Baumgartner et al. (2004), Kakasevski et al. (2008) und De Porto Alegre Muniz und De Moraes (2012) können schwerwiegende Probleme bei der Nutzung dieser Systeme nachweisen. Dies gilt sowohl für Lehrende als auch für Lernende. Generell ziehen die Testerinnen und Tester die Stabilität der Systeme einer großen Funktionsvielfalt vor.

7.2 Evaluiertes System⁴

Das in dieser Dissertation evaluierte System ist InfoChange⁵. InfoChange lässt sich vor allem in zwei Kategorien einordnen: *Groupware-Systeme (GW)* und *Lern-Management-Systeme (LMS)*. Zum einen können Termine, Dateien, Nachrichten und Informationen unter den Nutzerinnen und Nutzern getauscht und veröffentlicht werden. Dies entspricht dem Groupware-Charakter. Zum anderen bietet InfoChange die Möglichkeit, dass die Lehrkräfte mit ihren Schülern interagieren. Dies beinhaltet die Bereitstellung von (digitalen) Materialien, sowohl rezeptiv als auch interaktiv als auch die Möglichkeit des Testens. Dabei können beliebige Distanzen und Zeitverschiebungen (Asynchronitäten) überwunden werden (Ellis et al., 1991).

Der Aufbau von InfoChange orientiert sich an Schul- und Klassenräumen. Das System wird von mehreren Schulen parallel eingesetzt, wobei jede teilnehmende Schule über einen Schulraum verfügt. Dieser dient als Übersicht für alle Nutzerinnen und Nutzer der Schule (Lehrkräfte, Schülerinnen und Schüler, zugelassene Externe), listet alle Klassenräume auf und enthält generelle Informationen zu dem Schulraum (Informationen zur Schule, Kon-

²<http://moodle.de/> [Aufgerufen am 22.07.2015]

³<http://www.webct.com/> [Aufgerufen am 22.07.2015]

⁴In diesem Kapitel entstammen Teile aus der Publikation des DFG-Projekts (Welling et al., 2014).

⁵Es handelt sich um anonymisierte Namen, analog zu Welling et al. (2014).

taktdaten des Raumadministrators, Anzahl der Klassenräume, Anzahl der Mitglieder etc.). Der Raum ist allen angemeldeten Mitgliedern der Schule zugänglich. Neben Lehrerinnen und Lehrern, Schülerinnen und Schülern zählen dazu Externe (Eltern, Schulpersonal). Die Freischaltung des Zugangs erfolgt durch einen der schulinternen Administratoren. Administratorrechte besitzen die Mitglieder der Schulleitung und der schulinterne Administrator, wobei dieser als hauptverantwortlicher Administrator fungiert. Alle weiteren Räume in der Schule werden Klassenräume genannt. Ein Klassenraum muss nicht einer real existierenden Klasse zugeordnet sein. Der Begriff wird von den Entwicklern des Systems vorgegeben. Innerhalb des Schulraumes befinden sich über 100 Klassenräume. Die Nutzung der Klassenräume lässt sich grob in drei Gruppen unterteilen: Bei den ersten beiden Gruppen handelt es sich um Räume, die einer bestimmten Klasse zugeordnet oder für bestimmte Kurse innerhalb einer Klasse oder eines Jahrgangs angelegt wurden (z. B. „Informatik Klasse 10“). Diese Räume sind ausschließlich für pädagogische Aufgaben gedacht und werden wie ein Lern-Management-System (LMS) verwendet. Diese sind für die folgenden Analysen uninteressant.

The screenshot shows the 'InfoChange' website interface. At the top is a navigation menu with tabs: Home, Ankündigungen, Termine, Materialien, Diskussionen, Personen, Gruppen, Fächer, and an RSS button. The main content area is titled 'Überblick (Klassenraum)'. It features three main sections: 'Ankündigungen (4 gültige von 224)', 'Termine (50 heute und in der Zukunft von 553)', and 'Materialien (5 aus den letzten 7 Tagen von 552)'. Each section contains a list of items with columns for 'Datum' and 'Bearbeiter/Ersteller?'. To the right of the main content are three sidebars: 'Suche in Raum' with a search button, 'Rauminformationen' showing 'Aktive Personen (7 Tage)', 'Neue Beiträge', and 'Seitenaufrufe', 'Schlagwörter', and 'Kategorien' with a list of categories like Archiv, Computernutzung, FAQ, Konferenzen, Prüfungen, Unterricht, Schule, Verschiedenes, and Zeugnisse.

Abbildung 7.4: Einstiegsseite von InfoChange

Die letzte Gruppe der Klassenräume dient der Kommunikation der Lehrkräfte oder Schülergruppen untereinander (z. B. „Fachkonferenz Englisch“). Das proportionale Verhältnis der drei Raumgruppen beträgt etwa 30 Prozent zu 60 Prozent zu 10 Prozent. Die Klassenräume sind nur definierten Teilgruppen zugänglich und werden von dem Raumbesitzer verwaltet. Ein Raum für die Kommunikation der Lehrkräfte untereinander ist von zentralem Interesse für die weiteren Analysen. Bei diesem handelt es sich um das „Infozentrum“. Diesen Raum gibt es in physischer Form auch in der Schule. Wie sein analoges Pendant steht der Raum allen Lehrkräften zur Verfügung und nimmt für die Kommunikation innerhalb des Kollegiums eine zentrale Rolle ein, da unter anderem die

Schulleitung diesen Raum für die Weitergabe von Informationen und Materialien nutzt. Von dieser werden alle Termine in den Kalender des Raums eingestellt und zusätzlich als herunterladbare Datei (im PDF-Format) in den Materialien angeboten. Daneben ist dies der Raum, über den unter den Lehrkräften die meisten Materialien getauscht werden. Die Logfile-Analyse mit „realen“ Lehrerinnen und Lehrern bezieht sich auf diesen Raum.

Nach dem Einloggen gelangt die Nutzerin beziehungsweise der Nutzer automatisch auf die Startseite, welche eine Übersicht über aktuelle Ankündigungen, Termine, Materialien, Diskussionen, Personen, Gruppen und Fächer ermöglicht (vgl. Abbildung 7.4). Neben den Zusammenfassungen der einzelnen Rubriken, die vor allem aktuelle Änderungen anzeigen, stehen eigene Übersichtsseiten der entsprechenden Rubriken zur Verfügung. Diese bieten einen umfassenderen Überblick über die jeweiligen Ankündigungen, Termine etc. als die Kurzzusammenfassungen auf der Eingangsseite und sind über die Links in der Navigationsleiste zu erreichen (ganz oben in Abbildung 7.4 dargestellt).

Abbildung 7.5: Materialiensammlung von InfoChange

Ein besonderes Merkmal von InfoChange ist die flache Struktur des Systems. Im Gegensatz zu einer geläufigen Verzeichnisstruktur (Daten/Dateien in Ordnern) verzichtet InfoChange fast gänzlich auf eine hierarchische Anordnung der Einträge. In der Übersichtsseite der Materialien (vgl. Abbildung 7.5) ist der linke und mittlere Bereich für eine Darstellung der Materialien vorgesehen. Die Auflistung erfolgt standardmäßig anhand der aktuellsten Änderungen, kann aber auch nach Titel oder Bearbeiterin beziehungsweise Bearbeiter sortiert werden. Sie ist unabhängig vom Kontext (Fach, Jahrgang etc.). An einzelne Einträge angehängte Dateien (Dokumente, Bilder, Whiteboard-Dateien etc.) werden mittels eines entsprechenden Symbols dargestellt. Beim Aufrufen des Eintrags wird die Nutzerin beziehungsweise der Nutzer auf eine Detailseite geleitet, wo der Eintrag genauer beschrieben ist und die eventuell angefügten Dateien mit Dateinamen sowie die zugeord-

neten Schlagwörter und Kategorien angezeigt werden. Eine Eingrenzung der Materialien erfolgt anhand der Kategorien (rechts unten), Schlagwörter (rechts mittig) oder über die Suche. Zudem ist es möglich, Materialien einem bestimmten Fach zuzuordnen, das über ein Listenmenü zugänglich ist (über den Schlagwörtern). Die Fächer und Gruppen sind als Drop-down-Liste dargestellt, die Kategorien in einer Baumstruktur und Schlagwörter als „Tag Cloud“ (Lohmann et al., 2009). Häufig verwendete Schlagwörter werden größer und prägnanter dargestellt. Zur Eingrenzung der Materialien können die Lehrkräfte entweder eine der genannten Möglichkeiten nutzen oder verschiedene Filterfunktionen kombinieren. Andere Ablagemöglichkeiten der Materialien (Ordnerstruktur o. ä.) sind nicht möglich. Die Einteilung der Gruppen und die Zuordnung zu Fächern folgt keinem erkennbaren Muster. Es existieren vor allem Gruppen zu den Themen Fachentwicklungen, Schulentwicklungen und zu didaktischen Schwerpunkten einzelner Fächer. Die Zuordnung von Lehrkräften zu einzelnen Gruppen kann über die Administratoren des Raumes oder über die Nutzerinnen und Nutzer selbst erfolgen. Neue Einträge erhalten in den Übersichtsseiten eine entsprechende Kennzeichnung ([neu]). Gleiches trifft auf Änderungen in den Einträgen zu ([geändert]). InfoChange verfügt über eine umfangreiche Kommentarfunktion. Alle Einträge können kommentiert werden. Ein Kommentar zu einem Eintrag wird in den Übersichtsseiten kenntlich gemacht ([neue Anmerkung]). Die Übersichten über Fächer, Gruppen und Personen sowie die Kategorien und Schlagwörter dienen primär zur Orientierung und Kategorisierung des Inhalts. Einträge können miteinander verknüpft werden. Die Verknüpfung erfolgt händisch und kann bei der Erstellung des Eintrags oder zu einem späteren Zeitpunkt erfolgen. Eine Einschränkung ist nicht gegeben, sodass neben den erwähnten Kategorien, Schlagwörtern, Gruppen und Fächern auch Einträge untereinander verlinkt werden können. Die Verlinkungen sind bidirektional, sodass bei der Verknüpfung von zwei Einträgen in den detaillierten Ansichten beider Einträge der jeweils andere Eintrag angezeigt wird. In der Übersichtsseite der Termine (vgl. Abbildung 7.6) steht neben der Listenansicht eine Kalenderansicht zur Verfügung. Diese orientiert sich an den üblichen Darstellungsweisen (Tages-, Wochen- und Monatsansicht). Die einzige Einschränkung bei diesen Ansichten ist, dass dort keine Kategorien und Schlagwörter verwendet werden können. Veränderte und neu erstellte Items sind neben der Anzeige über das Webinterface per *Really-Simple-Syndication* (RSS-Feed; Kantel, 2007) abrufbar. Dieses Format bietet die Möglichkeit, Änderungen zu abonnieren und mittels externer Programme verfolgen zu können. Dazu zählen Dienste im Internet (beispielsweise Feedly⁶), Desktopprogramme, Smartphones, Tablets und Browser. Der Vorteil für die Nutzerin beziehungsweise den Nutzer liegt darin, dass die Änderungen im Hintergrund geladen und präsentiert werden. Neue Termine können auf diesem Wege nicht erstellt oder bestehende Termine geändert werden.

⁶<https://feedly.com/> [Aufgerufen am 10.07.2015]

Abbildung 7.6: Terminübersicht von InfoChange

7.3 Datenerfassung und Datenaufbereitung

Bevor die Logfiles aus dem Produktivsystem für weitere Analysen verwendet werden können, müssen diese wie in Kapitel 5.2.1 beschrieben aufbereitet werden. Dabei wird das beschriebene Verfahren mit *Data Cleaning*, *User Identification* und *Session Identification* verfolgt. Bei der Path Completion wird auf Grund der fehlenden Referenzeinträge der Logfiles ein eigens entwickeltes Verfahren eingesetzt. Die Logfiles von InfoChange können über zwei Wege erstellt werden. Zum einen können die Logfiles des Webservers verwendet werden. Dabei wird das Extended Logfile Format (vgl. Listing 4.2 auf Seite 15) verwendet. Zum anderen verfügt das System über die Möglichkeit, eigene Protokolle in der Datenbank anzulegen. Dort werden neben der IP-Adresse, der User-Agent-Angabe, der Zeitangabe, der URL und der Aktion (POST oder GET) die interne ID der Nutzerin beziehungsweise des Nutzers, die aufgerufene Rubrik und die Funktion sowie die interne ID des Eintrags angegeben. Die zusätzlichen Informationen werten die Datenbank-Logfiles stark auf, sodass diese als Basis dienen. InfoChange ist nur für registrierte Nutzerinnen und Nutzer zugänglich, was eine Entfernung von Crawlern und Bots überflüssig macht. Diese haben keinen Zugriff auf den geschlossenen Teil des Systems. Es werden nur die Zugriffe auf die Seiten protokolliert, was zur Folge hat, dass keine irrelevanten Bilder und Skriptdateien herausgefiltert werden müssen. Die Möglichkeit, Kalendereinträge per iCal und neue Einträge per RSS zu abonnieren, werden herausgefiltert. Diese Abfragen entstehen nicht durch aktives Eingreifen der Nutzerinnen und Nutzer, sondern finden automatisiert im Hintergrund statt. Die Logfiles geben keine Auskunft über eventuell aufgetretene Fehler (Aufruf nicht existenter Seiten, Softwarefehler), sodass davon ausgegangen wird, dass alle Zugriffe erfolgreich und problemlos waren.

Die Logfiles speichern eine eindeutige und persistente Nutzer-ID als separates Feld, was einer eindeutigen Nutzeridentifizierung entspricht. Eine Identifizierung der einzelnen Nutzerinnen und Nutzer erfolgt durch die Methode der Nutzerregistrierung (vgl. Kapitel 5.2.1). Ein Zugriff auf das System ist nur nach vorherigem Einloggen möglich. Dies erlaubt eine eindeutige Identifizierung der Nutzerinnen und Nutzer.

Die Trennung der einzelnen Besuche orientiert sich an der zweiten Heuristik (*h2*). Es ist bekannt, dass es bei InfoChange nach 30 Minuten Inaktivität zu einem automatischen Ausloggen kommt. Alle Besuche werden nach 30 Minuten getrennt. Daneben wird in den Logfiles das manuelle Ausloggen aus dem System protokolliert. Verlässt eine Nutzerin oder ein Nutzer das System auf manuellem Wege, wird ebenfalls von dem Ende des Besuchs ausgegangen. Erfolgt ein weiterer Besuch innerhalb des Intervalls von 30 Minuten, wird dieser als neuer Besuch gezählt.

Bei der Vervollständigung der fehlerhaften Pfade konnte nicht auf den Algorithmus mit der Nutzung des Referers zurückgegriffen werden, da die Logfiles keinen Referer enthalten. Für die weitere Analyse wurde ein graphenbasiertes Verfahren entwickelt. Bei der Vervollständigung der Daten von InfoChange werden Ansätze aus der Graphentheorie (Jansen, 2006; Wasserman und Faust, 1994) und die Seitenstruktur verwendet, um die unvollständigen Pfade zu identifizieren und anschließend zu vervollständigen. Teilweise orientiert sich dieser Ansatz dabei an den Ausführungen von Cooley et al. (1999). Das Vorgehen besteht aus mehreren Schritten. Zunächst wird auf Basis der tatsächlichen Seitenstruktur ein Graph erstellt. Eine Verlinkung in InfoChange ist immer bilateral. Es handelt sich daher um einen ungerichteten Graphen. In einem zweiten Schritt kann mittels des Graphen ermittelt werden, ob die entsprechenden Pfade in den Logfiles möglich sind. Existiert zwischen zwei Knoten (in diesem Fall Einträgen) keine Verknüpfung, muss der Pfad als unvollständig angesehen und vervollständigt werden. Handelt es sich dabei um einen vollständigen Pfad, so wird die entsprechende Kante zwischen den beiden Knoten gewichtet. Alle Kanten haben zu Beginn des Graphen eine Gewichtung von eins, vollständige Pfade zwischen zwei Knoten erhöhen die Gewichtung um eins. Nach diesem Durchlauf werden alle Gewichtungen der Kanten gedreht, d. h. der kleinste Wert wird mit dem größten Wert getauscht usw. Alle als unvollständig identifizierten Pfade werden durch den gedreht gewichteten Graphen vervollständigt. Dazu wird durch den Algorithmus zur Findung des kürzesten Pfads nach Dijkstra (1959) versucht, einen vollständigen Pfad über verknüpfte Knoten (Einträge) herzustellen. Der Algorithmus zieht dabei den Pfad des „geringsten Widerstands“, sprich der geringsten Gewichtung, vor. Es wird davon ausgegangen, dass häufig genutzte Pfade am optimalsten sind. Mit diesem Ansatz wurde versucht, alle unvollständigen Pfade zu erneuern. Insgesamt konnten über die Hälfte aller unvollständigen Pfade wiederhergestellt werden.

Die Länge der einzelnen Besuche ist zu diesem Zeitpunkt noch nicht weiter betrachtet worden. Um später die Bewegungen in dem System genauer zu analysieren, sollten nur Daten mit ausreichenden Bewegungen vorhanden sein. Besuche mit weniger als drei Aktionen werden von der weiteren Analyse ausgeschlossen. Der vorhandene Datensatz wird als Grundlage für die deskriptiven Analysen und die spätere Kombination herangezogen.

7.4 Nutzung des Systems⁷

Die zur Verfügung gestellten Logfiles der „realen“ InfoChange-Installation beziehen sich auf den Zeitraum von März 2011 bis März 2012. In den Zeiträumen von Mitte April 2011 bis Mitte Mai 2011 und Mitte September 2011 sind die Logfiles nicht komplett. Für die übrigen Zeiträume sind vollständige Daten vorhanden. Die Analysen beziehen sich auf eine Datenbasis von 303 Tagen. Es war während der Feldphase möglich, einen Blick in den Schulraum und den untersuchten Klassenraum der Schule zu werfen. Neben den Eindrücken konnte ein ungefährender Überblick über die Anzahl der Dateien, Materialien, Termine und Ankündigungen erhalten werden. In dem System befanden sich zum Ende des Erhebungszeitraumes rund 650 Materialien, 650 Termine und 275 Ankündigungen. In dem Untersuchungszeitraum sind etwa 160 Materialien, 320 Termine und 100 Ankündigungen neu erstellt worden. Die Zahl der einzelnen hochgeladenen Dateien ist deutlich höher, da jedem der drei genannten Einträge eine oder mehrere Dateien angehängt werden können. Insbesondere bei den Materialien kommt dies häufiger vor. In der Summe befinden sich über 3.000 Dateien in dem SIS.⁸

Während des Erhebungszeitraumes wurde das System von 136 verschiedenen Nutzerinnen und Nutzern genutzt. Diese tätigten etwa 3.100 Besuche und knapp 35.600 Aktionen in dem System. Das bedeutet, dass sich jede Nutzerin und jeder Nutzer etwas mehr als 22 Mal eingeloggt und pro Besuch 11,5 Aktionen durchgeführt hat. Die Nutzungsintensität der einzelnen Nutzerinnen und Nutzer variiert stark, wie in Abbildung 7.7 ersichtlich. Dort ist die Anzahl der Besuche (x-Achse) und die Anzahl der Zugriffe (y-Achse) abgetragen. Jeder Punkt steht für eine Nutzerin beziehungsweise einen Nutzer. Je dunkler die Färbung des Punktes, desto höher ist die durchschnittliche Zahl der Zugriffe pro Besuch. Besuche und Zugriffe sind nicht unabhängig voneinander, da jeder Besuch mindestens drei Aktionen beinhalten muss. Die Korrelation zwischen den beiden Variablen ist mit einem Wert von 0,976 fast perfekt.

Zunächst fällt auf, dass es drei Nutzerinnen und Nutzer gibt, die über die meisten Besuche und Zugriffe verfügen. Sie verfügen über 264 und je 315 Besuche und führen dabei knapp 4.096, 3.412 und 4.539 Aktionen durch. Die Aktionen pro Besuch liegen dagegen im Mittelfeld. Bei einem Gesamtzeitraum von 303 Tagen bedeutet dies, dass jeder dieser drei Nutzerinnen und Nutzer im Durchschnitt etwa einmal täglich in dem System eingeloggt war. Durch die Gruppendiskussionen und teilnehmenden Beobachtungen (Welling et al., 2014) wurde ersichtlich, dass Mitglieder der Schulleitung das System sehr intensiv nutzen. Es kann davon ausgegangen werden, dass es sich bei diesen drei Lehrkräften um Mitglieder der Schulleitung und den Administrator handelt. Der Abstand zu den anderen Nutzerinnen und Nutzern beträgt über 150 Besuche und 2.500 Zugriffe. Um eine Verzerrung der Analysen durch diese Nutzerinnen und Nutzer zu vermeiden, werden sie von der weiteren Analyse ausgeschlossen. Durch das Entfernen sinkt die Anzahl der Besuche auf knapp 2.200 und die Anzahl der Aktionen beläuft sich auf 23.550.

⁷In diesem Kapitel entstammen Teile aus der Publikation des DFG-Projekts (Welling et al., 2014).

⁸Diese sehr hohe Anzahl an Dateien führt dazu, dass anstatt der einzelnen Dateien stets von dem übergeordneten Eintrag gesprochen wird.

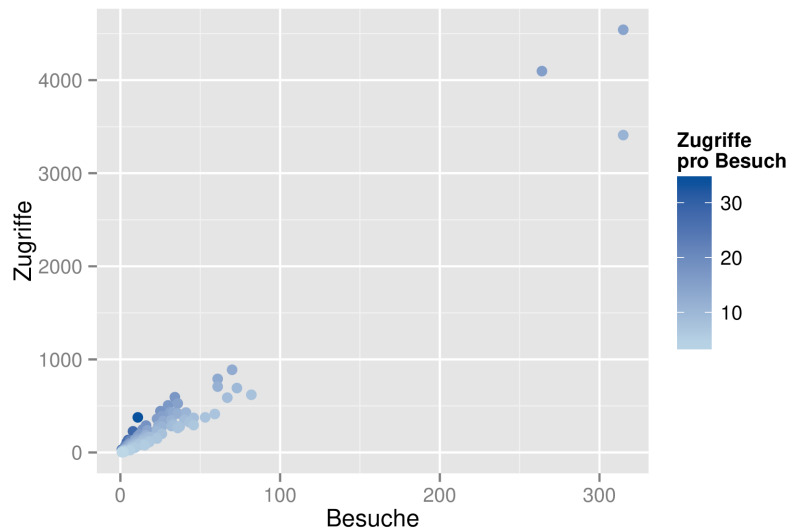


Abbildung 7.7: Besuche und Aktionen aller Nutzerinnen und Nutzer

Die Verteilung der weiteren Nutzerinnen und Nutzer fällt nach dem Entfernen der drei Nutzerinnen und Nutzer homogener aus (vgl. Abbildung 7.8). Es lassen sich grob Gruppen identifizieren. Eine Gruppe aus sechs intensiven Nutzerinnen und Nutzern. Diese haben mehr als 60 Besuche und 500 Zugriffe. Daneben gibt es eine Gruppe von 32 Nutzerinnen und Nutzern mit einer mittleren Nutzungsintensität. Diese beläuft sich auf 20 bis 60 Besuche und mehr als 250 Zugriffe. Die Fälle, die jeweils weniger als 250 Zugriffe und mehr als 20 Besuche beziehungsweise weniger als 20 Besuche und mehr als 20 Zugriffe aufweisen, werden der Gruppe ebenfalls zugeschlagen. Die Mehrzahl der Nutzerinnen und Nutzer wird der dritten und inaktivsten Gruppe zugeordnet. Weniger als 20 Besuche bedeutet, dass sich die Nutzerinnen und Nutzer dieser Gruppe weniger als zweimal pro Monat eingeloggt haben. Ein großer Teil dieser Gruppe weist weniger als zehn Besuche auf, was weniger als einem Besuch pro Monat entspricht.

Trotz der hohen Anzahl an weniger aktiven Nutzerinnen und Nutzern nimmt die Nutzung des Systems in der zweiten Hälfte des Untersuchungszeitraums leicht zu (vgl. Abbildung 7.9, in dunkelblau).⁹ Von Anfang März 2011 bis Ende Juli 2011 gibt es kaum Tage, an denen mehr als zehn Besuche zu verzeichnen sind. Anfang August 2011 steigen die Zugriffe stark an, erreichen ein Allzeithoch mit fast 55 Besuchen und bleiben in den folgenden anderthalb Monaten auf einem Niveau von bis zu 30 Besuchen pro Tag. Danach pendeln sich die Zugriffe bei etwa bis zu 20 Besuchen an prägnanten Tagen ein. Neben dem leicht steigenden Trend ist die Anzahl der Besuche durch das Schuljahr geprägt. In den Ferienzeiten (März 2011, Juli 2011, Oktober 2011 und Ende 2011) sind kaum Besuche zu verzeichnen. Die meisten Besuche ereignen sich Anfang August um den Beginn des neuen Schuljahres. Die Anzahl der Zugriffe verläuft wechselhaft. Auf einen starken Abfall an Besuchen erfolgt ein starker Anstieg. Bei den Abfällen handelt es sich um Wochenenden. Insbesondere am Sonntag und Montag kommt es zu den meisten Zugriffen. Diese Zahl nimmt im Laufe der Woche ab und erreicht am Samstag den tiefsten Stand. Im Vergleich

⁹Die Rohdaten finden sich im Anhang in Tabelle B.1.

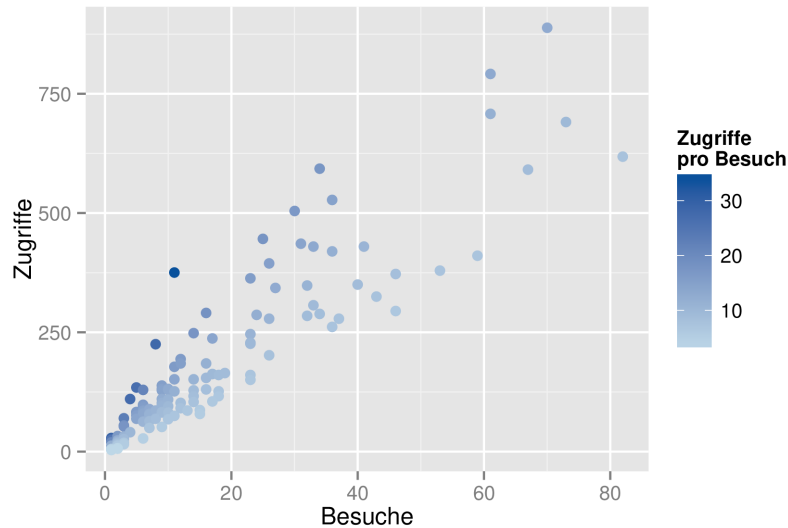


Abbildung 7.8: Besuche und Aktionen ausgewählter Nutzerinnen und Nutzer

dazu sind die Besuche mit allen Nutzerinnen und Nutzern in gelb dargestellt. Die Anzahl der Besuche ist höher, verläuft aber nach demselben Muster. Die Korrelation zwischen den Besuchen der beiden Gruppen ist mit einem Wert von 0,95 fast perfekt. Dies lässt den Schluss zu, dass sich die Besuche der drei sehr intensiven Nutzerinnen und Nutzer sehr gleichmäßig auf die Tage der anderen Nutzerinnen und Nutzer verteilen und eine Entfernung keine Verzerrung der Zugriffe nach Tagen verursacht.

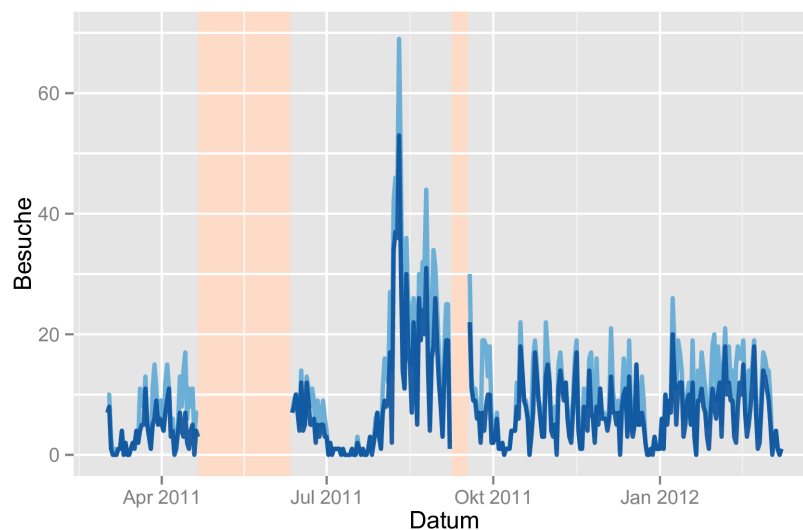


Abbildung 7.9: Besuche im Zeitverlauf

Die Nutzung innerhalb InfoChanges findet primär in den Rubriken Ankündigungen, Materialien und Terminen sowie der Startseite statt. Sie werden von den Nutzerinnen und Nutzern unterschiedlich intensiv genutzt (in absteigender Reihenfolge): Materialien (mit deutlichem Abstand), Ankündigungen, Termine und Startseite (vgl. Tabelle 7.1). Nach jedem Login werden die Nutzerinnen und Nutzer automatisch auf die Startseite

verwiesen. Dieser Umstand kann die Daten verzerren und dazu führen, dass die Startseite überproportional bewertet wird. Der Aufruf der Startseite im ersten Zugriff wird deshalb entfernt.¹⁰ Die Nutzerinnen und Nutzer kommen im Rahmen eines Besuchs häufiger auf die Startseite zurück, sodass diese nicht als Ausgangspunkt gewertet werden kann. Daneben findet eine intensive Nutzung der Suchfunktion statt. Weniger Zugriffe entfallen auf die Benutzer(-verwaltung), die Gruppen, die Fächer und die Diskussionen. Insbesondere die geringe Nutzung der Diskussionen ist verwunderlich, stellt sie ein Mittel zur Kommunikation dar.

Tabelle 7.1: Zugriffe auf die Rubriken

Rubriken	Zugriffe
Materialien	12.853
Ankündigungen	2.673
Termine	1.867
Suche	1.746
Startseite	1.746
Nutzer	1.025
Gruppen	655
Fächer	648
Diskussionen	337

Neben den Zugriffen auf die einzelnen Rubriken ist interessant, was genau die Nutzerinnen und Nutzer in den einzelnen Rubriken für Aktionen durchführen. Für eine visuelle Ansicht der Zugriffe auf die einzelnen Rubriken bietet sich die Darstellung als Treemap (Shneiderman, 1992) an. Bei dieser Darstellung (vgl. Abbildung 7.10) ist die Fläche eines Bereiches mit der Anzahl der Zugriffe verknüpft: Je größer die Fläche, desto höher die Anzahl der Zugriffe. Der Vorteil der Darstellung liegt darin, dass sich die Bereiche weiter unterteilen lassen. *Übersicht* protokolliert einen Zugriff auf die Gesamtübersicht einer Rubrik (vgl. Abbildung 7.5), *Herunterladen* den Download einer Datei, *Bearbeiten* das Bearbeiten oder Anlegen eines Eintrags (Ankündigung, Material etc.), *Details* für die detaillierte Ansicht eines Eintrags, *Schlagwort* die Eingrenzung durch ein oder mehrere Schlagwörter und *Kategorie* die Eingrenzung über die Kategorien. Anhand der Logfiles kann nicht unterschieden werden, ob ein Eintrag neu erstellt oder bearbeitet wird, sodass diese beiden Aktionen unter *Bearbeiten* zusammengefasst werden.

Zunächst bestätigt die Darstellungsform die bisherigen Ergebnisse: Die Nutzung der einzelnen Bereiche variiert stark und der Fokus liegt vor allem auf dem Bereich der Materialien. Dieser Bereich füllt etwa die Hälfte der gesamten Fläche aus. Danach folgen die Rubriken der Ankündigungen, der Termine, der Suche und der Startseite. Fast alle Rubriken sind farblich unterteilt. Die Ausnahme bildet die Startseite. Sie verzeichnet lediglich Zugriffe auf die Übersicht, da diese keine anderen Funktionen bietet. In den weniger intensiv aufgerufenen Bereichen der Benutzer(-verwaltung), der Gruppen, der Fächer und der Diskussionen entfällt jeweils etwa die Hälfte aller Zugriffe auf die Übersicht beziehungsweise auf die Detailseiten. Die Nutzung der Bearbeitungsfunktion ist gering, Zugriffe auf Kategorien und Schlagwörter konnten nicht protokolliert werden.

¹⁰Handelt es sich bei dem ersten Zugriff um eine andere Rubrik, wird der Zugriff nicht entfernt.

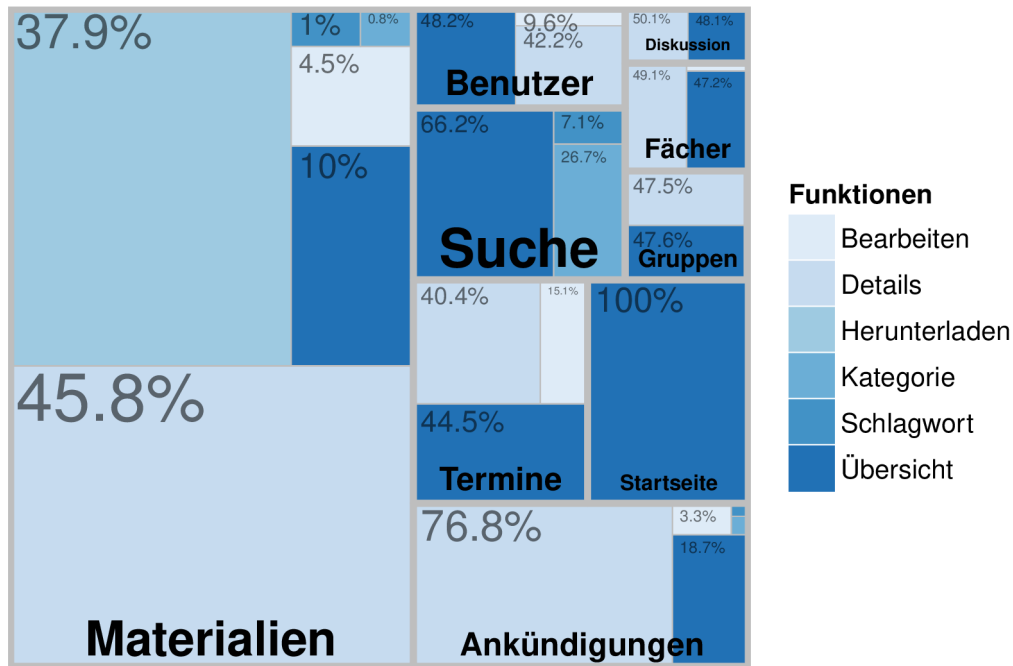


Abbildung 7.10: Treemap der Zugriffe auf Rubriken und Funktionen (Prozentzahlen beziehen sich auf die Zugriffe innerhalb der Rubrik; Funktionen mit weniger als 50 Zugriffen nicht angezeigt)

In den drei stark genutzten Rubriken sieht dies ähnlich aus. Bei den Ankündigungen entfallen etwa drei Viertel der Zugriffe auf die Detailansichten der Einträge und etwa ein Viertel auf die Übersicht. Dies lässt sich damit erklären, dass in der Übersicht lediglich die Überschriften der Beiträge aufgelistet werden und die Lehrerinnen und Lehrer für weitere Informationen gezwungen sind, den Eintrag aufzurufen. Das Bearbeiten der Einträge fällt mit drei Prozent der Zugriffe gering aus. Das Kommentieren einer Ankündigung wird nicht als Bearbeiten protokolliert, sondern ebenfalls als Detailansicht. Dies ist ein Grund für die geringe Anzahl an Bearbeitungen. Die Verwendung von Schlagwörtern und Kategorien ist kaum messbar. Die Ankündigungen liegen in chronologischer Reihenfolge vor. Dadurch ist die Suche nach neuen oder aktuellen Ankündigungen nicht nötig.

Bei den Terminen steht die Nutzung der Übersicht im Vordergrund. Etwa 45 Prozent aller Zugriffe dieser Rubrik entfallen auf die Übersichtsseite. Es ist davon auszugehen, dass der Grund darin liegt, dass die Übersichtsseite mehrere Kalenderansichten (Tag, Woche, Monat, Liste) bietet und diese Ansichten ausreichend für den Informationsbedarf der Lehrkräfte sind. Etwa 40 Prozent aller Zugriffe entfallen auf die Detailseiten der Termine. Das Aufrufen ist nötig, um alle Informationen eines Termins anzeigen zu lassen und um weitere Funktionen wie das Zusage zu einem Termin zu nutzen. Eine Zusage wird nicht als Bearbeitung protokolliert, sondern als weiteren Aufruf der Detailseite. Die restlichen 15 Prozent der Zugriffe entfallen auf das Editieren und Anlegen von Terminen. Dies ist der höchste Wert aller Rubriken.

In der Rubrik der Materialien können neben den bisherigen fünf Funktionen auch Materialien hoch- und heruntergeladen werden. Das Hochladen findet während des Bearbeitens statt und wird nicht separat protokolliert. Das Herunterladen wird als eigener Eintrag in den Logfiles vermerkt und entsprechend ausgewiesen. Die Dateien können über drei verschiedene Wege aufgerufen werden: Auf der Startseite werden die neuesten Materialien angezeigt und die angehängten Dateien als verlinktes Symbol neben der Überschrift angezeigt (vgl. Abbildung 7.4). Lädt eine Nutzerin oder ein Nutzer über diesen Weg eine Datei herunter, wird dies ebenfalls als Nutzung des Bereichs der Materialien protokolliert. Eine ausführlichere Übersicht befindet sich auf der Übersichtsseite der Materialien (vgl. Abbildung 7.5). Zuletzt können die einzelnen, einem Material zugeordneten Dateien über die Detailansicht des übergeordneten Eintrags angezeigt werden. Aus diesem Grund hängt der Download einer Datei jeweils mit dem Aufruf der Startseite, der Übersichtsseite der Materialien oder der Detailseite des Eintrags zusammen. In dem Bereich der Materialien erfolgen fast die Hälfte aller Zugriffe auf die Detailansicht, knapp 40 Prozent auf das Herunterladen von Dateien und knapp 10 Prozent auf die Übersichtsseite. Der Anteil der Bearbeitungen ist mit dem Anteil der Ankündigungen vergleichbar und liegt bei etwa 4 Prozent. Es ist davon auszugehen, dass die Lehrerinnen und Lehrer hauptsächlich über die Detailansicht auf die Dateien zugreifen.

Auffällig ist, dass die Termine deutlich mehr Bearbeitungen verzeichnen als die Materialien. Dies lässt sich damit erklären, dass während des Untersuchungszeitraumes etwa die doppelte Anzahl neuer Termine im Vergleich zu Materialien und Ankündigungen erstellt wurden. Zugriffe auf die Schlagwörter und Kategorien sind kaum messbar. Die letzten beiden Bereiche der Startseite und der Suche bilden Ausnahmen. Bei der Startseite entfallen alle Zugriffe auf die Übersicht, da es keine weiteren Funktionen gibt. In der Rubrik der Suche entfallen zwei Drittel aller Zugriffe auf die Startseite. Dabei handelt es sich um die Anzeige von Ergebnissen, da diese als Zugriff auf die Übersichtsseite protokolliert werden. In dieser Rubrik ist die Nutzung von Kategorien und Schlagwörtern deutlich höher (etwa 27 Prozent resp. sieben Prozent). Ein Zugriff auf die beiden Funktionen wird auf zwei Arten protokolliert. Zum einen können bereits bestehende Suchergebnisse weiter eingeschränkt werden. Werden auf der Startseite Schlagwörter und Kategorien ausgewählt, erfolgt dies intern auch über die Suchfunktion und wird dementsprechend in dieser Rubrik protokolliert. Erfolgt eine Auswahl von Schlagwörtern und Kategorien dagegen innerhalb einer Rubrik (Ankündigungen, Materialien, Termine), wird der Zugriff in der entsprechenden Rubrik protokolliert. Eine Eingrenzung durch Kategorien und Schlagwörter scheint über die Startseite und zur weiteren Eingrenzung von Suchergebnissen zu erfolgen und weniger innerhalb der einzelnen Rubriken.

Bisher wurde zwischen Zugriffen und Besuchen unterschieden. Für einen tieferen Einblick in die Nutzung macht es Sinn, beide Elemente gemeinsam zu betrachten. Theoretisch kann ein Besuch beliebig viele Zugriffe beinhalten. In der Praxis ist diese Definition letztlich von der Trennung einzelner Besuche nach 30 Minuten Inaktivität abhängig. Zudem soll analysiert werden, wie sich die Zugriffe auf die verschiedenen Rubriken im Verlauf der Besuche verändern. In Abbildung 7.11 wurde der Zugriff auf die Rubriken über den gesamten Zeitraum aggregiert, was eine differenzierte Betrachtung erleichtert. Auf der

horizontalen Achse sind die ersten 30 Zugriffe abgetragen, auf der vertikalen Achse alle 2.197 erfolgten Besuche. Zugriffe auf die verschiedenen Rubriken sind farblich dargestellt. Da nicht alle Besuche bis zu 30 Zugriffe beinhalten, sind die Säulen mit zunehmender Länge der Besuche kleiner. Dies erlaubt nicht nur, die Zugriffe auf die separaten Bereiche zu erläutern, sondern die durchschnittliche Länge der Besuche genauer zu spezifizieren.

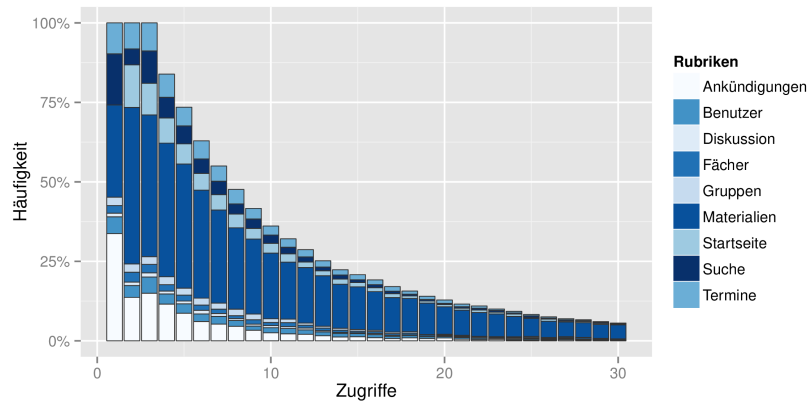


Abbildung 7.11: Verlauf der Besuche (bis zum 30. Zugriff)

Auffällig ist zunächst, dass die Anzahl der Zugriffe pro Besuch schnell abnimmt. Haben alle Nutzerinnen und Nutzer mindestens drei Zugriffe gemacht, fällt die Zahl ab dem vierten Zugriff stark ab. Weniger als 50 Prozent aller Nutzerinnen und Nutzer machen mehr als acht Zugriffe pro Besuch. Die Kurve flacht danach ab (weniger als zwei Drittel aller Nutzerinnen und Nutzer machen mehr als 11 Zugriffe pro Besuch), sinkt aber stetig weiter. 30 Zugriffe erreichen nur etwa fünf Prozent aller Besuche. InfoChange wird von vielen Nutzerinnen und Nutzern nur für kurze Besuche genutzt. Ein Blick auf die erfolgten Zugriffe bestätigt zunächst die Ergebnisse der Treemap: Die meisten Zugriffe erfolgen in den Bereichen der Ankündigungen, der Termine, der Materialien, der Suche und der Startseite. Es werden einige Auffälligkeiten sichtbar. Im ersten Zugriff entfallen keine Zugriffe auf die Startseite. Dies ist nicht verwunderlich, da diese Zugriffe entfernt wurden. Die übrigen Zugriffe scheinen aufgrund von direkten Links zu erfolgen. Dafür erfolgen überproportional viele Zugriffe auf die Rubrik der Ankündigungen. Das Interesse an den Ankündigungen nimmt im weiteren Verlauf der Besuche ab. Sie scheinen zu Beginn eines Besuchs von gesteigertem Interesse zu sein. Im ersten Zugriff erfolgen keine Zugriffe auf die Suchfunktion. Die ersten Zugriffe auf diese Rubrik erfolgen ab dem zweiten Zugriff bis zum vierten Zugriff. Danach nimmt die Nutzung dieser Rubrik proportional ab. Im zweiten Zugriff erfolgen die meisten Zugriffe auf die Materialien. Die Nutzung dieser Rubrik bleibt in dem weiteren Verlauf der Besuche hoch. Gleichzeitig kommt es zu ersten Aufrufen der Startseite. Die Intensität der Nutzung nimmt ab dem dritten Zugriff ab und ist vergleichbar mit der Suchfunktion. Im dritten Zugriff erfolgen erneut viele Zugriffe auf die Materialien, Ankündigungen, Startseite und Suche. Interessant ist, dass sich die Zugriffe auf die Termine in den ersten drei Zugriffen konstant verhalten. Anschließend nimmt die Nutzung proportional im Vergleich zu den anderen Rubriken ab. In den ersten drei Zugriffen wird ersichtlich, dass die Rubriken der Benutzer(-verwaltung), der

Diskussionen, Fächer und Gruppen von untergeordnetem Interesse sind. Ab dem vierten Zugriff sind keine Auffälligkeiten mehr feststellbar und die Zugriffe auf die Rubriken nehmen proportional ab.

Daraus kann gefolgert werden, dass die meisten Besuche von InfoChange relativ kurz sind und vor allem die ersten Zugriffe eine zentrale Bedeutung für die Nutzerinnen und Nutzer haben. Eine Erklärung ist, dass bei den Besuchen von InfoChange zu einem Großteil nach Änderungen geschaut wird, da diese in den Übersichten gekennzeichnet werden (Welling et al., 2014). Längere Besuche würden nach diesem Verständnis nur dann auftreten, wenn sich eine Nutzerin beziehungsweise ein Nutzer beispielsweise intensiver mit Materialien im System auseinandersetzen würde oder neue Materialien erstellt. Insgesamt wird ersichtlich, dass die generelle Nutzung von InfoChange im Untersuchungsverlauf eher gering einzustufen ist. Zwar ist die Aktivität leicht ansteigend, aber insbesondere die Anzahl und Länge der Besuche ist gering. Drei Viertel aller Nutzerinnen und Nutzer loggen sich weniger als zweimal pro Monat in das System ein, und weniger als 50 Prozent aller Nutzerinnen und Nutzer führen pro Besuch mehr als acht Aktionen durch. Die Zugriffe selber beschränken sich vor allem auf die Bereiche der Materialien, der Ankündigungen, der Termine, der Suche und der Startseite. Die Möglichkeiten, Diskussionen zu führen, werden kaum in Erwägung gezogen beziehungsweise durchgeführt. Die Rubriken der Benutzer(-verwaltung), der Gruppen und der Fächer sind ebenfalls von geringer Bedeutung.

8 Nutzertests als qualitative Datenquelle

Im Anschluss an die Logfile-Analyse wurden die Nutzertests durchgeführt. Bei der Entwicklung, Durchführung und Auswertung der Tests wurde möglichst genau nach den Vorgaben in Kapitel 6.3.1 vorgegangen. Insgesamt konnten 12 Testpersonen gewonnen werden. Zwei Tests fanden in einer laborähnlichen Situation statt, zehn weitere Tests wurden in einer Bremer Berufsbildenden Schule im Lehrerzimmer durchgeführt (Onsite-Testing). Es kam primär die Methode des lauten Denkens zum Einsatz. Zudem wurde der Bildschirm im Hintergrund aufgenommen und die Nutzertests im Backend von InfoChange protokolliert. Anschließend erfolgte die Auswertung. Die Ergebnisse der Nutzertests sind ausschlaggebend für die spätere Kombination von Logfiles und Usability-Problemen.

8.1 Entwicklung der Tests/Erstellung des Testplans

Die inhaltliche Ausarbeitung der Nutzertests wurde primär durch zwei Quellen unterstützt. Zum einen wurden die deskriptiven Analysen der Logfiles herangezogen (vgl. Kapitel 7.4). Daneben konnte durch die Verbindung zum DFG-Projekt „Mediatisierte Schulen“ auf Gruppendiskussionen und Interviews zurückgegriffen werden, in denen die Nutzerinnen und Nutzer detaillierter die Verwendung des SIS beschrieben (Welling et al., 2014). Auf Basis dieser Informationen wurde ein Testplan entwickelt. Dieser wird im Folgenden dargelegt.

8.1.1 Gründe und Ziele für den Nutzertest/Forschungsfragen

Ziel dieser Nutzertests ist eine summative Evaluation des Lern-Management-Systems (LMS) InfoChange. Das Produkt wird daneben als Schulinformationssystem (SIS) eingesetzt. Die Evaluation wird mit quantitativen und qualitativen Methoden durchgeführt.

In dem DFG-Projekt *Mediatisierte Organisationswelten in Schulen* wurde in mehreren Gruppendiskussionen und Interviews über das System gesprochen (Welling et al., 2014). In diesen Gesprächen wurden eine Reihe an Unzulänglichkeiten des Systems von Seiten der Lehrkräfte benannt. Daneben wurde eine Logfile-Analyse des Systems durchgeführt (vgl. Kapitel 7.4), die einen tieferen Einblick in die Nutzungssituation ermöglicht. Die Ergebnisse der beiden Methoden dienen als Quellen für die Aufgaben der Nutzertests.

Während des DFG-Forschungsprojekts wurde die in der Schule verwendete InfoChange-Installation kritisiert. Die Kritik zielt auf die vorgegebene Struktur des Systems. Durch die fehlende Ordnerstruktur seien Dateien schwer aufzufinden. Das Auffinden von Materialien sollte im Idealfall durch die Nutzung der Suche sowie Schlagwörter und Kategorien erfolgen. Dabei scheint es sich nicht um die vorherrschende Medienpraxis von Lehrerinnen

nen und Lehrern zu handeln. Das System wird teilweise mit der Redewendung „Kraut und Rüben“ (Welling et al., 2014, 272ff.) oder als „schlechte Plattform“ (Welling et al., 2014, 272f) beschrieben. Die hohe Flexibilität, die die flache Hierarchie des Systems bietet, wird wenig oder nicht wahrgenommen. In Verbindung mit dem Punkt der Unübersichtlichkeit steht ein weiterer Kritikpunkt. Das System sei nicht „intuitiv“ (Welling et al., 2014, 289f). Gleichzeitig wird die Usability insgesamt kritisiert (Welling et al., 2014, S. 274). Daneben gibt es Lehrerinnen und Lehrer, die diese Kritik nicht teilen und auf eine längere Einarbeitungszeit verweisen. Sobald man das System verstanden habe, sei der Umgang nicht mehr so kompliziert wie dargestellt. Diese Lehrkräfte arbeiten scheinbar häufiger mit dem System und loben die Möglichkeit, Materialien auszutauschen (Welling et al., 2014, 268f). Die Nutzung von weiteren zentralen Elementen wie den Ankündigungen und Terminen wird nur von der Schulleitung explizit thematisiert (Welling et al., 2014, S. 247). In weiteren Interviews und Gruppendiskussionen werden diese Rubriken nicht erwähnt. Performanz und Verfügbarkeit wurden nur indirekt kritisiert. Die Verfügbarkeit von frei zugänglichen Computern ist in der untersuchten Schule eingeschränkt. Es existiert zudem kein flächendeckendes WLAN-Netz, über welches die Lehrkräfte mit ihren privaten Endgeräten („Bring your own device“) auf das System zugreifen könnten.

Die Logfile-Analyse desselben Systems bestätigt die Ergebnisse der Nutzung zunächst (vgl. Kapitel 7.4). Die Hälfte aller Aktionen entfallen auf die Rubrik der Materialien. Es werden vor allem Detailseiten von Einträgen angeschaut und einzelne Dateien heruntergeladen. Danach folgen mit einigem Abstand Zugriffe auf die Rubriken der Ankündigungen und der Termine. Bei den Ankündigungen werden hauptsächlich die Detailseiten der Einträge aufgerufen (was abgegebene Kommentare mit einbezieht), bei den Terminen ist die Übersichtsseite präserter. In dieser Rubrik werden die meisten Einträge erstellt oder bearbeitet. Die Suche nimmt eine Sonderstellung bei den Zugriffen auf die Kategorien und Schlagwörter ein. Es handelt sich um die einzige Rubrik, bei der die Anzahl einen Wert von über einem Prozent überschreitet. Viele dieser Zugriffe dürften in Verbindung mit der Startseite stehen, da die Nutzung der beiden Funktionen auf der Startseite auf die Suche verweist. Die Zugriffe auf die Rubriken der Benutzer(-verwaltung), Gruppen, Fächer und Diskussionen sind deutlich geringer. Diese sind von geringerem Interesse für die späteren Nutzertests. Bei den drei erstgenannten Rubriken verwundert dies nicht weiter, da es sich um unterstützende und verweisende Seiten handelt (sogenannte „Auxiliary Pages“; Cooley et al., 1999). Sie verfügen über wenig relevanten Inhalt und bieten Hilfen bei dem Auffinden von Einträgen über die entsprechenden Rubriken. Die geringe Nutzung der Diskussionen verwundert. Es handelt sich um eine der intensivsten Kommunikationsmöglichkeiten des Systems. Die geringe Nutzung könnte zwei Gründe haben. Zum einen findet eine stärkere Nutzung der Ankündigungen statt. Ankündigungen können ebenfalls kommentiert werden, sodass es sich um eine Substitution der Diskussionen handeln könnte. Zum anderen wird von den Lehrkräften während der Gruppendiskussionen die Bedeutung der privaten E-Mail-Adresse als Kommunikationsmedium hervorgehoben (Welling et al., 2014, 314ff.). Es könnte sich ebenfalls um eine Substitution handeln.

Ausgehend von den Ergebnissen des Forschungsprojekts und der Logfile-Analyse soll durch die Nutzertests eine *summative Evaluation* erfolgen. Das verwendete System

liegt in einer fertig entwickelten Version vor. Die Nutzertests folgen keiner spezifischen Forschungsfrage, sondern sollen die Gebrauchstauglichkeit des Systems testen. Die eigentliche Forschungsfrage ist, ob sich die Logfiles der Nutzertests mit Logfiles aus einem realen System verknüpfen lassen, um die Ergebnisse der Tests auf die Gesamtheit aller Nutzerinnen und Nutzer übertragbar zu machen.

8.1.2 Merkmale der Testpersonen

Die Grundgesamtheit aller potenziellen Nutzerinnen und Nutzer von InfoChange stellen alle Lehrerinnen und Lehrer sowie Schülerinnen und Schüler dar. Die Logfiles aus dem Produktivsystem beziehen sich auf einen Raum, der ausschließlich von Lehrkräften zur internen Kommunikation verwendet wird. Das schließt Schülerinnen und Schüler als Zielgruppe aus. Lehrerinnen und Lehrer stellen daher die Grundgesamtheit dar.

Die „realen“ Logfile-Daten stammen aus einer integrierten Gesamtschule. Für eine inhaltlich sinnvolle Kombination der Daten sollten die rekrutierten Lehrkräfte im Sekundärbereich unterrichten. Die unterrichteten Fächer sind in den „realen“ Logfiles stark differenziert. Informationen über Geschlecht und Alter sind darüber hinaus nicht bekannt. Die 136 aktiven Nutzerinnen und Nutzern in den „realen“ Logfiles (vgl. Kapitel 7.4) entsprechen mehr als den zum Zeitpunkt der Schulbesuche etwa 110 in der Schule tätigen Lehrkräften (Welling et al., 2014). Da sich die Logfiles über einen Zeitraum von etwa einem Jahr erstrecken, ist mit einer Fluktuation der aktiven Lehrkräfte (inkl. Referendaren) in der Schule zu rechnen. Zudem werden ehemalige Lehrkräfte nicht automatisch vom Administrator gesperrt, sondern können sich nach dem Wechsel der Schule oder im Falle einer Pensionierung weiter im LMS anmelden. Die Zahl der Nutzerinnen und Nutzer dürfte damit dem kompletten aktiven Kollegium entsprechen. Daher ist hinsichtlich der Merkmale Alter, Geschlecht und der unterrichteten Fächer eine möglichst heterogene Gruppe an Lehrerinnen und Lehrern für die Nutzertests zu rekrutieren.

Die Durchführung der Nutzertests wird in einem Lern-Management-System (LMS) stattfinden. Die Testsituation dürfte für die Lehrerinnen und Lehrer eher ungewohnt sein, aber gleichzeitig sollten möglichst keine Verzerrungen und Einflüsse vorhanden sein. Zusätzlich kann die Durchführung der Nutzertests in einem unbekanntem System zu zusätzlichen Belastungen und Stressfaktoren führen. Es ist davon auszugehen, dass den Lehrkräften das LMS InfoChange unbekannt ist. Verschiedene LMS verfügen in der Regel allerdings über dieselben Funktionen und weisen eine ähnliche Benutzeroberfläche auf. Um die Irritationen und Eingewöhnung der Lehrerinnen und Lehrer möglichst gering zu halten, sollen die Testpersonen über Vorkenntnisse mit LMS oder verwandten Produkten verfügen. Dadurch kann die Doppelbelastung des unbekanntem Systems und der ungewohnten Testsituation unter Umständen abgebaut werden. Das Auswahlkriterium ist, dass entsprechende Anwendungen benutzt werden. Eine Nutzungsintensität wird nicht vorausgesetzt, da in den „realen“ Logfiles ein breites Spektrum hinsichtlich der Nutzungsintensität vorhanden ist (vgl. Kapitel 7.4).

Die Teilnehmerinnen und Teilnehmer sollen anhand folgender Merkmale rekrutiert werden:

- Lehren an der Sekundarstufe (Gymnasium, Realschule, Hauptschule, Gesamtschule oder Berufsbildende Schule).
- Nutzen Computerplattformen zur Vorbereitung des Unterrichts und zur Kommunikation mit Kolleginnen und Kollegen. Häufigkeit der Nutzung spielt keine Rolle.
- Alter: egal
- Geschlecht: egal
- Unterrichtete Fächer: egal

Insgesamt sollen zehn Nutzertests durchgeführt werden. Die Rekrutierung der Lehrerinnen und Lehrer erfolgte anfangs selbstselektiv über einen Flyer (siehe Anhang C.1), der vom Landesinstitut für Schule (LIS) der Stadt Bremen an in Frage kommende Schulen verteilt wurde. Bei diesen Schulen handelt es sich um Leuchtturmschulen, die aktuell eine weitere Lernplattform als zentrales Tool einführen. In dem Flyer werden die Kriterien zur Rekrutierung der Testpersonen aufgezählt. Gleichzeitig wird ein kleines Dankeschön (Incentive) versprochen.¹ Diese Rekrutierungsmethode erwies sich als wenig erfolgreich. Trotz mehrerer Nachfassaktionen konnten nur zwei Mitarbeiterinnen und Mitarbeiter aus dem LIS und eine weitere Lehrkraft aus den betreffenden Schulen rekrutiert werden. Daher wurden die Nutzertests von einer laborähnlichen Situation in ein *Onsite Testing* umgewandelt. In einer Berufsschule mit technischem Schwerpunkt wurden weitere Nutzertests im Lehrerzimmer durchgeführt. Diese fanden nicht während der Pausen, sondern während der Unterrichtszeiten statt, teilweise auch in einem separaten Besprechungsraum. Der technische Schwerpunkt bietet den Vorteil, dass tendenziell eher computeraffine Lehrerinnen und Lehrer rekrutiert werden. Eine größere Nähe zur Informatik sollte sich positiv auf die Qualität der Nutzertests auswirken. Die Aufgabenstellungen und eingesetzten Methoden blieben identisch. Auf die Entwicklung von Nutzerprofilen wurde verzichtet, da die Eingrenzung der Testerinnen und Tester auf Lehrkräfte in einer Sekundarstufe eine hinreichende Profildefinition darstellt.

8.1.3 Methodologie

Die Datenerfassung der Nutzertests ist vorrangig qualitativer Natur. Alle Testerinnen und Tester werden aufgefordert, laut zu denken („Think-Aloud“). Gleichzeitig wird die Interaktion mit dem System per Screencapturing festgehalten. Anhand dieser Daten sollen später die Usability-Probleme erkannt und herausgearbeitet werden. Da wenige Informationen über die Usability-Probleme in InfoChange vorhanden sind, wird bei der Auswertung eine *induktive Kategorienbildung* (Mayring, 2010; Mayring und Gläser-Zikuda, 2008) angewandt.

Neben den qualitativen Daten werden quantitative Daten gesammelt. Dies ist zum einen durch die Logfile-Analyse bedingt. Neben dem Verhalten der Testerinnen und Tester wird der Zeitpunkt (Timestamp) jeder Aktion festgehalten. Dadurch können einfache Perfor-

¹Dabei handelt es sich um einen Gutschein im Wert von 5 Euro für Amazon, Thalia oder Saturn. Der Gutschein wurde den teilnehmenden Testpersonen nach Abschluss aller Nutzertests zugesandt.

mance Measures wie die benötigte Zeit pro Aufgabe berechnet werden. Die Kombination von Zeitmessung und lautem Denken ist fehleranfällig (vgl. Kapitel 6.3.2), weshalb die Zeitmarken nicht direkt erfasst werden, sondern nachträglich über die Screencapturing-Videos extrahiert werden. Nach der Durchführung des Nutzertests wird den Teilnehmerinnen und Teilnehmern ein Fragebogen (siehe Kapitel C.3) vorgelegt. Dieser enthält eine übersetzte SUS-Skala (Brooke, 1996) und eine AttrakDiff-Skala (Hassenzahl et al., 2003). Zudem werden demografische Merkmale der Testpersonen abgefragt (Alter, Geschlecht, Dienstalter, Rolle in der Schule, Schulform, unterrichtete Fächer). Die demografischen Merkmale werden für eine deskriptive Beschreibung der Testpersonen herangezogen. Gleichzeitig wird sichergestellt, dass es sich um ein repräsentatives Sample hinsichtlich des Alters und der unterrichteten Fächer handelt.

8.1.4 Testumgebung und Testausrüstung

Die Nutzertests wurden an zwei verschiedenen Orten durchgeführt. Zunächst wurde für die Nutzertests der Besprechungsraum des Instituts für Informationsmanagement Bremen GmbH (ifib) eingeplant. Dieser war für die Zeit der Tests reserviert, sodass auch die Begrüßung und das Debriefing vor Ort stattfinden konnten. Die Wahl fiel auf diesen Raum, weil er einem guten Kompromiss zwischen einer Laborsituation und einer realistischen Arbeitssituation entspricht. Der Raum ist spartanisch eingerichtet. Im Raum befinden sich mehrere Stühle und Tische, ein Whiteboard, ein Beamer und eine Leinwand. An der Stirnseite des Raumes befindet sich die Leinwand. An der rückwärtigen Seite hängt eine stilisierte Version des Institutslogos. Die Seitenwände werden zum einen von dem Whiteboard bedeckt, auf der anderen Seite befindet sich eine Fensterfront. Die Testpersonen sitzen mit dem Rücken zum Whiteboard. Es befinden sich einige Grünpflanzen in dem Raum. Diese Testsituation dürfte die Testteilnehmer wenig ablenken. Der Raum ist nicht so steril wie ein Usability-Labor und dürfte die Testerinnen und Tester weniger an eine Testsituation erinnern. Auffällige Elemente wie Einwegspiegel oder Kameras fehlen. Die Nutzertests werden an einem Bildschirm (19 Zoll) mit Maus und Tastatur durchgeführt. Alle drei Peripheriegeräte sind an einen Laptop angeschlossen. Dieser nimmt den Ton über ein externes Mikrofon (lautes Denken) und den Bildschirm (Screencapturing) auf. In dem Raum sind Netzkabel vorhanden, was eine stabile Verbindung zum Referenzsystem garantiert.

Nach der schleppenden Rekrutierung durch die Selbstselektion per Flyer wurden die weiteren Nutzertests vor Ort in einer Schule durchgeführt. Dort erfolgte eine wesentlich effektivere und schnellere Rekrutierung von Lehrkräften. Die Nutzertests wurden im zentralen Lehrerzimmer der teilnehmenden Berufsschule durchgeführt, teilweise auch in dem angrenzenden Besprechungsraum. Das Lehrerzimmer ist ca. 70 m² groß und verfügt über mehrere Tischgruppen sowie eine Computerecke. An den Wänden hängen zum Teil Aushänge. Der größere Teil der Wände wird von Postfächern und weiteren Schränken beziehungsweise Regalen eingenommen. An einer Seite befindet sich eine durchgehende Fensterfront. Die Nutzertests wurden in einer Ecke des Raumes an einer separaten Tischgruppe durchgeführt. Zweimal konnte auf einen angrenzenden Bespre-

chungsraum ausgewichen werden. Die Nutzertests wurden während der Unterrichtszeit durchgeführt, da sich zu diesem Zeitpunkt nur wenige Lehrerinnen und Lehrer in dem Raum aufhielten und der Geräuschpegel gering war. Für die Nutzertests kam dasselbe technische Equipment zum Einsatz wie bei den Tests im ifib. Es wurde auf einen separaten Bildschirm verzichtet und alle Tests am Laptop durchgeführt. Die Internetverbindung wurde über einen UMTS-Stick hergestellt. Während der Tests kam es zu keinem Ausfall der Verbindung.

Die Tests fanden in einer Referenzinstallation von InfoChange statt. Diese befindet sich auf einem eigenen Server und hat keinerlei Verbindungen zu der Installation des Produkktivsystems. Es handelt sich um die Version 7.5.4. Die Version wurde ab dem 29. August 2011 bereitgestellt und dürfte damit während des längsten Protokollzeitraums verwendet worden sein. Vorherige Versionen stammten aus den Entwicklungszweigen 7.5.x und 7.3.x. In allen Versionen wurden primär Fehler (Bugs) behoben. Keiner dieser Fehler steht in direkter Verbindung zu den Aufgaben oder den gefundenen Problemen. Teilweise wurden in den Versionen neue Funktionen hinzugefügt oder überarbeitet. Diese wurden in den einzelnen Aufgaben nicht verwendet. Um die Referenzinstallation so realistisch wie möglich zu gestalten, wurde in dem System ein „Lehrerzimmer“ genannter Raum mit 20 fiktiven Nutzerinnen und Nutzern angelegt. Die Kennungen wurden verwendet, um fiktive Inhalte in dem System zu erstellen. Insgesamt wurden 51 Ankündigungen, 54 Termine, 19 Materialien, 17 Fächer, 16 Diskussionen und 9 Gruppen angelegt. Alle Einträge wurden mit 173 Schlagwörtern und 37 Kategorien referenziert. Für die Nutzertests wurden separate Nutzerkennungen verwendet.

8.1.5 Aufgabenliste des Nutzertests

Die deskriptive Logfile-Analyse und die Ergebnisse des DFG-Projekts haben gezeigt, dass es verschiedene Schwerpunkte bei der Nutzung von InfoChange gibt. Darunter fallen insbesondere die Materialien, die Termine, die Ankündigungen und das Auffinden von Einträgen über die Suche oder die Schlagwörter beziehungsweise Kategorien. Diese Schwerpunkte werden bei den Aufgaben für die Testerinnen und Tester erneut aufgegriffen und in die entsprechenden Aufgaben integriert. Die Aufgaben zu den Materialien beinhalten das Herunterladen beziehungsweise Anschauen von Dateien eines Eintrags und das Anlegen eines neuen Materials inklusive der Vergabe von Schlagwörtern und Kategorien. In der Rubrik der Termine sollte einem bestehenden Termin zugesagt, ein bestimmter Termin in der Vergangenheit gefunden und ein eigener Termin angelegt werden. In beiden Rubriken sind die Testerinnen und Tester gefordert, sowohl die Übersichtsseite, die Detailseite und die Funktion zum Erstellen eines Eintrages zu nutzen. Das Auffinden des Eintrages in der Vergangenheit kann manuell über ein „Durchklicken“ durch die Termine erfolgen oder über die Nutzung der integrierten Suchfunktion. Die dritte häufig genutzte Rubrik waren die Ankündigungen. Die Testerinnen und Tester sollen eine Ankündigung ihrer Wahl kommentieren, was sowohl die Nutzung der Übersichtsseite, der Detailseite und die Nutzung der Kommentarfunktion mit einschließt. Zuletzt soll die Suchfunktion genutzt werden. Neben der Suche nach einem bestimmten Termin sollen

alle Einträge zu einem bestimmten Schlagwort angezeigt werden. Die Nutzung der Kategorien fällt nach den Logfiles zwar deutlich höher aus, ist bei der Darstellung unter den Schlagwörtern angeordnet (vgl. Abbildung 7.4) und daher schwierig zu finden. Um eine Verzerrung der Nutzertests durch diese Hürde bei der erstmaligen Nutzung des Systems zu vermeiden, ist die Verwendung von Schlagwörtern vorgesehen. Die Nutzung der Startseite ist durch die Struktur von InfoChange implizit implementiert. Durch die geringe Nutzung der Rubriken der Benutzer(-verwaltung), der Gruppen, der Fächer und der Diskussionen werden diese nicht in die Aufgaben integriert. Während der Tests können allerdings Zugriffe auf die Rubriken erfolgen. Aus diesen Erkenntnissen und Entscheidungen wurden sieben Testaufgaben entwickelt. Die Aufgaben wurden den Testpersonen in schriftlicher Form vorgelegt. Die Aufgaben sind:

1. Lisa Lehrer hat im März einen Termin für den Projekttag des Jg. 6 erstellt. Finden Sie diesen und sagen Sie zu.
2. Erstellen Sie einen eigenen Termin mit den folgenden Eigenschaften:
 - Titel: Klausur Englisch Jg. 12
 - Datum: 17.03.2015, 10:00-12:00 Uhr
 - Ort: Raum C2.15
3. Sehen Sie sich die Beispielklausuren von Frau Kuhn in Deutsch für den Jahrgang 11 an.
4. Legen Sie ein neues Material mit der Beispielklausur im Fach Mathematik für den 11. Jahrgang vom Desktop (Beispielklausur_Mathematik_Jg11.pdf) an. Das Material soll die Schlagwörter „Jahrgang 11“, „Mathematik“, „Klausuren“ und der Kategorie „Unterricht“ und der Unterkategorie „Arbeitsbögen“ zugeordnet sein.
5. Kommentieren Sie eine Ankündigung Ihrer Wahl.
6. Finden Sie das Datum des Kinobesuchs des Films „Stauffenberg“ heraus.
7. Lassen Sie sich alle Einträge mit dem Schlagwort „Jahrgang 12“ anzeigen.

Aus den Aufgaben werden keine wie in der Literatur empfohlenen Szenarien entwickelt. Die als Testerinnen und Tester rekrutierten Lehrkräfte arbeiten ebenfalls mit entsprechenden Systemen und dürften daher keine thematische Einordnung in die Aufgaben benötigen. Gleichzeitig sollen die Nutzertests durch den Verzicht kurz gehalten und überflüssiger Input vermieden werden.

8.1.6 Rolle des Moderators

Bei dem Moderator und dem Testleiter in den Nutzertests handelt es sich um dieselbe Person. Während der Nutzertests verhält sich der Moderator so passiv wie möglich, um die Ergebnisse nicht zu verfälschen. Ein Eingreifen ist gestattet, wenn die Testerinnen und Tester nicht weiterkommen, sollte aber nicht zu früh erfolgen, sondern nach einiger Wartezeit. Die Testerinnen und Tester sollen die Möglichkeit erhalten, selbst auf die Lösung des Problems zu kommen oder weitere Probleme zu identifizieren. Wenn möglich, sind weitere Nachfragen auf die Zeit nach dem Nutzertest zu verschieben, um keine Unterbrechung herbeizuführen und den Nutzertest so störungsfrei wie möglich zu halten. Dies gilt insbesondere bei Rückfragen zu dem Ziel beziehungsweise dem Hintergrund des

Nutzertests. Nachfragen, die sich direkt auf die Aufgabe beziehen, sind zu beantworten, um eine Bearbeitung der Aufgabe zu gewährleisten. Es soll darauf geachtet werden, dass die Beantwortung der Nachfragen keine Informationen zur erfolgreichen Bewältigung der Aufgabe enthält. Alle Interaktionen zwischen Testperson und Moderator werden schriftlich festgehalten und in den Protokollen der Nutzertests vermerkt. Dies gilt für weitere Anmerkungen während der Debriefing-Phase. Anmerkungen während der Nutzertests werden durch die Audioaufnahme mit erfasst. Während der Nutzertests vor Ort kommt dem Moderator die Aufgabe zu, die Ablenkung für die Testperson möglichst gering zu halten. Dazu zählen weitere interessierte Lehrerinnen und Lehrer, die sich im Lehrerzimmer aufhalten.

8.1.7 Testablauf

Die Nutzertests lassen sich in drei Teile gliedern. Die Gesamtdauer soll sich auf etwa 40 bis 50 Minuten belaufen. Die Nutzertests werden den größten Zeitanteil einnehmen. Vorher sind eine Begrüßung und Einführung geplant. Nach den Tests erfolgt das Debriefing und die Testerinnen und Tester werden gebeten, den vorbereiteten Fragebogen auszufüllen.

1. *Begrüßung und Briefing* (10 Minuten):

Die Testpersonen werden vom Moderator begrüßt. Der Testleiter erläutert das weitere Vorgehen und die angewendeten Methoden (lautes Denken, Screencapturing, Logfile-Erhebung im Hintergrund). Die Erfassung der Logfiles als Verknüpfung zu den anderen Daten ist hervorzuheben und die Technik des lauten Denkens ausführlich zu erklären. Es ist davon auszugehen, dass diese den Testpersonen nicht bekannt ist. Weitere Fragen zur Studie und den Ergebnissen sollten möglichst nach dem Test besprochen werden. Bevor die Nutzertests starten, übergibt der Testleiter den Zettel mit den Login-Daten, dem Einstieg in das System und den Testfragen (siehe Anhang C.2).

2. *Durchführung des Nutzertests* (20 Minuten):

Im nächsten Schritt führen die Teilnehmerinnen und Teilnehmer den Nutzertest durch. Der Moderator sitzt neben ihnen und beobachtet den Prozess. Interessante und relevante Ereignisse sollten protokollarisch festgehalten werden. Ein Eingreifen soll nach Möglichkeit vermieden werden und ist zu protokollieren.

3. *Debriefing und Fragebogen* (10 Minuten):

Nach dem Nutzertest haben die Testpersonen die Möglichkeit, weitere Anmerkungen zu machen. Inhaltliche Angaben werden vom Moderator protokolliert. Ab diesem Zeitpunkt können weitere Angaben zu den weiteren Arbeiten im Rahmen der Dissertation gemacht werden. Zudem sollten die Teilnehmerinnen und Teilnehmer den Fragebogen ausfüllen. Abschließend wird nach dem Incentive gefragt werden.

8.2 Auswertung der Tests

Insgesamt wurden 12 Nutzertests durchgeführt. Drei Nutzertests wurden im Besprechungsraum des ifibs und neun Nutzertests in der teilnehmenden Berufsschule durchgeführt. Zwei der Testpersonen stammen aus dem Landesinstitut für Schule (LIS) des Landes Bremen. Beide sind in die Einführung eines LMS an Bremer Schulen involviert, weshalb ihnen ein Wissen über das Feld zugeschrieben wird. Diese Nutzertests wurden zunächst als Pretests geplant. Im Rahmen der Pretests wurden keine Probleme hinsichtlich der Aufgabenstellungen und Testfragen festgestellt. Bei der Auswertung der Nutzertests konnten keine wesentlichen Unterschiede zu den anderen Testerinnen und Testern festgestellt werden. Daher werden die beiden Nutzertests mit in die Analyse einfließen.

Die 12 Testpersonen verteilen sich unterschiedlich hinsichtlich der demografischen Merkmale Lebensalter und Geschlecht. Rekrutiert wurden fünf Testerinnen und sieben Tester. Die Altersverteilung ist heterogen (vgl. Tabelle 8.1). Es konnte keine Testerin beziehungsweise kein Tester jünger als 35 Jahre rekrutiert werden. Die anderen Altersspannen verteilen sich gleichmäßig.

Tabelle 8.1: Alter der Testerinnen und Tester

Alter	Anzahl
bis 29 Jahre	0
30-34 Jahre	0
35-39 Jahre	3
40-44 Jahre	1
45-49 Jahre	3
50-54 Jahre	2
55+ Jahre	3

Die Verteilung nach dem Dienstalalter ist homogener. Neun Testpersonen haben ein Dienstalalter zwischen fünf und 14 Jahren, drei Testerinnen und Tester sind seit über 25 Jahren im Lehrerberuf (beziehungsweise im Landesinstitut) beschäftigt. Die Unterschiede in den unterrichteten Fächern sind ebenfalls heterogener (vgl. Tabelle 8.2). Die Fächer Wirtschaft und Informatik werden von jeweils sechs Testerinnen und Testern unterrichtet. Drei Lehrerinnen und Lehrer unterrichten Mathematik. Die Verteilungen nach (Dienst-) Alter, Geschlecht und Unterrichtsfächern werden als hinreichend heterogen definiert. Eine Kombination mit den „realen“ Logfiles ist auf Basis des Samples möglich.

Alle rekrutierten Testerinnen und Tester haben den Nutzertest abgeschlossen und den anschließenden Fragebogen ausgefüllt. Die Dauer der Nutzertests schwankt zwischen achteinhalb Minuten und knapp 25 Minuten. Der Durchschnitt liegt bei knapp 14 Minuten und ist sechs Minuten geringer als ursprünglich geplant. Drei der 12 Testerinnen und Tester benötigen 20 Minuten und mehr (vgl. Tabelle B.4).

Neben der Gesamtdauer der Tests unterscheidet sich die benötigte Bearbeitungszeit der einzelnen Aufgaben voneinander. Diese Unterschiede sind in Abbildung 8.1 als Boxplot (Tukey, 1977) dargestellt. Es wird ein Boxplot für die Dauer (y-Achse) jeder Aufgabe (x-Achse) berechnet. Ein Boxplot visualisiert mehrere statistische Kennwerte. Die „Box“ spannt sich zwischen dem unteren Quantil und dem oberen Quantil auf. Das untere

Tabelle 8.2: Unterrichtete Fächer der Testerinnen und Tester

Fächer	Anzahl
Wirtschaft ¹	6
Informatik ²	6
Mathematik	3
Sport	2
Politik	2
Spanisch	1
Englisch	1
Elektrotechnik	1
Deutsch	1

Die Testpersonen aus dem LIS entfallen.

¹ Inkl. Wirtschaftslehre, Geschäftsprozesse

² Inkl. IT-Systeme, Programmierung, Informationsverarbeitung, PC-Anwendung (Excel)

Quantil ist als 25-Prozent-Grenze definiert, d. h. 25 Prozent aller Werte sind kleiner oder gleich dem unteren Quantil. In dem Boxplot von Aufgabe 1 entspricht das untere Quantil der Zeitangabe von einer Minute und 16 Sekunden und bedeutet, dass drei Testerinnen und Tester 1:16 Minuten oder weniger für diese Aufgabe benötigt haben. Das obere Quantil wird als 75-Prozent-Grenze definiert. In den Werten von Aufgabe 1 entspricht das obere Quantil der Zeitangabe von drei Minuten und 27 Sekunden. Die gesamte Box bildet den Interquantiabschabstand ab. In diesem befinden sich die mittleren 50 Prozent aller Werte. In Aufgabe 1 entspricht dies allen Werten, die zwischen einer Dauer von 1:16 Minuten und 3:27 Minuten liegen. Der Interquantiabschabstand beträgt eine Minute und 16 Sekunden. In der Box ist der Median als Linie dargestellt. Der Median stellt den mittleren Wert (von klein nach groß sortiert) dar. In Aufgabe 1 beläuft sich dieser auf 2 Minuten und 11 Sekunden. Der Vorteil des Medians gegenüber dem arithmetischen Mittel ist der geringe Einfluss von Ausreißern. Mit sinkender Fallzahl nimmt der Vorteil zu, da Ausreißer bei einer geringeren Fallzahl einen höheren Einfluss haben. Anhand der Position des Medians innerhalb der Box kann die Schiefe der Verteilung abgelesen werden. In Aufgabe 2 ist der Median am unteren (oder linken) Ende der Box. Dies ist ein Zeichen, dass die Verteilung rechtsschief ist, da 50 Prozent der Fälle auf Höhe des Medians oder darunter liegen. Neben der Box werden die sogenannten Whisker dargestellt. Diese dürfen eine maximale Länge (nach Tukey, 1977) von anderthalb Quantiabschabständen haben. Alle Fälle, die außerhalb der Whisker liegen, werden als Ausreißer gewertet und als Punkt dargestellt. Ausreißer kommen in den Aufgaben 3, 5, 6 und 7 vor. In allen anderen Aufgaben bewegen sich die benötigten Zeiten für die Aufgaben innerhalb der Whisker.

Die mittlere Bearbeitungszeit (Median) der Aufgaben variiert. Die Zeit der ersten beiden Aufgaben liegt bei etwa zwei Minuten. Die benötigte Zeit der Aufgaben 3, 5 und 6 ist halb so lang. Die Aufgabe 7 liegt mit einer Minute und 30 Sekunden zwischen diesen beiden Polen. Am längsten ist die Bearbeitungszeit der vierten Aufgabe. Das Anlegen eines eigenen Materials ist sowohl von der Schwierigkeit als auch vom Umfang (u. a. durch das Zuweisen von Schlagwörtern und Kategorien) höher einzustufen. Bei der Aufgabe gibt es die größte Streuung der Werte. Die Verteilung innerhalb der Box ist rechtsschief. Eine ähnlich hohe Streuung der Werte liegt nur bei der ersten Aufgabe vor. Eine mögliche Erklärung ist, dass die Testerinnen und Tester unterschiedlich lange für einen ersten

Überblick über das System benötigen. Bei allen anderen Aufgaben liegt die Streuung der mittleren 50 Prozent bei etwa einer Minute, bei den Aufgaben 6 und 7 liegen die mittleren 50 Prozent innerhalb einer Spanne von etwa 40 Sekunden. Über alle Aufgaben hinweg gibt es fünf Ausreißer. Diese stammen von den Testerinnen und Testern, die kumuliert am längsten für den Nutzertest benötigen. Es kann geschlossen werden, dass es Unterschiede in der Bearbeitungszeit gibt, die bis auf die erste und vierte Aufgabe eher gering sind. Bei den Werten sollte bedacht werden, dass die Verwendung des lauten Denkens einen Einfluss auf die Zeiten haben könnte.

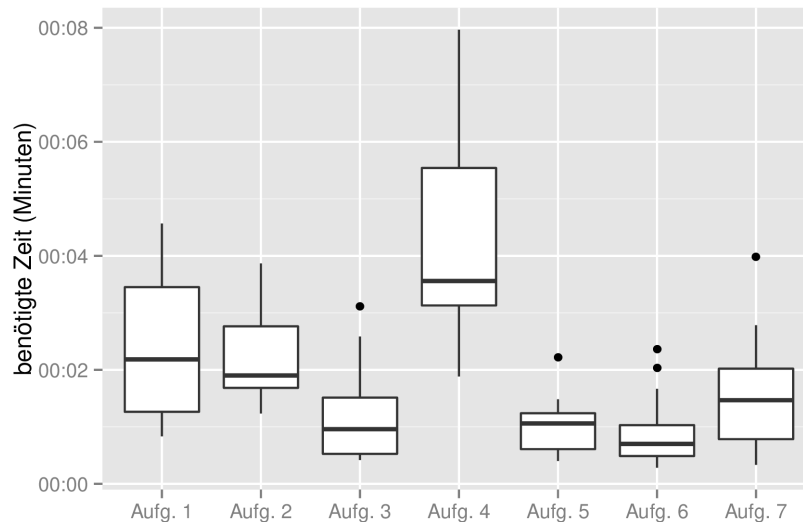


Abbildung 8.1: Benötigte Zeit pro Aufgabe

Die Ergebnisse der ausgefüllten Fragebögen zeigen, dass die Usability von den Testerinnen und Testern überwiegend positiv gesehen wird (vgl. Abbildung 8.2). Alle Fragen der SUS-Skala wurden von den Testpersonen mehrheitlich positiv oder sehr positiv beantwortet. Die durchschnittliche *SUS-Score* über alle Testerinnen und Tester hinweg liegt bei einem Wert von 72,6, was nach Bangor et al. (2009) einer akzeptablen bis guten Bewertung der Usability entspricht. Bei den gedrehten Fragen (2, 4, 6, 8, 10) fällt die Antwort entsprechend negativ oder sehr negativ aus. Die höchste Ablehnung erhalten die Fragen 4, 6, 8 und 10. Den Aussagen stimmen jeweils mehr als 70 Prozent der Testerinnen und Tester „nicht“ und „gar nicht“ zu. Die Frage 2, die negativ gestellt war, lehnten 66 Prozent der Testpersonen ab beziehungsweise voll ab. Die Fragen 1, 3, 5, 7 und 9 werden überwiegend positiv beantwortet. Die Zustimmung („stimme zu“ und „stimme voll zu“) liegt zwischen 50 und 60 Prozent. Die Antwortskala sieht eine neutrale Antwortmöglichkeit vor, weshalb die Werte höher zu bewerten sind. Der Vergleich zwischen den positiv und negativ formulierten Aussagen zeigt, dass bei den negativ formulierten Aussagen die volle Ablehnung höher ist als die volle Zustimmung bei den positiv formulierten Aussagen. Dies ist konform zur Forschung in der empirischen Sozialforschung (vgl. u. a. Kamoen, 2012). 60 Prozent der Testerinnen und Tester gaben nach dem Nutzertest an, InfoChange „gerne häufig benutzen“ (Frage 1) zu wollen. Nur zwei Testerinnen und Tester würden das System eher nicht häufiger nutzen. Die Hälfte der Testpersonen gaben an, dass das

System „einfach zu benutzen“ (Frage 3) sei und die Frage nach der „Umständlichkeit“ (Frage 8) wurde von 83 Prozent der Testerinnen und Tester abgelehnt. Die Ablehnung der Umständlichkeit ist stärker als die Zustimmung zur Einfachheit. Diese Zustimmung beziehungsweise Ablehnung wird durch die Fragen 4, 7, 9 und 10 untermauert. Diese Aussagen fragen nach der Lernbarkeit des Systems, der Sicherheit im Umgang mit dem System (Frage 9) und die benötigte Hilfestellung durch eine weitere Person (Frage 4). Die Lernbarkeit bezog sich nicht auf die eigene Person (Frage 10), sondern auch auf die Lernbarkeit durch Dritte (Frage 7). 75 Prozent der Testpersonen widersprechen der Aussage, dass sie „eine Menge lernen“ mussten, bevor sie das System verwenden können. Fünf von ihnen lehnen die Aussage stark ab. 60 Prozent der Testerinnen und Tester stimmen der Aussage zu, dass Dritte den Umgang mit dem System „sehr schnell lernen“ können. Die Hälfte der Testerinnen und Tester bejaht die Aussage, dass sie sich „bei der Benutzung des Systems sehr sicher“ fühlen, wobei eine Testperson der Aussage voll zustimmt. Drei Viertel aller Testerinnen und Tester glauben nicht, dass sie „Hilfe einer (medien-)technisch versierten Person“ benötigen. Bei dieser Aussage ist die stärkste Ablehnung („stimme gar nicht zu“) messbar. Sechs Testpersonen stimmen der Aussage gar nicht zu. Gleichzeitig werden die Aussagen über das Vorhandensein von Inkonsistenzen (Frage 6) und eine unnötige Komplexität (Frage 2) abgelehnt. Jeweils mindestens zwei Drittel der Testerinnen und Tester verneinen die Aussagen, jeweils fünf Testpersonen lehnen die Aussage voll ab. Hinzu kommt, dass sieben Testerinnen und Tester der Meinung sind, dass die Funktionen „gut integriert“ (Frage 5) sind.

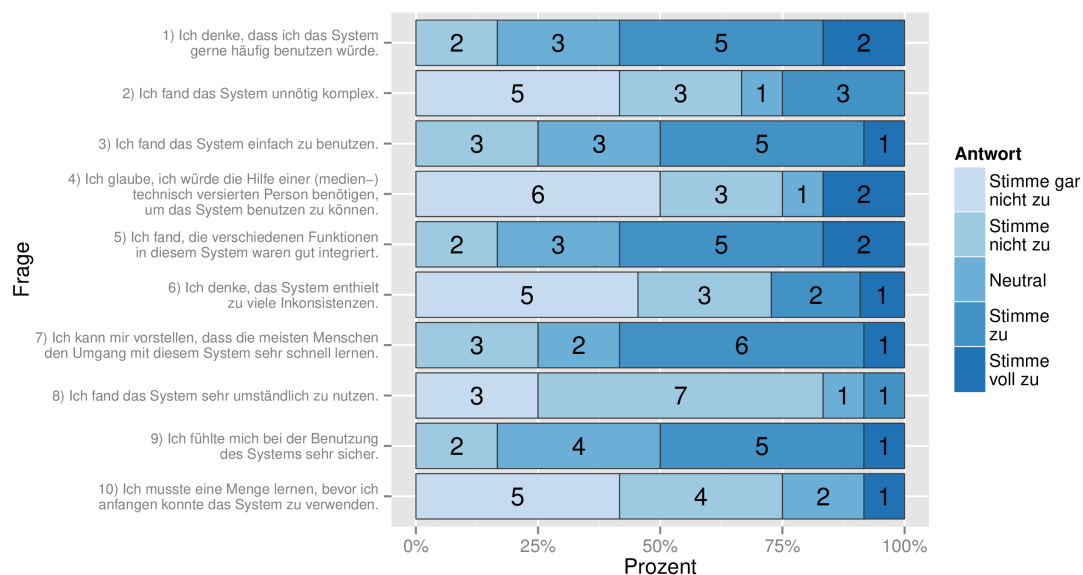


Abbildung 8.2: Antworten der SUS-Skala (in Prozent; die Zahlen geben die Häufigkeit je Antwortkategorie an; Die Antwortkategorien verlaufen von „stimme gar nicht zu“ (1) bis „stimme voll zu“ (5))

Vor der genauen Beschreibung der gefundenen Usability-Probleme lassen sich folgende Punkte festhalten. Zunächst ist die Stichprobe der Lehrerinnen und Lehrer hinsichtlich des Geschlechts, des Alters und der unterrichteten Fächer heterogen. Die benötigten Zeiten

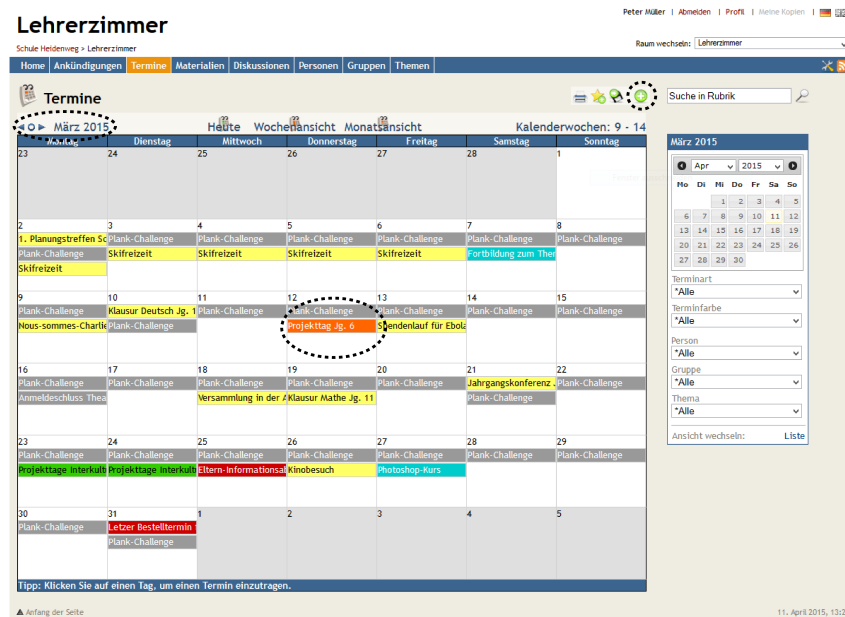


Abbildung 8.3: Usability-Probleme in der Terminübersicht
oben links ist die Navigation für den Monatswechsel markiert; oben rechts ist die Schaltfläche für das Anlegen eines neuen Termins markiert; mittig ist der gesuchte Termin markiert

für die Nutzertests streuen zwischen neun und 25 Minuten. In den meisten Aufgaben finden sich nur geringe Unterschiede in den Zeiten. Die Unterschiede werden eher durch Ausreißer hervorgerufen. In zwei Aufgaben streuen die Bearbeitungszeiten stärker. Eine der Aufgaben ist die Aufgabe 1, bei der sich die Testerinnen und Tester in dem System zurechtfinden mussten. Bei der zweiten Aufgabe handelt es sich um die komplexeste und umfangreichste Aufgabe des gesamten Nutzertests. Zuletzt kann durch die Auswertung der SUS-Skala des Fragebogens bestätigt werden, dass die Mehrzahl der Testpersonen InfoChange eine positive Usability zuschreibt.

Detaillierte Problembeschreibung

Die Tests fanden zum größten Teil im Februar 2015 statt. In allen sieben Aufgaben stießen die Testerinnen und Tester auf Probleme hinsichtlich der Usability. In diesem Kapitel erfolgt die detaillierte Beschreibung der gefundenen Probleme, wobei vor allem auf Screenshots zurückgegriffen wird. Eine Einordnung und Bewertung der Probleme erfolgt im nächsten Kapitel.

Probleme in Aufgabe 1 (Terminzusage)

Die Übersichtsseite der Termine (vgl. Abbildung 8.3) zeigt den März 2015 an. Die Navigation zwischen den Monaten funktioniert größtenteils problemlos, teilweise haben die Testerinnen und Tester Schwierigkeiten, den Monat zu wechseln. Die Monatsübersicht zeigt alle Termine der jeweiligen Tage an. Es besteht die Möglichkeit, über die einzelnen Termine zu hovern (siehe Abbildung C.1).² In dem sich darauf öffnenden Infofenster werden weitere Informationen (Beginn, Ende, Ort etc.) angezeigt. Das Auffinden des richtigen

²Zusätzliche Abbildungen finden sich im Anhang C.4 ab Seite 204.

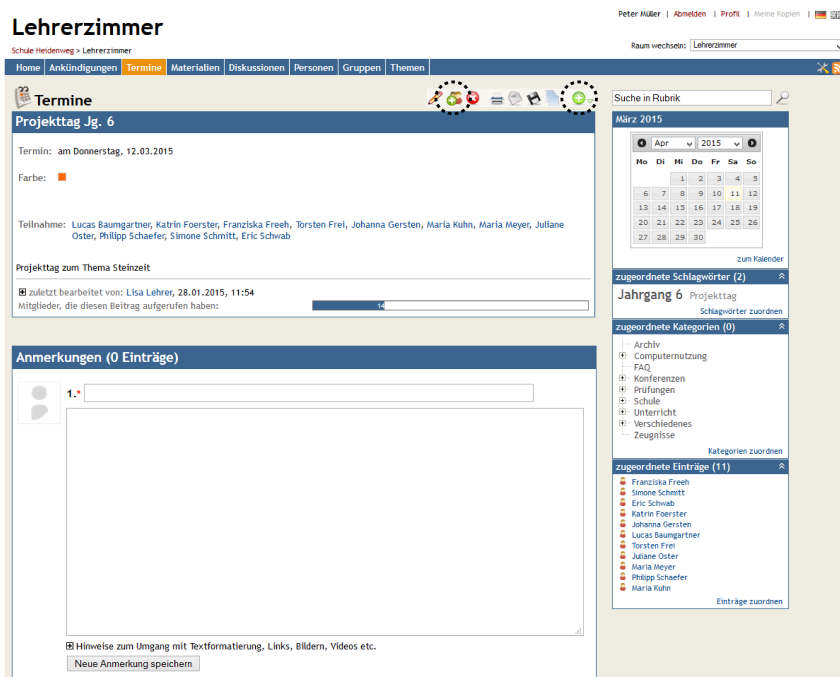


Abbildung 8.4: Usability-Probleme bei der Zusage eines Termins
 mittig ist die gesuchte Schaltfläche für die Terminzusage markiert; rechts daneben
 ist die Schaltfläche für das Anlegen eines neuen Eintrags markiert

Termins bereitet einigen Testerinnen und Testern Probleme, da in der Übersicht die Titel der Termine angezeigt werden, aber nicht die Namen der Erstellerin beziehungsweise des Erstellers (gleiches trifft auf die Übersicht in der Startseite zu). Die erste Aufgabe der Nutzertests war die Zusage zu einem Termin (*Projekttag Jg. 6*, erstellt von *Lisa Lehrer*, in Abbildung 8.3 markiert).

In Abbildung 8.4 ist die Detailansicht dargestellt. Neben Problemen, den richtigen Termin zu finden, stellt die Aktion der Zusage zu dem Termin die Testerinnen und Tester vor ein Problem. Im Gegensatz zu Groupware-Systemen wie Google Mail/Calendar, Apple iCloud oder Microsoft Exchange bietet InfoChange keine Möglichkeit, andere Teilnehmerinnen und Teilnehmer einzuladen. Eine Zusage muss von der jeweiligen Nutzerin beziehungsweise dem Nutzer in dem entsprechenden Termin manuell durchgeführt werden. Zwei Testpersonen merken diesen Umstand an, da sie eine Einladung zu dem Termin erwartet hätten. Auf der Detailseite des Termins werden die Namen der zugesagten Nutzerinnen und Nutzer (in diesem Fall die anderen Testerinnen und Tester) ausgeführt. Die eigene Zusage muss über den Button in der Symbolleiste erfolgen (in der Abbildung markiert). Dies stellt ein Problem für die Mehrzahl der Testerinnen und Tester (7 von 12) dar.

Probleme in Aufgabe 2 (Termin anlegen)

Das Anlegen eines neuen Termins kann auf zwei verschiedene Arten geschehen: Durch einen Klick auf den entsprechenden Tag in der Terminübersicht (vgl. Abbildung 8.3) gelangt die Nutzerin beziehungsweise der Nutzer in die Eingabemaske. Alternativ kann die Eingabemaske durch einen Klick auf das grüne Plus-Symbol (oben rechts) in der Symbolleiste aufgerufen werden. Das grüne Plus-Symbol findet sich auf der Startseite, auf

The screenshot shows the 'Lehrerzimmer' interface for creating a new appointment. The main form is titled 'Neuen Termin erstellen'. It has several input fields: 'Titel' (highlighted with a dashed circle), 'Beginn' (with sub-fields for 'Datum*' and 'Uhrzeit', both highlighted with dashed circles), 'Ende' (with sub-fields for 'Datum' and 'Uhrzeit', both highlighted with dashed circles), and 'Ort'. Below these are 'Kalenderoptionen' (with a checked box for 'Farben und Serientermine') and a 'Beschreibung' text area. At the bottom of the form are buttons for 'Daten suchen', 'Hochladen' (highlighted with a dashed circle), and 'Abbrechen'. There are also radio buttons for 'Terminart' (with 'privat' selected) and 'Bearbeitbar' (with 'Für alle bearbeitbar' selected). A 'Pflichtfelder' section at the very bottom contains 'Termin speichern' and 'Abbrechen' buttons. On the right side, there is a search bar 'Suche im Raum' and three sections for assigning 'Schlagwörter', 'Kategorien', and 'Einträge', each with a 'zuordnen' button.

Abbildung 8.5: Usability-Probleme in der Eingabemaske für neue Termine

oben links ist die Beschriftung des Titelfeldes markiert; darunter sind die Felder für das Beginn- und das Enddatum sowie die Anfangs- und die Endzeit markiert; rechts sind die Schaltflächen für das Zuordnen von Schlagwörtern beziehungsweise Kategorien markiert; unten ist die Schaltfläche für das Hochladen eines Anhangs markiert

den Übersichtsseiten und auf den Detailseiten (vgl. auch Abbildungen 8.4 und 8.6). Einige Testerinnen und Tester finden die Darstellungsform wenig eingängig und wünschen eine beschriftete Schaltfläche. Es kam während der Tests häufiger vor, dass die Testerinnen und Tester die *Alt Tags* der Buttons benutzen (vgl. Abbildung C.2). Das Anlegen neuer Einträge ist den Testerinnen und Testern teilweise auf Anhieb geläufig. Zwei Testpersonen hatten Probleme beim Auffinden der Eingabemaske. In der Eingabemaske eines neuen Termins (vgl. Abbildung 8.5) stoßen die Testerinnen und Tester vor allem bei der Eingabe des Datums und der Zeiten auf Probleme. In der Aufgabenbeschreibung war gefordert, einen Termin am 17.03.2015 von 10 Uhr bis 12 Uhr zu erstellen. Ein Pflichtfeld ist die Eingabe einer Datumsangabe in das Feld *Beginn*. Bleiben Datums- und Zeitangaben leer, interpretiert das System den Termin als ganztägig. Die Eingabe kann per Tastatureingabe oder über einen einblendbaren Kalender erfolgen (vgl. Abbildung C.3). Dieses Feld stellt keine Testperson vor Probleme. Daneben befindet sich das Feld für die Uhrzeit. Trotz der Bezeichnung „Uhrzeit“ ist die Startzeit gemeint, da es sich in der *Beginn*-Zeile befindet. Das stellt fünf Testerinnen und Tester vor Probleme. Eine Eingabe von „10:00 - 12:00“ oder „10-12“ wird von dem System mit einer Fehlermeldung beantwortet. Richtig ist, im oberen „Uhrzeit“-Feld „10:00“ und im unteren Feld „12:00“ einzutragen. Das untere *Datum*-Feld ist optional solange der Termin am selben Tag endet. Die Mehrzahl der Testerinnen und Tester trägt dasselbe Datum wie in dem oberen Feld ein. Der einblendbare Kalender übernimmt die Angabe des oberen Datums nicht, sondern zeigt den aktuellen Tag an. Ein Titel wird trotz der Auffälligkeit durch einen hohen Kontrast von zwei Testpersonen nicht eingegeben. Der Grund sind möglicherweise die invertierten Farben. Zwei Testpersonen

The screenshot shows a web application interface for a teacher's room. The main content area displays a list of materials with columns for 'Titel', 'bearbeitet', and 'bearbeitet von'. A red dashed box highlights a specific material entry. The right sidebar contains a search bar and a list of categories. The top navigation menu includes 'Home', 'Ankündigungen', 'Termine', 'Materialien', 'Diskussionen', 'Personen', 'Gruppen', and 'Themen'.

Abbildung 8.6: Usability-Probleme in der Materialübersicht
in der Liste sind das gesuchte Material und der Name der Erstellerin markiert;
oben rechts ist die Schaltfläche für das Anlegen eines neuen Materials markiert

verwechseln die Schaltfläche „Hochladen“ mit der Schaltfläche „Termin anlegen“. Solange keine Datei angefügt ist, gibt das System keine Rückmeldung bei der Verwendung der Schaltfläche.

Probleme in Aufgabe 3 (Material herunterladen/anschauen)

Die Übersichtsseite der Materialien listet alle verfügbaren Materialien auf. Die Materialien werden gleichwertig gelistet. Die Möglichkeit, Ordner oder eine ähnliche Baumstruktur anzulegen, existiert nicht. In der Aufgabenbeschreibung ist angegeben, die Beispielklausuren für das Fach Deutsch, die von Frau Kuhn hochgeladen wurden, anzuschauen beziehungsweise herunterzuladen. Der gesuchte Eintrag ist in Abbildung 8.6 markiert. In der Übersichtsseite ist in der dritten Spalte der Name der Erstellerin beziehungsweise des Erstellers angegeben (ebenfalls markiert).

Der Großteil der Testerinnen und Tester löst die Aufgabe nach dem vorgesehenen Weg über die Übersichtsseite der Materialien. Drei Testerinnen und Tester stoßen bei der Lösung der Aufgabe auf Probleme. Anstatt sich auf die Übersichtsseite der Materialien zu begeben, wählen die Testpersonen den Weg über die Rubrik der Personen. In dem Referenzsystem befinden sich über 50 Nutzerinnen und Nutzer. Die Auflistung aller Übersichtsseiten zeigt standardmäßig 20 Einträge an (vgl. Abbildung 8.7). Weitere Informationen werden im rechten Bereich der Seite angezeigt. Dort befindet sich die Möglichkeit umzublättern und die Einstellung für die Anzahl der gezeigten Einträge. Keine der drei erwähnten Testerinnen und Tester findet die Möglichkeit zum Umblättern.

Eine Testperson versucht das Material über die Rubrik der Gruppen zu finden. Es existieren keine Gruppen mit der Zuordnung des Faches Deutsch oder des 11. Jahrgangs (vgl. Abbildung C.4). Die Testperson wählt den Weg über die Menüleiste, um auf die Ma-

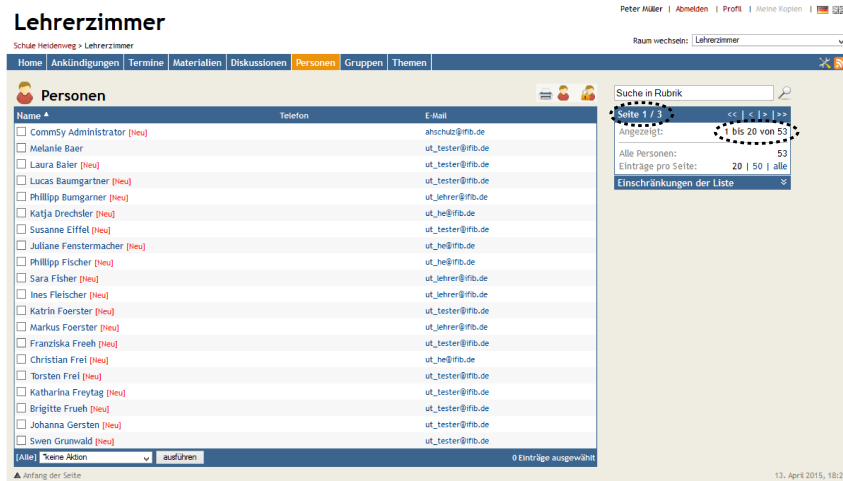


Abbildung 8.7: Usability-Probleme bei der Übersichtsseite der Personen
oben rechts ist die Angabe der aktuellen Seite markiert; schräg darunter ist die Anzahl der aktuell angezeigten Personen markiert

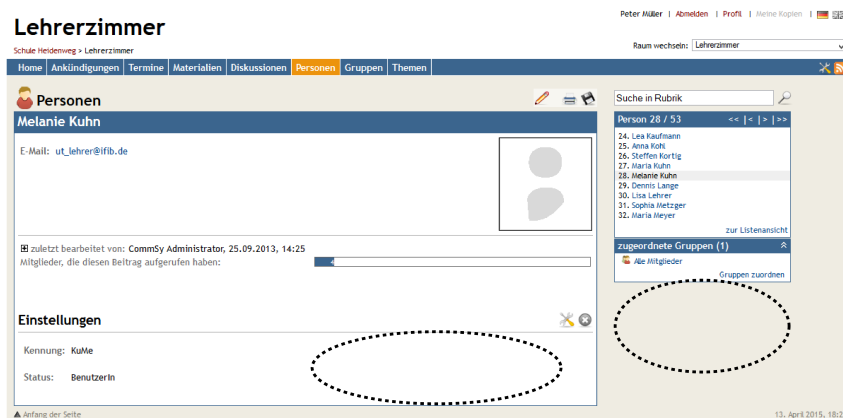


Abbildung 8.8: Usability-Problem auf der Detailseite der Personen
markiert sind die Felder, wo Aktivitäten und verlinkte Einträge erwartet wurden

aterialien zu gelangen. Die anderen beiden Testerinnen und Tester erbitten Hilfe durch den Moderator, der sie auf die Möglichkeiten im rechten Bereich hinweist. Durch diese Hilfe gelangen beide Testpersonen zu der Detailseite von Melanie Kuhn (vgl. Abbildung 8.8). Die Detailseite gibt weiterführende Informationen zu der Lehrerin an. Die Informationen können von den Lehrkräften selber editiert werden und umfassen E-Mail-Adresse, Telefonnummern, Adresse, einen Freitext, weitere Informationen und technische Details wie der Status der Nutzerin beziehungsweise des Nutzers in dem System (Benutzerin beziehungsweise Benutzer, Moderatorin beziehungsweise Moderator, Administratorin beziehungsweise Administrator) oder die Nutzerkennung. Im rechten Bereich der Seite sind alphabetisch die nächsten Nutzerinnen und Nutzer angezeigt und die Zugehörigkeit zu Gruppen, sofern vorhanden. Beide Testpersonen haben auf der Detailseite eine Auflistung der Aktivitäten und angelegter Einträge erwartet (jeweils markiert).

Eine der beiden Testpersonen klickt auf das Lupen-Symbol neben dem Suchfeld. Dadurch sucht sie nach dem Begriff „Suche in Rubrik“ innerhalb der Personen. Die Suchanfra-

ge gibt keine Ergebnisse zurück. Eine weitere Eingrenzung der Suche durch die Auswahl des Schlagwortes „Deutsch“ aus der Tag Cloud im rechten Bereich der Seite listet keine Suchergebnisse auf.

Probleme in Aufgabe 4 (Material anlegen)

Die Aufgabe bereitet den Testerinnen und Testern die größten Schwierigkeiten. Das deutet sich bereits in der quantitativen Auswertung der Bearbeitungsdauer an (vgl. Abbildung 8.1). Die Bearbeitung der Aufgabe dauerte im Durchschnitt (Median) länger als die anderen Aufgaben, zudem sind in dem Boxplot sowohl die Box als auch die Whisker größer beziehungsweise länger. Neue Einträge können über das grüne Plus-Symbol in der Toolbar angelegt werden (vgl. Abbildung 8.6). Vier Testpersonen hatten Probleme, diesen Button zu finden. Dies sind zwei Testerinnen und Tester mehr als bei der vergleichbaren Aktion in Aufgabe 2. In der Eingabemaske der Materialien (vgl. Abbildung 8.9) ist das Hochladen der Datei und die Zuordnung von Schlagwörtern und Kategorien problematisch für die Testerinnen und Tester. Sieben Testpersonen sind auf Probleme beim Hochladen der Datei gestoßen. Problematisch ist, dass die Schaltflächen für das Hochladen gräulich sind. Das führt neben einem verminderten Kontrast zu Irritationen, da angegraute Schaltflächen mit inaktiven Funktionen assoziiert werden. Daneben ist der vorgesehene Ablauf des Hochladens unüblich. Zunächst muss mittels der „Datei(en) suchen“-Schaltfläche eine oder mehrere Datei(en) des lokalen Computers ausgewählt werden. Der Vorgang geschieht über das Standard-Dialogfeld des Betriebssystems. Anschließend wird das Material eingeblendet und ein Ladebalken angezeigt. Das Hochladen beginnt nicht automatisch. Gespeichert wird die Datei erst, wenn die Testperson auf die darunter liegende Schaltfläche „Hochladen“ klickt oder direkt danach das Material anlegt beziehungsweise bei Änderungen speichert. In der Aufgabenstellung sollen Schlagwörter und Kategorien zugeordnet werden. Durch diese zusätzlichen Schritte wurde das Material mehrmals nicht erfolgreich hochgeladen. Die Testerinnen und Tester werden nicht informiert beziehungsweise gewarnt. Zweimal wurde eine fehlende Ordnerstruktur bemängelt. Jeweils sieben Testpersonen hatten Probleme bei der Zuordnung der vorgegebenen Schlagwörter („Jahrgang 11“, „Mathematik“, „Klausuren“) und Kategorien („Unterricht“, Unterkategorie „Arbeitsbögen“). Beide Probleme weisen verschiedene Facetten auf, so dass von jeweils drei Problemen bei den Schlagwörtern und Kategorien gesprochen werden muss. Jeweils vier Testpersonen hatten Probleme, die Schaltfläche für die Zuordnung der Schlagwörter beziehungsweise Kategorien zu finden. Diese befindet sich im rechten Bereich der Eingabemaske (markiert). Beim Aufruf der Eingabemaske sind keine Schlagwörter und Kategorien zugeordnet. Dies wird durch die Zahl in Klammern („0“) angezeigt. Bei den Schlagwörtern steht zudem „keine“ in dem Feld. Um eine Zuordnung vorzunehmen, muss auf den Link „Schlagwörter zuordnen“ beziehungsweise „Kategorien zuordnen“ geklickt werden. Anschließend öffnet sich ein neues Dialogfenster. Eine Testperson hatte bei dem Auffinden beider Links Probleme, bei den jeweils drei anderen Testerinnen und Testern handelt es sich um verschiedene Personen.

Der Link der Zuordnung der Schlagwörter öffnet ein überlagertes Dialogfenster (vgl. Abbildung 8.10). In dem Fenster werden alle Schlagwörter alphabetisch in drei Spalten

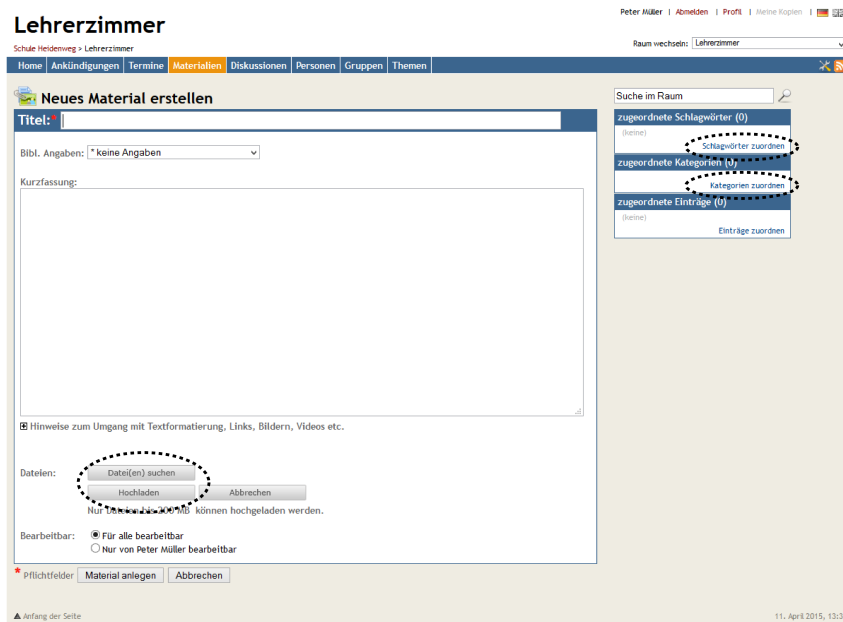


Abbildung 8.9: Usability-Probleme in der Eingabemaske für neue Materialien
rechts sind die Links für die Zuordnungen zu Schlagwörtern beziehungsweise Kategorien markiert; unten sind die Schaltflächen für das Hochladen von Dateien markiert

aufgelistet. Die Auflistung erfolgt spaltenweise. Neben jedem Schlagwort befindet sich ein Auswahlkästchen zur Zuordnung. Durch die hohe Zahl an Schlagwörtern (173) erstreckt sich die Liste über mehrere Bildschirme. Eine Suchfunktion zum Auffinden oder Eingrenzen der Schlagwörter ist nicht vorhanden. Das zwingt die Testerinnen und Tester zum Scrollen. Abschließend muss die Zuordnung durch Bestätigen der Schaltfläche „Schlagwörter zuordnen“ (unten links) abgeschlossen werden. Durch die Größe der Liste ist die Schaltfläche nicht sofort sichtbar. Zwei Testerinnen und Tester haben fälschlicherweise auf das „X“ in der oberen rechten Ecke des Fensters geklickt. Das schließt das Dialogfenster und bricht die Zuordnung ab, unabhängig davon, ob bereits Schlagwörter zugeordnet wurden oder nicht. Die Testperson wird nicht informiert beziehungsweise vor einem Informationsverlust gewarnt. Einer Testperson war es nicht möglich, alle Schlagwörter zuzuordnen.

Die Zuordnung der Kategorien erfolgt in einem ähnlichen Dialogfenster (vgl. Abbildung 8.11). Sie muss analog zu den Schlagwörtern durch einen Klick auf die Schaltfläche „Kategorien zuordnen“ unten links bestätigt werden. Ein Klick auf das „X“ oben rechts bricht die Zuordnung ab. Eine Testperson klickt auf das „X“ und bricht die Zuordnung ab anstatt diese zu speichern. Die Testerin beziehungsweise der Tester wird nicht vor einem möglichen Informationsverlust gewarnt. Die Kategorien können in einer Baumstruktur angelegt werden, die sich prinzipiell unbegrenzt weiter verschachteln lässt. Anfangs wird nur die oberste Ebene eingeblendet. Die Struktur lässt sich durch Klicks auf die Plus- und Minus-Symbole erweitern beziehungsweise reduzieren. Neben den Einträgen befinden sich Auswahlkästchen zur Auswahl einer Kategorie. Eine Suche innerhalb der Kategorien ist nicht vorhanden. Zum erfolgreichen Abschluss der Aufgabe muss die Kategorie



Abbildung 8.10: Usability-Probleme bei der Zuweisung von Schlagwörtern
 oben rechts ist die Schaltfläche zum Verlassen der Zuordnung ohne vorheriges Speichern markiert; unten links ist die Schaltfläche zum Speichern der zugeordneten Schlagwörter markiert



Abbildung 8.11: Usability-Probleme bei der Zuweisung von Kategorien
 die beiden linken Markierungen zeigen die Schaltflächen zum Erweitern beziehungsweise Verkleinern der Baumstruktur; die beiden rechten Markierungen zeigen die auszuwählenden Kategorien

„Unterricht“ erweitert werden, um die Unterkategorie „Arbeitsbögen“ auswählen zu können. Zwei Testpersonen gelingt dies nicht. Sie wählen nur die obere Kategorie aus. Das Auffinden der Unterkategorie gelingt ihnen erst nach einer Hilfestellung des Moderators.

Probleme in Aufgabe 5 (Ankündigung kommentieren)

In der Aufgabe 5 war es den Testerinnen und Testern freigestellt, eine Ankündigung ihrer Wahl zu kommentieren. In InfoChange können alle Ankündigungen, Termine und Materialien mit Anmerkungen versehen werden. Diese befinden sich unter den Angaben des Eintrags auf der Detailseite (vgl. Abbildung 8.12). Bereits erfolgte Anmerkungen werden dem Eintrag angefügt, so dass die Nutzerinnen und Nutzer eventuell unter Umständen scrollen müssen, um zur Eingabemaske zu gelangen. Die Anmerkung besteht aus einem Titel (erforderlich) und einer Nachricht.

Die Aufgabenstellung implizierte, die Übersichtsseite der Ankündigungen aufzurufen (vgl. Abbildung C.5) und dort eine Ankündigung auszuwählen. Eine Testperson hat einen Eintrag ihrer Wahl durch das Auswahlkästchen markiert und im Drop-Down-Menü im unteren Bereich der Übersicht nach einer Option zur Anmerkung gesucht. Eine weitere Testperson gibt keinen Titel, sondern einen Nachrichtentext ein. Sie wird auf eine weitere Seite mit einer Fehlermeldung geleitet, auf der darauf hingewiesen wird, dass das Titel-Feld ausgefüllt werden muss (vgl. Abbildung C.6). Die Bezeichnung des Titels ist rot gefärbt. Vor der blauen Hintergrundfarbe wirkt der Kontrast schwach, so dass dies der Testperson nicht auffällt. Häufiger haben die Testerinnen und Tester Probleme, die richtige Eingabemaske zu finden. Zwei Testpersonen brauchen länger, um die Eingabemaske im unteren Teil der Seite zu finden. Scheinbar verknüpfen sie die Anmerkungen nicht mit der Kommentarfunktion. Zwei Testerinnen und Tester suchen in der Toolbar über dem Eintrag nach einer Möglichkeit Anmerkungen hinzuzufügen. Eine Testperson wählt das „Stift“-Symbol aus, wodurch sie in die Bearbeitungsmaske des Eintrags gelangt. Eine weitere Testperson klickt auf das grüne Plus-Symbol und gelangt in die Eingabemaske für einen neuen Eintrag.

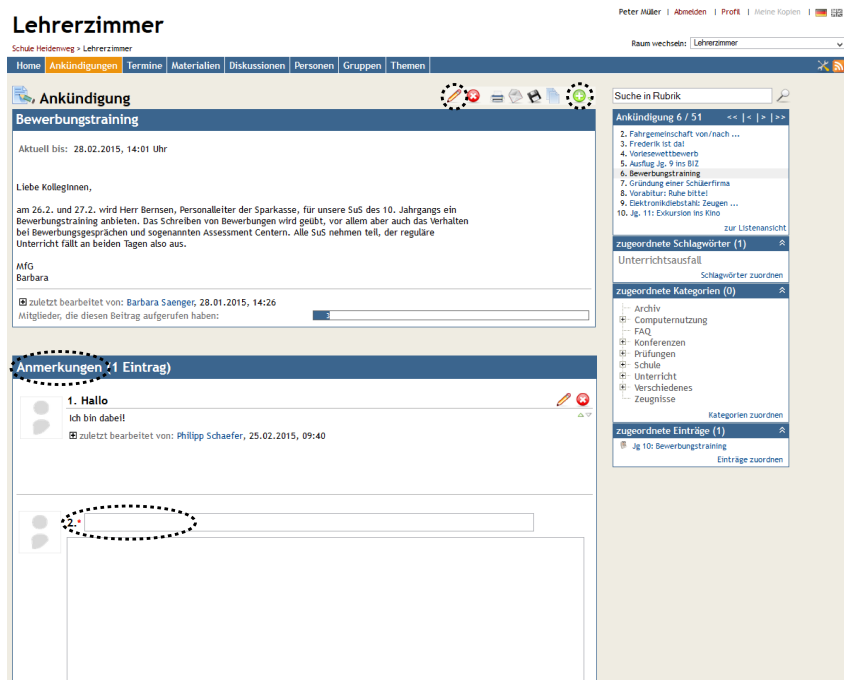


Abbildung 8.12: Usability-Probleme bei der Kommentierung einer Ankündigung
 die Markierung oben mittig zeigt die Schaltfläche zum Ändern der Ankündigung;
 die Markierung oben rechts zeigt die Schaltfläche zum Anlegen eines neuen Eintrags;
 die Markierung links mittig zeigt den Abschnittsbeginn der Anmerkungen;
 die unterste Markierung zeigt das Feld der Überschrift einer neuen Anmerkung

Probleme in Aufgabe 6 (Termin finden)

In der sechsten Aufgabe sollten die Testerinnen und Tester einen bestimmten Termin („Kinobesuch Stauffenberg“) finden. Der lag in der Vergangenheit (November 2014), sodass die Testpersonen auf die Suchfunktion angewiesen waren. Der optimale Lösungsweg ist die Suche über die Startseite (vgl. Abbildung 8.15) oder eine kontextbezogene Suche innerhalb der Termine (vgl. Abbildung C.7). In der Suche über den gesamten Raum werden neben der erwähnten Ankündigung der Termin und zusätzlich bereitgestellte Materialien angezeigt. Neben dem Termin werden Datum mit Uhrzeit und Ort angezeigt. Die Daten werden auch bei der kontextbezogenen Suche über die Rubrik der Termine angezeigt. Eine Testperson fand die Suchfunktion nicht und suchte den Termin händisch, indem sie manuell die Monatsansichten durchsucht hat. Es handelt sich um einen der beiden Ausreißer der Aufgabe. Alle anderen Testerinnen und Tester verwenden die Suchfunktion. Die Suche wird automatisch in den jeweiligen Kontext (die Rubrik) eingeteilt. Befindet sich die Testperson auf der Startseite, erfolgt eine Suche über den gesamten Raum. Die Beschriftung im Suchfeld „Suche im Raum“ weist darauf hin (vgl. linke Seite in Abbildung 8.13). Befindet sich die Testerin beziehungsweise der Tester in einer Rubrik (Ankündigungen, Termine, Materialien), ist die Suche automatisch in den entsprechenden Kontext eingebunden (vgl. rechte Seite in Abbildung 8.13).

Das stellt für drei Testerinnen und Tester ein Problem dar. Sie befinden sich nach der Aufgabe 5 in der Rubrik der Ankündigungen und verwenden die Suchfunktion. Das schränkt die Suche auf die Rubrik Ankündigungen ein (vgl. Abbildung 8.14). In



Abbildung 8.13: Veränderungen im Suchfeld zwischen Startseite und Rubriken

den Ergebnissen wird auf die Einschränkung im rechten Bereich hingewiesen („Rubrik: Ankündigungen“). Der Titel der Seite lautet allerdings „Suche im Raum“. Die Inkonsistenz der Beschriftung ist irreführend. Erschwerend kommt hinzu, dass es zu dem Termin eine Ankündigung gibt. Die Ergebnisseite zeigt rechts neben der Ankündigung ein Datum an. Es handelt sich *nicht* um das Datum des Termins, sondern um das Erstelldatum der Ankündigung. Die Spalte der Datumsangabe ist nicht beschriftet. Keine der drei Testpersonen fällt die Einschränkung auf. Sie werden vom Moderator darauf hingewiesen.

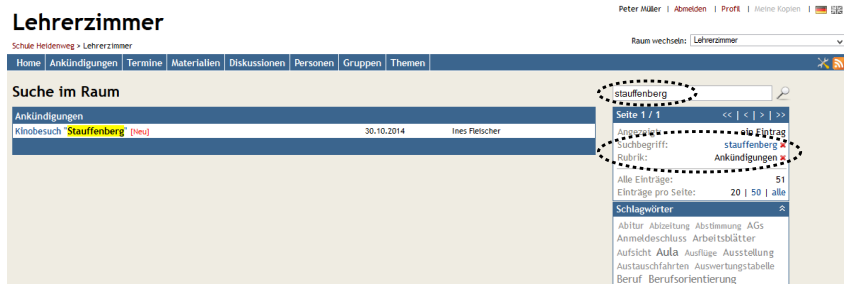


Abbildung 8.14: Suche Stauffenberg über Ankündigungen
die Markierung oben rechts zeigt den gewählten Suchbegriff (stauffenberg); die Markierung darunter zeigt den gewählten Suchbegriff (stauffenberg) und die Rubrik (Ankündigungen), in der die Suche stattfindet



Abbildung 8.15: Suche Stauffenberg
die Markierung oben rechts zeigt den gewählten Suchbegriff (stauffenberg); die Markierung darunter zeigt den gewählten Suchbegriff (stauffenberg), über den die Suche stattfindet

Eine Testperson klickt vor der Eingabe des Suchbegriffs auf das Lupen-Symbol. Dadurch sucht sie nach dem Begriff „Suche in Rubrik“, was keine Ergebnisse liefert. Es erfolgt kein Hinweis auf einen nicht eingegebenen Suchbegriff. Nach einem Hinweis des Moderators korrigiert sie den Suchbegriff und findet den richtigen Eintrag.

Probleme in Aufgabe 7 (Verwendung Schlagwörter)

Die letzte Aufgabe des Nutzertests beinhaltete die Verwendung von Schlagwörtern. Bedingt durch die flache Struktur von InfoChange kommt Schlagwörtern und Kategorien eine entscheidende Rolle bei der Strukturierung des Inhalts zu. In den Produktiv-Logfiles

ist die Verknüpfung der Einträge mit Schlagwörtern und Kategorien sehr umfangreich. Die umfangreiche Verschlagwortung und Kategorisierung wurde im Referenzsystem übernommen. In der Aufgabe sollten alle Einträge mit dem Schlagwort „Jahrgang 12“ angezeigt werden. Die Lösung der Aufgabe ist in Abbildung 8.16 abgebildet. Auf der Startseite müssen die Testpersonen in der Tag Cloud auf den Begriff „Jahrgang 12“ klicken. Die interne Suche gibt dann alle Einträge der Rubriken Ankündigungen, Termine und Nachrichten, die mit dem Schlagwort „Jahrgang 12“ verknüpft sind, aus.

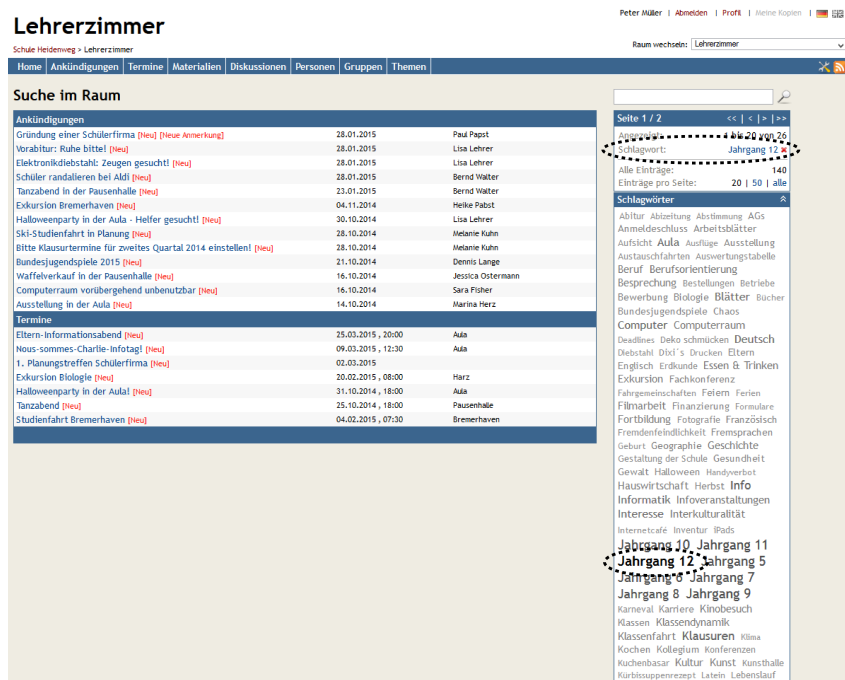


Abbildung 8.16: Auflistung der Einträge mit Schlagwort die Markierung oben rechts zeigt das gewählte Schlagwort (Jahrgang 12); die Markierung unten rechts zeigt das gewählte Schlagwort schwarz hinterlegt

Die Hälfte aller Testerinnen und Tester stieß bei der Lösung der Aufgabe auf Probleme, häufig mehrfach. Sechs Testpersonen suchen zunächst nach dem Schlagwort. Die interne Suchmaschine durchsucht nur die 7 Einträge der Rubriken Ankündigungen, Materialien, Termine, Diskussionen und Gruppen. Personen werden implizit mit in die Suche einbezogen, wenn sie einen Eintrag angelegt oder bearbeitet haben. Die Suchfunktion trennt Suchbegriffe automatisch bei Leerzeichen. Sucht die Testerin beziehungsweise der Tester nach „Jahrgang 12“, werden alle Einträge aufgelistet, die entweder den Begriff „Jahrgang“ oder den Begriff „12“ enthalten. Dies schließt zum Beispiel die Gruppe „Jahrgang 11“ oder Termine mit einer Anfangszeit um „12:30“ mit ein (siehe Abbildung 8.17).

Die Ergebnissseite ermöglicht es, die Suche weiter zu filtern. Dies kann durch die Auswahl eines oder mehrerer Schlagwörter beziehungsweise Kategorien erfolgen. Vier der sechs Testerinnen und Tester grenzen die Suche durch die Auswahl des Schlagwortes „Jahrgang 12“ aus der Tag Cloud ein (vgl. Abbildung 8.18). Prinzipiell handelt es sich um die richtige Handlung. Die Aufgabe ist allerdings nicht gelöst, da die Liste durch den Suchbegriff und das Schlagwort eingegrenzt ist. Alle Testpersonen müssen von dem Moderator auf diese Eingrenzung aufmerksam gemacht werden. Sie wird im rechten

Abbildung 8.17: Suche nach Schlagwort
 die Markierung oben rechts zeigt den genutzten Suchbegriff (Jahrgang 12); die Markierung darunter zeigt das gewählte Schlagwort (Jahrgang 12), über das die Suche erfolgt

Bereich der Zusammenfassung angezeigt. Durch das Löschen des Suchbegriffs kann die Aufgabe erfolgreich abgeschlossen werden.

Zwei der sechs Testpersonen übersehen die Tag Cloud. Sie suchen nach dem Schlagwort und finden dieses nicht in den Suchergebnissen. Sie grenzen die Suche danach nicht weiter ein, sondern verlassen die Suche wieder.

8.3 Bewertung der gefundenen Usability-Probleme

Während der Nutzertests sind die Testerinnen und Tester auf insgesamt 42 Probleme in InfoChange gestoßen. In jeder der sieben gestellten Aufgaben wurden Probleme entdeckt. Die Anzahl der gefundenen Fehler je Aufgabe variiert (vgl. Tabelle 8.3). Insbesondere in den Aufgaben 1, 2, 4 und 5 stießen die Testpersonen auf viele Probleme. Die hohe Anzahl in den Aufgaben 1 und 4 ist nicht überraschend. Allen Testerinnen und Testern war das System vorher unbekannt, sodass mit einer gewissen Eingewöhnungszeit in der ersten Aufgabe zu rechnen ist. Die vierte Aufgabe war die vom Umfang am aufwendigsten, da neben der Beschreibung des Materials eine Datei hochgeladen werden musste und der Eintrag mit Schlagwörtern und Kategorien verknüpft werden sollte. In der zweiten Aufgabe sollten die Testpersonen einen neuen Termin anlegen. Diese Aufgabe war ebenfalls umfangreich und ließ viele Möglichkeiten für Probleme zu. In Aufgabe 5 sollte eine Ankündigung kommentiert werden. Die Testerinnen und Tester stießen auf sechs Probleme.

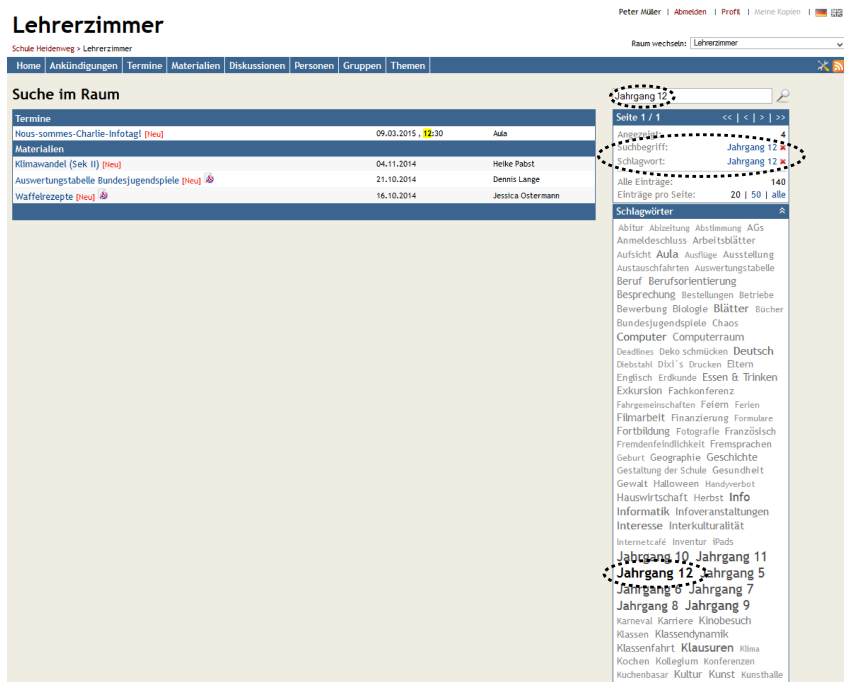


Abbildung 8.18: Suche nach Schlagwort und Rubrik
 die Markierung oben rechts zeigt den genutzten Suchbegriff (Jahrgang 12); die Markierung darunter zeigt das gewählte Schlagwort (Jahrgang 12) und den Suchbegriff (Jahrgang 12) über den die Suche stattfindet; die Markierung unten rechts zeigt das gewählte Schlagwort schwarz hinterlegt

Tabelle 8.3: Gefundene Probleme je Aufgabe

Aufgabe	Probleme	schwerwiegende Probleme
1	9	1
2	8	2
3	4	1
4	9	4
5	6	0
6	3	1
7	3	2

Vor der weiteren Verwendung der Ergebnisse sollten die Probleme hinsichtlich ihrer Reichweite und Schwere gewichtet werden (vgl. Kapitel 6.3.1). Nicht alle Probleme sind ähnlich gravierend. Die Bestimmung der Schwere eines Problems erfolgt durch die Häufigkeit des Auftretens, analog zu Barnum (2011, S.263f). Um als schwer genug für die weitere Betrachtung eingeschätzt zu werden, muss ein Problem bei mindestens 25 Prozent aller Testerinnen und Tester auftreten. Alle anderen Probleme werden von der weiteren Analyse ausgeschlossen. Das Vorgehen bietet zwei Vorteile. Zum einen ist sichergestellt, dass für die Analyse der problematischen Pfadmerkmale genügend Daten vorhanden sind. Zum anderen können vereinzelt Probleme, die durch die Nervosität oder Ungeduld der Testerinnen und Tester ausgelöst wurden, herausgefiltert werden. Darunter fällt zum Beispiel die Testperson, die in Aufgabe 5 auf das „Lupen“-Symbol in der Suche klickt und nach dem Begriff „Suche in Rubrik“ sucht. Anschließend wird für die gefilterten Aufgaben die Reichweite ermittelt. Globale Probleme sind als problematischer einzuschätzen als lokale Probleme.

Von den 42 gefundenen Usability-Problemen bleiben nach der Bestimmung der Schwere 11 schwerwiegende Probleme übrig (vgl. Abbildung 8.19). Darunter fallen ein Problem in Aufgabe 1, zwei Probleme in Aufgabe 2, ein Problem in Aufgabe 3, vier Probleme in Aufgabe 4, ein Problem in Aufgabe 6 und zwei Probleme in Aufgabe 7. In Aufgabe 5 befindet sich kein schwerwiegendes Problem.

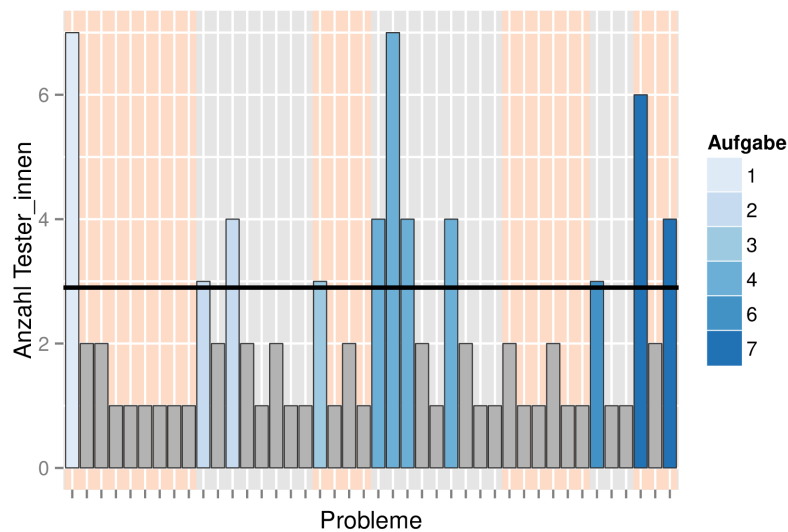


Abbildung 8.19: Schwerwiegende Probleme

Das schwerwiegende Problem der *ersten Aufgabe* ($UP_{A_1P_1}$) ist ein lokales Problem. Es handelt sich um das Finden des „Zusagen“-Buttons auf der Detailseite der Termine (vgl. Abbildung 8.4). Das Problem tritt ausschließlich auf dieser Seite auf: Andere Einträge verfügen nicht über die Funktion. In *Aufgabe 2* wurden zwei Probleme als schwerwiegend identifiziert. Bei beiden Problemen handelt es sich um lokale und globale Probleme. Drei Testpersonen „stolpern“ über das Feld für das Enddatum ($UP_{A_2P_1}$) und vier Testerinnen und Tester bereitet das gewünschte Eingabeformat der Zeit Probleme ($UP_{A_2P_3}$). Das Problem der Eingabe des Enddatums kommt ausschließlich in der Rubrik der Termine

vor. Zeitangaben kommen auch in anderen Rubriken vor. Beim Erstellen von neuen Ankündigungen können diese bis zu einem von der Erstellerin beziehungsweise Ersteller bestimmten Zeitpunkt als aktuell definiert werden. Daher ist das Zeitformat als globales Problem zu verstehen. Es ist der Kategorie des Feedbacks zuzuordnen, da den Testerinnen und Testern keine Hilfe bezüglich des Eingabeformats angezeigt wird. In der *dritten Aufgabe* wurde ein schwerwiegendes Problem entdeckt ($UP_{A_3P_1}$). Bei der Suche nach dem Material nutzen drei Testerinnen und Tester den Weg über die Personenseiten. Alle können die Detailseite der gesuchten Person (Frau Kuhn) nicht finden. Die Auflistung erstreckt sich über mehrere Seiten. Die Schaltflächen zum Weiterblättern befinden sich im oberen rechten Bereich der Seite und nicht wie üblich am Anfang oder am Ende der Auflistung. Es handelt sich um ein globales Problem, da es bei allen Auflistungen vorkommt. Das Problem fällt dabei unter die Kategorie Navigation. Die vier schwerwiegenden Probleme in *Aufgabe 4* sind globale Probleme. Sieben Testerinnen und Tester haben Probleme bei dem Hochladen der Datei des neuen Materials ($UP_{A_4P_2}$). Die Möglichkeit, Dateien an einen Eintrag anzuhängen, besteht auch in den Rubriken Ankündigungen, Termine, Diskussionen und Themen. Das Problem ist als global einzustufen und ist in die Kategorien der Informationsdarstellung oder des Feedbacks einzusortieren. In die erste Kategorie, weil die Informationsdarstellung suggeriert, dass die Datei bereits hochgeladen wurde. Gleichzeitig fehlt das Feedback, dass der Prozess des Hochladens noch nicht abgeschlossen ist. Ein weiteres Problem besteht für die Testerinnen und Tester darin, dass sie die Schaltfläche (das grüne Plus-Symbol) für das Anlegen eines neuen Materials nicht finden ($UP_{A_4P_1}$). Die Schaltfläche wird für das Anlegen neuer Einträge in jeder Rubrik verwendet. In der zweiten Aufgabe haben zwei Testpersonen das Problem, die entsprechende Schaltfläche für das Anlegen eines neuen Termins zu finden. Dies ist ein globales Problem, welches der Kategorie der Informationsdarstellung zugeordnet wird. Vier Testpersonen haben Probleme bei der Zuordnung von Schlagwörtern ($UP_{A_4P_3}$) und Kategorien ($UP_{A_4P_6}$), insbesondere den Link zum Zuordnen der Schlagwörter beziehungsweise der Kategorien zu finden. Der Link befindet sich im rechten Bereich der Seite. Die Probleme beziehen sich nicht auf Vorgehensweisen innerhalb der Funktionen. Beide Probleme sind als global einzustufen, da neue Einträge in den Rubriken Ankündigungen, Termine und Diskussionen mit Schlagwörtern und Kategorien versehen werden können. Sie lassen sich der Kategorie Navigation zuordnen. In *Aufgabe 6* konnte ein schwerwiegendes Problem identifiziert werden ($UP_{A_6P_1}$). Drei Testerinnen und Tester haben statt des Termins die separate Ankündigung aufgerufen. In der Ankündigung werden Ort, Datum und Uhrzeit genannt. Damit können zwar alle benötigten Informationen aufgerufen werden, allerdings müssen separate Ankündigungen erstellt werden. Bei fast allen anderen Terminen können die Informationen nicht über diesen Weg beschafft werden. Problematisch ist die Verwendung der Suchfunktion bei allen drei Testpersonen. Sie befinden sich nach der vorherigen Aufgabe in den Ankündigungen und suchen automatisch innerhalb dieser Rubrik. Dies wird durch den Text „Suche in Rubrik“ (vgl. Abbildung 8.13) deutlich. In den Ergebnissen ist die Eingrenzung der Suche auf eine Rubrik (in diesem Fall Ankündigungen) durch die Überschrift (vgl. Abbildung 8.14) und in der detaillierten Beschreibung der Suchbegriffe im rechten Bereich der Ergebnisseite ersichtlich. Das Problem ist global, da die Eingrenzung

der jeweiligen Rubrik automatisch erfolgt. Es sind Probleme der Kategorie der Informationsdarstellung durch die nicht deutliche Darstellung der Eingrenzung und der Kategorie des Feedbacks durch die nicht vorhandene Rückmeldung über die Eingrenzung. Die letzten beiden Probleme betreffen die Suchfunktion und wurden in der *Aufgabe 7* identifiziert. Sechs Testpersonen suchen nach dem in der Aufgabenstellung geforderten Schlagwort anstatt dieses im rechten Bereich der Seite auszuwählen ($UP_{A_7P_1}$, vgl. Abbildung 8.16). Schlagwörter und Kategorien werden nicht mit in die Suche einbezogen. Dies ist nicht ersichtlich. Es handelt sich um ein globales Problem und betrifft die Kategorien der Informationsdarstellung und des Feedbacks. Bei dem zweiten schwerwiegenden Problem der Suchfunktion haben vier Testerinnen und Tester ein Problem mit der Darstellung der Suchergebnisse ($UP_{A_7P_3}$). Sie grenzen die Suche mit dem Schlagwort und dem Schlagwort als Suchbegriff doppelt ein (vgl. Abbildung 8.17). Das Problem ist die Darstellung der beiden Begriffe. Sie ist für die Testpersonen nicht sofort eindeutig. Mehrere Testerinnen und Tester haben angenommen, dass sie eine neue Suche beginnen anstatt die vorhandenen Ergebnisse weiter einzugrenzen. Dieses Problem tritt in der Suchfunktion auf und ist lokal.

Tabelle 8.4: Beschreibung schwerwiegende Probleme

Bezeichnung	Aufgabe	Problem	Beschreibung
$UP_{A_1P_1}$	1	1	Finden des „Zusage“-Buttons
$UP_{A_2P_1}$	2	1	Eingabe des Enddatums bei Terminen
$UP_{A_2P_3}$	2	3	Format Zeitangabe
$UP_{A_3P_1}$	3	1	Fehlende Verlinkung auf der Detailseite der Benutzer_innen
$UP_{A_4P_1}$	4	1	Finden der Eingabemaske neuer Materialien
$UP_{A_4P_2}$	4	2	Hochladen einer Datei
$UP_{A_4P_3}$	4	3	Zuordnen von Schlagwörtern
$UP_{A_4P_6}$	4	6	Zuordnen von Kategorien
$UP_{A_6P_1}$	6	1	Auffinden von Termin durch Ankündigung
$UP_{A_7P_1}$	7	1	Suche nach Schlagwort
$UP_{A_7P_3}$	7	3	Kombination von Suche nach Schlagwort und Schlagwort

9 Kombination von Nutzertests und Logfiles

In dem letzten Schritt erfolgt die Kombination von den Produktiv-Logfiles mit den Ergebnissen der Nutzertests. Zunächst werden beide Logfiles in ein Format gebracht, das einen Vergleich ermöglicht. Die Darstellung erfolgt über eine *Transaction Matrix* (Liu, 2011). Diese Art der Darstellung ist im Web Mining verbreitet und bietet eine einfache und gleichzeitig flexible Speicherung der Informationen. Die Umwandlung der Daten von Vektoren in die Matrizen ist mit einem Informationsverlust verbunden. In der Matrix werden die summierten Zugriffe abgebildet. Der Informationsverlust ist zu untersuchen und zu minimieren. Die Umwandlung muss für beide Datenquellen (Produktivsystem und Referenzsystem) durchgeführt werden. Die Logfiles der beiden Systeme müssen nicht aneinander angepasst werden, da beide aus demselben Anwendung stammen. Anschließend erfolgt die Kombination der Daten durch ein Distanzmaß inklusive anschließender Evaluation.

9.1 Umwandlung der Logfiles in eine Transaction Matrix

Die Logfiles müssen für den Vergleich in ein anderes Format gebracht werden. Tabelle 9.1 zeigt einen beispielhaften Besuch aus den Produktiv-Logfiles. Die Tabelle enthält nur relevante Spalten. Aus der Tabelle lassen sich die Besuche über verschiedene Vektoren abbilden. Der Vektor mit dem höchsten Informationsgrad basiert auf den internen IDs der Seiten und der Verweildauer auf den einzelnen Seiten.¹ Ein beispielhafter Besuch sieht als Vektor wie folgt aus:

107 (2s) ⇒ 101 (39s) ⇒ 2694203 (19s) ⇒ 100 (5s) ⇒ 102 (2s) ⇒ 101 (23s) ⇒ 201 (29s) ⇒ 101 (22s) ⇒ 100 (18s)

Bei der Darstellungsweise gehen kaum Informationen verloren. Den Informationen kann der Inhalt der Seiten nicht entnommen werden. Die Übersichtsseiten, das Erstellen beziehungsweise das Bearbeiten von Einträgen und die Startseite haben eine dreistellige Kennung. Die ID der anderen Einträge gibt keine Auskunft über die Rubrik des Eintrags. Ferner ist beispielsweise bei den Materialien nicht ersichtlich, ob der Eintrag angeschaut (Detailansicht) oder eine Datei heruntergeladen wurde. Die obige Darstellung macht nur Sinn, wenn sie mit zusätzlichen Informationen erweitert wird:

107; Ankündigungen; Übersicht (2s) ⇒ 101; Termine; Übersicht (39s) ⇒ 2694203; Termine; Detailansicht (19s) ⇒ 100; Startseite; Übersicht (5s) ⇒

¹Die Verweildauer auf der letzten Seite kann nicht gemessen werden und wird mit dem Durchschnitt der Verweildauer auf den anderen Seiten berechnet.

Tabelle 9.1: Beispiel eines Logfiles (ausgewählte relevante Spalten)

timestamp	request	method	user_id	item_id	module	fct	param	time_spend
25.09.2011 13:04:23	/index.php?cid=501266&mod=home&fct=index	GET	928415	100	home	index		26
25.09.2011 13:04:49	/index.php?cid=501266&mod=announcement&fct=index	GET	928415	107	announcement	index		2
25.09.2011 13:04:51	/index.php?cid=501266&mod=date&fct=index	GET	928415	101	date	index		39
25.09.2011 13:05:30	/index.php?cid=501266&mod=date&fct=detail&iid=2694203&mode=private&month=20111001&presentation_mode=2	GET	928415	2694203	date	detail	iid=2694203&mode=private&month=20111001&presentation_mode=2	19
25.09.2011 13:05:49	/index.php?cid=501266&mod=home&fct=index	GET	928415	100	home	index		5
25.09.2011 13:05:54	/index.php?cid=501266&mod=material&fct=index	GET	928415	102	material	index		2
25.09.2011 13:05:56	/index.php?cid=501266&mod=date&fct=index	GET	928415	101	date	index		23
25.09.2011 13:06:19	/index.php?cid=501266&mod=date&fct=edit&iid=NEW&day=1&month=20111001&year=2011&week=1316998800&presentation_mode=1&time=0&modus_from=calendar	GET	928415	201	date	edit	iid=NEW&day=1&month=20111001&year=2011&week=1316998800&presentation_mode=1&time=0&modus_from=calendar	29
25.09.2011 13:06:48	/index.php?cid=501266&mod=date&fct=index&browse=week&week=1317603600&presentation_mode=1	GET	928415	101	date	index	browse=week&week=1317603600&presentation_mode=1	22
25.09.2011 13:07:10	/index.php?cid=501266&mod=home&fct=index	GET	928415	100	home	index		18*

102; Materialien; Übersicht (2s) \Rightarrow 101; Termine; Übersicht (23s) \Rightarrow
 201; Termine; Bearbeiten (29s) \Rightarrow 101; Termine; Übersicht (22s) \Rightarrow
 100; Startseite; Übersicht (18s)

In diesem Fall handelt es sich um die Information bezüglich der Rubrik und der Funktion. Die zusätzlichen Informationen lassen eine genauere Untersuchung der Besuche zu. Bei der Bestimmung der Gebrauchstauglichkeit sind generische Merkmale von gesteigertem Interesse. Der Inhalt eines bestimmten Eintrags ist von geringerem Interesse, da der Aufbau der Seiten stets identisch ist. Jede Übersichtsseite, Detailseite oder Editierseite ist gleich aufgebaut. Unterschiede bestehen zwischen den Rubriken. Innerhalb der Rubriken kommen ist der Aufbau der Einträge bei Beibehaltung der Funktionen stets gleich. Die Information der verbrachten Zeit pro Seite ist ambivalent zu betrachten. Zwar erhöhen die Zeiten den Informationsgehalt des Vektors. Mit den Angaben können beispielsweise Besuche auf Basis der verbrachten Zeit verglichen werden. Für die Verwendung in der Kombination sprechen allerdings zwei negative Argumente. Zum einen sind die Nutzerinnen und Nutzer der Produktiv-Logfiles sehr heterogen hinsichtlich ihrer Aktivität in InfoChange (vgl. Abbildung 7.8). Die Logfiles geben keinen Aufschluss darüber, ob sich die Nutzerinnen und Nutzer sinnvoll nach diesem Merkmal trennen beziehungsweise kombinieren lassen. Zum anderen weist die Usability-Forschung darauf hin, dass die Verwendung der Methode des lauten Denkens einen Einfluss auf die Bearbeitung der Aufgaben hat (Dumas und Loring, 2008; Kraemer und Ummelen, 2004; van den Haak et al., 2003). Die Zeit verlängert sich durch die Anwendung der Methode. Aus diesen Gründen werden die Seiten-ID und die verbrachte Zeit pro Seite von den weiteren Analysen ausgeschlossen. Der Beispielvektor des Besuches sieht wie folgt aus:

Ankündigungen; Übersicht \Rightarrow Termine; Übersicht \Rightarrow Termine; Detailansicht
 \Rightarrow Startseite; Übersicht \Rightarrow Materialien; Übersicht \Rightarrow Termine; Übersicht
 \Rightarrow Termine; Bearbeiten \Rightarrow Termine; Übersicht \Rightarrow Startseite; Übersicht

Problematisch bei dem weiteren Vorgehen mit den Vektoren ist die unterschiedliche Länge. Um zwei oder mehr Vektoren miteinander vergleichen zu können, werden diese in eine *Transaction Matrix* (oder auch *Transaction Data*; vgl. Chitrea und Thanamani, 2011; Liu, 2011; Mobasher, Dai, Luo und Nakagawa, 2002; Pal et al., 2002; Wang et al., 2009) überführt. Jede Zeile einer Matrix steht für einen Besuch, jede Spalte für eine Eigenschaft, in diesem Fall die Rubrik oder die Funktion. Bei der Überführung der Vektoren in eine *Transaction Matrix* geht die Reihenfolge der Seitenaufrufe verloren. Der Verlust wird als verkraftbar gewertet, da die Reihenfolge bei der Berechnung der Ähnlichkeiten keine Rolle spielt. Die Spalten werden nach der Anzahl der Zugriffe auf die Rubriken und Funktionen hochgezählt. Die einzelnen Werte der Vektoren enthalten zwei Informationen (Rubrik, Funktion), was zwei verschiedene *Transaction Matrizen* ermöglicht. Eine Matrix separiert Rubrik und Funktion (*Rub-Funk-Matrix*), in der zweiten Matrix werden beide Informationen zusammen gespeichert (*Komb-Matrix*). Der Informationsgehalt der *Komb-Matrix* ist größer, da die jeweilige alle Kombination aus Rubrik und Funktion als Spalten abgebildet werden. Die *Rub-Funk-Matrix* verfügt dagegen über eine höhere Generalisierbarkeit und ist kleiner (weniger Spalten). Die Gefahr einer „dünnbesetzten“ (engl. *Sparse Matrix*; Golub und van Loan, 1996) Matrix ist geringer. In den Tabellen 9.2 und 9.3 ist der Beispielvektor

abgebildet. Zusätzlich wurde ein weiterer Besuch abgebildet.² Alle Besuche werden vor der weiteren Bearbeitung durch die Gesamtlänge des Besuchs (Aktionen) normiert. Die Daten sind dadurch prozentual zu interpretieren und lassen einen einfacheren Vergleich zu.

Tabelle 9.2: Darstellung als Rub-Funk-Matrix

A	G	B	D	F	M	St	Su	T	D	B	Ü	H	K	S
0,11	0,00	0,00	0,00	0,00	0,11	0,22	0,00	0,56	0,11	0,11	0,78	0,00	0,00	0,00
0,00	0,12	0,00	0,00	0,00	0,88	0,00	0,00	0,00	0,50	0,00	0,19	0,31	0,00	0,00

Abkürzungen erster Teil: A = Ankündigungen; B = Benutzer(-verwaltung); D = Diskussionen; F = Fächer; G = Gruppen; M = Materialien; St = Startseite; Su = Suche; T = Termine
 Abkürzungen zweiter Teil: B = Bearbeiten; D = Detailansicht; K = Kategoriensuche; S = Schlagwortsuche; Ü = Übersichtsseite

Tabelle 9.3: Darstellung als Komb-Matrix

A_B	A_D	A_K	A_S	A_Ü	B_B	B_D	B_Ü	D_B	D_D	D_K	D_Ü
0,00	0,00	0,00	0,00	0,11	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F_B	F_D	F_Ü	G_B	G_D	G_Ü	M_B	M_D	M_H	M_K	M_S	M_Ü
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,11
0,00	0,00	0,00	0,00	0,06	0,06	0,00	0,44	0,31	0,00	0,00	0,13
St_Ü	Su_K	Su_S	Su_Ü	T_B	T_D	T_Ü					
0,22	0,00	0,00	0,00	0,11	0,11	0,33					
0,00	0,00	0,00	0,00	0,00	0,00	0,00					

Abkürzungen erster Teil: A = Ankündigungen; B = Benutzer(-verwaltung); D = Diskussionen; F = Fächer; G = Gruppen; M = Materialien; St = Startseite; Su = Suche; T = Termine
 Abkürzungen zweiter Teil: B = Bearbeiten; D = Detailansicht; K = Kategoriensuche; S = Schlagwortsuche; Ü = Übersichtsseite

Die Komb-Matrix hat im Vergleich zu der Rub-Funk-Matrix eine höhere Anzahl an Spalten, da die Kombinationsmöglichkeiten zahlreicher sind. Eine Folge ist die höhere Anzahl der leeren Felder (mit null Zugriffen). Durch die Trennung von Rubrik und Funktion ist die Zeilensumme der Rub-Funk-Matrix stets zwei (entspricht 200 Prozent), da die Summe der Zugriffe im Beispiel auf die Rubriken und Funktionen gebildet wird. Die Abbildung des ersten Vektors zeigt, dass jeweils elf Prozent der Zugriffe auf die Rubriken der Ankündigungen (Spalte A) und Materialien (Spalte M) entfallen. 22 Prozent aller Zugriffe entfallen auf die Startseite (Spalte St) und am Häufigsten hält sich die Nutzerin beziehungsweise der Nutzer in der Rubrik der Termine auf (56 Prozent; Spalte T). Von den angebotenen Funktionen nutzt die Nutzerin beziehungsweise der Nutzer zu fast vier Fünfteln primär die Übersichten. Mit jeweils elf Prozent werden die Funktionen der

²Es handelt sich um den folgenden Vektor:

Gruppen; Übersicht ⇒ Gruppen; Detailansicht ⇒ Materialien; Übersicht ⇒ Materialien; Detailansicht ⇒ Materialien; Detailansicht ⇒ Materialien; Detailansicht ⇒ Materialien; Herunterladen ⇒ Materialien; Detailansicht ⇒ Materialien; Herunterladen ⇒ Materialien; Übersicht ⇒ Materialien; Detailansicht ⇒ Materialien; Herunterladen ⇒ Materialien; Detailansicht ⇒ Materialien; Detailansicht ⇒ Materialien; Detailansicht ⇒ Materialien; Herunterladen

Detailansicht und des Bearbeitens genutzt. Insgesamt bleiben acht der 15 Spalten leer beziehungsweise weisen keine Zugriffe auf. In der Komb-Matrix ist die Darstellung der Zugriffe differenzierter. Jeweils elf Prozent der Zugriffe entfallen auf die Übersichtsseite der Ankündigungen (Spalte A_Ü), die Übersichtsseite der Materialien (Spalte M_Ü), das Erstellen oder Bearbeiten eines Termins (Spalte T_B) und die Detailansicht der Termine (Spalte T_D). Die Startseite wird mit 22 Prozent der Zugriffe häufiger besucht (Spalte St_Ü). Die Nutzerin beziehungsweise der Nutzer bewegt sich am Häufigsten in der Übersichtsseite der Termine (33 Prozent; Spalte T_Ü). In der Darstellung wird ersichtlich, dass sich Zugriffe auf die Rubriken der Ankündigungen, Materialien und der Startseite auf die Übersichtsseite beschränken. Dieser Detailgrad ist in der Rub-Funk-Matrix nicht vorhanden. Zudem kann ermittelt werden, dass es sich bei dem neu angelegten beziehungsweise bearbeiteten Eintrag um einen Termin handelt. Diese Schlussfolgerung ist nicht bei der Rub-Funk-Matrix möglich. Der Nachteil ist, dass 25 der 31 Spalten leer sind beziehungsweise keine Zugriffe aufweisen. Sechs Spalten verfügen über entsprechende Werte. In der Rub-Funk-Matrix sind es sieben Spalten.

Die Abbildung des zweiten Vektors erfolgt analog. Der Besuch fokussiert den Bereich der Materialien. In der Rub-Funk-Matrix verweisen 88 Prozent aller Zugriffe auf die Rubrik (Spalte M). Die restlichen 12 Prozent der Zugriffe bewegen sich in der Rubrik der Gruppen (Spalte G). Die Hälfte der Zugriffe ruft die Detailseite eines Eintrags auf (Spalte D), knapp 20 Prozent der Zugriffe entfallen auf die Funktion der Detailansicht und etwa 30 Prozent aller Zugriffe der Nutzerin beziehungsweise des Nutzers laden eine Datei herunter. Der Vergleich mit der Komb-Matrix offenbart erneut mehr Details. Die Zugriffe auf die einzelnen Rubriken teilen sich im Vergleich zum ersten Vektor auf verschiedene Funktionen auf. In der Rubrik der Gruppen verteilen sich jeweils sechs Prozent der Zugriffe auf die Detailansicht (Spalte G_D) und die Übersichtsseite (Spalte G_Ü). In der Rubrik der Materialien bewegt sich die Nutzerin beziehungsweise der Nutzer auf den Detailseiten (44 Prozent; Spalte M_D) und lädt Dateien herunter (31 Prozent; Spalte M_H). Die restlichen 13 Prozent der Zugriffe entfallen auf die Übersichtsseite der Materialien (Spalte M_Ü). Wie beim ersten Vektor ist die Komb-Matrix dünn besetzt.

Die beiden aufgeführten Besuche stammen aus den Produktiv-Logfiles. Im weiteren Verlauf dieses Kapitels werden Besuche aus den Produktiv-Logfiles mit den Logfiles der Nutzertests kombiniert. Die Ähnlichkeit zwischen zwei Besuchen beziehungsweise den daraus extrahierten Vektoren, wird anhand der Kosinus-Ähnlichkeit (vgl. Gleichung 5.2.2) bestimmt. Bei der Berechnung der beiden Matrizen ergeben sich Unterschiede in der Bestimmung der Ähnlichkeit. Die beiden Besuche werden anhand der Rub-Funk-Matrix mit einer Ähnlichkeit von 0,275 bestimmt. Bei der Verwendung der Komb-Matrix liegt die Ähnlichkeit nach dem Kosinus bei 0,054. Bei einem Wert von 0 kann von keiner Ähnlichkeit gesprochen werden (vergleichbar mit einem 90°-Winkel). Der Wert 1 steht für die vollständige Ähnlichkeit (vergleichbar mit einem 0°-Winkel). In der Rub-Funk-Matrix weisen drei Spalten Werte beider Vektoren auf (Spalten M, D und Ü). Dies entspricht etwa der Hälfte aller besuchten Rubriken und Funktionen der beiden Vektoren. In der Komb-Matrix existiert nur eine Spalte, auf die beide Vektoren zugreifen (Spalte M_Ü). Um diese Erkenntnis generalisieren zu können, wurden aus den 2.197 Besuchen der

Produktiv-Logfiles 100 Vergleiche von zwei zufällig ausgewählten Besuchen durchgeführt. Die ausführliche Darstellung findet sich im Anhang in Tabelle B.3.

Die 100 Vergleiche³ lassen folgenden Schluss zu. In der Rub-Funk-Matrix beträgt die Ähnlichkeit im Durchschnitt 0,64 (sd = 0,23). Der Durchschnitt der Komb-Matrix ist geringer. Er liegt bei 0,41 (sd = 0,29). Der Unterschied zwischen den beiden Mittelwerten ist höchstsignifikant.⁴ Die Unterschiede können auf die gesamten Besuche übertragen werden. In der Tabelle wird ein Nachteil der Rub-Funk-Matrix ersichtlich. Zieht man beispielsweise die Vergleiche in Zeile 12 oder 15 heran (auf der zweiten Seite der Tabelle), werden dort zwei Besuche miteinander verglichen, die *keine* gleichen Rubriken und Funktionen aufrufen. In der Komb-Matrix werden die Ähnlichkeiten mit 0 bestimmt. Die Rub-Funk-Matrix gibt eine Ähnlichkeit von 0,306 beziehungsweise 0,39 aus. In dem Vergleich 12 werden in beiden Besuchen Übersichtsseiten aufgerufen. Im Vergleich 15 werden Übersichtsseiten und Detailansichten genutzt. In beiden Vergleichen sind die Rubriken, in welchen die Funktionen eingesetzt werden, unterschiedlich. Dies lässt vermuten, dass die Rub-Funk-Matrix zum Überschätzen der Ähnlichkeiten neigt. In der Abbildung 9.1 ist der Unterschied in einem Histogramm dargestellt. Es wird ersichtlich, dass die Rub-Funk-Matrix eine stärkere Ähnlichkeit betont. 44 der 100 Vergleiche haben eine Ähnlichkeit von 0,75 oder mehr. In der Komb-Matrix haben 15 Vergleiche eine Ähnlichkeit von 0,75 oder mehr. Etwa ein Drittel aller Vergleiche sind kleiner oder gleich 0,25. Zwischen diesen Extremen sind die Ähnlichkeiten gleichverteilt. Bei der Rub-Funk-Matrix steigen die Fallzahlen parallel zur Kosinus-Ähnlichkeit. Dies spricht dafür, die Komb-Matrix als Grundlage für die Transaction Matrix zu betrachten.

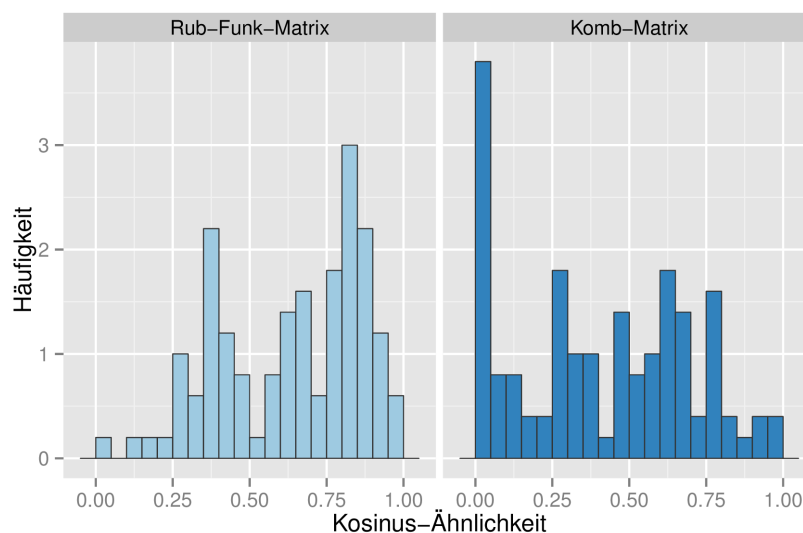


Abbildung 9.1: Vergleich der Ähnlichkeiten der Transaction Matrix

³Der beispielhafte Vergleich der beiden Vektoren befindet sich in der Tabelle in Zeile 18.

⁴Durch einen Wilcoxon-Rang-Test ermittelt (Hollander et al., 2013; Siegel, 2001). Vorherige Tests mit dem Shapiro-Wilk-Test (Royston, 1982a,b) haben ergeben, dass bei beiden Variablen keine Normalverteilung vorliegt.

9.2 Vom Problempfad zum problematischen Pfadmerkmal

Die Umwandlung der Besuche und der daraus resultierenden Pfade in eine Transaction Matrix muss für die Produktiv-Logfiles der InfoChange-Installation und für das Referenzsystem erfolgen. Im vorherigen Kapitel wurde die Komb-Matrix als optimale Transaction Matrix bestimmt. Verglichen mit der Rub-Funk-Matrix bietet die Komb-Matrix eine größere Detailtiefe und neigt zu einer Gleichverteilung der Kosinus-Ähnlichkeiten.

Während die Umwandlung bei den Produktiv-Logfiles direkt erfolgen kann, müssen die Logfiles der Nutzertests vorher weiter aufgearbeitet werden. In Kapitel 8.3 wurden diejenigen Probleme als schwerwiegend definiert, die mindestens drei Testerinnen und Testern auffielen (vgl. Tabelle 9.4). Für jedes der elf schweren Probleme liegen zwischen drei und sieben *Problempfade* vor. Diese Problempfade werden in eine Transaction Matrix überführt. Der Prozess ist analog zu dem im vorherigen Kapitel beschriebenen Vorgehen. Nach der Umwandlung wird nicht mehr von Problempfaden, sondern von *problematischen Pfadmerkmalen* gesprochen. Die Matrizen enthalten mehrere Zeilen, je eine Zeile für jede Testperson, die das Problem identifiziert hat. An diesem Punkt bieten sich zwei Vorgehensweisen an. Zum einen können die individuellen problematischen Pfadmerkmale mit den Produktiv-Logfiles verglichen werden. Das hätte den Vorteil, dass Unterschiede zwischen einzelnen problematischen Pfadmerkmalen eines Problems detaillierter herausgearbeitet werden könnten. Allerdings würden die bisherigen Analysen der gefundenen Probleme dann nicht mehr ausreichen. Dort wurde das gefundene Usability-Problem beschrieben, ohne auf eventuelle Besonderheiten der einzelnen Testerinnen und Tester einzugehen. Diese Vorgehensweise entsprach dem Ziel des Nutzertests, allgemeine Probleme zu identifizieren. Diesem Vorgehen folgend werden die drei bis sieben problematischen Pfadmerkmale für jedes Problem kombiniert. Das Ergebnis sind *generische problematische Pfadmerkmale* (pPM_{A_i,P_i}). Die Kombination erfolgt durch die Berechnung des Mittelwertes aller Spalten der Matrix. Die generischen problematischen Pfadmerkmale enthalten dann weiterhin die prozentuale Verteilung der Zugriffe auf die einzelnen Kombinationen von Rubrik und Funktion (Komb-Matrix, vgl. Abbildung 9.2).

Wichtig für die Kombination zwischen den Logfiles des Testsystems und des Produktivsystems ist der Umstand, dass sich die generischen Pfadmerkmale der Usability-Tests unterscheiden. Eine zu hohe Ähnlichkeit der einzelnen generischen Pfadmerkmale würde dazu führen, dass Pfadmerkmale der „realen“ Logfiles multiplen Usability-Problemen zugeordnet werden. Die Unterschiedlichkeit der generischen Pfadmerkmale ist eine Voraussetzung für die erfolgreiche Anwendung des Verfahrens (vgl. Abbildung 3.3). In Abbildung 9.2 sind die Zugriffe der generischen problematischen Pfadmerkmale der schwerwiegenden Usability-Probleme dargestellt. Auf der X-Achse sind die verschiedenen schwerwiegenden Usability-Probleme dargestellt. Auf der Y-Achse werden die einzelnen Rubriken und Funktionen abgetragen. Zur höheren Übersichtlichkeit wurden die Bezeichnungen abgekürzt. Die Legende befindet sich in der Bildunterschrift. Zudem sind die einzelnen Aufgaben per vertikalem Strich voneinander getrennt. Bei den schwerwiegenden Usability-Problemen derselben Aufgabe ist mit einer größeren Übereinstimmung der generischen problematischen Pfadmerkmale zu rechnen, da die

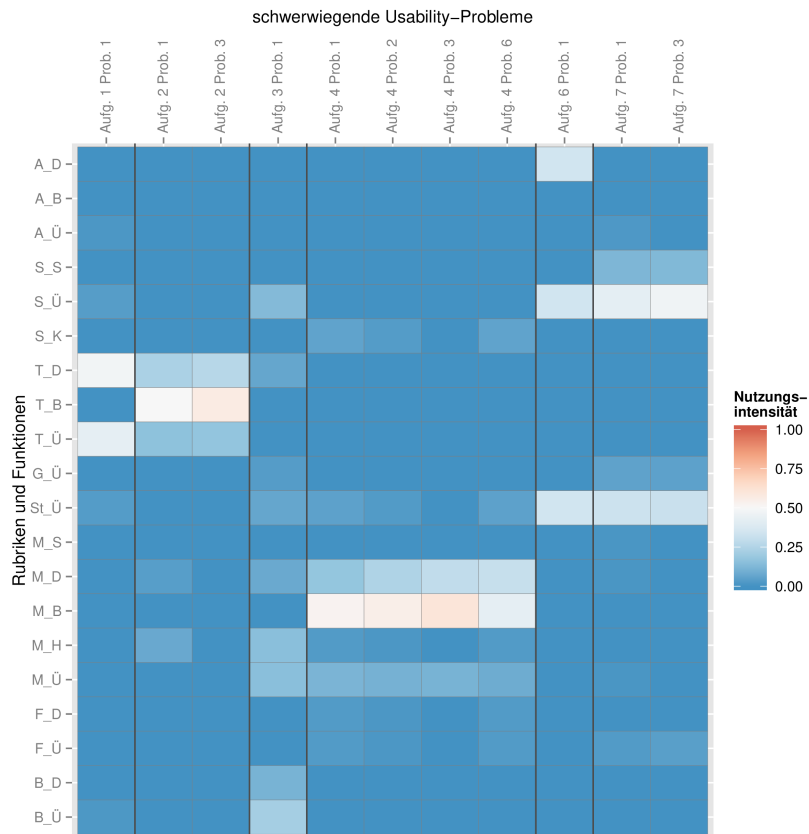


Abbildung 9.2: Zugriffe der generischen problematischen Pfadmerkmale auf Rubriken und Funktionen

Abkürzungen erster Teil: A = Ankündigungen; B = Benutzer(-verwaltung); F = Fächer; G = Gruppen; M = Materialien; St = Startseite; Su = Suche; T = Termine
 Abkürzungen zweiter Teil: B = Bearbeiten; D = Detailansicht; K = Kategoriensuche; S = Schlagwortsuche; Ü = Übersichtsseite

Testerinnen und Tester sich in ähnlichen Rubriken aufgehalten haben dürften. Diese Vermutung wird für die Aufgaben 2, 4 und 7 bestätigt. Die Testerinnen und Tester bewegen sich bei der zweiten Aufgabe (Erstellen eines neuen Termins) vor allem in der Rubrik der Termine, wobei der Schwerpunkt jeweils in der Bearbeiten-Funktion liegt. Die generischen Pfadmerkmale des ersten Problems der zweiten Aufgabe (Eingabe des Enddatums bei Terminen) tangieren leicht die Rubrik der Materialien. Dadurch ergibt sich ein Unterschied zwischen den beiden Problemen. Bei den schwerwiegenden Usability-Problemen der Aufgaben 4 (Anlegen eines neuen Materials) und 7 (Anzeigen aller Einträge mit dem Schlagwort „Jahrgang 12“) sind ähnliche Auffälligkeiten. Die Testerinnen und Tester, die die entsprechenden Probleme hatten, bewegen sich primär in der entsprechenden Rubrik. Dennoch sind leichte Unterschiede in der Intensität der Nutzung der Rubriken und Funktionen sichtbar.

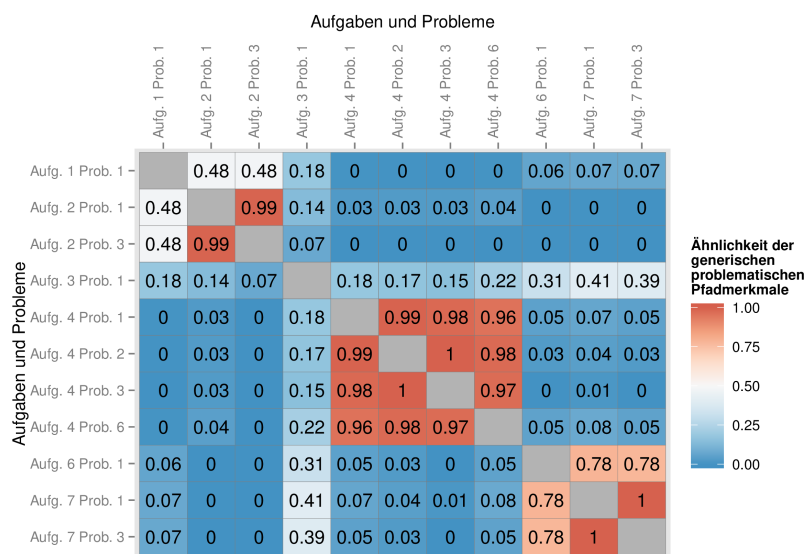


Abbildung 9.3: Ähnlichkeit der generischen problematischen Pfadmerkmale

Um diese Unterschiede genauer zu untersuchen, wurde in einem weiteren Schritt die Ähnlichkeit der generischen problematischen Pfadmerkmale untereinander berechnet (vgl. Abbildung 9.3). Die Ähnlichkeiten wurden mittels der Kosinus-Ähnlichkeit bestimmt. In der Abbildung sind die generischen problematischen Pfadmerkmale der Aufgaben auf der x-Achse und der y-Achse abgetragen. Die Abbildung ist diagonal gespiegelt. Die Färbung der Zellen gibt die Ähnlichkeit der jeweiligen beiden Probleme wieder. Eine dunklere Färbung entspricht einer höheren Ähnlichkeit. Zusätzlich wird die Kosinus-Ähnlichkeit in den Zellen angegeben. Die Abbildung bestätigen die Ergebnisse der Abbildung 9.2. Innerhalb der Aufgaben ähneln sich die generischen problematischen Pfadmerkmale sehr stark (Aufgabe 2, 4 und 7). Dieser Effekt ist erwartungskonform. Die Ähnlichkeiten zwischen den einzelnen Aufgaben sind sich dagegen deutlich geringer. Zwischen den generische problematischen Pfadmerkmalen der Aufgaben 1 und 2 beziehungsweise 6 und 7 beträgt die Ähnlichkeit jeweils 0,48 respektive 0,78. Auffällig ist ebenfalls die Ähnlichkeit des generischen problematischen Pfadmerkmals des ersten Problems der dritten Aufgabe mit den generischen problematischen Pfadmerkmalen der Aufgaben 6

und 7. Die Ähnlichkeit liegt allerdings bei maximal 0,41.

Diese Ergebnisse sind erfreulich. Positiv sind vor allem die deutlichen Unterschiede zwischen den generischen problematischen Pfadmerkmalen der schwerwiegenden Usability-Probleme insgesamt. Die Unterschiede sind bei den Aufgaben 1, 2, 3 und 4 deutlich sichtbar und setzen sich von den anderen Aufgaben ab. Dies wird durch die berechneten Kosinus-Ähnlichkeiten bestätigt. Die generischen Pfadmerkmale der Aufgaben 6 und 7 sind im Bezug auf die Übersichtsseite der Schlagwörter (S_Ü) und die Übersichtsseite der Startseite (St_Ü) ähnlicher. In dem ersten Problem der sechsten Aufgabe (Auffinden eines Termins durch Ankündigung) sind häufigere Zugriffe auf die Detailansicht der Ankündigungen (A_D) verzeichnet. Die Testerinnen und Tester der beiden Probleme der Aufgabe 7 greifen nicht auf diese Seiten zu, sondern bewegen dafür in der Schlagwortsuche der Schlagwörter (S_S) und der Übersichtsseite der Gruppen (G_Ü). Diese Zugriffe befinden sich nicht in den generischen Pfadmerkmalen des ersten Problems der Aufgabe 6. Die Kosinus-Ähnlichkeit ist mit einem Wert von 0,78 ebenfalls vergleichsweise hoch. Eine ähnliche Überschneidung gibt es zwischen den generischen problematischen Pfadmerkmalen der Aufgaben 1 und 2. Bei allen drei Problemen sind Aufrufe auf die Übersichtsseiten der Termine (T_Ü) und die Detailseiten der Termine (T_D) ausgeprägt. Die Intensität ist in der ersten Aufgabe höher. Zudem greifen die Testerinnen und Tester der beiden Probleme der zweiten Aufgabe (Anlegen eines eigenen Termins) intensiv auf die Bearbeitungsfunktion der Termine (T_B) zu. Dies ist bei dem ersten Problem der ersten Aufgabe (Finden des „Zusage“-Buttons) nicht der Fall. Die Kosinus-Ähnlichkeit zwischen den generischen problematischen Pfadmerkmalen der Aufgaben 1 und 2 ist mit einem Wert von 0,48 eher gering.

Dies bedeutet, dass die generischen Pfadmerkmale der schwerwiegenden Probleme hinreichend differenziert sind, um für die weitere Kombination herangezogen zu werden. Sie erfüllen damit eine wichtige Voraussetzung für die weitere Verwendung. Einzig die generischen problematischen Pfadmerkmale der Aufgaben 6 und 7 sind sich ähnlicher und sollten mit Vorsicht betrachtet werden..

Tabelle 9.4: Bewertung der schwerwiegenden Usability-Probleme

Aufgabe	Problem	Schwere*	Reichweite	Kategorie	Ähnlichkeit
1	1	7	lokal	-	0,914
2	1	3	lokal	-	0,849
2	3	4	global	Informationsdarstellung/Feedback	0,890
3	1	3	global	Navigation	0,691
4	1	4	global	Informationsdarstellung/Feedback	0,729
4	2	7	global	Informationsdarstellung	0,792
4	3	4	global	Navigation	0,777
4	6	4	global	Navigation	0,612
6	1	3	global	Informationsdarstellung	1,000
7	1	6	global	Informationsdarstellung/Feedback	0,928
7	3	4	lokal	-	0,928

*Gemessen an der Anzahl der Tester_innen.

Neben der Berechnung des Mittelwertes wird die Ähnlichkeit der einzelnen *problematischen Pfadmerkmale* mit den *generischen problematischen Pfadmerkmalen* berechnet. Dies erfolgt auf Basis der Kosinus-Ähnlichkeit (vgl. Gleichung 5.2.2). Die Ergebnisse sind in

der Tabelle 9.4 zu finden. Vier der elf gebildeten *generischen problematischen Pfadmerkmale* weisen eine sehr hohe Ähnlichkeit mit den separaten *problematischen Pfadmerkmalen* auf. Bei den *problematischen Pfadmerkmalen* des Problems der Zusage zu einem Termin ($pPM_{A_1P_1}$), des Problems des Suchens des Termins über die Ankündigungen ($pPM_{A_6P_1}$), des Problems der Suche nach dem Schlagwort ($pPM_{A_7P_1}$) und des Problems der Kombination von Suche und Schlagwort ($pPM_{A_7P_3}$) ist die Ähnlichkeit mit den *generischen problematischen Pfadmerkmalen* größer als 0,9. Die *generischen problematischen Pfadmerkmale* der schwerwiegenden Probleme der zweiten Aufgabe ($pPM_{A_2P_1}$, $pPM_{A_2P_3}$) weisen eine hohe Ähnlichkeit mit den *generischen Pfadmerkmalen* auf. Bei dem Problem der Eingabe des Enddatums und dem Problem der Definition der Zeitangabe bemisst sich die Ähnlichkeit auf über 0,8. Die Hälfte aller *generischen problematischen Pfadmerkmale* weisen damit eine hohe beziehungsweise sehr hohe Ähnlichkeit mit den *problematischen Pfadmerkmalen* auf, aus denen sie gebildet wurden. Drei der vier gefundenen Probleme in Aufgabe 4 kann eine akzeptable Anpassung zugesprochen werden. Die *generischen problematischen Pfadmerkmale* der Probleme des Auffindens der Eingabemaske ($pPM_{A_4P_1}$), dem Hochladen einer Datei ($pPM_{A_4P_2}$) und dem Zuordnen der Schlagwörter ($pPM_{A_4P_3}$) weisen eine Ähnlichkeit zwischen 0,729 und 0,792 auf. Die Ähnlichkeit der *problematischen Pfadmerkmale* mit den *generischen problematischen Pfadmerkmalen* bei dem Problem der Zuordnung der Kategorien ($pPM_{A_4P_6}$) liegt mit einer Ähnlichkeit von 0,612 niedriger. Dies könnte als Hinweis gewertet werden, dass sich die individuellen *problematischen Pfadmerkmale* der Testerinnen und Tester voneinander unterscheiden, ohne dabei allzu überlappend in den *generischen problematischen Pfadmerkmalen* zu sein. Das schwerwiegende Problem in der dritten Aufgabe ($UP_{A_3P_1}$, fehlende Verlinkung zwischen den Detailseiten der Benutzer und deren angelegten Einträgen) wird mittelmäßig durch die *generischen problematischen Pfadmerkmale* repräsentiert. Die Kosinus-Ähnlichkeit liegt bei 0,691. Die beiden letztgenannten *generischen problematischen Pfadmerkmale* ($pPM_{A_4P_6}$, $pPM_{A_3P_1}$) sollten bei weiteren Betrachtungen vorsichtig interpretiert werden.

Dennoch liegen die internen Kosinus-Ähnlichkeiten der einzelnen *generischen problematischen Pfadmerkmale* in der Regel höher als die Unterschiede zwischen den Aufgaben. Die Ähnlichkeiten zwischen den Aufgaben (vgl. Abbildung 9.3) sind durchgehend geringer als die Ähnlichkeiten der einzelnen *problematischen Pfadmerkmale* der einzelnen Usability-Probleme (vgl. Tabelle B.3). Davon ausgenommen sind die *generischen problematischen Pfadmerkmale*, die aus der gleichen Aufgabe stammen. Dieses Ergebnis lässt den Schluss zu, dass die interne Validität der *generischen problematischen Pfadmerkmale* nicht verletzt ist und diese die einzelnen *problematischen Pfadmerkmale* der Testerinnen und Tester aus den Nutzertests akzeptabel bis sehr gut abbilden.

9.3 Kombination von Pfadmerkmalen

In diesem letzten Schritt der Kombination werden die Logfiles des Produktivsystems mit den Logfiles der Nutzertests zusammengeführt. Ziel dieses Schrittes ist es, herauszufinden, wie häufig sich die *generischen problematischen Pfadmerkmale* in den Produktiv-Logfiles finden lassen. Dadurch lassen sich detailliertere Informationen zu der Schwere der Pro-

bleme finden. Beide Datensätze werden als *Komb-Matrix* aufgebaut, um möglichst viele Informationen zu erhalten. Mittels der *Kosinus-Ähnlichkeit* wird die Übereinstimmung zwischen beiden Matrizen berechnet. Der Vergleich erfolgt für jedes schwerwiegende Problem. Die untere Grenze der Ähnlichkeiten ist zunächst unbekannt. Bei der Erstellung der generischen problematischen Pfadmerkmale belief sich die geringste Kosinus-Ähnlichkeit auf 0,612. Daher wird eine Kosinus-Ähnlichkeit von 0,6 als unterste Grenze bei der Identifizierung von problematischen Pfadmerkmalen in den Produktiv-Logfiles festgelegt. Diese Grenze wird in 0,05er-Schritten bis zu einem maximalen Wert von 0,95 erhöht.

9.3.1 Übersicht über die Zuordnungen

Tabelle 9.5: Identifizierte generische problematische Pfadmerkmale in den Produktiv-Logfiles

Aufgabe	Problem	Ähnlichkeit								
		0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	
1	1	128	110	89	65	58	39	30	18	
2	1	63	54	46	39	31	27	25	14	
2	3	60	52	44	39	30	27	25	13	
3	1	177	88	24	10	4	2	0	0	
4	1	106	86	68	58	47	37	18	12	
4	2	118	101	80	63	54	43	25	11	
4	3	116	102	82	64	56	40	24	8	
4	6	155	131	116	93	75	56	36	15	
6	1	303	243	152	101	51	36	14	6	
7	1	211	172	130	105	67	45	19	2	
7	3	205	164	131	101	66	42	17	5	

Beläuft sich die Kosinus-Ähnlichkeit auf 0,6, werden alle generischen problematischen Pfadmerkmale in den Besuchen der Produktiv-Logfiles gefunden (vgl. Tabelle 9.5 und Abbildung 9.4). Die Anzahl variiert zwischen den Aufgaben und den Problemen. Viele Zuordnungen finden sich in den Aufgaben 6 und 7. In der Aufgabe 6 wird das erste Problem ($UP_{A_6P_1}$) bei 303 Besuchen des Produktivsystems identifiziert. Der Datensatz besteht aus 2.197 Besuchen. Wird die Kosinus-Ähnlichkeit von 0,6 festgesetzt, bedeutet dies, dass bei knapp 14 Prozent aller Besuche ähnliche generische problematische Pfadmerkmale wie die der Testerinnen und Tester gefunden werden. Dies kann als Anhaltspunkt für das Auftreten des Problems (Auffinden des Termins über die Ankündigungen) interpretiert werden. Den beiden schwerwiegenden Problemen der Aufgabe 7 ($UP_{A_7P_1}$, Suche nach dem Schlagwort; $UP_{A_7P_3}$, Kombination von Schlagwort und Suchbegriff) werden 211 beziehungsweise 205 Pfadmerkmale aus dem Produktivsystem zugeordnet. Dies entspricht einem Wert zwischen neun beziehungsweise zehn Prozent. Die generischen problematischen Pfadmerkmale des ersten Problems der dritten Aufgabe ($pPM_{A_3P_1}$, fehlende Verlinkung in den Detailseiten der Benutzer) werden ebenfalls häufig identifiziert (177 Mal identifiziert, acht Prozent). Danach folgen in absteigender Reihenfolge das sechste Problem der vierten Aufgabe ($UP_{A_4P_6}$, Zuordnen der Kategorien, 155 Identifizierungen, sieben Prozent), das erste Problem der ersten Aufgabe ($UP_{A_1P_1}$, Auffinden des Zusage-Buttons, 128 Mal identifiziert, knapp sechs Prozent), das zweite Problem der vierten Aufgabe ($UP_{A_4P_2}$, Probleme

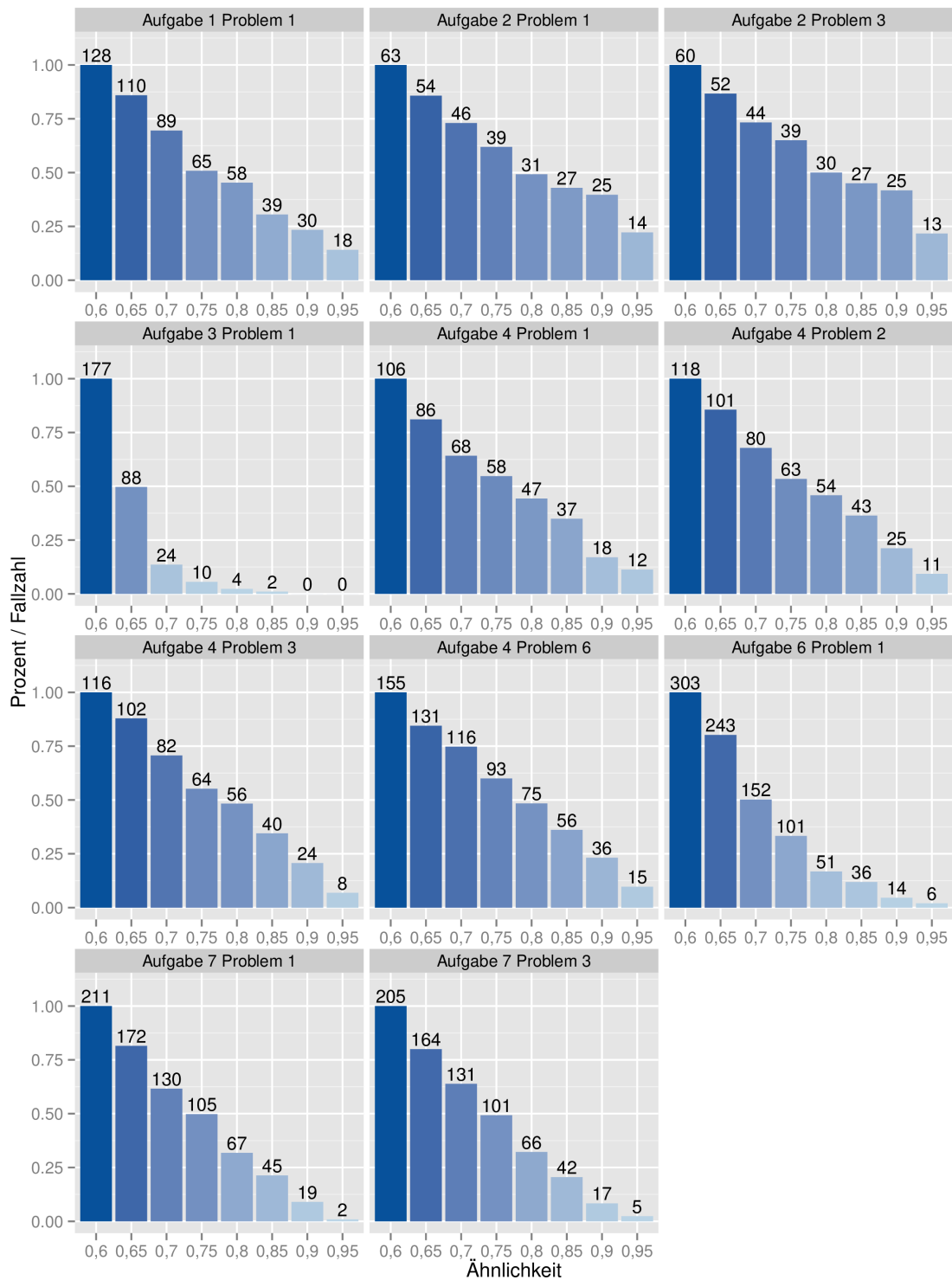


Abbildung 9.4: Identifizierte generische problematische Pfadmerkmale in den Produktiv-Logfiles

beim Hochladen einer Datei, 118 Identifizierungen, fünf Prozent), das dritte Problem der vierten Aufgabe ($UP_{A_4P_3}$, Zuordnen der Schlagwörter, 116 Mal identifiziert, fünf Prozent) und das erste Problem der vierten Aufgabe ($UP_{A_4P_1}$, Auffinden der Eingabemaske, 106 Identifizierungen, knapp fünf Prozent). Die geringsten Identifizierungen weisen die beiden Probleme der zweiten Aufgabe ($UP_{A_2P_1}$, Eingabe des Enddatums; $UP_{A_2P_3}$, Format der Zeitangabe) mit jeweils 63 beziehungsweise 60 zugeordneten Pfadmerkmalen auf. Die Unterschiede bestehen vor allem zwischen den einzelnen Aufgaben. Die Probleme derselben Aufgabe werden ähnlich häufig identifiziert. Die Unterschiede innerhalb der Probleme der Aufgaben 2, 4 und 7 sind marginal. Die Entdeckung ist nicht verwunderlich, da sich die identifizierten Nutzerinnen und Nutzer zwangsläufig innerhalb derselben Aufgaben bewegen. Trotzdem sind kleine Unterschiede zu erkennen. Dies deutet darauf hin, dass die 12 Nutzertests sich hinsichtlich der problematischen Pfadmerkmale unterscheiden.

Bei einer leicht höheren minimalen Ähnlichkeit der problematischen Pfadmerkmale (0,65, vgl. Tabelle 9.5 und Abbildung 9.4) sinkt die Zahl der identifizierten Pfadmerkmale in den Produktiv-Logfiles. Dies ist erwartungskonform, da eine genauere Übereinstimmung der Daten eine höhere Ähnlichkeit der Matrizen voraussetzt. Die Anzahl der identifizierten Pfadmerkmale sinkt zwischen den einzelnen Problemen überwiegend ähnlich. Die bei einer Kosinus-Ähnlichkeit von 0,6 sehr häufig identifizierten Probleme ($UP_{A_6P_1}$, $UP_{A_7P_1}$, $UP_{A_7P_3}$) weisen auch bei einer Übereinstimmung von 0,65 die höchsten Fallzahlen auf (243, 172, 164). Bei allen drei Problemen sinkt die Zahl um etwa 20 Prozent. Bei den anderen Problemen sinkt die Zahl der identifizierten Pfadmerkmale um ca. 15 Prozent. Die einzige Ausnahme stellt das erste Problem der dritten Aufgabe dar ($UP_{A_3P_1}$). Dieses wurde bei einer Ähnlichkeit von 0,6 sehr häufig identifiziert (177 Übereinstimmungen). Bei einer Ähnlichkeit von 0,65 sinkt die Zahl um die Hälfte auf 88 Identifizierungen. Es handelt sich um das Problem der fehlenden Verlinkung auf den Detailseiten der Benutzerinnen und Benutzer. Eventuell liegt dies an der „Diversität“ der Pfadmerkmale. Bei der Bildung der generischen problematischen Pfadmerkmale lag die Ähnlichkeit der einzelnen problematischen Pfadmerkmale unter 0,7.

Die stark sinkende Fallzahl des ersten Problems der dritten Aufgabe ($UP_{A_3P_1}$) setzt sich bei steigender Kosinus-Ähnlichkeit fort. Bei einer Ähnlichkeit von 0,7 fällt die Anzahl der identifizierten Pfadmerkmale in den Produktiv-Logfiles von 88 auf 24. Dies entspricht einem Rückgang von über 70 Prozent. Bei weiter steigender Ähnlichkeit kommt es zu einem hohen Rückgang der Übereinstimmungen. Bei einer Ähnlichkeit von 0,75 lassen sich die generischen problematischen Pfadmerkmale noch zehnmal in den Produktiv-Logfiles identifizieren. Das entspricht einem Rückgang von über 50 Prozent. Bei einer Kosinus-Ähnlichkeit von 0,8 liegt die Fallzahl der Übereinstimmungen bei vier, bei einer Ähnlichkeit von 0,85 lassen sich noch zwei Besuche der Produktiv-Logfiles zuordnen. Ab einer Ähnlichkeit von 0,9 finden sich keine generischen problematischen Pfadmerkmale mehr.

Die Fallzahlen der anderen Probleme sinken ebenfalls bei steigender Ähnlichkeit. Der Rückgang ist allerdings nicht so stark wie bei dem ersten Problem der dritten Aufgabe. Bis zu einer Ähnlichkeit von einschließlich 0,8 liegt der Rückgang der Fallzahlen zwischen 15 und 30 Prozent. Der Rückgang ist bei anfangs häufig gefundenen generischen proble-

matischen Pfadmerkmalen stärker. So sinkt die Fallzahl des ersten Problems der sechsten Aufgabe ($UP_{A_6P_1}$) von 303 bei einer Ähnlichkeit von 0,6 auf 51 bei einer Ähnlichkeit von 0,8. Der Rückgang der Identifizierungen der beiden Probleme der siebten Aufgabe ($UP_{A_7P_1}$, $UP_{A_7P_3}$) ist ebenfalls stark (von 211 auf 67 beziehungsweise von 205 auf 66). Die hohen Rückgänge der Fallzahlen sind bei den anderen schwerwiegenden Problemen nicht zu finden, mit Ausnahme des ersten Problems der dritten Aufgabe ($UP_{A_3P_1}$). Diejenigen Probleme, die bei einer geringen Ähnlichkeit häufig in den Produktiv-Logfiles identifiziert werden konnten, sind bei den hohen Ähnlichkeiten von mindestens 0,9 kaum aufzufinden. Der Rückgang bei den anderen Problemen ist geringer. Dies führt dazu, dass diese Probleme bei einer hohen Ähnlichkeit häufiger identifiziert werden als diejenigen bei einer geringen Übereinstimmung. Die Ähnlichkeit der problematischen Pfadmerkmale mit den generischen problematischen Pfadmerkmalen hat darauf keinen Einfluss. Die Korrelation nach Spearman (Claus et al., 2004) zwischen der Ähnlichkeit und dem Rückgang in Prozent liegt bei 0,13.

Die verschiedenen Ähnlichkeiten lassen daher verschiedene Lösungen zu. Die Ergebnisse mit den Ähnlichkeiten 0,6 und 0,65 erscheinen aufgrund der hohen Fallzahlen zu universal gewählt. Es erscheint schwer vorstellbar, dass bspw. bei knapp 14 Prozent aller Besuche ein Problem mit dem Auffinden von Terminen über die Ankündigungen ($UP_{A_6P_1}$) auftaucht. Gleiches gilt für die häufige Identifizierung der generischen problematischen Pfadmerkmale bei Problemen mit der Suchfunktion ($pPM_{A_7P_1}$, $pPM_{A_7P_3}$). Liegt die Kosinus-Ähnlichkeit (0,9 und 0,95) dagegen zu hoch, sinken die Fallzahlen sehr stark ab. Bei einer Ähnlichkeit von 0,9 liegen die höchsten Identifizierungen bei 30 beziehungsweise 36 zugeordneten Pfadmerkmalen der Produktiv-Logfiles. Dies entspricht etwa anderthalb Prozent aller Besuche. Das Produktivsystem wurde von 133 verschiedenen Nutzerinnen und Nutzern besucht. In den Nutzertests wurden Probleme als schwerwiegend definiert, die von mindestens drei Testerinnen und Testern identifiziert wurden. Da in den Nutzertests eine vergleichbare Zielgruppe (Lehrerinnen und Lehrer) rekrutiert wurde, ist schwer vorstellbar, dass diese Probleme aufgrund der Stichprobe der Testpersonen so häufig identifiziert wurden. Hinzu kommt, dass die Nutzerinnen und Nutzer des Produktivsystems in der Mehrzahl weniger als 20 Besuche in den 303 Tagen der Datenerfassung hatten. Ein Effekt durch die Eingewöhnung und das Entwickeln von Workarounds bei Problemen dürfte daher auch kein Kriterium im Vergleich mit den Testerinnen und Testern sein. Die festzulegende Kosinus-Ähnlichkeit sollte daher zwischen 0,7 und 0,85 liegen.

9.3.2 Festlegung der Kosinus-Ähnlichkeit

Für die endgültige Auswahl der Kosinus-Ähnlichkeit muss ein Kompromiss zwischen Anzahl und Eindeutigkeit der Pfadmerkmale gefunden werden. Eine zu hohe Ähnlichkeit führt zu wenig gefundenen problematischen Pfadmerkmalen. Ist die Ähnlichkeit dagegen zu gering gewählt, scheint die Kombination beliebig zu werden. In einem ersten Schritt wurde die Ähnlichkeit bereits auf einen Wert zwischen 0,7 und 0,85 eingegrenzt. Um eine zu starke Generalisierbarkeit zu vermeiden, muss die optimale Lösung auf der einen Seite eine möglichst hohe Fallzahl bieten, und auf der anderen Seite müssen die

generalisierten problematischen Pfadmerkmale vor allem innerhalb der gleichen Aufgabe Besuche aus den Produktiv-Logfiles identifizieren. Dies wird durch eine tabellarische und eine grafische Ansicht der Identifizierungen sichtbar gemacht. Für jede angegebene Kosinus-Ähnlichkeit (0,7, 0,75, 0,8, 0,85) wird eine Tabelle und eine Grafik erstellt.⁵

Die Tabellen listen die gefundenen Pfadmerkmale der schwerwiegenden Probleme in den Produktiv-Logfiles auf. Beispielsweise sind in Tabelle 9.6 bei dem ersten Problem von Aufgabe 1 ($UP_{A_1P_1}$) 89 Besuche aus dem Produktivsystem den generischen problematischen Pfadmerkmalen zugeordnet worden. In einer weiteren Spalte wird angegeben, ob die Besuche, die dem Problem zugeordnet wurden, auch anderen Problemen zugeordnet wurden. Eine hohe Zahl bedeutet, dass mit denselben Besuchen aus dem Produktivsystem mehrere generische problematische Pfadmerkmale verknüpft werden. Grundsätzlich ist die Zahl der mehrfachen Identifizierungen so gering wie möglich zu halten, da eine hohe Anzahl an mehrfachen Identifizierungen die Trennschärfe der Zuordnungen verringert und die Ergebnisse möglicherweise verzerrt. In der Tabelle 9.6 sind von den 89 Besuchen der Produktiv-Logfiles 34 Besuche auch anderen Problemen zugeordnet worden. Dies führt dazu, dass die Anzahl der mehrfach zugeordneten Besuche aus den Produktiv-Logfiles größer als die Anzahl der gefundenen generischen Pfadmerkmale in der Aufgabe selber sein kann. Sichtbar wird dies an der zweiten Zeile der Tabelle 9.6. Bei einer Kosinus-Ähnlichkeit von 0,7 werden 46 Besuche aus dem Produktivsystem dem ersten Problem der zweiten Aufgabe ($UP_{A_2P_1}$) zugeordnet. Einige (oder alle) 46 Besuche werden weitere 61 Mal anderen Problemen zugeordnet. Weniger schwerwiegend ist, wenn die mehrfachen Identifizierungen innerhalb der gleichen Aufgabe erfolgen. Die Darstellung in der Transaction Matrix unterscheidet die aufgerufenen Rubriken und die genutzten Funktionen. Schwerwiegende Probleme aus derselben Aufgabe müssen folglich eine höhere Kosinus-Ähnlichkeit untereinander aufweisen als Aufgaben, bei denen in unterschiedlichen Rubriken und mit anderen Funktionen (Material hochladen vs. Termin zusagen) genutzt wurden. Die Anzahl der mehrfachen identifizierten Pfadmerkmale wird nach Identifizierungen innerhalb derselben Aufgaben beziehungsweise zwischen verschiedenen Aufgaben getrennt. In zwei weiteren Spalten folgt Angabe in Prozentzahlen, was eine einfachere Vergleichbarkeit ermöglicht. In der Tabelle 9.6 bedeutet dies, dass in dem ersten Problem der ersten Aufgabe ($UP_{A_1P_1}$) alle 34 mehrfach zugeordneten Besuche aus den Produktiv-Logfiles in anderen Aufgaben vorkommen und nicht in derselben Aufgabe. Dies stellt ein unbefriedigendes Ergebnis dar. Von den 64 mehrfach zugeordneten Besuchen des ersten Problems in Aufgabe 2 ($UP_{A_2P_1}$) werden 44 Besuche anderen Problemen derselben Aufgabe zugeordnet und 17 Besuche Problemen in anderen Aufgaben zugeordnet. Dies entspricht einem prozentualen Verhältnis von 72 zu 28 und stellt ein wesentlich besseres Ergebnis dar als bei einer geringeren Kosinus-Ähnlichkeit. Die Anzahl der mehrfachen Pfadmerkmale sollte möglichst klein sein. Eine vollständige Eliminierung der mehrfachen Zuordnungen ist vermutlich nicht realisierbar, da dieselben Testpersonen teilweise mehrere Probleme einer Aufgabe aufgedeckt haben. Dies bedeutet, dass die Pfade deshalb eine höhere Ähnlichkeit haben dürften.

⁵Die Tabellen und Grafiken für die vorher ausgeschlossenen Kosinus-Ähnlichkeiten sind im Anhang D aufgeführt.

Tabelle 9.6: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,7

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	89	34	0	34	0%	100%
Aufgabe 2	Problem 1	46	61	44	17	72%	28%
Aufgabe 2	Problem 3	44	61	44	17	72%	28%
Aufgabe 3	Problem 1	24	5	0	5	0%	100%
Aufgabe 4	Problem 1	68	200	200	0	100%	0%
Aufgabe 4	Problem 2	80	224	224	0	100%	0%
Aufgabe 4	Problem 3	82	224	224	0	100%	0%
Aufgabe 4	Problem 6	116	228	228	0	100%	0%
Aufgabe 6	Problem 1	152	139	0	139	0%	100%
Aufgabe 7	Problem 1	130	195	125	70	64%	36%
Aufgabe 7	Problem 3	131	197	125	72	63%	37%
		962	1568	1214	354	77%	23%

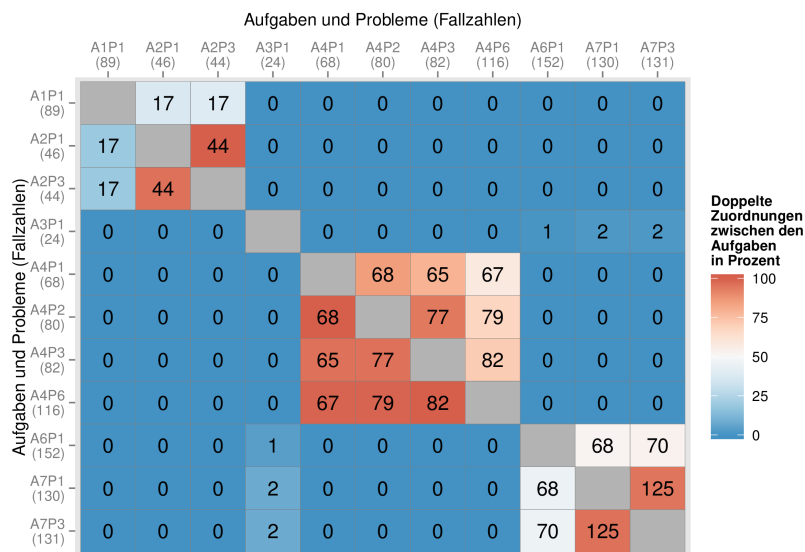


Abbildung 9.5: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,7

Die mehrfach identifizierten generischen Pfadmerkmale der Produktiv-Logfiles werden analog zu den Tabellen als Grafik dargestellt. Dies geschieht für alle gewählten Kosinus-Ähnlichkeiten. Die Grafiken übernehmen die Informationen aus den Tabellen, ermöglichen aber gleichzeitig die Aufteilung der mehrfachen Pfadmerkmale genauer zu bestimmen. Dazu werden auf den beiden Achsen die Aufgaben (A_x) und Probleme (P_x) abgetragen. Die Zahlen in Klammern stehen für die Anzahl an identifizierten Besuchen aus den Produktiv-Logfiles. Es handelt sich *nicht* um die mehrfach identifizierten Besuche. Die Werte in Abbildung 9.5 auf beiden Achsen stimmen mit den Werten der dritten Spalte in Tabelle 9.6 überein. Die Felder der sich ergebenden Matrix (ähnlich eines „Dot Plots“; Gibbs und McIntyre, 1970) werden mit den mehrfachen Besuchen in den jeweils zwei verglichenen Problemen gefüllt, indem die Zahl der dritten Spalte in Tabelle 9.6 auf die verschiedenen Zellen verteilt wird. Dadurch zeigen die Zellen detaillierter an, wie sich die mehrfachen Identifizierungen verteilen. Es handelt sich daher nur noch um *doppelte Identifizierungen*. Die Grafik wird spaltenweise interpretiert. Das erste Problem der ersten Aufgabe ($UP_{A_1P_1}$) wird in 89 Besuchen des Produktivsystems identifiziert. Nach Tabelle 9.6 sind 34 der Besuche mehrfach identifiziert worden. Der Abbildung 9.5 kann entnommen werden, dass bei den beiden schwerwiegenden Problemen in der Aufgabe 2

($UP_{A_2P_1}$, $UP_{A_2P_3}$) jeweils 17 Besuche doppelt identifiziert werden (erste Spalte, zweite und dritte Reihe). Neben der Anzahl der doppelten Fälle gibt die Färbung der Zellen an, um wie viel Prozent der Gesamtanzahl der identifizierten Besuche es sich handelt. Die Zahl bezieht sich bewusst *nicht* auf die mehrfach gefundenen generischen problematischen Pfadmerkmale. Dadurch ist es möglich, die Anzahl der mehrfach zugeordneten Besuche ins Verhältnis mit allen identifizierten Besuchen zu setzen und die Ähnlichkeit zwischen zwei schwerwiegenden Problemen zu beziffern. Die Farbskala reicht von grün (keine doppelt identifizierten Besuche) über gelb (50 Prozent doppelt identifizierte Besuche) bis rot (100 Prozent doppelt identifizierte Besuche). In Abbildung 9.5 ist zu sehen, dass die jeweils 17 doppelten Besuche der Produktiv-Logfiles knapp 20 Prozent aller identifizierten Besuche des ersten Problems der ersten Aufgabe ($UP_{A_1P_1}$) ausmachen. Die beiden Probleme stammen aus der zweiten Aufgabe ($UP_{A_2P_1}$, $UP_{A_2P_3}$), weshalb in Tabelle 9.6 in der letzten Spalte 100 Prozent in unterschiedlichen Aufgaben abgetragen sind. Das erste Problem der zweiten Aufgabe ($UP_{A_2P_1}$) wurde 46 Besuchen aus dem Produktivsystem zugeordnet. Die Besuche wurden 61 Mal mindestens doppelt anderen schwerwiegenden Problemen zugeordnet. In der Abbildung 9.5 kann abgelesen werden, dass 17 Besuche dem ersten Problem der ersten Aufgabe ($UP_{A_1P_1}$) und 44 Probleme dem dritten Problem derselben Aufgabe ($UP_{A_2P_3}$) zugeordnet wurden. Durch die Färbung der Zellen ist ersichtlich, dass die 17 doppelten Besuche etwa 40 Prozent aller dem zweiten Problem der zweiten Aufgabe zugeordneten Besuche entsprechen. Die Übereinstimmung mit dem dritten Problem derselben Aufgabe liegt mit über 95 Prozent sehr hoch. Da die Besuche aus derselben Aufgabe stammen, kann dies vernachlässigt werden.

Tabelle 9.7: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,75

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	65	14	0	14	0%	100%
Aufgabe 2	Problem 1	39	45	38	7	84%	16%
Aufgabe 2	Problem 3	39	45	38	7	84%	16%
Aufgabe 3	Problem 1	10	0	0	0	0%	0%
Aufgabe 4	Problem 1	58	170	170	0	100%	0%
Aufgabe 4	Problem 2	63	182	182	0	100%	0%
Aufgabe 4	Problem 3	64	182	182	0	100%	0%
Aufgabe 4	Problem 6	93	184	184	0	100%	0%
Aufgabe 6	Problem 1	101	73	0	73	0%	100%
Aufgabe 7	Problem 1	105	137	100	37	73%	27%
Aufgabe 7	Problem 3	101	136	100	36	74%	26%
		738	1168	994	174	85%	15%

Tabelle 9.8: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,8

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	58	4	0	4	0%	100%
Aufgabe 2	Problem 1	31	32	30	2	94%	6%
Aufgabe 2	Problem 3	30	32	30	2	94%	6%
Aufgabe 3	Problem 1	4	0	0	0	0%	0%
Aufgabe 4	Problem 1	47	137	137	0	100%	0%
Aufgabe 4	Problem 2	54	153	153	0	100%	0%
Aufgabe 4	Problem 3	56	155	155	0	100%	0%
Aufgabe 4	Problem 6	75	153	153	0	100%	0%
Aufgabe 6	Problem 1	51	28	0	28	0%	100%
Aufgabe 7	Problem 1	67	80	66	14	82%	18%
Aufgabe 7	Problem 3	66	80	66	14	82%	18%
		539	854	790	64	93%	7%

Tabelle 9.9: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,85

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfach Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	39	0	0	0	0%	0%
Aufgabe 2	Problem 1	27	27	27	0	100%	0%
Aufgabe 2	Problem 3	27	27	27	0	100%	0%
Aufgabe 3	Problem 1	2	0	0	0	0%	0%
Aufgabe 4	Problem 1	37	98	98	0	100%	0%
Aufgabe 4	Problem 2	43	112	112	0	100%	0%
Aufgabe 4	Problem 3	40	106	106	0	100%	0%
Aufgabe 4	Problem 6	56	110	110	0	100%	0%
Aufgabe 6	Problem 1	36	13	0	13	0%	100%
Aufgabe 7	Problem 1	45	47	40	7	85%	15%
Aufgabe 7	Problem 3	42	46	40	6	87%	13%
		394	586	560	26	96%	4%

Der Vergleich der Tabellen 9.6, 9.7, 9.8 und 9.9 zeigt zunächst, dass mit steigender Kosinus-Ähnlichkeit die identifizierten generischen problematischen Pfadmerkmale abnehmen. Insgesamt werden bei einer Kosinus-Ähnlichkeit von 0,7 (vgl. Tabelle 9.6) 962 Besuche den generischen problematischen Pfadmerkmalen zugeordnet. Die Zahl sinkt auf 738 bei einer Kosinus-Ähnlichkeit von 0,75 (vgl. Tabelle 9.7) und fällt weiter (539 zugeordnete Besuche bei einer Ähnlichkeit von 0,8, vgl. Tabelle 9.8 und 394 zugeordnete Besuche bei einer Kosinus-Ähnlichkeit von 0,85, vgl. Tabelle 9.9). Dieser Zusammenhang wurde bereits in Tabelle 9.5 dargestellt. Mit steigender Ähnlichkeit nimmt parallel die Anzahl der mehrfach identifizierten Besuche ab. Bei einer Kosinus-Ähnlichkeit von 0,7 ist der Faktor von mehrfachen Identifizierungen zu gefundenen Pfadmerkmalen 1,63 (962 zu 1568 Besuchen) und fällt weiter auf den Faktor 1,58 (Kosinus-Ähnlichkeit 0,75 und 0,8) beziehungsweise 1,49 (Kosinus-Ähnlichkeit 0,85). Gleichzeitig nehmen die mehrfach identifizierten Besuche zwischen den Aufgaben ab und finden zunehmend innerhalb der gleichen Aufgabe statt. Über alle Kosinus-Ähnlichkeiten hinweg ist die mehrfache Zuordnung von Besuchen bei den Problemen der Aufgaben 4 und 7 höher als die Anzahl der dem schwerwiegenden Problem zugeordneten Besuche. Die mehrfachen Identifizierungen der schwerwiegenden Probleme der Aufgabe 4 finden nur innerhalb der eigenen Aufgabe statt. In der Aufgabe 7 sind die mehrfachen Zuordnungen der Besuche vornehmlich innerhalb der eigenen Aufgabe (64 beziehungsweise 63 Prozent in Tabelle 9.6, 73 beziehungsweise 74 Prozent in Tabelle 9.7, jeweils 82 Prozent in Tabelle 9.8, 85 beziehungsweise 87 Prozent in Tabelle 9.9). Die Gesamtprozentzahl der mehrfachen Zuordnungen innerhalb der gleichen Aufgaben steigt von 77 Prozent bei einer Kosinus-Ähnlichkeit von 0,7 (vgl. Tabelle 9.6), über 85 Prozent (Kosinus-Ähnlichkeit von 0,75, vgl. Tabelle 9.7) und 93 Prozent (Kosinus-Ähnlichkeit von 0,8, vgl. Tabelle 9.8) bis auf 96 Prozent bei einer Ähnlichkeit von 0,85 (vgl. Tabelle 9.9).

Ein Überblick über die Abbildungen 9.5, 9.6, 9.7 und 9.8 zeigt ein ähnliches Bild. Die mehrfachen Zuordnungen neigen zu einer „Blockbildung“. Die schwerwiegenden Probleme der Aufgaben 1 und 2, der Aufgaben 6 und 7 sowie die Probleme innerhalb der Aufgabe 4 zeigen doppelte Identifizierungen. Die generischen Pfadmerkmale des schwerwiegenden Problems in Aufgabe 1 ($UP_{A_1P_1}$, Finden des „Zusage“-Buttons) ordnen jeweils Besuche den schwerwiegenden Pfadmerkmalen des ersten Problems ($UP_{A_2P_1}$, Eingabe des Enddatums bei Terminen) und des dritten Problems ($UP_{A_2P_3}$, Format der Zeitangabe) der zweiten Aufgabe zu. Gleiches trifft bei den anderen Kombinationen der drei

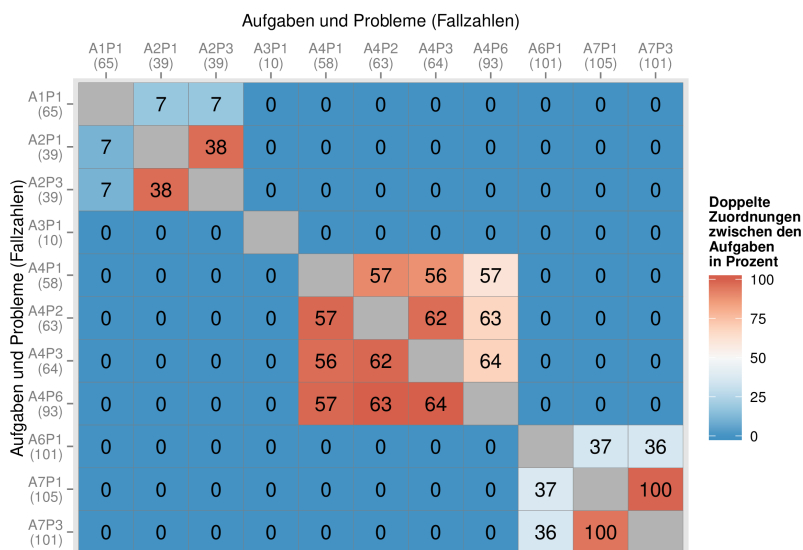


Abbildung 9.6: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,75

Probleme zu. Da sich die Nutzerinnen und Nutzer bei allen drei Problemen innerhalb derselben Rubrik (Termine) aufhalten, ist eine Überlappung nicht verwunderlich. Die doppelte Identifizierung ist innerhalb der Aufgabe 2 höher als zwischen Aufgabe 1 und Aufgabe 2. Dort treten zwischen 95 und 100 Prozent aller identifizierten Besuche doppelt auf. Mit zunehmender Kosinus-Ähnlichkeit zwischen den generischen problematischen Pfadmerkmalen und den Produktiv-Logfiles nimmt die doppelte Zuordnung zwischen den Aufgaben 1 und 2 ab, bis bei einer Ähnlichkeit von 0,85 keine Doppelungen mehr vorkommen. Ein weiterer „Block“ ist bei allen Problemen der Aufgabe 4 zu finden. Die Anzahl der doppelten Zuordnungen liegt zwischen 50 und 100 Prozent. Alle Probleme stammen aus derselben Aufgabe, weshalb dies als nicht gravierend gewertet wird. Der letzte „Block“ sind mehrfache Zuordnungen von Besuchen der Produktiv-Logfiles zu den schwerwiegenden Problemen der Aufgaben 6 und 7. Es zeigen sich Parallelen zu den Aufgaben 1 und 2. Dem schwerwiegenden Problem der Aufgabe 6 ($UP_{A_6P_1}$, Auffinden eines Termins durch Ankündigungen) und den schwerwiegenden Problemen der Aufgabe 7 ($UP_{A_7P_1}$, Suche nach dem Schlagwort; $UP_{A_7P_3}$, Kombination von Suche und Schlagwort) werden doppelte Besuche zugeordnet. Die Anzahl innerhalb der Aufgabe 7 ist höher und liegt zwischen knapp 90 und 100 Prozent. Zwischen den Aufgaben sind die doppelten Zuordnungen geringer (max. 46 Prozent) und nehmen mit steigender Kosinus-Ähnlichkeit ab. Das schwerwiegende Problem der Aufgabe 3 ($UP_{A_3P_1}$, Fehlende Verlinkung auf den Detailseiten der Benutzerinnen und Benutzer) wird nicht mehrfach zugeordnet. Einzige Ausnahme ist bei einer Kosinus-Ähnlichkeit von 0,7. Zwei Besuche aus den Produktiv-Logfiles werden ebenfalls den beiden schwerwiegenden Problemen der Aufgabe 7 zugeordnet. Ein Besuch wird dem schwerwiegenden Problem in Aufgabe 6 zugeordnet. Mit steigender Kosinus-Ähnlichkeit erfolgen die doppelten Identifizierungen nicht mehr.

Dies lässt den Schluss zu, dass eine Ähnlichkeit von über 0,75 gewählt werden sollte. Primär kann dies mit der Anzahl der mehrfachen Zuordnungen begründet werden. Bei

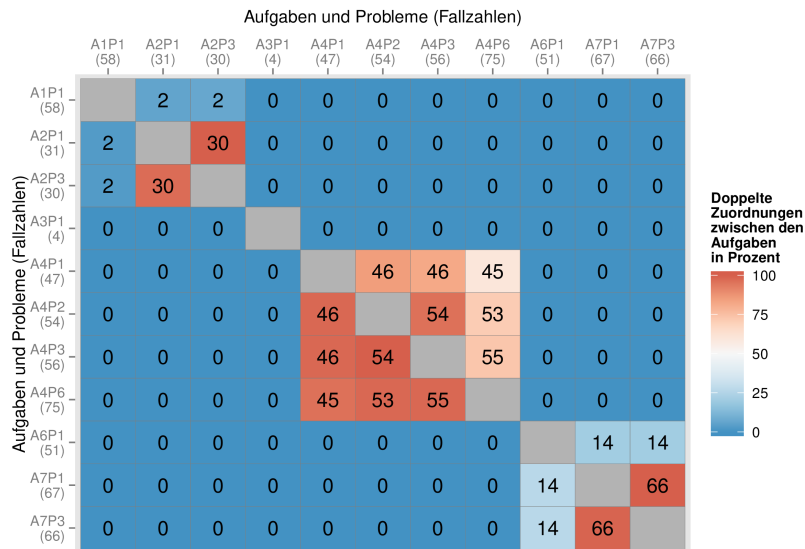


Abbildung 9.7: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,8

den Kosinus-Ähnlichkeiten von 0,7 und 0,75 erfolgen mehr als zehn Prozent der mehrfachen Zuordnungen außerhalb der gleichen Aufgaben (vgl. Tabellen 9.6 und 9.7). Bei einer Kosinus-Ähnlichkeit von 0,7 werden zudem Besuche, die dem schwerwiegenden Problem der dritten Aufgabe ($UP_{A_3P_1}$) zugeordnet werden, ebenfalls bei den schwerwiegenden Problemen der sechsten und siebten Aufgabe zugeordnet (vgl. Tabelle 9.6).

Der Vergleich der Zuordnungen der Besuche aus dem Produktiv-System zu den Pfadmerkmalen der Testerinnen und Tester bei den Kosinus-Ähnlichkeiten von 0,8 und 0,85 zeigt, dass die Anzahl der zugeordneten Besuche um ein Viertel abnimmt, wenn die Ähnlichkeit erhöht wird (539 zu 394). Der Faktor der mehrfachen Zuordnungen sinkt von 1,58 (854 zu 539) auf 1,49 (586 zu 394). Die Anzahl der Zuordnungen bei anderen Aufgaben sinkt dagegen nur leicht von sieben Prozent auf vier Prozent, liegt aber unter zehn Prozent. Bei beiden Kosinus-Ähnlichkeiten ist die Verteilung der Zuordnungen der Besuche aus dem Produktivsystem auf die verschiedenen Probleme ähnlich. Das sechste Problem der Aufgabe 4 ($UP_{A_4P_6}$, Zuordnen der Kategorien) lässt sich 75-fach beziehungsweise 56-fach in den Logfiles identifizieren. Ebenfalls häufig werden die beiden schwerwiegenden Probleme der Aufgabe 7 ($UP_{A_7P_1}$, Suche nach dem Schlagwort; $UP_{A_7P_3}$, Kombination von Schlagwort und Suche) in den Produktiv-Logfiles entdeckt. Bei einer steigenden Kosinus-Ähnlichkeit fällt die Anzahl der zugeordneten Besuche um etwa ein Drittel ab. Das erste Problem der ersten Aufgabe ($UP_{A_1P_1}$, Finden des „Zusage“-Buttons) wird bei einer Kosinus-Ähnlichkeit von 0,8 58 Besuchen zugeordnet. Steigt die Ähnlichkeit auf 0,85, sinkt die Zuordnung zu dem schwerwiegenden Problem auf 39 ab. Dies entspricht einem Rückgang um 33 Prozent. Der größte Rückgang ist bei dem ersten Problem der Aufgabe 3 ($UP_{A_3P_1}$, Fehlende Verlinkungen auf der Detailseite der Benutzerinnen und Benutzer) zu beobachten. Die Zuordnungen fallen von vier Besuchen auf zwei Besuche.

Ein Vergleich der Abbildungen 9.7 und 9.8 zeigt, dass die prozentuale doppelte Zuordnung der Besuche besser wird. Dies trifft primär auf die Aufgabe 4 zu. Die sechs Zellen unter der Diagonalen werden „gelber“ bei einer Kosinus-Ähnlichkeit von 0,85, was

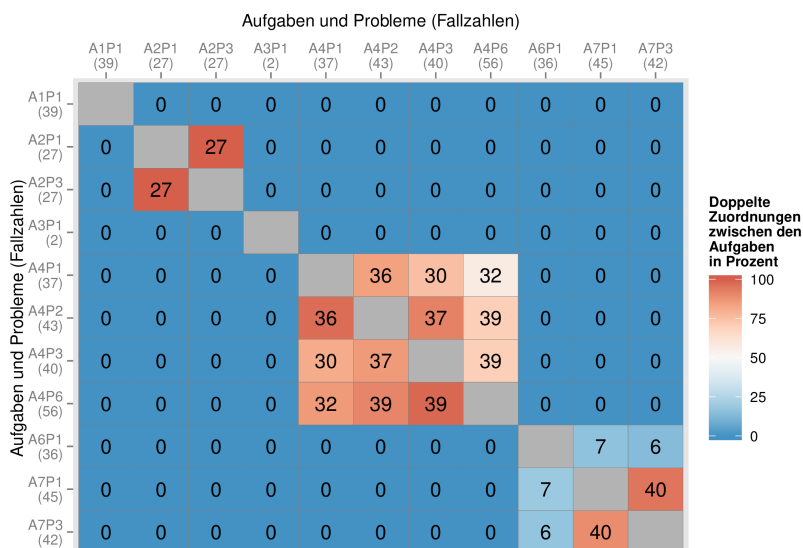


Abbildung 9.8: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,85

einer geringeren doppelten Zuordnung in Prozent entspricht. Oberhalb der Diagonalen verändert sich die Färbung der Zellen kaum. In dem „Block“ der Aufgaben 1 und 2 findet sich ab einer Ähnlichkeit von 0,85 keine doppelte Zuordnung von Besuchen zwischen den Aufgaben mehr. Bei einer Ähnlichkeit von 0,8 wurden jeweils zwei Besuche der Produktiv-Logfiles beiden schwerwiegenden Problemen der zweiten Aufgabe ($UP_{A_2P_1}$, $UP_{A_2P_3}$) und vice versa zugeordnet. Dies entspricht zwischen drei und sechs Prozent der Fälle. Innerhalb der Aufgabe 2 bleibt die doppelte Zuordnung mit 100 Prozent und knapp darunter hoch, kann aber weiterhin vernachlässigt werden. Problematischer ist, dass beiden Problemen in Aufgabe 2 jeweils fast dieselbe Anzahl an Besuchen zugeordnet wird. Der Algorithmus hat scheinbar Probleme, die generischen problematischen Pfadmerkmale adäquat zu trennen. Ein ähnlicher Sachverhalt kann innerhalb der siebten Aufgabe beobachtet werden. Bei beiden schwerwiegenden Problemen ($UP_{A_7P_1}$, Suche nach dem Schlagwort; $UP_{A_7P_3}$, Kombination von Suche und Schlagwort) ist die doppelte Zuordnung der Pfade mit 89 bis 95 Prozent hoch. Eindeutige Pfade kommen allerdings in beiden Problemen vor. Die weitere Erhöhung der Kosinus-Ähnlichkeit um den Wert 0,05 (vgl. Abbildung D.3 im Anhang) verbessert die Zuordnung in beiden „Blöcken“ wenig. Sie führt vielmehr zu einer starken Verringerung der Fallzahlen. Im Vergleich zwischen den drei schwerwiegenden Problemen der Aufgaben 6 und 7 werden Besuche aus den Produktiv-Logfiles doppelt zugeordnet. Diese liegen prozentual zwischen 22 und 27 Prozent bei einer Kosinus-Ähnlichkeit von 0,8 und zwischen 15 und 20 Prozent bei einer Kosinus-Ähnlichkeit von 0,85.

Aufgrund dieser Ergebnisse wird eine Kosinus-Ähnlichkeit von 0,85 festgelegt. Diese bietet den besten Kompromiss aus einer genauen Zuordnung der Besuche aus den Produktiv-Logfiles und gleichzeitig eine hohe Fallzahl. Eine höhere Kosinus-Ähnlichkeit liefert zwar noch eindeutigere Zuordnungen mit weniger mehrfach zugeordneten Besuchen (vgl. Abbildungen D.3 und D.4 im Anhang). Allerdings sinkt die Fallzahl bei einer zu hohen Spezifizierung weiter ab. Eine geringere Kosinus-Ähnlichkeit führt zu einer

größeren Doppelung der Zuordnungen, insbesondere bei einer Ähnlichkeit unter 0,7 (vgl. Abbildungen D.1 und D.2 im Anhang). Zudem werden mehr als der Hälfte aller Besuche Usability-Probleme zugeschrieben. Dies erscheint als zu hoch.

Die gewählte Kosinus-Ähnlichkeit von 0,85 besagt, dass es bei knapp 18 Prozent aller 2.197 Besuchen zu Problemen hinsichtlich der Usability gekommen sein *kann*. Die Verteilung der schwerwiegenden Probleme zwischen den Aufgaben ist homogen. Einzig das erste Problem der dritten Aufgabe ($UP_{A_3P_1}$, fehlende Verlinkungen auf der Detailseite der Benutzerinnen und Benutzer) wird nur zweimal durch die generischen problematischen Pfadmerkmale in den Produktiv-Logfiles identifiziert. Häufiger ist die Identifizierung der Probleme in den Aufgaben 1, 4 und 6. Im Hinblick auf die Arbeitsschwerpunkte im Produktivsystem (Ankündigungen, Materialien und Termine) ist die höhere Anzahl an Zuordnungen verständlich. Die große Anzahl an Problemen bei der Nutzung der Schlagwörter ($UP_{A_7P_1}$, Suche nach Schlagwort; $UP_{A_7P_3}$, Kombination von Suche nach Schlagwort und Schlagwort) überrascht dagegen. In dem Produktivsystem fällt die Nutzung der Schlagwörter gering aus (vgl. Abbildung 7.10). Dieser auf den ersten Blick offensichtliche Mismatch kann gleichzeitig ein Hinweis darauf sein, dass viele Nutzerinnen und Nutzer bei der Nutzung auf Probleme stoßen, da die Ergebnisse durch die hohe Anzahl an Identifikationen besonders betont werden.

9.3.3 Einordnung der Ergebnisse

Das Ergebnis der Kombination der beiden Logfile-Daten stellt ein erfreuliches Ergebnis dar. Zunächst können mittels der gewählten Kosinus-Ähnlichkeit von 0,85 den gefundenen Usability-Problemen entsprechende Pfade aus den „realen“ Logfiles zugeordnet werden. Dies lässt drei Schlüsse zu. Zum einen ist die Darstellung der Pfade über die Transaction Matrix ein möglicher Weg zur Aufbereitung der Logfiles. Auf diesem Weg geht die Reihenfolge der Zugriffe durch die Nutzerinnen und Nutzer beziehungsweise die Testerinnen und Tester verloren, dennoch werden alle schwerwiegenden Probleme durch die anschließende Bestimmung per Kosinus-Ähnlichkeit in den Logfiles identifiziert. Zum anderen ist die Methode differenziert genug, um Besonderheiten in den Logfiles herauszuarbeiten. Dies zeigt sich beispielsweise an den gefundenen schwerwiegenden Problemen bei der Nutzung der Schlagwörter ($UP_{A_7P_1}$, Suche nach Schlagwort; $UP_{A_7P_3}$, Kombination von Suche nach Schlagwort und Schlagwort). Bei einer Kosinus-Ähnlichkeit von 0,85 werden die Probleme jeweils 45 Mal und 42 Mal in den „realen“ Logfiles identifiziert. Häufiger wird nur das sechste Problem der vierten Aufgabe ($UP_{A_4P_6}$, Zuordnen von Kategorien) zugeordnet. Angesichts der geringen Anzahl der Zugriffe auf diese Funktion in dem Produktivsystem (vgl. Kapitel 7.4) erscheint die Anzahl zunächst unverhältnismäßig hoch. Die wahrscheinlichere Vermutung lautet dagegen, dass in dem Produktivsystem eine hohe Anzahl an Nutzerinnen und Nutzern ebenfalls Usability-Probleme beim Umgang mit den Schlagwörtern haben. Die Besonderheiten gehen dabei nicht mit einer zu hohen Generalisierbarkeit einher. Bei einer Kosinus-Ähnlichkeit von 0,85 sind doppelte Zuordnungen ausschließlich in den gleichen Aufgaben zu finden. Die Ähnlichkeit ist daher spezifisch genug.

Bei der Bewertung des Vorgehens, der Methode und der Ergebnisse ist es erfreulich, dass sich das gewählte Vorgehen, die Kombination von Usability-Problemen aus Nutzertests und Logfile-Analyse aus einem Produktivsystem als praktikabel erwiesen hat. Dadurch lässt sich auch die zentrale Forschungsfrage dieser Dissertation abschließend beantworten. Die zentrale Frage war, ob sich Logfile-Analysen und Nutzertests zu besserer Messung der Usability kombinieren lassen. Um die Frage insgesamt beantworten zu können, erscheint es sinnvoll, diese in mehrere Teile zu zergliedern. Optimal ist es dabei, analog zu der Vorgehensweise aufzusplitten (vgl. Kapitel 3). In der ersten Phase (vgl. Abbildung 3.3) war die zentrale Überlegung, dass verschiedene Testerinnen und Tester während der Nutzertests über verschiedene Probleme „stolpern“. Weiterhin liegt die Annahme zu Grunde, dass die dabei anfallenden Logfiles unterschiedlich sind. Dies bedeutet, dass sich die herausgearbeiteten Pfade der Probleme beziehungsweise daraus gewonnenen Pfadmerkmale der Probleme unterschiedlich sind. In der zweiten Phase wurden die in Phase eins gewonnenen problematischen Pfadmerkmale mit „realen“ Logfiles kombiniert, um die Häufigkeit der gefundenen Usability-Probleme bestimmen zu können. Dies impliziert zudem, dass sich die problematischen Pfadmerkmale soweit unterscheiden, dass es zu keiner großen Generalisierung der Zuordnungen kommt.

In den Nutzertests fielen den Testerinnen und Testern verschiedene Usability-Probleme auf. Die Annahme, dass Usability-Probleme nutzerspezifisch sind, kann damit bestätigt werden. Sie bestätigt zudem die bisherigen Erkenntnisse der Usability-Literatur (Barnum, 2011; Dumas und Fox, 2008; Dumas und Redish, 1999; Rubin und Chisnell, 2008), die ebenfalls von derselben Annahme ausgehen. Die bisherigen Anstrengungen, die benötigte Anzahl an Testpersonen festzulegen (vgl. u. a. Faulkner, 2003; Nielsen, 1994b; Nielsen und Landauer, 1993; Tullis und Albert, 2008; Virzi, 1990, 1992) unterstreichen die Unterschiedlichkeit der Testerinnen und Tester. Gleiches gilt für die Logfiles des Testsystems. Wie in Abbildung 9.2 ersichtlich, unterschieden sich die problematischen Pfadmerkmale der schwerwiegenden Probleme untereinander. Da diese auf den Logfiles aufbauen, kann davon ausgegangen werden, dass die Nutzung des Systems nutzerspezifisch, aufgabenspezifisch *und* problemspezifisch ist. Gleichzeitig konnte gezeigt werden, dass die individuellen Pfadmerkmale von einem Usability-Problem fast immer ähnlich beziehungsweise sehr ähnlich sind (vgl. Tabelle 9.4). Die durchschnittliche Ähnlichkeit der separaten problematischen Pfadmerkmale mit den generischen problematischen Pfadmerkmalen wies bis auf das sechste Problem von Aufgabe 6 sowie das erste Problem von Aufgabe 3 eine Kosinus-Ähnlichkeit von mindestens 0,7 auf. Bei sechs der elf schwerwiegenden Probleme lag die Ähnlichkeit bei mindestens 0,8. Die generischen problematischen Pfadmerkmale sind auf der einen Seite spezifisch genug, um sich von den anderen Pfadmerkmalen zu unterscheiden. Auf der anderen Seite bilden sie die individuellen problematischen Pfadmerkmale adäquat ab.

Damit konnten die beiden zentralen Annahmen der ersten Phase bestätigt werden. Unterschiedliche Testerinnen und Tester finden unterschiedliche Usability-Probleme. Die Interaktion mit dem LMS ist dabei ebenfalls nutzer- und problemspezifisch. Durch die Bestätigung der beiden Annahmen kann gefolgert werden, dass sich einzelnen Usability-Problemen jeweils spezifische problematische Pfadmerkmale zuordnen lassen. Die proble-

matischen Pfadmerkmale bilden die Usability-Probleme adäquat und individuell ab.

Die Kombination der generischen problematischen Pfadmerkmale der schwerwiegenden Usability-Probleme mit den Logfiles aus dem Produktivsystem war die zweite Phase des empirischen Vorgehens (vgl. Abbildung 3.4). Ziel war es herauszufinden, wie häufig die entdeckten Usability-Probleme bei einer größeren Nutzerbasis vorkommen können. Die Identifikation in den „realen“ Logfiles bietet einen Hinweis darauf, dass die Nutzerin beziehungsweise der Nutzer ebenfalls auf das Usability-Problem gestoßen sein könnte. Für die Kombination wurden Logfiles desselben LMS über einen Zeitraum von 300 Tagen verwendet (vgl. Kapitel 7.4). Um eine Kombination zu ermöglichen wurden die „realen“ Logfiles ebenfalls in eine Komb-Matrix überführt. Die Verbindung von den generischen problematischen Pfadmerkmalen mit den „realen“ Pfadmerkmalen fand mittels der Kosinus-Ähnlichkeit statt. Da keine Vorerfahrungen in Bezug auf die minimale Ähnlichkeit vorlagen, musste diese manuell festgelegt werden. Die endgültige Ähnlichkeit musste dabei zwei Anforderungen erfüllen: Sie sollte weder zu generisch noch zu spezifisch sein. Als Gradmesser für die optimale Kosinus-Ähnlichkeit wurden die Anzahl der Identifizierungen in den „realen“ Logfiles und die Anzahl der mehrfachen Identifizierungen herangezogen. Es erschien unwahrscheinlich, dass bei 3.100 Besuchen innerhalb der 300 Tage keine Nutzerin beziehungsweise kein Nutzer ebenfalls über die schwerwiegenden Probleme „gestolpert“ wäre. Auf der anderen Seite wäre eine Zuordnung von der Hälfte aller Besuche zu viel gewesen. Als besseres Maß für die Kosinus-Ähnlichkeit stellte sich die Anzahl der mehrfachen Zuordnungen heraus. Wie in Abbildung 9.2 ersichtlich, unterscheiden sich die generischen problematischen Pfadmerkmale deutlich voneinander. Die gewählte Kosinus-Ähnlichkeit sollte daher deutlich zwischen den Aufgaben trennen. Aus diesem Grund wurde geschaut, wie häufig mehrfache Pfadmerkmale in den Produktiv-Logfiles identifiziert werden. Eine hohe Anzahl der mehrfachen Identifizierungen spricht für eine hohe Generalisierung. Die mehrfache Identifizierung innerhalb einer Aufgabe kann dagegen aufgrund der ähnlichen generischen problematischen Pfadmerkmale vorkommen (vgl. erneut Abbildung 9.2). Sie sollten aber so klein wie möglich gehalten werden.

Die gewählte Kosinus-Ähnlichkeit von 0,85 stellt bei den gewählten Vorgaben die optimale Lösung dar. Allen generischen problematischen Pfadmerkmalen können Pfadmerkmale aus den *realen* Logfiles zugeordnet werden (vgl. Tabelle 9.5). Insgesamt werden 394 Besuche zugeordnet (vgl. Tabelle 9.9). Das bedeutet, dass es bei etwa 13 Prozent aller Besuche zu Problemen hinsichtlich der Usability gekommen sein *kann*. Insgesamt wurden 586 Pfadmerkmale der „realen“ Logfiles mindestens zweifach den problematischen Pfadmerkmalen zugeordnet. Erfreulicherweise erfolgten dabei 96 Prozent der mehrfachen Zuordnungen innerhalb der gleichen Aufgabe. Die verbleibenden vier Prozent entfallen ausschließlich auf die Aufgaben 6 und 7. Gemessen an der Anzahl der Identifizierungen liegt die Anzahl der mehrfachen Identifizierungen bei 15 Prozent und darunter. Die gewählte Kosinus-Ähnlichkeit entspricht aber der optimalen Lösung, was die Kriterien der Anzahl der Identifizierungen und der mehrfachen Zuordnungen betrifft.

Damit kann auch die zweite Phase des Vorgehens erfolgreich abgeschlossen werden. Die Kombination der in Phase 1 generierten generischen problematischen Pfadmerkmale der

schwerwiegenden Probleme konnten erfolgreich mit Logfiles aus einem Produktivsystem verknüpft werden. Dies bedeutet, dass die gefundenen Usability-Probleme hinsichtlich ihrer Häufigkeit (und damit abgeleitet ihrer Schwere) mit einer deutlich breiteren Datenbasis verknüpft werden konnten. Gleichzeitig wurde die gewählte Kosinus-Ähnlichkeit mit Bedacht gewählt und kritisch evaluiert, damit diese weder zu spezifisch noch zu generisch gewählt wurde. Dabei wurde neben der Anzahl der Zuordnungen auf die mehrfachen Zuordnungen geachtet und diese als weitere Kennzahl mit in die Analyse aufgenommen. *Insgesamt kann die Forschungsfrage, ob sich Logfile-Analysen und Nutzertests zu besserer Messung der Usability kombinieren lassen, bejaht werden.* Auf Basis der Ergebnisse lässt sich darauf schließen, dass insbesondere die gefundenen Usability-Probleme bei der Zusage eines Termins (Aufgabe 1 Problem 1), beim Anlegen neuer Materialien (alle Probleme der Aufgabe 4) sowie bei der Verwendung von Schlagwörtern (Aufgabe 7 Problem 1 und 2) häufiger in den „realen“ Logfiles auftreten. Insbesondere die Probleme der Aufgabe 7 sind interessant, da die Nutzung der Rubrik und der Funktionen in dem Produktivsystem eher gering ausfallen. Die gefundenen Probleme beim Anschauen von Materialien (Aufgabe 3 Problem 1, Fehlende Verlinkung auf den Detailseiten der Benutzerinnen und Benutzer) können dagegen vernachlässigt werden, da das Problem in den Nutzertests häufiger auftritt als in den „realen“ Logfiles.

10 Fazit

In dieser Dissertation wurde eine neue Methode zur Usability-Evaluierung entwickelt und getestet. Die Idee war, Nutzertests mit Logfile-Analysen zu kombinieren. Der hauptsächliche Vorteil sollte vor allem die breitere Datenbasis durch die Kombination der beiden Verfahren sein. Gleichzeitig konnte mit den Logfiles eine in webbasierten Umgebungen meistens vorhandene Datenquelle erschlossen und genutzt werden. Daneben bietet die Methode eine Kombination von quantitativen und qualitativen Methoden. Dazu wurden zunächst die theoretischen Grundlagen der Logfile-Analyse und der Usability aufgearbeitet. Dies sollte eine Einordnung der beiden maßgeblich verwendeten Verfahren innerhalb des Fachgebiets der Informatik erlauben. Parallel wurden digitale Kooperationssysteme, insbesondere Lern-Management-Systeme (LMS) und Groupware-Systeme vorgestellt. Sowohl die Logfile-Analyse als auch die Nutzertests wurden in einem solchen System vorgenommen. Im Rahmen des DFG-Projekts „Mediatisierte Organisationswelten in Schulen“ wurde die Veränderung in der Kommunikation zwischen Lehrerinnen und Lehrern durch digitale Medien untersucht. Die Rolle des Autors innerhalb des Projekts war die Analyse der digitalen Kommunikation und Kooperation durch Logfile-Analysen. Durch die Logfile-Analysen und die Erkenntnis, dass das LMS InfoChange Probleme hinsichtlich der Benutzerführung aufweist, ist die Idee zu dieser Dissertation entstanden.

Die Fragestellung wurde in drei empirischen Phasen untersucht und beantwortet. Zunächst wurden vorhandene Logfiles aus einer „realen“ Installation des LMS untersucht und analysiert. Es wurde ersichtlich, dass die Lehrkräfte vor allem Materialien über das System austauschen, gemeinsam Termine verwalten und Ankündigungen veröffentlichen. Intensivere kooperative Funktionen wie Diskussionsforen werden dagegen kaum genutzt. Dies gilt auch für die geringe Anzahl an Kommentaren in den Ankündigungen.

Anschließend wurden in einer Referenz-Installation des Systems Nutzertests durchgeführt. Die Teilnehmerinnen und Teilnehmer waren hauptsächlich Lehrkräfte einer Berufsbildenden Schule in Bremen. Der Großteil der Nutzertests fand in dem Lehrerzimmer der Schule statt. Drei weitere Tests wurden in einer laborähnlichen Situation durchgeführt. In den Nutzertests sollten die Teilnehmerinnen und Teilnehmer sieben Aufgaben in InfoChange ausführen. Die Aufgaben wurden primär aus den Logfile-Analysen und sekundär aus den Ergebnissen der teilnehmenden Beobachtungen und Gruppendiskussionen des DFG-Projekts gewonnen. Insgesamt wurden über 40 Usability-Probleme identifiziert. Insbesondere beim Hochladen neuer Materialien, dem Erstellen eines Termins und der Benutzung der Suchfunktion wurden zahlreiche Usability-Probleme entdeckt. Elf Probleme wurden von mindestens drei Testerinnen und Testern gefunden. Die Schwelle von drei Testerinnen und Testern galt als Kriterium für ein schwerwiegendes Problem. Nur schwerwiegende Probleme wurden in die weitere Analyse mit einbezogen. Die elf

Probleme verteilen sich auf sechs der sieben Aufgaben. Insbesondere bei dem Erstellen neuer Materialien und Termine sowie bei der Nutzung der Schlagwörter kam es zu Schwierigkeiten hinsichtlich der Usability. Die Mehrzahl der Probleme sind global. Das bedeutet, dass sie nicht nur in der Aufgabe auftreten, sondern auch in anderen Kontexten und Nutzungsszenarien von InfoChange vorkommen können.

In der dritten empirischen Phase wurden beide Datenquellen miteinander kombiniert. Dabei handelt es sich um die eigentliche Innovation dieser Dissertation. Die Nutzertests wurden im Hintergrund protokolliert und somit eine zweite Quelle an Logfiles erschlossen. Nach einer Aufbereitung der Logfiles der Nutzertests wurden beide Datenbestände in ein einheitliches und vergleichbares Format (Transaction Matrix) überführt. Mittels der Kosinus-Ähnlichkeit wurde statistisch ermittelt, wie häufig sich die generischen problematischen Pfadmerkmale der Nutzertests in den Pfadmerkmalen des Produktivsystems identifizieren lassen. Da bisher keine ähnlichen Verfahren bekannt waren, war die minimale Ähnlichkeit unbekannt und musste zunächst festgelegt werden. Es wurde ein Wert (0,85) gewählt, der die Ergebnisse nicht zu generisch werden lässt und gleichzeitig nicht zu spezifisch ist. Die Anzahl der mehrfachen Zuordnungen von Pfadmerkmalen des Produktivsystems zu den generischen problematischen Pfadmerkmalen der Nutzertests ist bei der gewählten Ähnlichkeit akzeptabel. Die mehrfachen Identifizierungen kommen zu über 95 Prozent innerhalb einer Aufgabe vor und erstrecken sich kaum über mehrere Aufgaben hinweg. Die Herleitung der endgültigen Ähnlichkeit erfolgte mittels deskriptiver Tabellen und Visualisierungen der doppelten Zuordnungen.

Die zentrale Forschungsfrage kann durch die gelungene Kombination erfolgreich beantwortet werden. Der daraus resultierende Vorteil wirkt sich positiv auf die Nutzertests aus. Bei diesen handelt es sich um eine qualitative Methode. Sie ermöglichen ein tiefes und umfassendes Verständnis der gefundenen Probleme, sind aber aufwendig durchzuführen, insbesondere was den zeitlichen Aufwand (Vorbereitung, Durchführung und Auswertung) angeht. Die Schwere eines gefundenen Usability-Problems wird in der Auswertung der Nutzertests entweder nach der Hartnäckigkeit, der Häufigkeit oder nach eigenen Kriterien festgelegt. Durch die Kombination mit den Logfiles ist es möglich, die Nutzertests auf eine breitere Basis zu stellen und zu schauen, wie häufig die gefundenen Probleme bei anderen Nutzerinnen und Nutzern vorkommen. Dies erlaubt eine quantitative und objektive Bewertung der Schwere der gefundenen Usability-Probleme. Die Ergebnisse haben gezeigt, dass es Unterschiede in den Logfiles gibt, die sich nicht durch die reine Verteilung der Zugriffe erklären lassen. Dazu zählen vor allem die Probleme bei der Nutzung der Suchfunktion in Verbindung mit den Schlagwörtern.

Die entwickelte Methode vergrößert das Wissen von mehreren Teilbereichen der Informatik und verwandter Disziplinen. Das Kernthema ist die Usability. Durch die Verknüpfung von Logfile-Analyse und Nutzertests wird das Methodenspektrum der Usability-Evaluationen erweitert. Methoden-Triangulationen sind in dem Bereich der Usability-Forschung bisher wenig verbreitet. Zwar werden häufig Fragebögen und Nutzertests miteinander kombiniert, allerdings werden die Fragebögen häufig nur als additive Methode eingesetzt. Eine Verknüpfung von quantitativen und qualitativen Methoden wie in

dieser Dissertation ist bisher nicht im Bereich der Usability-Forschung bekannt. Einzig ein Verfahren nutzt Logfile-Daten, um den Ablauf eines Cognitive Walkthrough vorzubereiten. Eine derart starke Verzahnung von Logfile-Analyse und Nutzertests ist zum jetzigen Zeitpunkt unbekannt. Daneben zeigt die Dissertation, dass Logfile-Analysen über tiefer gehende Informationen als reine Protokolldaten verfügen. Allerdings ist der Mehrwert des Verfahrens isoliert betrachtet geringer als die durch die Kombination mit einer anderen Methode. Diese Erkenntnis sollte für die Bereiche des Data Minings beziehungsweise der Wissensentdeckung in Datenbanken (KDD) übernommen werden. Dort herrscht häufig eine starke Fokussierung auf die Daten als einzige Erkenntnisquelle. Die Verknüpfung zeigt, dass die Erklärungskraft der Logfile-Analyse erheblich zunimmt, wenn eine Methoden-Triangulation angewendet wird. Dies trägt auch dazu bei, dass der finale Schritt in der Logfile-Analyse, die Interpretation, umfangreicher und fundierter erfolgen kann. Die Ergebnisse von Usability-Evaluationen sollen primär die Bedienbarkeit von Anwendungen verbessern. Eine höhere Bedienbarkeit schlägt sich häufig in einer höheren Akzeptanz der Anwendungen nieder. Dies wird unter anderem durch das TAM-Modell erklärt. Eine hohe Bedienbarkeit und damit einhergehende Akzeptanz der Anwendung ist auch während der Softwareentwicklung sowie der Softwareweiterentwicklung zu beachten. Darüber hinaus ist die Usability eng mit der Human-Computer-Interaction (HCI) verwandt.

In dieser Dissertation wurde die Kombination von Logfile-Analyse und Nutzertests zur summativen Evaluation eines LMS verwendet. Nach der Durchführung der Nutzertests und der Aufdeckung der gefundenen Probleme sollten in einem nächsten Schritt die dringendsten Probleme durch die Entwicklerinnen und Entwickler behoben werden. Zu Zwecken der Qualitätssicherung können nach dem Beheben der Probleme weitere Logfiles aus Produktivsystemen erhoben und erneut mit den Daten der Nutzertests verglichen werden. Sollten die Probleme nicht mehr in den neuen Logfiles entdeckt werden, kann dies als Indiz für eine erfolgreiche Behebung des Problems beziehungsweise der Probleme gewertet werden. Dabei handelt es sich um die in dem Vorgehen dargestellte *Phase 2+x*. In der Weiterentwicklung stellt dies in den Augen des Autors die primäre Anknüpfung an die Dissertation dar. Denkbar sind daneben zwei weitere Anwendungsmöglichkeiten zur Verstetigung, in denen die Methode in dem Bereich der Softwareentwicklung zum Einsatz kommen könnte.

Die generischen problematischen Pfadmerkmale der schwerwiegenden Probleme können längerfristig gespeichert werden. Im Zuge der stetigen Weiterentwicklung der Anwendung kann weiterhin auf diese zurückgegriffen werden und anhand von neuen Logfiles auf eventuell vorhandene Usability-Probleme geschlossen werden. Sollten weitere Nutzertests durchgeführt werden, kann über den Aufbau einer internen Datenbank mit generischen problematischen Pfadmerkmalen nachgedacht werden. Sobald neue Logfiles verfügbar sind, kann eine Kombination mit den generischen problematischen Pfadmerkmalen in der Datenbank durchgeführt werden. Das Vorgehen ist nur solange einsetzbar, wie der Aufbau beziehungsweise die Struktur der Anwendung nicht verändert wird. Eine Veränderung der Benutzeroberfläche führt meist zu einer veränderten Interaktion. Bereits vorhandene generische problematische Pfadmerkmale beziehen sich dagegen

auf die vorherigen Darstellungsformen. Die weitere Verwendung dürfte daher zu einer Verfälschung der Ergebnisse führen und sollte nicht weiter verwendet werden. Wird der technische Aufbau der Anwendung verändert, führt dies unter Umständen dazu, dass die generischen problematischen Pfadmerkmale nicht mehr zu den neuen Pfadmerkmalen der „realen“ Logfiles passen. Dies macht eine Anwendung ebenfalls unmöglich.

Bisher wurden die generischen problematischen Pfadmerkmale nur mit einer Installation des Produktivsystems kombiniert. Denkbar ist auch, eine Kombination mit mehreren Installationen der Anwendung durchzuführen. Der benötigte Aufwand kann dabei vernachlässigt werden, insbesondere weil diesem ein deutlich höherer Erkenntnisgewinn gegenüberstehen kann. So kann die Häufigkeit genauer anhand von mehreren Installationen bestimmt werden. Gleichzeitig werden eventuell besondere Nutzungsweisen und Nutzungsschwerpunkte der einzelnen Installationen aufgedeckt. Diese können bei überregional oder auch supranational eingesetzten Anwendungen weitere Hinweise auf kulturelle Besonderheiten hinsichtlich der Usability liefern.

Die Kombination der Logfile-Daten aus dem Testsystem und dem Produktivsystem wurde über eine Transaction Matrix realisiert. Eine Überführung in dieses Datenformat geht mit einem Informationsverlust einher. Die aktuelle Form der Transaction Matrix behält so viele Informationen wie möglich, ohne allzu generisch zu werden. Weder die genaue Reihenfolge der Seitenaufrufe noch die verbrachte Zeit auf den einzelnen Seiten wurden mit in den Vergleich aufgenommen. Das Ausschließen der Zeit wurde mit dem negativen Effekt des lauten Denkens auf die Bearbeitungsdauer der Aufgaben begründet. Die Reihenfolge der Aufrufe würde dazu führen, dass die Transaction-Matrix nicht mehr in ihrer bisherigen Form aufbereitet werden könnte. Allerdings könnte die Einbeziehung der Reihenfolge bei komplexeren Anwendungen oder Testszenarien mehr Sinn machen. Die hier verwendete Komb-Matrix führte zu einer fast perfekten Trennung zwischen den Aufgaben. Gleichzeitig ist die Trennung der generischen problematischen Pfadmerkmale innerhalb der Aufgaben ausbaufähig. Insbesondere bei der Aufgabe, einen Termin zu erstellen, werden den beiden schwerwiegenden Problemen dieselben Besuche aus den Produktiv-Logfiles zugeordnet, obwohl beide Probleme von unterschiedlichen Testerinnen und Testern gefunden wurden. Hier könnte die Berücksichtigung der Reihenfolge der Zugriffe eventuell zu einer besseren Trennung führen.

In weiteren Schritten kann die Kombination von Logfiles und Nutzertests methodologisch weiter ausgebaut werden. Im Anschluss an die Kombination könnten die Ergebnisse weitergehend statistisch untersucht werden. Aufgrund der geringen Zahl an schwerwiegenden Problemen wurde beispielsweise von einer Cluster-Analyse abgesehen. Bei einer größeren Anzahl an problematischen Pfadmerkmalen könnten ähnliche Pfadmerkmale zu Clustern gruppiert werden. Ein anderer Ansatz wäre die vorhandenen Logfiles der Produktivsysteme vor der Kombination zu clustern. Dies entspräche dem ursprünglich angedachten, jedoch verworfenen Vorgehen. Bei einer größeren Datenbasis oder der Kombination von Logfiles aus mehreren Installationen der Anwendung macht die vorherige Clusteranalyse trotz des Verwerfens eventuell dennoch Sinn.

Daneben könnte die Kombination auch mit anderen Distanzmaßen als der Kosinus-Ähnlichkeit durchgeführt werden. Ein systematischer Vergleich der verschiedenen Di-

stanzmaße zur Ähnlichkeitsbestimmung von Logfile-Daten liegt nach Kenntnis des Autors bisher noch nicht vor.

Die Nutzerinnen und Nutzer des Produktivsystems führen durchschnittlich weniger als elf Aktionen pro Besuch durch. Die Komb-Matrix verfügt über 31 Spalten und wurde bereits möglichst generisch gewählt. Der genaue Inhalt der Seite, der in den Logfiles durch die eindeutige Seiten-ID dargestellt wird, ist von nachrangigem Interesse, da in dem LMS ein festgelegtes Design vorgeschrieben ist. Dies führt dazu, dass alle Ankündigungen, Materialien, Termine etc. dasselbe Aussehen haben. Der Inhalt einer einzelnen Seite war von geringem Interesse. Trotz dieser Generalisierung können die Besuche im Durchschnitt nicht alle Spalten der Matrix füllen. Eine „leere“ Matrix könnte zu einer Verzerrung der Kombination führen und ist bisher nicht untersucht worden. Hier liegt ein möglicher zweiter Grund, die Transaction Matrix mit einer anderen Darstellungsmethode zu vergleichen. Unklar ist bislang außerdem, ob die unterschiedlichen Längen der Besuche einen Einfluss auf die Zuordnung der generischen problematischen Pfadmerkmale zu den Pfadmerkmalen der „realen“ Logfiles haben.

Zuletzt stellt die hier vorgestellte Methode nur eine Einschätzung der Häufigkeit der gefundenen Usability-Probleme der Nutzertests. In den Analysen wurde deshalb stets davon gesprochen, wie viele Nutzerinnen und Nutzer ebenfalls dasselbe Usability-Problem haben *können*. Bisher liegen keine weiteren Erkenntnisse vor, wie hoch die Aussagekraft der Kombination ist. Dieses Problem lässt sich wahrscheinlich nicht methodologisch berechnen, sondern wird in der Praxis vor allem auf Erfahrungswerten beruhen. Daher sollte die Methode in Zukunft häufiger und auch in anderen Anwendungen angewandt werden, um neben der Aussagekraft weitere methodologische Fortschritte zu erzielen. Insgesamt stellt das in dieser Dissertation vorgestellte Verfahren Usability-Forschern ein nützliches Tool zur Verfügung, um mit einem geringen Mehraufwand die Aussagekraft der Nutzertests signifikant zu steigern. Da gerade in webbasierten Umgebungen eine Protokollierung mindestens auf Basis des Webservers standardmäßig aktiviert ist, kann eine bereits zur Verfügung stehende Datenquelle additiv genutzt werden.

A Heuristiken

A.1 Heuristiken von Nielsen

Die originale Liste wurde von Molich und Nielsen (1990) entwickelt. Der Punkt „Help and documentation“ wurde 1991 hinzugefügt (Nielsen, 1993).

- Simple and natural dialogue
- Speak the users' language
- Minimize the users' memory load
- Consistency
- Clearly marked exits
- Shortcuts
- Precise and constructive error messages
- Prevent errors
- Help and documentation

Das überarbeitete Set stammt aus einer Faktoranalyse von 249 Usability-Problemen (Nielsen, 1994a):

- Visibility of system status: The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
- Match between system and the real world: The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural logical order.
- User control and freedom: Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
- Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
- Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.
- Recognition rather than recall: Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
- Flexibility and efficiency of use: Accelerators – unseen by the novice user – may often speed up the interaction for the expert user to such an extent that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

- Aesthetic and minimalist design: Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
- Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- Help and documentation: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

A.2 Heuristiken von Muller et al. (1998)

1. System status

The system keeps users informed about what is going on through appropriate feedback within a reasonable time.

2. Task sequencing

Users can select and sequence tasks (when appropriate), rather than the system taking control of the user's actions. Wizards are available but are optional and under user control.

3. Emergency exits

Users can easily find „emergency exits“ if they choose system functions by mistake (emergency exits allow the user to leave the unwanted state without having to go through an extended dialogue. Users can make their own decisions (with clear information and feedback) regarding the costs of exiting current work. They can access undo and redo operations.

4. Flexibility and efficiency of

Accelerators are available to experts, but are unseen by the novice. Users are able to tailor frequent actions. Alternative means of access and operation are available for users who differ from the „average“ user (e.g., in physical or cognitive ability, culture, language, etc.).

5. Match between system and the real world

The system speaks the user's language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Messages are based on the user's real world, making information appear in a natural and logical order.

6. Consistency and standards

Each word, phrase, or image in the design is used consistently, with a single meaning. Each interface object or computer operation is always referred to using the same consistent word, phrase, or image. Follow the conventions of the delivery system or platform.

7. Recognition rather than recall

Objects, actions, and options are visible. The user does not have to remember information from one part of the dialogue to another. Instructions for use of the system

are visible or easily retrievable whenever appropriate.

8. Aesthetic and minimalistic design

Dialogs do not contain information that is irrelevant or rarely needed (extra information in a dialog competes with the relevant units of information and diminishes their relative visibility).

9. Help and documentation

The system is intuitive and can be used for the most common tasks without documentation. Where needed, documentation is easy to search, supports a user task, lists concrete steps to be carried out, and is sized appropriately to the user's task. Large documents are supplemented with multiple means of finding their contents (tables of contents, indexes, searches, etc.).

10. Help users recognize, diagnose, and recover from errors

Error messages precisely indicate the problem and constructively suggest a solution. They are expressed in plain (users') language (no codes). Users are not blamed for the errors.

11. Error prevention

Even better than good error messages is a careful design that prevents a problem from occurring in the first place. Users' „errors“ are anticipated, and the system treats the „error“ as either a valid input or an ambiguous input to be clarified.

12. Skills

The system supports, extends, supplements, or enhances the user's skills, background knowledge, and expertise. The system does not replace them. Wizards support, extend, or execute decisions made by users.

13. Pleasurable and respectful interaction with the user

The user's interactions with the system enhance the quality of her or his experience. The user is treated with respect. The design reflects the user's professional role, personal identity, or intention. The design is aesthetically pleasing - with an appropriate balance of artistic as well as functional value.

14. Quality work

The system supports the user in delivering quality work to her or his clients (if appropriate). Attributes of quality work include timeliness, accuracy, aesthetic appeal, and appropriate levels of completeness.

15. Privacy

The system helps the user to protect personal or private information - belonging to the user or to his clients.

A.3 Heuristiken von Shneiderman (eight golden rules; 2010)

1. Strive for consistency

Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; consistent color, layout, capitalization, fonts, and others should be employed throughout. Exceptions should be comprehensible and limited in number.

2. Cater to universal usability

Recognize the needs of diverse users and design for plasticity, facilitating transformation of content. Consider novice-expert differences, age ranges, disabilities, and technology diversity. Add features for novices - for example explanations-, and features for experts - for example shortcuts or faster pacing.

3. Offer informative feedback

Provide system feedback for every user action. For frequent and minor actions, the response can be modest, whereas for infrequent and major actions, the response should be more substantial. Consider the visual presentation of showing changes of the objects of interest.

4. Design dialogs to yield closure

Sequences of actions should be organized into groups with a beginning, middle, and end. Provide informative feedback at the completion of a group of actions.

5. Prevent errors

Design the system such that users cannot make serious errors. In case of an error, the system should offer simple, constructive, and specific actions for recovery. Erroneous actions should leave the system state unchanged, or instructions about restoring the state should be provided.

6. Permit easy reversal of actions

As much as possible, actions should be reversible to encourage users' exploration of unknown options. Units of reversibility could be single actions, a data-entry task, or a complete group of actions, such as entry of a name and address block.

7. Support internal locus of control

Users should feel that they are in charge of the interface and that it responds to their actions. Avoid surprising interface actions, tedious data-entry sequences, difficulties in finding necessary information, and inability to produce the action desired. Make users initiators of actions rather than just responders.

8. Reduce short-term memory load

Human information processing is constrained -7 plus/minus 2 chunks of information are believed to be rememberable. Therefore keep the interface simple and multiple-page displays consolidated. Also reduce window-motion frequency and allow for sufficient training time for codes, mnemonics, and sequences of actions. Where appropriate, provide online access to command-syntax forms, abbreviations, codes, and other necessary information.

A.4 Heuristiken von Sarodnick und Brau (2011)

1. Task adequacy (Aufgabenangemessenheit)

All functionalities required for solving the tasks have to be present within the system. They have to be designed in a way, that they support and relieve the user when performing routine tasks.

2. Process adequacy (Prozessangemessenheit)

The system should be optimized to enable the solving of actual tasks in typical

environments, it should be related to the higher goal of the actual process, and it should be tailored to the qualification and experiences of the real users.

3. Capability of self-description (Selbstbeschreibungsfähigkeit) The system status should be provided in a consistent and immediate way. The user should be able to choose the level of detail of the system status information.
4. Controllability (Steuerbarkeit)
The user should be able to control the dialog, and should have the possibility to use several input assistances, or to quit the system without data loss.
5. Conformance with users' expectations (Erwartungskonformität) The information should conform to system- and platform-specific concepts. For similar tasks, the dialogs should also be similar, and should be displayed at their expected position.
6. Error tolerance (Fehlertoleranz)
Error messages should be expressed clearly. They should contain, for example, information about the type and the context of the error. The user should be informed about irreversible actions.
7. System- and data-safety (System- und Datensicherheit)
The system should always work stable and without data loss, even if users provide defective inputs, or under higher load.
8. Individual adjustability (Individualisierbarkeit)
The dialog system should be individually adjustable, conforming to the users' preferences, as long as it serves the users' effectiveness, efficiency and satisfaction and does not contradict the required technical or security-related constraints.
9. Learning conductiveness (Lernförderlichkeit)
Learning approaches, as for example Learning by Doing, should be supported through stepwise instructions or navigation aids.
10. Control through perception (Wahrnehmungssteuerung)
Interface layout should be minimalistic. Groupings, colors, and useful reduction of information, or similar, should be used in a way that the users' attention is directed to the relevant information.
11. Joy of use (Joy of Use)
Task sequences and graphical design of the system should avoid monotony and appear up to date, but also consider the necessary consistency. Metaphors should be used adequately and should match the context of usage.
12. Intercultural aspects (Interkulturelle Aspekte)
The system should be matching a defined user population in terms of, for example, their functional, organizational, or national culture.

A.5 Heuristiken von Mehlenbacher et al. (2005)

Tabelle A.1: Heuristiken nach Mehlenbacher et al. (2005)

Learner Background and Knowledge	
Accessibility	<p>Has the WBI been viewed on different platforms, browsers, and modem speeds?</p> <p>Is the site ADA compliant (e.g., red and yellow colors are problematic for visually-challenged users)?</p> <p>Have ISO-9000 standards been considered?</p>
Customizability and maintainability	<p>Does printing of the screen(s) require special configuration to optimize presentation and, if so, is this indicated on the site?</p> <p>Are individual preferences/sections clearly distinguishable from one another?</p> <p>Is manipulation of the presentation possible and easy to achieve?</p>
Error support and feedback	<p>Is a design solution possible that prevents a problem from occurring in the first place?</p> <p>When users select something does it differentiate itself from other unselected items?</p> <p>Do menu instructions, prompts, and error messages appear in the same place on each screen?</p>
Navigability and user movement	<p>Does the site clearly separate navigation from content?</p> <p>How many levels down can users traverse and, if more than three, is it clear that returning to their initial state is possible with a single selection?</p> <p>Can users see where they are in the overall site at all times?</p> <p>Do the locations of navigational elements remain consistent?</p> <p>Is the need to scroll minimized across screens and frames within screens?</p>
User control, error tolerance, and flexibility	<p>Are users allowed to undo or redo previous actions?</p> <p>Can users cancel an operation in progress without receiving an error message?</p> <p>Are multiple windows employed and, if so, can they be manipulated easily?</p>
Social Dynamics	
Mutual goals and outcomes	<p>Are learners rewarded for using the communication tools?</p> <p>Are communication tools provided that allow synchronous and asynchronous interaction?</p> <p>Do communication tools allow information revision, organization, and management?</p> <p>Are interactions organized around instructional objectives and task deliverables?</p>
Communication protocols	<p>Are instructions provided for engaging with other learners online?</p> <p>Are documents, resources, and task instructions shared across learners?</p>
Instructional Content	
Completeness	<p>Are levels clear and explicit about the „end“ or parameters of the site?</p> <p>Are there different „levels“ of use and, if so, are they clearly distinguishable?</p>
Examples and case studies	<p>Are examples, demonstrations, or case studies of user experiences available to facilitate learning?</p>

Fortsetzung auf der nächsten Seite

Tabelle A.1 – Fortsetzung

Readability and quality of writing	<p>Are examples divided into meaningful sections, e.g., overview, demonstration, explanation, and so on?</p> <p>Is the text in active voice and concisely written (> 4 < 15 words/sentence)?</p> <p>Are terms consistently plural, verb+object or noun+verb, etc., avoiding unnecessarily redundant words?</p> <p>Do field labels reside on the right of the fields they are closely related to?</p> <p>Does white space highlight a modular text design that separates information chunks from each other?</p> <p>Are bold and color texts used sparingly to identify important text (limiting use of all capitals and italics to improve readability)?</p>
Relationship with real-world tasks	<p>Can users understand the content of the information presented easily?</p> <p>Is terminology and labeling meaningful, concrete, and familiar to the target audience?</p> <p>Do related and interdependent functions and materials appear on the same screen?</p> <p>Is sequencing used naturally, if sequences of common events or narratives are expected?</p> <p>Does the site allow users to easily complete their transactions or selections?</p>
Interaction Display	
Aesthetic appeal	<p>Does the screen design appear minimalist (uncluttered, readable, memorable)?</p> <p>Are graphics or colors employed aesthetically?</p> <p>Are distractions minimized (e.g., movement, blinking, scrolling, animation, etc.)?</p>
Consistency and layout	<p>Does every screen display begin with a title/subject heading that describes contents?</p> <p>Is there a consistent icon design and graphic display across screens?</p> <p>Is layout, font choice, terminology use, color, and positioning of items the same throughout the site (< 4 of any of the above is usually recommended)?</p>
Typographic cues and structuring	<p>Does text employ meaningful discourse cues, modularization, chunking?</p> <p>Is information structured by meaningful labeling, bulleted lists, or iconic markers?</p> <p>Are legible fonts and colors employed?</p> <p>Is the principle of left-to-right placement linked to most-important to least-important information?</p>
Visibility of features and self-description	<p>Are objects, actions, and options visible?</p> <p>Do users have to remember information from one part of a dialogue to another?</p> <p>Are prompts, cues, and messages placed where users will be looking on the screen?</p> <p>Do text areas have „breathing space“ around them?</p> <p>Is white space used to create symmetry and to lead the eye in the appropriate direction?</p>

Instructor Activities

Fortsetzung auf der nächsten Seite

Tabelle A.1 – Fortsetzung

Authority and authenticity	Does the site establish a serious tone or presence? Are users reminded of the security and privacy of the site? Are humor or anthropomorphic expressions used minimally? Is direction given for further assistance if necessary?
Intimacy and presence	Is an overall tone established that is present, active, and engaging? Does the site act as a learning environment for users, not simply as a warehouse of unrelated links?
Environment and Tools	
Help and support documentation	Does the site support task-oriented help, tutorials, and reference documentation? Is help easy to locate and access on the site? Is the help table of contents or menu organized functionally, according to user tasks?
Metaphors and maps	Does the site use an easily recognizable metaphor that helps users identify tools in relation to each other, their state in the system, and options available to them?
Organization and information relevance	Is a site map available? Is the overall organization of the site clear from the majority of screens? Are primary options emphasized in favor of secondary ones?
Reliability and functionality	Do all the menus, icons, links, and opening windows work predictably across platforms? Have important interactive features and multimedia elements been tested across platforms and browsers?

A.6 Heuristiken von Reeves et al. (2002)

1. *Visibility of system status*: The e-learning program keeps the learner informed about what is happening, through appropriate feedback within reasonable time.
2. *Match between system and the real world*: The e-learning program's interface employs words, phrases and concepts familiar to the learner or appropriate to the content, as opposed to system-oriented terms. Wherever possible, the e-learning program utilizes real-world conventions that make information appear in a natural and logical order.
3. *Error recovery and exiting*: The e-learning program allows the learner to recover from input mistakes and provides a clearly marked "exit" to leave the program without requiring the user to go through an extended dialogue.
4. *Consistency and standards*: When appropriate to the content and target audience, the e-learning program adheres to general software conventions and is consistent in its use of different words, situations, or actions.
5. *Error prevention*: The e-learning program is designed to prevent common problems from occurring in the first place.
6. *Navigation support*: The e-learning program makes objects, actions, and options visible so that the user does not have to remember information when navigating from one part of the program to another. Instructions for use of the program are always visible or easily retrievable.

7. *Aesthetics*: Screen displays do not contain information that is irrelevant, and “bells and whistles” are not gratuitously added to the e-learning program.
8. *Help and documentation*: The e-learning program provides help and documentation that is readily accessible to the user when necessary. The help provides specific concrete steps for the user to follow. All documentation is written clearly and succinctly.
9. *Interactivity*: The e-learning program provides content-related interactions and tasks that support meaningful learning.
10. *Message Design*: The e-learning program presents information in accord with sound information-processing principles.
11. *Learning Design*: The interactions in the e-learning program have been designed in accord with sound principles of learning theory.
12. *Media Integration*: The inclusion of media in the e-learning program serves clear pedagogical and/or motivational purposes.
13. *Instructional Assessment*: The e-learning program provides assessment opportunities that are aligned with the program objectives and content.
14. *Resources*: The e-learning program provides access to all the resources necessary to support effective learning.
15. *Feedback*: The e-learning program provides feedback that is contextual and relevant to the problem or task in which the learner is engaged.

B Tabellen

B.1 Zugriffe auf das Produktivsystem im Zeitverlauf

Tabelle B.1: Zugriffe auf InfoChange im Zeitverlauf

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
02.03.2011	10	7
03.03.2011	10	8
04.03.2011	3	1
05.03.2011	0	0
06.03.2011	1	0
07.03.2011	0	0
08.03.2011	1	1
09.03.2011	1	1
10.03.2011	4	4
11.03.2011	0	0
12.03.2011	2	2
13.03.2011	0	0
14.03.2011	0	0
15.03.2011	1	1
16.03.2011	2	2
17.03.2011	2	1
18.03.2011	4	4
19.03.2011	2	2
20.03.2011	11	4
21.03.2011	9	5
22.03.2011	8	5
23.03.2011	13	11
24.03.2011	5	5
25.03.2011	3	3
26.03.2011	1	1
27.03.2011	11	5
28.03.2011	15	6
29.03.2011	11	9
30.03.2011	5	5
31.03.2011	9	6
01.04.2011	9	5
02.04.2011	4	4
03.04.2011	12	7
04.04.2011	15	9
05.04.2011	11	11
06.04.2011	4	3
07.04.2011	6	4
08.04.2011	1	0

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
09.04.2011	3	1
10.04.2011	5	4
11.04.2011	13	7
12.04.2011	8	4
13.04.2011	14	3
14.04.2011	17	7
15.04.2011	6	2
16.04.2011	11	1
17.04.2011	8	4
18.04.2011	11	5
19.04.2011	3	0
20.04.2011	7	4
21.04.2011	7	3
12.06.2011	7	7
13.06.2011	9	8
14.06.2011	10	10
15.06.2011	10	7
16.06.2011	5	4
17.06.2011	14	12
18.06.2011	5	4
19.06.2011	5	5
20.06.2011	13	12
21.06.2011	11	9
22.06.2011	8	5
23.06.2011	11	5
24.06.2011	9	6
25.06.2011	3	2
26.06.2011	9	5
27.06.2011	6	3
28.06.2011	5	5
29.06.2011	9	5
30.06.2011	6	3
01.07.2011	4	3
02.07.2011	0	0
03.07.2011	2	2
04.07.2011	0	0
05.07.2011	1	1
06.07.2011	1	1
07.07.2011	1	1
08.07.2011	1	1
09.07.2011	0	0
10.07.2011	1	1
11.07.2011	0	0
12.07.2011	0	0
13.07.2011	0	0
14.07.2011	0	0
15.07.2011	1	1
16.07.2011	0	0
17.07.2011	0	0
18.07.2011	3	2

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
19.07.2011	1	1
20.07.2011	0	0
21.07.2011	1	1
22.07.2011	0	0
23.07.2011	0	0
24.07.2011	1	1
25.07.2011	3	3
26.07.2011	1	0
27.07.2011	2	2
28.07.2011	6	4
29.07.2011	7	7
30.07.2011	3	3
31.07.2011	1	1
01.08.2011	11	7
02.08.2011	16	9
03.08.2011	14	8
04.08.2011	12	9
05.08.2011	27	17
06.08.2011	14	2
07.08.2011	42	34
08.08.2011	46	37
09.08.2011	42	36
10.08.2011	69	53
11.08.2011	47	31
12.08.2011	20	14
13.08.2011	11	11
14.08.2011	36	30
15.08.2011	26	18
16.08.2011	23	15
17.08.2011	15	7
18.08.2011	26	22
19.08.2011	23	13
20.08.2011	6	5
21.08.2011	30	26
22.08.2011	24	19
23.08.2011	32	24
24.08.2011	31	20
25.08.2011	44	31
26.08.2011	25	14
27.08.2011	5	4
28.08.2011	24	16
29.08.2011	34	18
30.08.2011	31	26
31.08.2011	24	18
01.09.2011	17	12
02.09.2011	9	8
03.09.2011	3	3
04.09.2011	18	13
05.09.2011	25	19
06.09.2011	25	19

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
07.09.2011	2	1
18.09.2011	30	22
19.09.2011	13	11
20.09.2011	9	9
21.09.2011	11	9
22.09.2011	8	6
23.09.2011	8	7
24.09.2011	3	2
25.09.2011	19	7
26.09.2011	19	4
27.09.2011	18	8
28.09.2011	13	10
29.09.2011	18	10
30.09.2011	7	2
01.10.2011	2	2
02.10.2011	3	3
03.10.2011	7	6
04.10.2011	2	1
05.10.2011	1	1
06.10.2011	2	2
07.10.2011	0	0
08.10.2011	1	1
09.10.2011	1	1
10.10.2011	1	1
11.10.2011	4	4
12.10.2011	4	4
13.10.2011	4	4
14.10.2011	12	8
15.10.2011	9	6
16.10.2011	22	18
17.10.2011	14	14
18.10.2011	9	9
19.10.2011	9	8
20.10.2011	6	6
21.10.2011	0	0
22.10.2011	6	3
23.10.2011	16	7
24.10.2011	19	17
25.10.2011	15	13
26.10.2011	9	9
27.10.2011	8	6
28.10.2011	3	3
29.10.2011	3	3
30.10.2011	22	13
31.10.2011	17	15
01.11.2011	13	12
02.11.2011	5	4
03.11.2011	4	3
04.11.2011	8	5
05.11.2011	1	0

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
06.11.2011	15	11
07.11.2011	17	14
08.11.2011	13	11
09.11.2011	10	9
10.11.2011	12	12
11.11.2011	7	6
12.11.2011	3	3
13.11.2011	6	2
14.11.2011	9	7
15.11.2011	16	13
16.11.2011	19	17
17.11.2011	5	4
18.11.2011	2	1
19.11.2011	1	1
20.11.2011	11	8
21.11.2011	10	7
22.11.2011	8	4
23.11.2011	16	14
24.11.2011	17	7
25.11.2011	8	6
26.11.2011	3	2
27.11.2011	16	8
28.11.2011	7	5
29.11.2011	11	11
30.11.2011	7	6
01.12.2011	7	6
02.12.2011	6	6
03.12.2011	5	4
04.12.2011	8	6
05.12.2011	21	13
06.12.2011	11	7
07.12.2011	8	7
08.12.2011	7	5
09.12.2011	9	7
10.12.2011	1	0
11.12.2011	9	8
12.12.2011	16	10
13.12.2011	13	11
14.12.2011	4	3
15.12.2011	19	13
16.12.2011	7	5
17.12.2011	4	3
18.12.2011	6	5
19.12.2011	15	15
20.12.2011	5	5
21.12.2011	8	5
22.12.2011	13	7
23.12.2011	9	5
24.12.2011	2	2
25.12.2011	0	0

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
26.12.2011	0	0
27.12.2011	1	1
28.12.2011	1	0
29.12.2011	3	3
30.12.2011	1	1
31.12.2011	1	1
01.01.2012	1	1
02.01.2012	6	6
03.01.2012	1	1
04.01.2012	10	10
05.01.2012	6	5
06.01.2012	9	9
07.01.2012	9	7
08.01.2012	26	20
09.01.2012	18	13
10.01.2012	11	5
11.01.2012	19	11
12.01.2012	17	12
13.01.2012	14	12
14.01.2012	6	3
15.01.2012	8	5
16.01.2012	10	9
17.01.2012	12	10
18.01.2012	5	5
19.01.2012	19	12
20.01.2012	6	3
21.01.2012	0	0
22.01.2012	14	9
23.01.2012	10	9
24.01.2012	17	11
25.01.2012	14	8
26.01.2012	11	7
27.01.2012	5	3
28.01.2012	5	1
29.01.2012	11	7
30.01.2012	18	11
31.01.2012	20	8
01.02.2012	9	6
02.02.2012	18	10
03.02.2012	13	12
04.02.2012	5	3
05.02.2012	15	12
06.02.2012	21	18
07.02.2012	12	10
08.02.2012	18	12
09.02.2012	12	9
10.02.2012	14	9
11.02.2012	4	3
12.02.2012	15	2
13.02.2012	18	8

Fortsetzung auf der nächsten Seite

Tabelle B.1 – Fortsetzung

Datum	alle Nutzer_innen	ausgewählte Nutzer_innen
14.02.2012	16	12
15.02.2012	16	10
16.02.2012	19	15
17.02.2012	8	7
18.02.2012	5	3
19.02.2012	14	5
20.02.2012	12	8
21.02.2012	15	12
22.02.2012	19	18
23.02.2012	10	7
24.02.2012	3	1
25.02.2012	3	2
26.02.2012	11	8
27.02.2012	17	14
28.02.2012	16	13
29.02.2012	14	11
01.03.2012	14	9
02.03.2012	9	5
03.03.2012	1	0
04.03.2012	3	3
05.03.2012	4	4
06.03.2012	1	1
07.03.2012	0	0
08.03.2012	1	1

B.2 Zugriffe auf die Rubriken und Funktionen des Produktivsystems

Tabelle B.2: Zugriffe auf die Rubriken und Funktionen (Detailansicht)

Rubriken	Funktionen	Zugriffe
Materialien	Details	5.882
Materialien	Herunterladen	4.865
Materialien	Übersicht	1.291
Materialien	Bearbeiten	584
Materialien	Schlagwort	131
Materialien	Kategorie	100
Ankündigungen	Details	2054
Ankündigungen	Übersicht	501
Ankündigungen	Bearbeiten	89
Ankündigungen	Kategorie	17
Ankündigungen	Schlagwort	12
Termine	Übersicht	831
Termine	Details	755
Termine	Bearbeiten	281
Suche	Übersicht	1.156
Suche	Kategorie	466
Suche	Schlagwort	124
Startseite	Übersicht	1.746
Benutzer	Übersicht	494
Benutzer	Details	433
Benutzer	Bearbeiten	98
Gruppen	Übersicht	312
Gruppen	Details	311
Gruppen	Bearbeiten	32
Fächer	Details	318
Fächer	Übersicht	306
Fächer	Bearbeiten	24
Diskussion	Details	169
Diskussion	Übersicht	162
Diskussion	Bearbeiten	4
Diskussion	Kategorie	2

Ähnlichkeit der Pfade

Tabelle B.3: Ähnlichkeit der Pfade

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
1	5	7	0,817	0,524	M_D ⇒ M_B ⇒ St_Ü ⇒ M_D ⇒ M_Ü	M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ A_D ⇒ A_D ⇒ M_H
2	20	4	0,863	0,697	A_D ⇒ M_D ⇒ M_H ⇒ A_D ⇒ St_Ü ⇒ A_D ⇒ M_H ⇒ St_Ü ⇒ F_D ⇒ T_D ⇒ F_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ G_D ⇒ M_D ⇒ M_Ü ⇒ A_Ü	M_D ⇒ M_D ⇒ M_H ⇒ M_Ü
3	3	12	0,402	0,136	M_Ü ⇒ M_D ⇒ M_H	Su_Ü ⇒ St_Ü ⇒ F_Ü ⇒ B_Ü ⇒ B_D ⇒ G_D ⇒ B_D ⇒ B_Ü ⇒ D_Ü ⇒ M_Ü ⇒ T_Ü ⇒ Su_Ü
4	4	46	0,791	0,469	M_D ⇒ A_D ⇒ M_H ⇒ St_Ü	M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_B ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_Ü ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_B ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B
5	16	25	0,378	0,033	M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_B ⇒ M_D ⇒ A_Ü ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_H	D_Ü ⇒ D_D ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ A_D ⇒ A_D ⇒ Su_K ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D ⇒ Su_K ⇒ T_D
6	5	3	0,818	0,6	M_Ü ⇒ F_Ü ⇒ F_D ⇒ St_Ü ⇒ Su_Ü	Su_Ü ⇒ Su_Ü ⇒ St_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
7	30	32	0,825	0,796	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	Su_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ M_S ⇒ M_- D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_S ⇒ M_D
8	21	6	0,874	0,787	A_D ⇒ M_D ⇒ M_H ⇒ F_Ü ⇒ M_Ü ⇒ M_D ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ Su_K ⇒ M_D ⇒ M_D ⇒ Su_Ü ⇒ St_Ü ⇒ Su_K ⇒ Su_K ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
9	5	3	0,452	0,218	T_Ü ⇒ M_Ü ⇒ T_Ü ⇒ T_D ⇒ A_D	A_D ⇒ M_D ⇒ M_H
10	3	3	0,507	0	A_D ⇒ M_D ⇒ M_H	B_D ⇒ B_D ⇒ G_D
11	5	10	0,39	0,327	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	Su_S ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ A_Ü ⇒ T_Ü ⇒ T_D ⇒ St_Ü ⇒ Su_S ⇒ St_Ü
12	8	6	0,306	0	A_D ⇒ M_H ⇒ B_Ü ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ A_D ⇒ M_H	Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü
13	9	16	0,275	0,054	A_Ü ⇒ T_Ü ⇒ T_D ⇒ St_Ü ⇒ M_Ü ⇒ T_Ü ⇒ T_B ⇒ T_Ü ⇒ St_Ü	G_Ü ⇒ G_D ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
14	10	7	0,842	0,296	Su_K ⇒ A_D ⇒ T_Ü ⇒ M_Ü ⇒ F_Ü ⇒ St_Ü ⇒ A_D ⇒ T_Ü ⇒ D_Ü ⇒ D_D	Su_Ü ⇒ A_D ⇒ Su_Ü ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D
15	4	8	0,39	0	T_Ü ⇒ T_D ⇒ T_Ü ⇒ T_D	B_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ B_D ⇒ St_Ü ⇒ B_Ü ⇒ B_D ⇒ B_Ü
16	12	11	0,591	0,311	A_D ⇒ T_Ü ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ St_Ü ⇒ B_Ü ⇒ B_D ⇒ B_Ü ⇒ St_Ü ⇒ A_D ⇒ St_Ü ⇒ M_Ü ⇒ M_B ⇒ Su_K

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
17	19	4	0,966	0,91	A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_K ⇒ T_Ü ⇒ T_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_D ⇒ M_H
18	6	3	0,601	0,471	B_Ü ⇒ M_Ü ⇒ M_D ⇒ A_Ü ⇒ A_D ⇒ M_H	Su_K ⇒ M_D ⇒ M_H
19	9	3	0,857	0,397	M_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_S ⇒ M_S ⇒ M_Ü ⇒ St_Ü	Su_Ü ⇒ M_D ⇒ M_H
20	26	29	0,699	0,5	M_Ü ⇒ M_S ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_S ⇒ M_Ü ⇒ M_S ⇒ M_Ü ⇒ M_- S ⇒ M_Ü ⇒ M_S ⇒ M_Ü ⇒ M_S ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_K ⇒ A_D ⇒ Su_K ⇒ Su_K ⇒ D_Ü ⇒ D_D ⇒ M_D ⇒ M_H ⇒ B_Ü ⇒ B_D ⇒ St_Ü ⇒ Su_K ⇒ M_D ⇒ D_Ü ⇒ D_D ⇒ St_Ü ⇒ Su_K ⇒ D_Ü ⇒ D_D ⇒ D_D ⇒ D_D
21	4	11	0,635	0,686	A_D ⇒ M_H ⇒ A_D ⇒ M_H	A_D ⇒ Su_K ⇒ A_D ⇒ M_H ⇒ T_Ü ⇒ A_D ⇒ B_Ü ⇒ B_D ⇒ G_D ⇒ F_Ü ⇒ D_Ü
22	44	3	0,725	0,267	G_Ü ⇒ G_D ⇒ G_D ⇒ G_D ⇒ G_D ⇒ G_Ü ⇒ G_D ⇒ G_Ü ⇒ B_Ü ⇒ G_Ü ⇒ G_D ⇒ B_- D ⇒ St_Ü ⇒ A_D ⇒ M_H ⇒ A_D ⇒ M_H ⇒ D_Ü ⇒ D_D ⇒ D_D ⇒ D_D ⇒ D_D ⇒ D_D ⇒ F_Ü ⇒ F_D ⇒ F_D ⇒ F_D ⇒ F_D ⇒ F_- D ⇒ F_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_D	A_D ⇒ D_D ⇒ A_D
23	6	8	0,824	0,612	M_D ⇒ M_D ⇒ A_D ⇒ M_H ⇒ M_Ü ⇒ M_S	M_Ü ⇒ F_Ü ⇒ F_D ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_Ü ⇒ St_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
24	29	7	0,859	0,774	$M_D \Rightarrow M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_{\ddot{U}} \Rightarrow M_D$ $\Rightarrow M_H \Rightarrow M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow$ $St_{\ddot{U}} \Rightarrow M_D \Rightarrow M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D$ $\Rightarrow M_{\ddot{U}} \Rightarrow M_D \Rightarrow M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow$ $St_{\ddot{U}} \Rightarrow Su_{\ddot{U}} \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_D$ $\Rightarrow M_H$	$Su_{\ddot{U}} \Rightarrow M_D \Rightarrow M_H \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}} \Rightarrow$ $M_D \Rightarrow M_H$
25	26	8	0,769	0,336	$M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D$ $\Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_{-}$ $D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow$ $M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow St_{\ddot{U}} \Rightarrow M_{\ddot{U}}$ $\Rightarrow M_D \Rightarrow M_D \Rightarrow M_H$	$A_D \Rightarrow M_H \Rightarrow A_D \Rightarrow M_H \Rightarrow M_{\ddot{U}} \Rightarrow M_K$ $\Rightarrow A_{\ddot{U}} \Rightarrow A_D$
26	15	17	0,632	0,617	$M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D$ $\Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow$ $M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H$	$M_{\ddot{U}} \Rightarrow G_{\ddot{U}} \Rightarrow M_{\ddot{U}} \Rightarrow A_{\ddot{U}} \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}}$ $\Rightarrow M_{\ddot{U}} \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}} \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}}$ $\Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow$ M_H
27	5	20	0,776	0,55	$Su_{\ddot{U}} \Rightarrow M_D \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}} \Rightarrow M_D$	$A_D \Rightarrow M_H \Rightarrow A_D \Rightarrow A_D \Rightarrow M_H \Rightarrow St_{\ddot{U}}$ $\Rightarrow Su_{\ddot{U}} \Rightarrow T_D \Rightarrow T_D \Rightarrow Su_{\ddot{U}} \Rightarrow M_D \Rightarrow$ $M_{\ddot{U}} \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow D_{\ddot{U}}$ $\Rightarrow D_D \Rightarrow M_D \Rightarrow M_H$
28	35	29	0,948	0,921	$Su_K \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}} \Rightarrow M_D \Rightarrow M_D \Rightarrow M_{-}$ $D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow$ $M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H$ $\Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_{-}$ $H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow$ $M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D$ $\Rightarrow M_H$	$A_{\ddot{U}} \Rightarrow T_{\ddot{U}} \Rightarrow M_{\ddot{U}} \Rightarrow M_K \Rightarrow M_D \Rightarrow Su_K$ $\Rightarrow M_D \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_{-}$ $D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow$ $M_D \Rightarrow M_H \Rightarrow M_D \Rightarrow M_H \Rightarrow Su_K \Rightarrow A_{\ddot{U}}$ $\Rightarrow A_D \Rightarrow M_H \Rightarrow Su_K \Rightarrow Su_K \Rightarrow A_D \Rightarrow$ $T_{\ddot{U}}$
29	3	3	0,189	0	$M_D \Rightarrow M_D \Rightarrow M_H$	$A_D \Rightarrow St_{\ddot{U}} \Rightarrow Su_{\ddot{U}}$
30	4	9	0,814	0,606	$A_D \Rightarrow M_{\ddot{U}} \Rightarrow M_D \Rightarrow M_H$	$M_D \Rightarrow M_H \Rightarrow M_{\ddot{U}} \Rightarrow F_{\ddot{U}} \Rightarrow F_D \Rightarrow M_{\ddot{U}}$ $\Rightarrow F_{\ddot{U}} \Rightarrow F_D \Rightarrow M_D$

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
31	7	6	0,802	0,527	A_D ⇒ M_H ⇒ T_D ⇒ T_Ü ⇒ St_Ü ⇒ M_D ⇒ M_H	Su_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D
32	10	5	0,667	0,478	A_Ü ⇒ T_Ü ⇒ M_Ü ⇒ St_Ü ⇒ F_Ü ⇒ A_Ü ⇒ A_D ⇒ G_Ü ⇒ St_Ü ⇒ M_D	D_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü
33	10	15	0,624	0,444	M_Ü ⇒ M_D ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H	Su_Ü ⇒ M_D ⇒ M_D ⇒ St_Ü ⇒ Su_Ü ⇒ M_- D ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü
34	9	3	0,576	0,361	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D	T_Ü ⇒ T_D ⇒ M_H
35	3	3	0,775	0,775	A_D ⇒ M_H ⇒ A_D	Su_Ü ⇒ A_D ⇒ M_H
36	20	4	0,597	0,562	A_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ T_Ü ⇒ A_Ü ⇒ A_D ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ D_D ⇒ D_B ⇒ St_Ü ⇒ A_B ⇒ T_D	A_D ⇒ M_H ⇒ M_D ⇒ M_H
37	27	3	0,779	0,655	A_D ⇒ M_D ⇒ M_D ⇒ Su_Ü ⇒ Su_S ⇒ M_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ St_Ü ⇒ M_Ü ⇒ M_B ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B	M_D ⇒ Su_K ⇒ M_D
38	4	13	0,702	0,289	M_D ⇒ A_Ü ⇒ A_D ⇒ M_H	Su_S ⇒ T_D ⇒ M_Ü ⇒ M_S ⇒ M_D ⇒ St_Ü ⇒ M_Ü ⇒ Su_S ⇒ T_D ⇒ M_Ü ⇒ M_S ⇒ M_D ⇒ M_H
39	8	20	0,846	0,711	A_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_K ⇒ Su_K ⇒ Su_K ⇒ A_D ⇒ A_D ⇒ A_D ⇒ M_H ⇒ A_D ⇒ Su_K ⇒ G_Ü ⇒ G_D ⇒ G_D ⇒ B_D ⇒ M_Ü ⇒ M_D
40	11	9	0,871	0,822	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_B ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ T_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
41	3	10	0,933	0,645	Su_Ü ⇒ A_D ⇒ M_Ü	T_D ⇒ Su_Ü ⇒ T_D ⇒ M_Ü ⇒ A_Ü ⇒ Su_Ü ⇒ A_D ⇒ Su_Ü ⇒ M_D ⇒ M_D
42	3	3	0,283	0	G_Ü ⇒ G_D ⇒ G_D	A_Ü ⇒ M_D ⇒ M_H
43	3	3	0,625	0,333	A_Ü ⇒ A_D ⇒ M_H	A_D ⇒ Su_K ⇒ M_D
44	11	36	0,816	0,391	M_Ü ⇒ M_D ⇒ F_Ü ⇒ F_D ⇒ M_Ü ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_B ⇒ M_D	A_D ⇒ St_Ü ⇒ M_D ⇒ M_H ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_- D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
45	10	4	0,446	0,098	Su_Ü ⇒ A_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H	T_Ü ⇒ T_D ⇒ St_Ü ⇒ A_D
46	13	10	0,576	0,332	M_Ü ⇒ M_D ⇒ M_B ⇒ M_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_Ü ⇒ M_D	G_Ü ⇒ G_B ⇒ G_D ⇒ M_Ü ⇒ G_Ü ⇒ G_D ⇒ G_B ⇒ M_Ü ⇒ M_B ⇒ D_Ü
47	8	8	0,949	0,791	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_Ü ⇒ T_Ü	A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ A_D
48	3	7	0,839	0,696	D_D ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_D ⇒ M_Ü ⇒ A_Ü ⇒ A_D ⇒ M_H
49	13	15	0,887	0,854	M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü ⇒ Su_K ⇒ Su_Ü ⇒ Su_K ⇒ Su_K ⇒ M_Ü
50	12	8	0,625	0,293	M_Ü ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ M_H	Su_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H
51	10	3	0,855	0,629	M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_D	Su_Ü ⇒ M_D ⇒ M_H

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
52	12	20	0,825	0,711	M_B ⇒ A_D ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ A_Ü ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H	M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
53	5	5	0,381	0	M_D ⇒ M_H ⇒ M_Ü ⇒ St_Ü ⇒ A_D	T_B ⇒ T_Ü ⇒ T_D ⇒ B_Ü ⇒ D_Ü
54	24	8	0,41	0,165	Su_S ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_S ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_S ⇒ M_D ⇒ M_D ⇒ M_H	A_D ⇒ St_Ü ⇒ T_Ü ⇒ St_Ü ⇒ M_Ü ⇒ St_Ü Ü ⇒ A_D ⇒ M_H
55	3	7	0,946	0,759	M_D ⇒ M_D ⇒ M_H	A_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
56	3	12	0,292	0,075	Su_Ü ⇒ A_D ⇒ A_D	F_B ⇒ Su_Ü ⇒ M_B ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
57	4	6	0,849	0,632	M_D ⇒ A_D ⇒ M_H ⇒ T_Ü	A_D ⇒ M_H ⇒ A_D ⇒ M_H ⇒ A_Ü ⇒ St_Ü
58	3	16	0,365	0	A_D ⇒ A_D ⇒ A_D	T_D ⇒ T_Ü ⇒ St_Ü ⇒ M_D ⇒ M_Ü ⇒ St_Ü ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_Ü ⇒ B_D
59	3	6	0,312	0	M_D ⇒ M_Ü ⇒ M_D	A_D ⇒ B_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü
60	23	4	0,434	0,173	Su_K ⇒ Su_K ⇒ St_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ M_D ⇒ M_Ü ⇒ M_K ⇒ M_K ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ Su_K ⇒ A_D
61	13	4	0,419	0,384	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_K ⇒ Su_K ⇒ Su_K ⇒ Su_K ⇒ Su_K ⇒ Su_K ⇒ M_D ⇒ M_D ⇒ M_H	A_B ⇒ A_D ⇒ St_Ü ⇒ Su_K

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
62	5	8	0,747	0,624	M_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_B	M_D ⇒ M_D ⇒ M_D ⇒ M_Ü ⇒ F_Ü ⇒ F_D ⇒ St_Ü ⇒ Su_Ü
63	5	53	0,685	0,694	Su_K ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H	Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
64	7	5	0,778	0,577	F_Ü ⇒ F_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ Su_Ü
65	13	16	0,299	0,037	M_D ⇒ M_D ⇒ M_H ⇒ M_B ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_Ü ⇒ A_Ü ⇒ M_Ü ⇒ T_Ü ⇒ T_D ⇒ T_B ⇒ T_B ⇒ T_D ⇒ T_B ⇒ T_B ⇒ T_Ü ⇒ T_D ⇒ T_B ⇒ T_B ⇒ T_D ⇒ T_Ü
66	6	31	0,65	0,298	A_D ⇒ M_H ⇒ A_D ⇒ M_H ⇒ A_D ⇒ A_D	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_- D ⇒ M_H ⇒ M_B ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ St_Ü ⇒ Su_Ü ⇒ A_D ⇒ St_Ü ⇒ M_D ⇒ M_H

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
67	3	122	0,929	0,832	Su_Ü ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_B ⇒ M_B ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_B ⇒ M_B ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_B ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_- D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ B_D ⇒ B_B ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ A_D ⇒ A_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ A_Ü
68	5	3	0,248	0	A_Ü ⇒ St_Ü ⇒ A_Ü ⇒ A_D ⇒ A_Ü	M_Ü ⇒ M_D ⇒ M_H
69	11	11	0,353	0,126	T_D ⇒ M_H ⇒ T_D ⇒ M_H ⇒ G_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	Su_K ⇒ A_D ⇒ Su_K ⇒ Su_K ⇒ Su_Ü ⇒ A_D ⇒ A_D ⇒ A_D ⇒ St_Ü ⇒ B_Ü ⇒ B_D

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
70	65	5	0,457	0,01	Su_K ⇒ A_D ⇒ M_H ⇒ Su_K ⇒ M_D ⇒ M_H ⇒ D_D ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_B ⇒ St_Ü ⇒ Su_Ü ⇒ A_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	F_Ü ⇒ F_D ⇒ F_D ⇒ T_D ⇒ M_Ü
71	30	3	0,468	0,079	A_D ⇒ M_H ⇒ T_Ü ⇒ B_Ü ⇒ B_D ⇒ B_B ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ B_D ⇒ B_Ü ⇒ B_D ⇒ F_Ü ⇒ F_D ⇒ F_Ü ⇒ F_D ⇒ F_D ⇒ M_Ü ⇒ M_K ⇒ M_K	A_D ⇒ M_H ⇒ St_Ü
72	9	4	0,484	0,297	G_Ü ⇒ G_D ⇒ B_D ⇒ St_Ü ⇒ T_B ⇒ St_Ü ⇒ Su_Ü ⇒ St_Ü ⇒ Su_Ü	A_D ⇒ M_H ⇒ St_Ü ⇒ A_D
73	10	3	0,125	0	M_D ⇒ M_B ⇒ T_Ü ⇒ M_Ü ⇒ M_D ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_B ⇒ M_D	Su_K ⇒ Su_K ⇒ A_D
74	7	4	0,799	0,655	M_D ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H	F_Ü ⇒ F_D ⇒ M_D ⇒ M_H
75	6	3	0	0	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	T_Ü ⇒ St_Ü ⇒ T_Ü
76	12	3	0,361	0,123	A_D ⇒ M_H ⇒ St_Ü ⇒ A_D ⇒ M_Ü ⇒ M_D ⇒ A_Ü ⇒ F_Ü ⇒ F_D ⇒ F_D ⇒ F_D ⇒ M_D	B_Ü ⇒ St_Ü ⇒ Su_Ü
77	5	3	0,696	0,258	M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ T_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
78	4	16	0,843	0,495	Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H	F_B ⇒ F_Ü ⇒ M_Ü ⇒ M_B ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_B ⇒ M_D ⇒ M_Ü
79	6	114	0,437	0,25	B_Ü ⇒ B_B ⇒ B_B ⇒ A_D ⇒ M_H ⇒ T_D	M_B ⇒ Su_K ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_- D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_- D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
80	24	4	0,854	0,459	Su_Ü ⇒ M_Ü ⇒ st_Ü ⇒ Su_Ü ⇒ st_Ü ⇒ Su_Ü ⇒ M_D ⇒ st_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ A_Ü ⇒ st_Ü ⇒ Su_Ü ⇒ A_D ⇒ B_Ü ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ st_Ü ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ st_Ü ⇒ T_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
81	6	7	0,918	0,763	A_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_Ü	Su_K ⇒ M_Ü ⇒ M_K ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
82	32	58	0,975	0,964	M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	Su_Ü ⇒ M_D ⇒ M_H ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ St_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ Su_Ü
83	12	26	0,994	0,984	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_- H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_D ⇒ M_H ⇒ A_D ⇒ M_H
84	7	6	0,32	0	A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_Ü ⇒ A_D ⇒ M_H	T_D ⇒ Su_Ü ⇒ B_Ü ⇒ Su_Ü ⇒ B_D ⇒ B_D
85	8	25	0,861	0,599	M_Ü ⇒ M_D ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_B ⇒ M_D	M_D ⇒ M_H ⇒ St_Ü ⇒ M_Ü ⇒ St_Ü ⇒ Su_- Ü ⇒ M_D ⇒ M_D ⇒ St_Ü ⇒ Su_Ü ⇒ A_D ⇒ A_D ⇒ A_D ⇒ A_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_D ⇒ A_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
86	3	10	0,601	0,456	A_D ⇒ A_D ⇒ M_H	A_D ⇒ M_H ⇒ Su_Ü ⇒ M_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_H ⇒ Su_Ü ⇒ M_D ⇒ M_H
87	3	10	0,289	0,109	B_Ü ⇒ St_Ü ⇒ Su_Ü	T_D ⇒ T_B ⇒ T_B ⇒ T_D ⇒ T_Ü ⇒ T_B ⇒ T_B ⇒ T_D ⇒ A_Ü ⇒ St_Ü

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
88	9	40	0,892	0,758	G_D ⇒ G_Ü ⇒ St_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	T_Ü ⇒ A_Ü ⇒ T_Ü ⇒ A_Ü ⇒ A_D ⇒ A_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_D ⇒ M_- H ⇒ A_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ A_- D ⇒ M_H ⇒ A_D ⇒ A_D ⇒ A_D ⇒ St_Ü ⇒ T_Ü ⇒ M_Ü ⇒ D_Ü ⇒ D_D ⇒ D_D ⇒ D_D
89	12	13	0,84	0,639	A_D ⇒ Su_K ⇒ T_Ü ⇒ St_Ü ⇒ T_Ü ⇒ St_- Ü ⇒ A_D ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ F_Ü ⇒ Su_Ü ⇒ F_D
90	12	4	0,772	0,231	T_Ü ⇒ St_Ü ⇒ M_Ü ⇒ M_S ⇒ M_D ⇒ M_H ⇒ M_S ⇒ M_D ⇒ M_S ⇒ M_S ⇒ St_Ü ⇒ Su_Ü	M_B ⇒ M_B ⇒ M_D ⇒ M_Ü
91	7	24	0,854	0,686	M_D ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	A_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ St_Ü ⇒ F_Ü ⇒ F_D ⇒ M_Ü ⇒ Su_Ü ⇒ M_Ü ⇒ M_B ⇒ M_Ü ⇒ M_Ü ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_H
92	4	3	0,354	0	Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H	A_D ⇒ St_Ü ⇒ T_Ü
93	5	9	0,837	0,484	A_D ⇒ Su_K ⇒ M_D ⇒ M_D ⇒ M_D	T_D ⇒ St_Ü ⇒ T_D ⇒ A_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H
94	3	23	0,769	0,596	A_D ⇒ A_D ⇒ M_H	A_D ⇒ A_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ D_Ü ⇒ D_- D ⇒ A_Ü ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ T_Ü ⇒ A_Ü ⇒ A_D ⇒ M_H
95	5	3	0,671	0,507	A_D ⇒ M_H ⇒ St_Ü ⇒ M_D ⇒ St_Ü	A_D ⇒ A_D ⇒ M_H
96	6	11	0,389	0,26	M_D ⇒ St_Ü ⇒ T_Ü ⇒ T_D ⇒ T_Ü ⇒ T_D	Su_K ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ St_Ü ⇒ Su_K ⇒ M_D ⇒ M_H

Fortsetzung auf der nächsten Seite

Tabelle B.3 – Fortsetzung

#	Länge V. 1	Länge V. 2	Rub-Funk-M.	Komb-M.	Vektor 1	Vektor 2
97	10	6	0,397	0	G_Ü ⇒ G_D ⇒ G_D ⇒ M_Ü ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H ⇒ M_D ⇒ M_H	B_Ü ⇒ Su_Ü ⇒ A_D ⇒ A_D ⇒ A_D ⇒ A_D
98	3	6	0,688	0	F_D ⇒ D_Ü ⇒ B_Ü	T_D ⇒ M_Ü ⇒ St_Ü ⇒ Su_Ü ⇒ A_Ü ⇒ G_Ü
99	4	17	0,657	0,381	M_D ⇒ G_D ⇒ B_D ⇒ B_D	Su_K ⇒ A_D ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_H ⇒ M_Ü ⇒ M_D ⇒ M_D ⇒ M_D ⇒ M_D ⇒ St_Ü ⇒ Su_Ü ⇒ M_D ⇒ M_D ⇒ M_H
100	5	9	0,35	0	T_Ü ⇒ T_D ⇒ T_D ⇒ T_Ü ⇒ M_Ü	M_D ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_B ⇒ M_D ⇒ M_B ⇒ M_D ⇒ St_Ü

Dauer der Aufgaben

Tabelle B.4: Bearbeitungsdauer der Aufgaben nach Testerinnen und Testern in Minuten
(Ausreißer kursiv markiert)

Test	Aufg. 1	Aufg. 2	Aufg. 3	Aufg. 4	Aufg. 5	Aufg. 6	Aufg. 7	Gesamt
1	03:20	02:25	00:40	06:50	01:13	00:37	<i>03:59</i>	19:04
2	00:50	01:45	01:39	03:18	01:14	00:49	00:54	10:29
3	02:24	02:43	02:35	03:28	00:27	00:30	01:17	13:24
4	01:12	03:30	00:25	05:00	01:15	00:27	01:35	13:24
5	02:38	01:32	00:33	02:37	00:24	00:27	00:20	08:31
6	03:48	01:48	01:13	03:39	00:57	00:42	00:22	12:29
7	01:17	01:17	00:42	03:18	00:46	01:40	02:23	11:23
8	04:34	02:54	<i>03:07</i>	07:58	<i>02:13</i>	<i>02:22</i>	01:42	24:50
9	01:58	02:00	00:27	02:13	00:37	00:48	01:54	09:57
10	00:53	01:14	01:22	01:53	00:35	00:17	00:26	06:40
11	01:39	01:44	00:25	05:11	01:10	00:42	01:21	12:12
12	03:55	03:52	01:28	06:37	01:29	<i>02:02</i>	02:47	22:10

C Usability

C.1 Flyer für Lehrkräfte



Lehrkräfte für Studie im Rahmen einer Dissertation gesucht

Ich möchte im Rahmen meiner Dissertation die Usability („Gebrauchstauglichkeit“) eines Schulinformations-Systems untersuchen und suche dafür interessierte Lehrkräfte von weiterführenden Schulen, die an einer Studie teilnehmen möchten. Die unterrichteten Fächer, das Alter und das Geschlecht spielen keine Rolle. Einzige Voraussetzung ist, dass Sie Computerplattformen (Netzwerklaufwerke, Dropbox, Moodle o.ä., Facebook etc.) zur Vorbereitung des Unterrichts und zur Kommunikation mit Kolleginnen und Kollegen nutzen. Wie häufig Sie die entsprechenden Plattformen einsetzen, ist dagegen nicht relevant.

Die Tests sollen zwischen Februar und Mitte März 2015 stattfinden und werden jeweils etwa 30 bis 45 Minuten in Anspruch nehmen. Die genaue Terminierung kann flexibel und individuell abgestimmt werden. Alle Teilnehmerinnen und Teilnehmer erhalten für Ihre Mühen ein kleines Dankeschön.

Bei Interesse oder weiteren Fragen stehe ich Ihnen gerne zur Verfügung.

Arne Hendrik Schulz

AG Informationsmanagement
Fachbereich 3
Universität Bremen

Am Fallturm 1 (Eingang F)
Raum: 3.79
28359 Bremen

Tel.: 0421 218 56585
E-Mail: ahschulz@uni-bremen.de
Web: www.informatik.uni-bremen.de/agim

***EXZELLENT.**

C.2 Handout an die Testerinnen und Tester

Testablauf

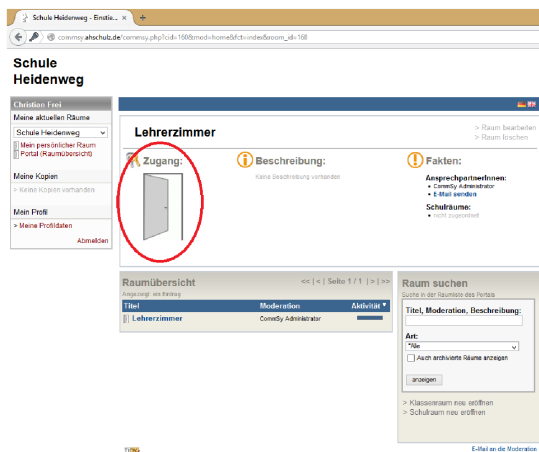
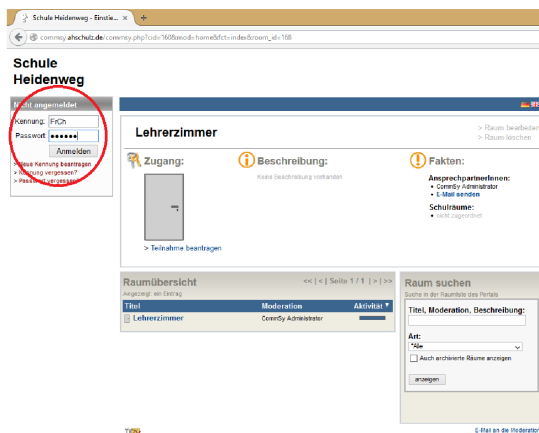
Zugang zum System

Für die Zeit des Usability Tests nehmen Sie die Rolle von Lea Kaufmann, einer Lehrkraft in der Schule Heidenweg. Vor Ihnen im Browser sollte bereits die Startseite des Schulinformationssystems geöffnet sein. Loggen Sie sich bitte zunächst mit den folgenden Zugangsdaten ein:

Kennung: KaLe

Passwort: annLea

Danach betreten Sie wie hier geschildert den Raum und fahren mit den Aufgaben fort.



Aufgaben

Bitte führen Sie die folgenden Aufgaben möglichst selbstständig durch. Sollten Sie größere Schwierigkeiten bekommen oder nicht mehr weiter wissen, hilft Ihnen der Testleiter gerne weiter. Bitte denken Sie auch daran, ihre Gedankengänge verbal zu artikulieren.

- Lisa Lehrer hat im März einen Termin für den Projekttag des Jg. 6 erstellt. Finden Sie diesen und sagen sie zu.
- Erstellen Sie einen eigenen Termin mit den folgenden Eigenschaften:
 - Titel: Klausur Englisch Jg. 12
 - Datum: 17.03.2015, 10:00-12:00 Uhr
 - Ort: Raum C2.15
- Sehen Sie sich die Beispielklausuren von Frau Kuhn in Deutsch für den Jahrgang 11 an.
- Legen Sie ein neues Material mit der Beispielklausur im Fach Mathematik für den 11. Jahrgang vom Desktop (Beispielklausur_Mathematik_Jg11.pdf) an. Das Material soll die Schlagwörter „Jahrgang 11“, „Mathematik“, „Klausuren“ und der Kategorie „Unterricht“ und der Unterkategorie „Arbeitsbögen“ zugeordnet sein.
- Kommentieren Sie eine Ankündigung Ihrer Wahl.
- Finden Sie das Datum des Kinobesuchs des Films „Stauffenberg“ heraus.
- Lassen Sie sich alle Einträge mit dem Schlagwort „Jahrgang 12“ anzeigen.

C.3 Fragebogen für die Testerinnen und Tester

1 Beurteilung der Usability

1.1 Aussagen über das System

Nachfolgend finden Sie einige Aussagen über das System. Bitte geben Sie an, inwiefern Sie diesen Aussagen zustimmen. 1 bedeutet hierbei, dass Sie dieser Aussage „gar nicht zustimmen“. 5, dass Sie „voll zustimmen“.

	Stimme gar nicht zu 1	2	3	4	Stimme voll zu 5
Ich denke, dass ich das System gerne häufig benutzen würde.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fand das System unnötig komplex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fand das System einfach zu benutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich glaube, ich würde die Hilfe einer (medien)technisch versierten Person benötigen, um das System benutzen zu können.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, das System enthielt zu viele Inkonsistenzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fand das System sehr umständlich zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühlte mich bei der Benutzung des Systems sehr sicher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1.2 Beurteilung des Systems

Nachfolgend finden Sie Wortpaare, mit deren Hilfe Sie die Beurteilung vornehmen können. Sie stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

Ein Beispiel:

sympathisch ○ ○ ○ ○ ⊙ ○ ○ unsympathisch

Diese Bewertung bedeutet, dass das Produkt eher sympathisch, aber noch verbesserungsbedürftig ist.

Denken Sie nicht lange über die Wortpaare nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt. Vielleicht passen einige Wortpaare nicht so gut auf das Produkt, kreuzen Sie aber trotzdem bitte immer eine Antwort an. Bei der Beantwortung geht es ausschließlich um Ihre persönlichen Ansichten, es gibt also keine richtigen oder falschen Antworten.

menschlich	○	○	○	○	○	○	○	technisch
isolierend	○	○	○	○	○	○	○	verbindend
angenehm	○	○	○	○	○	○	○	unangenehm
originell	○	○	○	○	○	○	○	konventionell
einfach	○	○	○	○	○	○	○	kompliziert
fachmännisch	○	○	○	○	○	○	○	laienhaft
hässlich	○	○	○	○	○	○	○	schön
praktisch	○	○	○	○	○	○	○	unpraktisch
sympathisch	○	○	○	○	○	○	○	unsympathisch
umständlich	○	○	○	○	○	○	○	direkt
stilvoll	○	○	○	○	○	○	○	stillos
voraussagbar	○	○	○	○	○	○	○	unberechenbar
minderwertig	○	○	○	○	○	○	○	wertvoll
ausgrenzend	○	○	○	○	○	○	○	einbeziehend
bringt mich den Leuten näher	○	○	○	○	○	○	○	trennt mich von Leuten
nicht vorzeigbar	○	○	○	○	○	○	○	vorzeigbar
zurückweisend	○	○	○	○	○	○	○	einladend
phantasielos	○	○	○	○	○	○	○	kreativ
gut	○	○	○	○	○	○	○	schlecht
verwirrend	○	○	○	○	○	○	○	übersichtlich
abstoßend	○	○	○	○	○	○	○	anziehend
mutig	○	○	○	○	○	○	○	vorsichtig
innovativ	○	○	○	○	○	○	○	konservativ
lahm	○	○	○	○	○	○	○	fesselnd
harmlos	○	○	○	○	○	○	○	herausfordernd
motivierend	○	○	○	○	○	○	○	entmutigend
neuartig	○	○	○	○	○	○	○	herkömmlich
widerspenstig	○	○	○	○	○	○	○	handhabbar

2 Einige Daten zu Ihrer Person

2.1 Bitte geben Sie Ihr Geschlecht an:

- Weiblich Männlich

2.2 Wie alt sind Sie?

- bis 29 Jahre 45 bis 49 Jahre
 30 bis 34 Jahre 50 bis 54 Jahre
 35 bis 39 Jahre 55 Jahre und älter
 40 bis 44 Jahre

2.3 Bitte geben Sie Ihr Dienstalter an:

- im Referendariat 15 bis 24 Dienstjahre
 1 bis 4 Dienstjahre 25 Dienstjahre und mehr
 5 bis 14 Dienstjahre

2.4 Welche Rolle nehmen Sie in Ihrer Schule ein?

- Schulleitung Erzieherin oder Erzieher
 Lehrerin oder Lehrer Sozialpädagogin oder Sozialpädagoge
 Lehrerin oder Lehrer für sonderpädagogische Förderung Sonstiges: _____

2.5 An was für einer Schule unterrichten Sie?

- Hauptschule Realschule
 Gymnasium Gesamtschule
 Anderes: _____

2.6 Welche Fächer unterrichten Sie (hauptsächlich)?

C.4 Zusätzliche Abbildungen

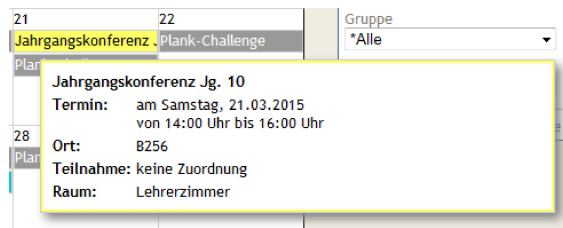


Abbildung C.1: Hover-Effekt bei Terminen

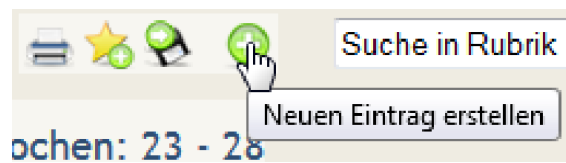


Abbildung C.2: Hover-Effekt bei Symbolen

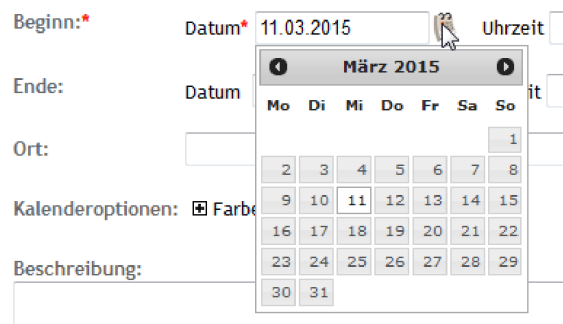


Abbildung C.3: Eingblendeter Kalender



Abbildung C.4: Übersichtsseite der Gruppen

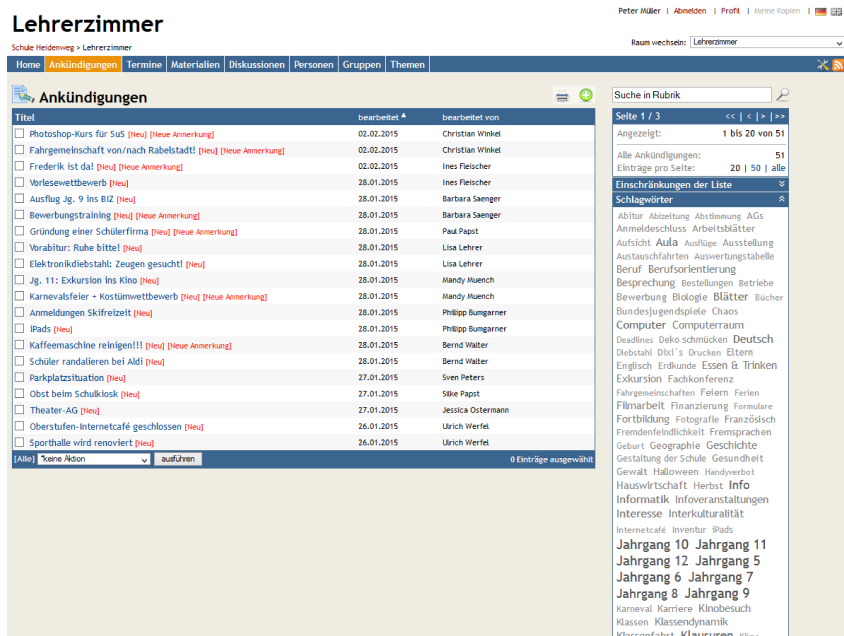


Abbildung C.5: Übersichtsseite der Ankündigungen

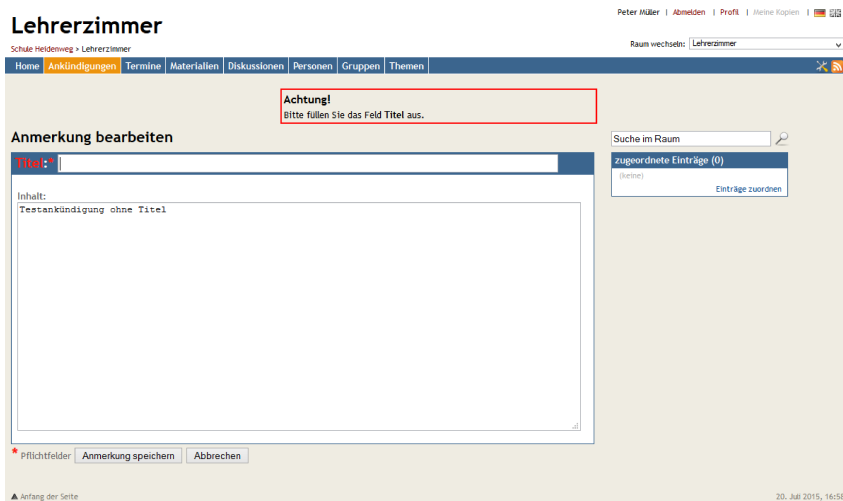


Abbildung C.6: Fehlermeldung Ankündigungen bei fehlendem Titel



Abbildung C.7: Suche Stauffenberg über Termine

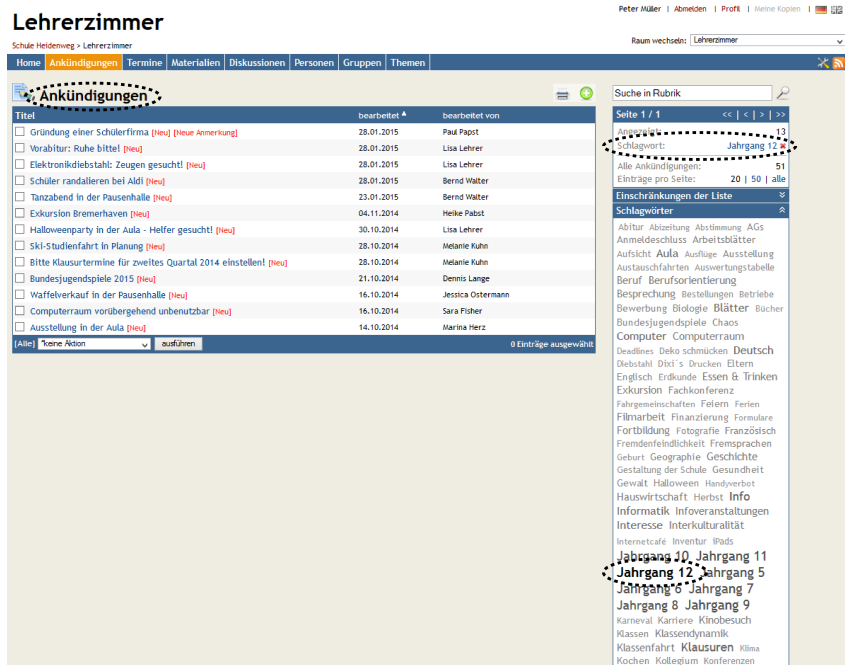


Abbildung C.8: Auflistung der Einträge mit Schlagwort in Rubrik

D Kombination

Tabelle D.1: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,6

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	128	72	0	72	0%	100%
Aufgabe 2	Problem 1	63	100	60	40	60%	40%
Aufgabe 2	Problem 3	60	98	60	38	61%	39%
Aufgabe 3	Problem 1	177	164	0	164	0%	100%
Aufgabe 4	Problem 1	106	323	317	6	98%	2%
Aufgabe 4	Problem 2	118	346	340	6	98%	2%
Aufgabe 4	Problem 3	116	343	337	6	98%	2%
Aufgabe 4	Problem 6	155	348	340	8	98%	2%
Aufgabe 6	Problem 1	303	340	0	340	0%	100%
Aufgabe 7	Problem 1	211	418	205	213	49%	51%
Aufgabe 7	Problem 3	205	412	205	207	50%	50%
		1642	2964	1864	1100	63%	37%

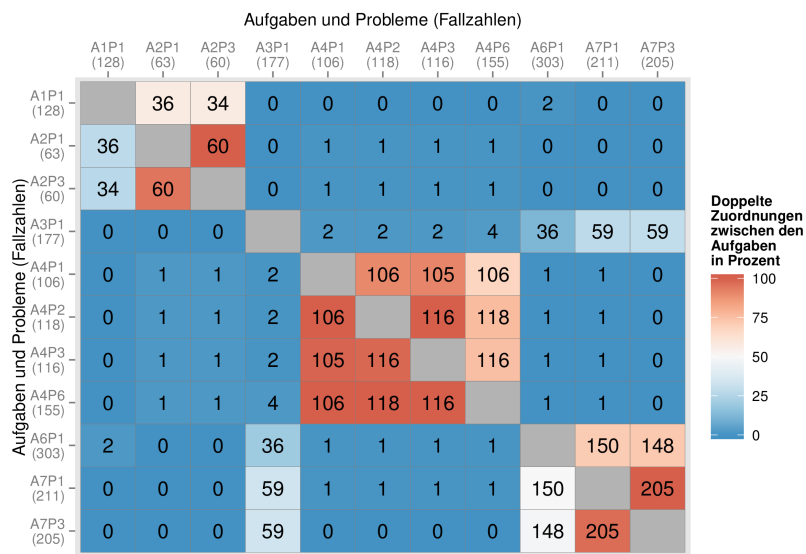


Abbildung D.1: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,6

Tabelle D.2: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,65

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	110	50	0	50	0%	100%
Aufgabe 2	Problem 1	54	81	51	30	63%	37%
Aufgabe 2	Problem 3	52	79	51	28	65%	35%
Aufgabe 3	Problem 1	88	33	0	33	0%	100%
Aufgabe 4	Problem 1	86	259	257	2	99%	1%
Aufgabe 4	Problem 2	101	289	287	2	99%	1%
Aufgabe 4	Problem 3	102	290	288	2	99%	1%
Aufgabe 4	Problem 6	131	288	286	2	99%	1%
Aufgabe 6	Problem 1	243	245	0	245	0%	100%
Aufgabe 7	Problem 1	172	298	161	137	54%	46%
Aufgabe 7	Problem 3	164	294	161	133	55%	45%
		1303	2206	1542	664	70%	30%

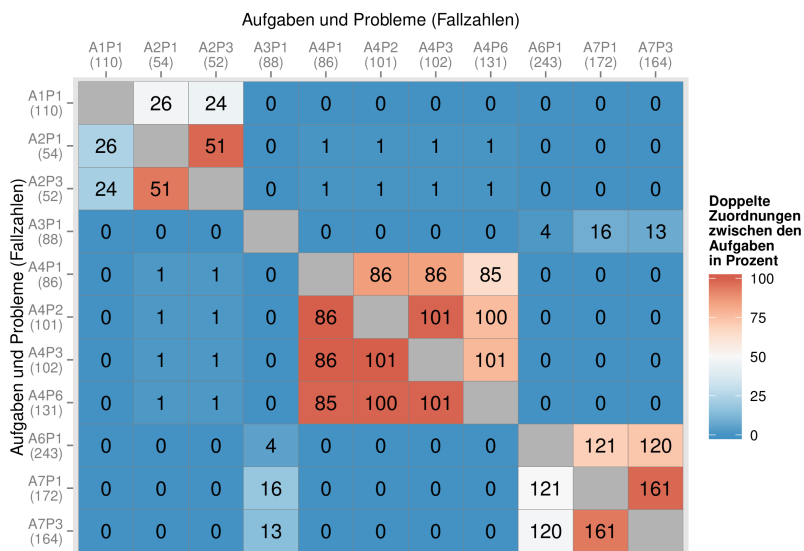


Abbildung D.2: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,65

Tabelle D.3: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,9

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	30	0	0	0	0%	0%
Aufgabe 2	Problem 1	25	25	25	0	100%	0%
Aufgabe 2	Problem 3	25	25	25	0	100%	0%
Aufgabe 3	Problem 1	0	0	0	0	0%	0%
Aufgabe 4	Problem 1	18	45	45	0	100%	0%
Aufgabe 4	Problem 2	25	61	61	0	100%	0%
Aufgabe 4	Problem 3	24	60	60	0	100%	0%
Aufgabe 4	Problem 6	36	56	56	0	100%	0%
Aufgabe 6	Problem 1	14	0	0	0	0%	0%
Aufgabe 7	Problem 1	19	16	16	0	100%	0%
Aufgabe 7	Problem 3	17	16	16	0	100%	0%
		233	304	304	0	100%	0%

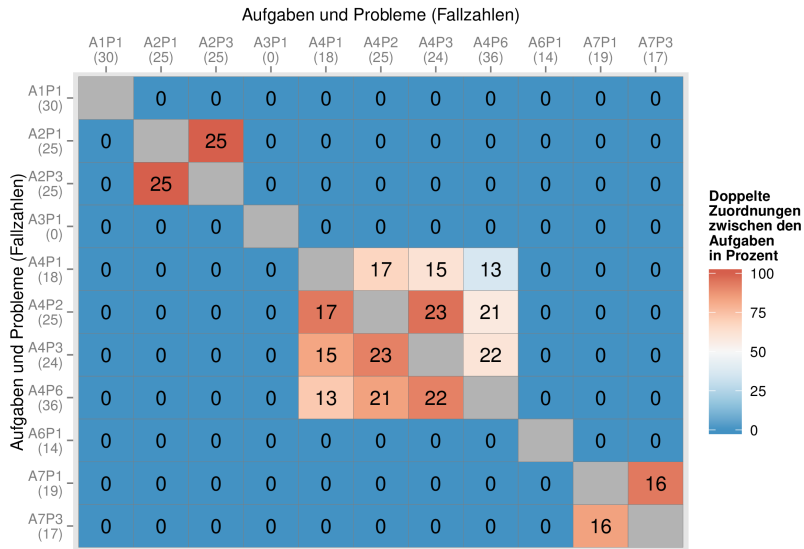


Abbildung D.3: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,9

Tabelle D.4: Mehrfach zugeordnete Pfade bei Ähnlichkeit von mindestens 0,95

Aufgabe	Problem	gefundene Pfadmerkmale	mehrfache Pfadmerkmale	davon in der gleichen Aufgabe	davon in anderen Aufgaben	Prozent in gleichen Aufgaben	Prozent in unterschiedlichen Aufgaben
Aufgabe 1	Problem 1	18	0	0	0	0%	0%
Aufgabe 2	Problem 1	14	13	13	0	100%	0%
Aufgabe 2	Problem 3	13	13	13	0	100%	0%
Aufgabe 3	Problem 1	0	0	0	0	0%	0%
Aufgabe 4	Problem 1	12	24	24	0	100%	0%
Aufgabe 4	Problem 2	11	24	24	0	100%	0%
Aufgabe 4	Problem 3	8	21	21	0	100%	0%
Aufgabe 4	Problem 6	15	15	15	0	100%	0%
Aufgabe 6	Problem 1	6	0	0	0	0%	0%
Aufgabe 7	Problem 1	2	2	2	0	100%	0%
Aufgabe 7	Problem 3	5	2	2	0	100%	0%
		104	114	114	0	100%	0%

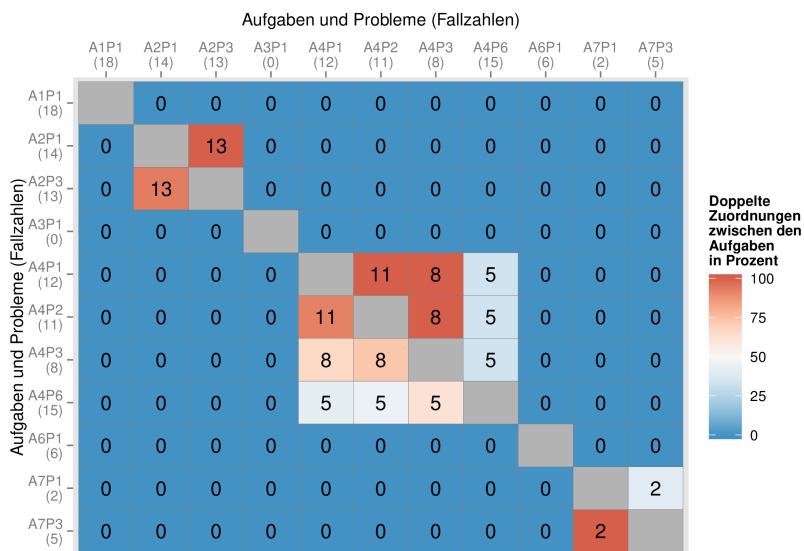


Abbildung D.4: Mehrfach zugeordnete Pfade bei einer Ähnlichkeit von mindestens 0,95

Literatur

- Adomavicius, Gediminas; Tuzhilin, Tuzhilin (2001): *Extending recommender systems: A multidimensional approach*. In: Proceedings of the International Joint Conference on Artificial Intelligence, Workshop on Intelligent Techniques for Web Personalization (ITWP2001). IJCAI '01. Seattle, WA: AAAI.
- Agarwal, Ritu; Prasad, Jayesh (1999): *Are Individual Differences Germane to the Acceptance of New Information Technologies?* Decision Sciences 30 (2), S. 361–391.
- Agrawal, Rakesh; Imieliski, Tomasz; Swami, Arun (1993): *Mining association rules between sets of items in large databases*. ACM SIGMOD Record 22 (2), S. 207–216.
- Agrawal, Rakesh; Srikant, Ramakrishnan (1994): *Fast Algorithms for Mining Association Rules in Large Databases*. In: Proceedings of the 20th International Conference on Very Large Data Bases. VLDB '94. San Francisco, CA: Morgan Kaufmann, S. 487–499.
- Agrawal, Rakesh; Srikant, Ramakrishnan (1995): *Mining sequential patterns*. In: Proceedings of the Eleventh International Conference on Data Engineering, 1995. Proceedings of the Eleventh International Conference on Data Engineering, 1995. Taipei, Taiwan: IEEE, S. 3–14.
- Alavi, Maryam; Leidner, Dorothy E. (2001): *Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues*. MIS Quarterly 25 (1), S. 107–136.
- Alpaydin, Ethem (2014): *Introduction to Machine Learning*. Cambridge, MA: MIT Press.
- Altman, Naomi S. (1992): *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*. The American Statistician 46 (3), S. 175–185.
- Altrichter, Herbert (2006): *Schulentwicklung. Widersprüche unter neuen Bedingungen*. Pädagogik 58 (3), S. 6–10.
- Anderson, Corin R. (2002): „A Machine Learning Approach to Web Personalization“. Diss. Seattle, WA: University of Washington.
- Ankerst, Michael; Breunig, Markus M.; Kriegel, Hans-Peter; Sander, Jörg (1999): *OPTICS: ordering points to identify the clustering structure*. In: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data. SIGMOD '99. New York, NY: ACM, S. 49–60.
- Ansari, Suhail; Kohavi, Ron; Mason, Llew; Zheng, Zijian (2001): *Integrating E-Commerce and Data Mining: Architecture and Challenges*. In: Proceedings of the 2001 IEEE International Conference on Data Mining. ICDM '01. Washington, DC: IEEE, S. 27–34.
- Ardito, Carmelo; Costabile, Maria F.; De Marsico, Marilena; Lanzilotti, Rosa; Levialdi, Stefano; Plantamura, Paola; Roselli, Teresa; Rossano, Veronica; Tersigni, Manuela (2004): *Towards Guidelines for Usability of e-Learning Applications*. In: Stary, Christian; Stephanidis, Constantine (Hrsg.): *User-Centered Interaction Paradigms for Universal Access in the Information Society*. Berlin: Springer, S. 185–202.
- Ardito, Carmelo; Costabile, Maria F.; De Marsico, Marilena; Lanzilotti, Rosa; Levialdi, Stefano; Roselli, Teresa; Rossano, Veronica (2006): *An approach to usability evaluation of e-learning applications*. Universal Access in the Information Society 4 (3), S. 270–283.
- Arroyo, Ernesto; Selker, Ted; Wei, Willy (2006): *Usability tool for analysis of web designs using mouse tracks*. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems. CHI EA '06. New York, NY: ACM, S. 484–489.
- Atterer, Richard; Schmidt, Albrecht (2007): *Tracking the interaction of users with AJAX applications for usability testing*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '07. New York, NY: ACM, S. 1347–1350.
- Atterer, Richard; Wnuk, Monika; Schmidt, Albrecht (2006): *Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction*. In: Proceedings of the 15th International Conference on World Wide Web. WWW '06. New York, NY: ACM, S. 203–212.

- Au, Fiora T. W.; Baker, Simon; Warren, Ian; Dobbie, Gillian (2008): *Automated usability testing framework*. In: Proceedings of the 9th Conference on Australasian User Interface. AUIC '08. Darlinghurst, Australia: Australian Computer Society, S. 55–64.
- Azevedo, Ana I. R. L.; Santos, Manuel Filipe (2008): *KDD, SEMMA and CRISP-DM: a parallel overview*. In: Abraham, Ajith A.; Weghorn, Hans (Hrsg.): Proceedings of the IADIS European Conference on Data Mining 2008. ECDM '08. Amsterdam, Netherlands: IADIS Press, S. 182–185.
- Backhaus, Klaus; Erichson, Bernd; Plinke, Wulff; Weiber, Rolf (2011): *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Baecker, Ronald M.; Grudin, Jonathan; Buxton, William A. S.; Greenberg, Saul (1995): *A historical and intellectual perspective*. In: Baecker, Ronald M.; Grudin, Jonathan; Buxton, William A. S.; Greenberg, Saul (Hrsg.): Human-computer Interaction. San Francisco, CA: Morgan Kaufmann, S. 35–47.
- Baglioni, M.; Ferrara, U.; Romei, A.; Ruggieri, S.; Turini, F. (2003): *Preprocessing and Mining Web Log Data for Web Personalization*. In: Cappelli, Amedeo; Turini, Franco (Hrsg.): AI*IA 2003: Advances in Artificial Intelligence. Berlin/Heidelberg: Springer, S. 237–249.
- Bailey, Robert W.; Allan, Robert W.; Raiello, P. (1992): *Usability testing vs. heuristic evaluation: A head-to-head comparison*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 36 (4), S. 409–413.
- Baker, Ryan S. J. D.; Yacef, Kalina (2009): *The state of educational data mining in 2009: A review and future visions*. Journal of Educational Data Mining 1 (1), S. 3–17.
- Balabanovic, Marko; Shoham, Yoav (1995): *Learning Information Retrieval Agents: Experiments with Automated Web Browsing*. In: Knoblock, Craig; Levy, Alon (Hrsg.): Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogenous, Distributed Resources. Menlo Park, CA: AAAI Press, S. 13–18.
- Balzert, Helmut; Liggesmeyer, Peter (2011): *Lehrbuch der Softwaretechnik. Entwurf, Implementierung, Installation und Betrieb*. Heidelberg: Spektrum.
- Banerjee, Arindam; Ghosh, Joydeep (2000): *Concept-based Clustering of Clickstream Data*. In: Proceedings of the 3rd International Conference on Information Technology. ITNG '00, S. 145–160.
- Bangor, Aaron; Kortum, Philip; Miller, James (2009): *Determining what individual SUS scores mean: Adding an adjective rating scale*. Journal of usability studies 4 (3), S. 114–123.
- Barnum, Carol (2011): *Usability Testing Essentials. Ready, Set...Test*. Burlington, MA: Morgan Kaufmann.
- Barnum, Carol; Bevan, Nigel; Cockton, Gilbert; Nielsen, Jakob; Spool, Jared; Wixon, Dennis (2003): *The "magic number 5": is it enough for web testing?* In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. CHI '03. New York, NY: ACM, S. 698–699.
- Bastien, Christian J. M.; Scapin, Dominique L. (1995): *Evaluating a user interface with ergonomic criteria*. International Journal of Human-Computer Interaction 7 (2), S. 105–121.
- Baumgartner, Peter; Häfele, Hartmut; Maier-Häfele, Kornelia (2004): *Lernplattformen im Feldtest*. In: Meister, Dorothee M.; Tergan, Sigmar-Olaf; Zentel, Peter (Hrsg.): Evaluation von E-Learning. Münster: Waxmann, S. 108–122.
- Bayir, Murat A.; Toroslu, Ismail H.; Cosar, Ahmet; Fidan, Guven (2008): *Discovering More Accurate Frequent Web Usage Patterns*. arXiv:0804.1409.
- Beck, Klaus (2015): *Kommunikationswissenschaft*. Konstanz: UVK Verlagsgesellschaft.
- Behrmann, Marlene; Zemel, Richard S.; Mozer, Michael C. (1998): *Object-based attention and occlusion: Evidence from normal participants and a computational model*. Journal of Experimental Psychology: Human Perception and Performance 24 (4), S. 1011–1036.
- Benoît, Gerald (2002): *Data Mining*. Annual Review of Information Science and Technology 36 (1), S. 265–310.
- Berendt, Bettina; Mobasher, Bamshad; Nakagawa, Miki; Spiliopoulou, Myra (2002): *The Impact of Site Structure and User Environment on Session Reconstruction in Web Usage Analysis*. In: Masand, Brij; Spiliopoulou, Myra; Srivastava, Jaideep; Zaiane, Osmar R. (Hrsg.): WEBKDD 2002 - Mining Web Data for Discovering Usage Patterns and Profiles. Berlin/Heidelberg: Springer, S. 159–179.

- Bergmann, Rainer; Garrecht, Martin (2008): *Organisation und Projektmanagement*. Heidelberg: Physica Verlag.
- Berkhin, Pavel (2002): *A Survey of Clustering Data Mining Techniques*. Technical Report. San Jose, CA: Accrue Software.
- Berland, Matthew; Baker, Ryan S.; Blikstein, Paulo (2014): *Educational Data Mining and Learning Analytics: Applications to Constructionist Research*. *Technology, Knowledge and Learning* 19 (1), S. 205–220.
- Bevan, Nigel; Macleod, Miles (1994): *Usability measurement in context*. *Behaviour & Information Technology* 13 (1), S. 132–145.
- Bezdek, James C.; Ehrlich, Robert; Full, William (1984): *FCM: The fuzzy c-means clustering algorithm*. *Computers & Geosciences* 10 (2), S. 191–203.
- Bias, Randolph G. (1994): *The Pluralistic Usability Walkthrough: Coordinated Empathies*. In: Nielsen, Jakob; Mack, Robert L. (Hrsg.): *Usability Inspection Methods*. New York, NY: John Wiley & Sons, S. 63–76.
- Bienkowski, Marie; Feng, Mingyu; Means, Barbara (2012): *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. Technical Report ED-04-CO-0040. Washington, DC: U.S. Department of Education, Office of Educational Technology. URL: <https://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf> (aufgerufen am 25.03.2014).
- Blackmon, Marilyn H.; Kitajima, Muneo; Polson, Peter G. (2005): *Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '05. New York, NY: ACM, S. 31–40.
- Blackmon, Marilyn H.; Polson, Peter G.; Kitajima, Muneo; Lewis, Clayton (2002): *Cognitive walkthrough for the web*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '02. Minneapolis, MN: ACM, S. 463–470.
- Blecken, Alexander; Bruggemann, Daniel; Marx, Wiebke (2010): *Usability Evaluation of a Learning Management System*. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences. HICSS '10. Washington, DC: IEEE Computer Society, S. 1–9.
- Bohnsack, Ralf (2003): *Rekonstruktive Sozialforschung: Einführung in qualitative Methoden*. Opladen: Leske + Budrich.
- Bojko, Agnieszka (2005): *Eye Tracking in User Experience Testing: How to Make the Most of It*. In: Proceedings of the 14th Annual Conference of the Usability Professionals Association. Montreal, Kanada, S. 1–9.
- Bojko, Agnieszka; Stephenson, Abigail (2005): *Supplementing conventional usability measures with eye movement data in evaluating visual search performance*. In: Proceedings of the 11th International Conference on Human-Computer Interaction. Las Vegas, NV: Lawrence Erlbaum Associates, S. 117–121.
- Bonchi, F.; Giannotti, F.; Gozzi, C.; Manco, G.; Nanni, M.; Pedreschi, D.; Renso, C.; Ruggieri, S. (2001): *Web log data warehousing and mining for intelligent web caching*. *Data & Knowledge Engineering* 39 (2), S. 165–189.
- Boren, Ted; Ramey, Judith (2000): *Thinking aloud: reconciling theory and practice*. *IEEE Transactions on Professional Communication* 43 (3), S. 261–278.
- Borges, José; Levene, Mark (2000): *Data Mining of User Navigation Patterns*. In: Masand, Brij; Spiliopoulou, Myra (Hrsg.): *Web Usage Analysis and User Profiling*. Berlin/Heidelberg: Springer, S. 92–112.
- Bortz, Jürgen; Döring, Nicola (2006): *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bounsaythip, Catherine; Rinta-Runsala, Esa (2001): *Overview of Data Mining for Customer Behavior Modeling*. Technical Report TTE1-2001-18. Espoo, Finnland: VTT Information Technology. URL: <http://www.vtt.fi/inf/julkaisut/muut/2001/customerprofiling.pdf> (aufgerufen am 26.07.2012).

- Bowers, Victoria A.; Snyder, Harry L. (1990): *Concurrent versus Retrospective Verbal Protocol for Comparing Window Usability*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 34 (17), S. 1270–1274.
- Brachman, Ronald J.; Anand, Tej (1996): *Advances in knowledge discovery and data mining*. In: Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy (Hrsg.): *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI, S. 37–57.
- Breiter, Andreas (2002): *Wissensmanagementsysteme in Schulen - wie bringe ich Ordnung ins Chaos?* medienpaedagogik online 2 (2).
- Breiter, Andreas; Averbeck, Ines; Welling, Stefan; Schulz, Arne Hendrik (2015): *Der Medienpass als Instrument zur Förderung von Medienkompetenz. Entwicklungsstand in Grundschulen in Nordrhein-Westfalen im Vergleich*. Düsseldorf: LfM.
- Breiter, Andreas; Lange, Angelina; Stauke, Emese, Hrsg. (2008): *School Information System and Data-based Decision-making/ Schulinformationssysteme und datengestützte Entscheidungsprozesse*. Frankfurt a. M.: Peter Lang.
- Brinck, Tom; Hofer, Erik (2002): *Automatically evaluating the usability of web sites*. In: CHI '02 Extended Abstracts on Human Factors in Computing Systems. CHI '02. New York, NY: ACM, S. 906–907.
- Brooke, John (1996): *SUS - A quick and dirty usability scale*. In: McClelland, Ian L.; Jordan, Patrick; Jordan, Jordan W. (Hrsg.): *Usability Evaluation in Industry*. London, UK: Taylor & Francis.
- Brooke, John (2013): *SUS: a retrospective*. Journal of Usability Studies 8 (2), S. 29–40.
- Brüsemeister, Thomas (2008): *Qualitative Forschung: ein Überblick*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Büchner, Alex G.; Mulvenna, Maurice D. (1998): *Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining*. ACM SIGMOD Record 27 (4), S. 54–61.
- Burger, Karyn; Jackson, Ellen (1997): *Usability Evaluation Techniques for Large-Scale Web Sites*. In: Howard, Steve; Hammond, Judy; Lindgaard, Gitte (Hrsg.): *Human-Computer Interaction INTERACT '97*. Dordrecht: Springer, S. 571–572.
- Burkart, Roland (2002): *Kommunikationswissenschaft: Grundlagen und Problemfelder. Umriss einer interdisziplinären Sozialwissenschaft*. Wien: Böhlau.
- Caliński, T.; Harabasz, J (1974): *A dendrite method for cluster analysis*. Communications in Statistics 3 (1), S. 1–27.
- Card, Stuart K.; Moran, Thomas P.; Newell, Allen (1983): *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.
- Carroll, John M. (2000): *Five reasons for scenario-based design*. Interacting with Computers 13 (1), S. 43–60.
- Carroll, John M. (2003): *Making Use: Scenario-Based Design of Human-Computer Interactions*. Cambridge, MA: MIT Press.
- Carstensen, Peter H.; Schmidt, Kjeld (1999): *Computer Supported Cooperative Work: New Challenges to Systems Design*. In: Itoh, Kenji (Hrsg.): *Handbook of Human Factors*. Tokio: Asakura Publishing, S. 619–636.
- Catledge, Lara D.; Pitkow, James E. (1995): *Characterizing browsing strategies in the World-Wide web*. Computer Networks and ISDN Systems 27 (6), S. 1065–1073.
- Chakrabarti, Soumen (2000): *Data mining for hypertext: a tutorial survey*. SIGKDD Explorations Newsletter 1 (2), S. 1–11.
- Chakrabarti, Soumen (2002): *Mining the Web: Discovering Knowledge from Hypertext Data*. San Francisco, CA: Morgan Kaufmann.
- Chakrabarti, Soumen; Dom, Byron E.; Kumar, S. Ravi; Raghavan, Prabhakar; Rajagopalan, Sridhar; Tomkins, Andrew; Gibson, David; Kleinberg, Jon (1999): *Mining the Web's link structure*. Computer 32 (8), S. 60–67.
- Chakrabarti, Soumen; Dom, Byron; Indyk, Piotr (1998): *Enhanced hypertext categorization using hyperlinks*. ACM SIGMOD Record 27 (2), S. 307–318.
- Chang, Cheng-Yue; Chen, Ming-Syan (2002): *A new cache replacement algorithm for the integration of web caching and prefetching*. In: Proceedings of the 11th International Conference on Information and Knowledge Management. CIKM '02. New York, NY: ACM, S. 632–634.

- Chang, Elizabeth; Dillon, Tharam S. (1997): *Automated Usability Testing*. In: Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction. INTERACT '97. London, UK: Chapman & Hall, S. 77–84.
- Charrad, Malika; Ghazzali, Nadia; Boiteau, Véronique; Niknafs, Azam (2014): *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set*. Journal of Statistical Software 61 (6), S. 1–36.
- Chen, Mao; LaPaugh, Andrea S.; Singh, Jaswinder P. (2002): *Predicting category accesses for a user in a structured information space*. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '02. New York, NY: ACM, S. 65–72.
- Chen, Yan; Huang, Lixian; Li, Lulu; Luo, Qi; Wang, Ying; Xu, Jing (2007): *The Experimental Approaches of Assessing the Consistency of User Interface*. In: Jacko, Julie (Hrsg.): Human-Computer Interaction. Interaction Design and Usability. Berlin/Heidelberg: Springer, S. 420–427.
- Chi, Ed H.; Rosien, Adam; Supattanasiri, Gesara; Williams, Amanda; Royer, Christiaan; Chow, Celia; Robles, Erica; Dalal, Brinda; Chen, Julie; Cousins, Steve (2003): *The bloodhound project: automating discovery of web usability issues using the InfoScout simulator*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '03. New York, NY: ACM, S. 505–512.
- Chin, John P.; Diehl, Virginia A.; Norman, Kent L. (1988): *Development of an instrument measuring user satisfaction of the human-computer interface*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '88. New York, NY: ACM, S. 213–218.
- Chitraa, V.; Thanamani, Antony S. (2011): *A Novel Technique for Sessions Identification in Web Usage Mining Preprocessing*. International Journal of Computer Applications 34 (9), S. 24–28.
- Chung, Chih-Hung; Pasquini, Laura A.; Allen, Jeff M.; Koh, Chang E. (2012): *A Review of Web-based Learning Management System Design on Higher Education*. In: E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. E-Learn '12. Chesapeake, VA: AACE, S. 1428–1436.
- Cichosz, Pawel (2015): *Data mining algorithms: explained using R*. Chichester, West Sussex, U.K.: John Wiley & Sons.
- Cios, Krzysztof J.; Pedrycz, Witold; Swiniarski, Roman W.; Kurgan, Lukasz A. (2007): *Data mining. A knowledge discovery approach*. New York, NY: Springer.
- Clauss, Günther; Finze, Falk R.; Partzsch, Lothar (2004): *Statistik Grundlagen: Für Soziologen, Pädagogen, Psychologen und Mediziner*. Frankfurt a. M.: Harri Deutsch.
- Cockton, Gilbert; Woolrych, Alan; Lavery, Darryn (2008): *Inspection-Based Evaluations*. In: Sears, Andrew; Jacko, Julie A. (Hrsg.): The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications. New York, NY: Lawrence Erlbaum, S. 1171–1190.
- Cohen, William; McCallum, Andrew; Quass, Dullan (2000): *Learning to understand the Web*. IEEE Data Engineering Bulletin 23 (3), S. 17–24.
- Constantine, Larry L.; Lockwood, Lucy A. D. (1999): *Software for Use: A Practical Guide to the Methods of Usage-Centered Design: A Practical Guide to the Models and Methods of Usage-centred Design*. Boston, MA [u.a.]: Addison Wesley.
- Cooley, Robert W. (2000): „Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data“. Diss. Minneapolis, Mn: University of Minnesota.
- Cooley, Robert W.; Mobasher, Bamshad; Srivastava, Jaideep (1997a): *Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns*. In: Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop. KDEX '97. Washington, DC: IEEE, S. 2–9.
- Cooley, Robert W.; Mobasher, Bamshad; Srivastava, Jaideep (1997b): *Web Mining: Information and Pattern Discovery on the World Wide Web*. In: 9th International Conference on Tools with Artificial Intelligence. ICTAI '97. Newport Beach, CA: IEEE, S. 558–567.
- Cooley, Robert W.; Mobasher, Bamshad; Srivastava, Jaideep (1999): *Data Preparation for Mining World Wide Web Browsing Patterns*. Knowledge and Information Systems 1 (1), S. 5–32.
- Cooper, Alan; Reimann, Robert; Cronin, David (2007): *About Face 3: The Essentials of Interaction Design*. Indianapolis, IN: John Wiley & Sons.

- Cortes, Corinna; Vapnik, Vladimir (1995): *Support-vector networks*. *Machine Learning* 20 (3), S. 273–297.
- Costabile, Maria F.; De Marsico, Maria; Lanzilotti, Rosa; Plantamura, Paola; Roselli, Teresa (2005): *On the Usability Evaluation of E-Learning Applications*. In: *Proceedings of the 38th Hawaii International Conference on System Sciences*. 38th Hawaii International Conference on System Sciences. HICSS '05. Washington, DC: IEEE.
- Costabile, Maria F.; Matera, Maristella (2001): *Guidelines for hypermedia usability inspection*. *IEEE MultiMedia* 8 (1), S. 66–69.
- Couper, Mick P. (2000): *Web Surveys*. *Public Opinion Quarterly* 64 (4), S. 464–494.
- Cover, Thomas M.; Hart, Peter E. (1967): *Nearest neighbor pattern classification*. *IEEE Transactions on Information Theory* 13 (1), S. 21–27.
- Davies, David L.; Bouldin, Donald W. (1979): *A Cluster Separation Measure*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1 (2), S. 224–227.
- Davis, Fred D. (1989): *Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology*. *MIS Quarterly* 13 (3), S. 319–340.
- De Leeuw, Edith D.; Hox, Joop J.; Dillman, Don A. (2009): *International handbook of survey methodology*. New York, NY [u.a.]: Psychology Press.
- De Porto Alegre Muniz, Maria I.; De Moraes, Anamaria (2012): *Usability issues in Learning Management Systems (LMS)*. *Work: A Journal of Prevention, Assessment and Rehabilitation* 41, S. 832–837.
- De Smet, Cindy; Bourgonjon, Jeroen; De Wever, Bram; Schellens, Tammy; Valcke, Martin (2012): *Researching instructional use and the technology acceptance of learning management systems by secondary school teachers*. *Computers & Education* 58 (2), S. 688–696.
- Dell, Robert F.; Roman, Pablo E.; Velasquez, Juan D. (2008): *Web User Session Reconstruction Using Integer Programming*. In: *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. WI-IAT '08*. Sydney, Australia: IEEE, S. 385–388.
- Department of Health and Human Services; General Services Administration (2006): *Research-based web design & usability guidelines*. Washington, D.C.: U.S. Dept. of Health und Human Services.
- Desgraupes, Bernard (2014): *clusterCrit: Clustering Indices*. URL: <http://CRAN.R-project.org/package=clusterCrit> (aufgerufen am 19.02.2015).
- Deza, Michel M.; Deza, Elena (2009): *Encyclopedia of Distances*. Berlin/Heidelberg: Springer.
- Diekmann, Andreas (2010): *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt.
- Dijkstra, Edsger W. (1959): *A note on two problems in connexion with graphs*. *Numerische Mathematik* 1 (1), S. 269–271.
- Dillman, Don (2000): *Mail and internet surveys: the tailored design method*. New York, NY [u.a.]: Wiley.
- Dix, Alan J.; Finlay, Janet E.; Abowd, Gregory D.; Beale, Russell (2004): *Human-Computer Interaction*. Harlow, UK [u.a.]: Pearson Education.
- Domenech, Jose M.; Lorenzo, Javier (2007): *A tool for web usage mining*. In: *Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning. IDEAL'07*. Berlin/Heidelberg: Springer, S. 695–704.
- Doubleday, Ann; Ryan, Michele; Springett, Mark; Sutcliffe, Alistair (1997): *A Comparison of Usability Techniques for Evaluating Design*. In: *Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques. DIS '97*. New York, NY: ACM, S. 101–110.
- Duchowsky, Andrew (2007): *Eye Tracking Methodology*. London: Springer.
- Duda, Richard O.; Hart, Peter E. (1973): *Pattern classification and scene analysis*. New York, NY [u.a.]: Wiley.
- Dumas, Joseph S. (2001): *Usability testing methods: Think aloud protocols*. In: Branaghan, Russell J. (Hrsg.): *Design by people for people: Essays on usability*. Chicago, IL: Usability Professionals' Association, S. 119–130.

- Dumas, Joseph S.; Fox, Jean E. (2008): *Usability Testing: Current Practice and Future Directions*. In: Sears, Andrew; Jacko, Julie A. (Hrsg.): *The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications*. New York, NY: Lawrence Erlbaum Associates, S. 1129–1149.
- Dumas, Joseph S.; Loring, Beth (2008): *Moderating Usability Tests: Principles and Practices for Interacting*. Burlington, MA: Morgan Kaufmann.
- Dumas, Joseph S.; Redish, Janice C. (1999): *A Practical Guide to Usability Testing*. Exeter, UK: Intellect Ltd.
- Dunn, J. C. (1974): *Well-Separated Clusters and Optimal Fuzzy Partitions*. *Journal of Cybernetics* 4 (1), S. 95–104.
- Eason, Ken D. (1984): *Towards the experimental study of usability*. *Behaviour & Information Technology* 3 (2), S. 133–143.
- Ebling, Maria R.; John, Bonnie E. (2000): *On the contributions of different empirical data in usability testing*. In: *Proceedings of the 3rd conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques*. DIS '00. New York, NY: ACM, S. 289–296.
- Eirinaki, Magdalini; Vazirgiannis, Michalis (2003): *Web mining for web personalization*. *ACM Transactions on Internet Technology* 3 (1), S. 1–27.
- Ellis, Clarence A; Gibbs, Simon J; Rein, Gail (1991): *Groupware: some issues and experiences*. *Communications of the ACM* 34 (1), S. 39–58.
- Esslimani, Ilham; Brun, Armelle; Boyer, Anna (2008): *Enhancing collaborative filtering by frequent usage patterns*. In: *Proceedings of the 1st International Conference on the Applications of Digital Information and Web Technologies*. Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First. ICADIWT '08. Ostrava, Czech Republic: IEEE, S. 180–185.
- Ester, Martin; Kriegel, Hans; Sander, Jorg; Xu, Xiaowei (1996): *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. In: Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama (Hrsg.): *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. KDD '96. Portland, OR: AAAI Press, S. 226–231.
- Estivill-Castro, Vladimir; Yang, Jianhua (2000): *Fast and Robust General Purpose Clustering Algorithms*. In: Mizoguchi, Riichiro; Slaney, John (Hrsg.): *PRICAI 2000 Topics in Artificial Intelligence*. Berlin/Heidelberg: Springer, S. 208–218.
- Etminani, Kobra; Akbarzadeh-T., Mohammad-R.; Yanehsari, Noorali Raeji (2009): *Web Usage Mining : users navigational patterns extraction from web logs using Ant-based Clustering Method*. In: *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference*. IFSA-EUSFLAT. Lissabon, Portugal: International Fuzzy Systems Association/European Society for Fuzzy Logic und Technology, S. 396–401.
- Etminani, Kobra; Rezaeian Delui, Amin; Raeji Yanehsari, Noorali; Rouhani, Modjtaba (2009): *Web usage mining: Discovery of the users navigational patterns using SOM*. In: *1st International Conference on Networked Digital Technologies*. NDT '09. Ostrava, Czech Republic: IEEE, S. 224–249.
- EU (2009): *Richtlinie 2009/136/EG des Europäischen Parlaments und des Rates*.
- Everitt, Brian S.; Landau, Sabine; Leese, Morven; Stahl, Daniel (2011): *Cluster analysis*. Chichester, West Sussex, U.K.: Wiley.
- Facca, Federico M.; Lanzi, Pier L. (2005): *Mining interesting knowledge from weblogs: a survey*. *Data & Knowledge Engineering* 53 (3), S. 225–241.
- Faulbaum, Frank; Stahl, Matthias; Wiegand, Erich, Hrsg. (2012): *Qualitätssicherung in der Umfrageforschung*. Wiesbaden: Springer. URL: <http://link.springer.com/10.1007/978-3-658-00515-3> (aufgerufen am 31. 07. 2015).
- Faulkner, Laura (2003): *Beyond the five-user assumption: Benefits of increased sample sizes in usability testing*. *Behavior Research Methods, Instruments, & Computers* 35 (3), S. 379–383.
- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996): *From Data Mining to Knowledge Discovery in Databases*. *AI Magazine* 17 (3), S. 37–54.

- Fenstermacher, Kurt D.; Ginsburg, Mark (2002): *Mining Client-Side Activity for Personalization*. In: Proceedings of the 4th IEEE International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems. WECWIS '02. Washington, DC: IEEE, S. 205.
- Few, Stephen (2009): *Now you see it: simple visualization techniques for quantitative analysis*. Oakland, CA: Analytics Press.
- Fielding, R.; Gettys, J.; Mogul, J.; Frystyk, H.; Masinter, L.; Leach, P.; Berners-Lee, T. (1999): *Hypertext Transfer Protocol. HTTP/1.1*. URL: <http://www.ietf.org/rfc/rfc2616.txt> (aufgerufen am 20.07.2012).
- Flick, Uwe (2007): *Qualitative Sozialforschung: Eine Einführung*. Reinbek bei Hamburg: rororo.
- Flick, Uwe (2008): *Triangulation: eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Franke, Nikolas (2011): *Crowdsourcing: innovative User-Netzwerke nutzen*. In: Albers, Sönke; Gassmann, Oliver (Hrsg.): *Handbuch Technologie- und Innovationsmanagement*. Wiesbaden: Gabler Verlag, S. 689–706.
- Freiberg, Martina; Baumeister, Joachim (2008): *A Survey on Usability Evaluation Techniques and an Analysis of their actual Application*. Report No. 450. Würzburg, Germany: University of Würzburg, Institute of Computer Science. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.142.6951> (aufgerufen am 23.10.2012).
- Friehs, Barbara (2003): *Wissensmanagement im schulischen Kontext: Lehrer und Lehrerinnen im Spannungsfeld organisatorischen Lernens*. Frankfurt a. M.: Peter Lang.
- Fu, Yongjian; Creado, Mario; Ju, Chunhua (2001): *Reorganizing web sites based on user access patterns*. In: Proceedings of the 10th International Conference on Information and Knowledge Management. CIKM '01. New York, NY: ACM, S. 583–585.
- Fujioka, Ryosuke; Tanimoto, Ryo; Kawai, Yuki; Okada, Hidehiko (2007): *Tool for detecting webpage usability problems from mouse click coordinate logs*. In: Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Design and Usability. HCI' 07. Berlin/Heidelberg: Springer, S. 438–445.
- Funke, Joachim; Anderson, John Robert (2013): *Kognitive Psychologie*. Berlin: Springer VS.
- García, Enrique; Romero, Cristóbal; Ventura, Sebastián; Calders, Toon (2007): *Drawbacks and solutions of applying association rule mining in learning management systems*. In: Proceedings of the International Workshop on Applying Data Mining in e-Learning. ADML '07. Crete, Greece, S. 13–22.
- Gediga, Günther; Hamborg, Kai-Christoph (2002): *Evaluation in der Software-Ergonomie*: Zeitschrift für Psychologie / Journal of Psychology 210 (1), S. 40–57.
- Gediga, Günther; Hamborg, Kai-Christoph; Düntsch, Ivo (1999): *The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems*. Behaviour & Information Technology 18 (3), S. 151–164.
- Gibbs, Adrian J.; McIntyre, George A. (1970): *The Diagram, a Method for Comparing Sequences*. European Journal of Biochemistry 16 (1), S. 1–11.
- Gläser, Jochen; Laudel, Grit (2010): *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Goldberg, Joseph H.; Stimson, Mark J.; Lewenstein, Marion; Scott, Neil; Wichansky, Anna M. (2002): *Eye tracking in web search tasks: design implications*. In: Proceedings of the 2002 Symposium on Eye tracking Research & Applications. ETRA '02. New York, NY: ACM, S. 51–58.
- Golub, Gene H.; van Loan, Charles F. (1996): *Matrix Computations*. Baltimore: Johns Hopkins University Press.
- Goodman, Elizabeth; Kuniavsky, Mike; Moed, Andrea (2012): *Observing the User Experience: A Practitioner's Guide to User Research*. Waltham, MA: Morgan Kaufmann.
- Gordon, A. D (1999): *Classification*. Boca Raton, FL: Chapman & Hall/CRC.
- Gould, John D.; Lewis, Clayton (1985): *Designing for Usability: Key Principles and What Designers Think*. Communications of the ACM 28 (3), S. 300–311.
- Gower, J. C. (1967): *A Comparison of Some Methods of Cluster Analysis*. Biometrics 23 (4), S. 623–637.
- Greller, Wolfgang; Drachler, Hendrik (2012): *Translating Learning into Numbers: A Generic Framework for Learning Analytics*. Educational Technology & Society 15 (3), S. 42–57.

- Gross, Tom; Koch, Michael (2007): *Computer-Supported Cooperative Work*. München: Oldenbourg.
- Grossman, David A.; Frieder, Ophir (2004): *Information Retrieval: Algorithms and Heuristics*. Dordrecht: Springer.
- Groves, Robert M.; Dillman, Don A.; Eltinge, John L.; Little, Roderick J. A., Hrsg. (2002): *Survey Nonresponse*. New York, NY: Wiley.
- Groves, Robert M.; Fowler, Floyd J.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor; Tourangeau, Roger (2009): *Survey Methodology*. Hoboken, NJ: Wiley.
- Grudin, Jonathan (1994): *Computer-supported cooperative work: history and focus*. *Computer* 27 (5), S. 19–26.
- Guan, Zhiwei; Lee, Shirley; Cuddihy, Elisabeth; Ramey, Judith (2006): *The Validity of the Stimulated Retrospective Think-aloud Method As Measured by Eye Tracking*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '06. New York, NY: ACM, S. 1253–1262.
- Guha, Sudipto; Rastogi, Rajeev; Shim, Kyuseok (1998): *CURE: an efficient clustering algorithm for large databases*. In: Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data. SIGMOD '98. New York, NY: ACM, S. 73–84.
- Guha, Sudipto; Rastogi, Rajeev; Shim, Kyuseok (1999): *ROCK: a robust clustering algorithm for categorical attributes*. In: Proceedings of the 15th International Conference on Data Engineering. 15th International Conference on Data Engineering. Sydney, Australia: IEEE, S. 512–521.
- Gupta, Vishal; Lehal, Gurpreet S. (2009): *A Survey of Text Mining Techniques and Applications*. *Journal of Emerging Technologies in Web Intelligence* 1 (1), S. 60–76.
- Halkidi, Maria; Batistakis, Yannis; Vazirgiannis, Michalis (2001): *On Clustering Validation Techniques*. *Journal of Intelligent Information Systems* 17 (2), S. 107–145.
- Hallam-Baker, Phillip M.; Behlendorf, Brian (1996): *Extended Log File Format*. URL: <http://www.w3.org/TR/WD-logfile.html> (aufgerufen am 26.07.2012).
- Hammon, Larissa; Hippner, Hajo (2012): *Crowdsourcing*. *Business & Information Systems Engineering* 4 (3), S. 163–166.
- Han, Hyoil; Elmasri, Ramez (2004): *Learning Rules for Conceptual Structure on the Web*. *Journal of Intelligent Information Systems* 22 (3), S. 237–256.
- Han, Jiawei; Kamber, Micheline; Pei, Jian (2012): *Data Mining: Concepts and Techniques*. Amsterdam [u.a.]: Morgan Kaufmann.
- Harty, Julian (2011): *Finding usability bugs with automated tests*. *Communications of the ACM* 54 (2), S. 44–49.
- Harvey, Craig M.; Koubek, Richard J.; Darisipudi, Ashok; Rothrock, Ling (2011): *Cognitive User Modeling*. In: Vu, Kim-Phuong L.; Proctor, Robert W. (Hrsg.): *Handbook of Human Factors in Web Design*. Boca Raton, FL: CRC Press, S. 85–105.
- Hassenzahl, Mark; Burmsiter, Michael; Koller, Franz (2003): *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In: Ziegler, Jürgen; Szwillus, G. (Hrsg.): *Mensch & Computer 2003. Interaktion in Bewegung*. Stuttgart/Leipzig: B.G. Teubner, S. 187–196.
- Hassler, Marco (2010): *Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren*. Heidelberg: Verlagsgruppe Hüthig-Jehle-Rehm.
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2011): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.
- Hatscher, Michael (2001): *Joy of use — Determinanten der Freude bei der Software-Nutzung*. In: Oberquelle, Horst; Oppermann, Reinhard; Krause, Jürgen (Hrsg.): *Mensch & Computer 2001*. Wiesbaden: Vieweg+Teubner Verlag, S. 445–446.
- Hegner, Marcus (2003): *Methoden zur Evaluation von Software*. IZ-Arbeitsbericht Nr. 29. Bonn: GESIS: Informations Zentrum-Sozialwissenschaften.
- Heinrich, Lutz J.; Heinzl, Armin; Riedl, René (2011): *Wirtschaftsinformatik. Einführung und Grundlegung*. Berlin/Heidelberg: Springer.
- Hennig, Christian (2014): *How Many Bee Species? A Case Study in Determining the Number of Clusters*. In: Spiliopoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hrsg.): *Data Analysis, Machine Learning and Knowledge Discovery*. Cham: Springer, S. 41–49.

- Henzinger, Monika (2000): *Link analysis in web information retrieval*. IEEE Data Engineering Bulletin 23 (3), S. 3–8.
- Hepp, Andreas (2011): *Medienkultur: Die Kultur mediatisierter Welten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Heyer, Gerhard; Quasthoff, Uwe; Wittig, Thomas (2006): *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. Herdecke: W3L-Verlag.
- Hickson, Ian; Berjon, Robin; Faulkner, Steve; Leithead, Travis; Doyle Navara, Erika; O'Connor, Edward; Pfeiffer, Silvia (2014): *HTML5. A vocabulary and associated APIs for HTML and XHTML. W3C Recommendation 28 October 2014*. URL: <http://www.w3.org/TR/html5/> (aufgerufen am 01.08.2015).
- Hipp, Jochen; Güntzer, Ulrich; Nakhaeizadeh, Gholamreza (2000): *Algorithms for Association Rule Mining - a General Survey and Comparison*. SIGKDD Explorations Newsletter 2 (1), S. 58–64.
- Hocking, Toby; Vert, Jean-Philippe; Bach, Francis; Joulin, Armand (2011): *Clusterpath: an Algorithm for Clustering using Convex Fusion Penalties*. In: Getoor, Lise; Scheffer, Tobias (Hrsg.): *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. ICML '11. New York, NY: ACM, S. 745–752.
- Hogan, Bernie (2008): *Analyzing Social Networks via the Internet*. In: Fielding, Nigel; Lee, Raymond M.; Blank, Grant (Hrsg.): *The SAGE handbook of online research methods*. Los Angeles: SAGE, S. 141–160.
- Holden, Richard J.; Karsh, Ben-Tzion (2010): *The Technology Acceptance Model: Its past and its future in health care*. Journal of Biomedical Informatics 43 (1), S. 159–172.
- Hollander, Myles; Wolfe, Douglas A.; Chicken, Eric (2013): *Nonparametric Statistical Methods*. Hoboken, NJ: Wiley.
- Hollingsed, Tasha; Novick, David G. (2007): *Usability inspection methods after 15 years of research and practice*. In: *Proceedings of the 25th Annual ACM International Conference on Design of Communication*. SIGDOC '07. New York, NY: ACM, S. 249–255.
- Holzinger, Andreas (2005): *Usability Engineering Methods for Software Developers*. Communications of the ACM 48 (1), S. 71–74.
- Höppner, Frank (2005): *Association Rules*. In: Maimon, Oded; Rokach, Lior (Hrsg.): *The Data Mining and Knowledge Discovery Handbook*. New York: Springer, S. 353–376.
- Hou, Jingyu; Zhang, Yanchun (2003): *Effectively finding relevant Web pages from linkage information*. IEEE Transactions on Knowledge and Data Engineering 15 (4), S. 940–951.
- Hubert, Lawrence; Schultz, James (1976): *Quadratic Assignment as a General Data Analysis Strategy*. British Journal of Mathematical and Statistical Psychology 29 (2), S. 190–241.
- Huiying, Zhang; Wei, Liang (2004): *An intelligent algorithm of data pre-processing in Web usage mining*. In: *5th World Congress on Intelligent Control and Automation. Fifth World Congress on Intelligent Control and Automation, 2004. WCICA '04*. Hangzhou, China: IEEE, 3119–3123 Vol.4.
- ISO (1998): *ISO 9241-11:1998. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit; Leitsätze*.
- ISO (2003): *ISO 14915-1:2002. Software-Ergonomie für Multimedia-Benutzungsschnittstellen - Teil 1: Gestaltungsgrundsätze und Rahmenbedingungen*.
- ISO (2006): *ISO 9241-110:2006. Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung*.
- ISO (2010): *ISO 9241-210:2010. Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*.
- Ivancsy, Renata; Kovacs, Ferenc (2006): *Clustering techniques utilized in web usage mining*. In: *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases. AIKED'06*. Stevens Point, WI: World Scientific, Engineering Academy and Society (WSEAS), S. 237–242.
- Ivory, Melody Y. (2001): *„An Empirical Foundation for Automated Web Interface Evaluation“*. Diss. Berkeley, CA: University of Berkeley.
- Ivory, Melody Y.; Hearst, Marti A. (2001): *The state of the art in automating usability evaluation of user interfaces*. ACM Computing Surveys 33 (4), S. 470–516.

- Jaccard, Paul (1912): *The distribution of the flora in the alpine zone*. *New Phytologist* 11 (2), S. 37–50.
- Jain, A. K.; Murty, M. N.; Flynn, P. J. (1999): *Data Clustering: A Review*. *ACM Computing Surveys* 31 (3), S. 264–323.
- Jansen, Dorothea (2006): *Einführung in die Netzwerkanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Jaspers, Monique W. M. (2009): *A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence*. *International Journal of Medical Informatics* 78 (5), S. 340–353.
- Jeffries, Robin; Desurvire, Heather (1992): *Usability testing vs. heuristic evaluation: was there a contest?* *SIGCHI Bulletin* 24 (4), S. 39–41.
- Jeffries, Robin; Miller, James R.; Wharton, Cathleen; Uyeda, Kathy (1991): *User interface evaluation in the real world: a comparison of four techniques*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '91. New York, NY: ACM, S. 119–124.
- Johansen, Robert (1991): *Teams for tomorrow [groupware]*. In: *Proceedings of the 24th Annual Hawaii International Conference on System Sciences*. *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences, 1991*. HICSS '91. Kauai, HI: IEEE, S. 521–534.
- John, Bonnie E. (2003): *Information Processing and Skilled Behavior*. In: Carroll, John M. (Hrsg.): *HCI models, theories, and frameworks. Toward a multidisciplinary science*. San Francisco, CA: Morgan Kaufmann, S. 55–101.
- John, Bonnie E.; Kieras, David E. (1996a): *The GOMS family of user interface analysis techniques: comparison and contrast*. *ACM Transactions on Computer-Human Interaction* 3 (4), S. 320–351.
- John, Bonnie E.; Kieras, David E. (1996b): *Using GOMS for user interface design and evaluation: which technique?* *ACM Transactions on Computer-Human Interaction* 3 (4), S. 287–319.
- John, Bonnie E.; Packer, Hilary (1995): *Learning and using the cognitive walkthrough method: a case study approach*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. New York, NY: ACM, S. 429–436.
- Kakasevski, Gorgi; Mihajlov, Martin; Arsenovski, Sime; Chungurski, Slavcho (2008): *Evaluating usability in learning management system moodle*. In: *Proceedings of the ITI 30th International Conference on Information Technology Interfaces*. Cavtat, Croatia: IEEE, S. 613–618.
- Kamoen, Naomi (2012): *Positive versus negative: a cognitive perspective on wording effects for contrastive questions in attitude surveys*. Utrecht, Netherlands: LOT.
- Kantel, Jörg (2007): *RSS und Atom - kurz & gut*. O'Reilly.
- Karahoca, Dilek; Karahoca, Adem (2009): *Assessing effectiveness of the cognitive abilities and individual differences on e-learning portal usability evaluation*. *Procedia - Social and Behavioral Sciences* 1 (1), S. 368–380.
- Karat, Claire-Marie; Campbell, Robert; Fiegel, Tarra (1992): *Comparison of empirical testing and walkthrough methods in user interface evaluation*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. New York, NY: ACM, S. 397–404.
- Kargupta, Hillol (2004): *Data mining: next generation challenges and future directions*. Menlo Park, CA: AAAI Press.
- Karypis, George; Han, Eui-Hong; Kumar, Vipin (1999): *CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling*. *IEEE Computer* 32 (8), S. 68–75.
- Kaufman, Leonard; Rousseeuw, Peter J. (2005): *Finding Groups in Data. An Introduction to Cluster Analysis*. Hoboken, NJ: Wiley.
- Keen, Peter G. W. (1981): *Information Systems and Organizational Change*. *Communications of the ACM* 24 (1), S. 24–33.
- Kelley, J. F. (1984): *An iterative design methodology for user-friendly natural language office information applications*. *ACM Transactions on Information Systems* 2 (1), S. 26–41.
- Kerres, Michael (2001): *Multimediale und telemediale Lernumgebungen: Konzeption und Entwicklung*. München: Oldenbourg.
- Kieras, David (2008): *Model-Based Evaluation*. In: Sears, Andrew; Jacko, Julie A. (Hrsg.): *The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications*. New York, NY: Lawrence Erlbaum Associates, S. 1191–1208.

- King, William R.; He, Jun (2006): *A meta-analysis of the technology acceptance model*. Information & Management 43 (6), S. 740–755.
- Kittur, Aniket; Chi, Ed H.; Suh, Bongwon (2008): *Crowdsourcing User Studies with Mechanical Turk*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '08. New York, NY: ACM, S. 453–456.
- Kiy, Alexander; Lucke, Ulrike (2014): *Learning-Analytics-Werkzeuge im Praxisvergleich*. In: Proceedings of DeLFI Workshops 2014 co-located with 12th e-Learning Conference of the German Computer Society, Freiburg, Germany, September 15, 2014. DeLFI 2014. CEUR-WS.org, S. 104–111.
- Kleinberg, Jon M.; Kumar, Ravi; Raghavan, Prabhakar; Rajagopalan, Sridhar; Tomkins, Andrew S. (1999): *The Web as a Graph: Measurements, Models, and Methods*. In: Asano, Takano; Imai, Hideki; Lee, D. T.; Nakano, Shin-ichi; Tokuyama, Takeshi (Hrsg.): Computing and Combinatorics. Berlin/Heidelberg: Springer, S. 1–17.
- Kohn, Martin (2011): *Schulentwicklung 2.0. Digitale Lern- und Arbeitswelten*. Weinheim: Beltz.
- Kort, Joke; de Poot, Henk (2005): *Usage analysis: combining logging and qualitative methods*. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. CHI '05. New York, NY: ACM, S. 2121–2122.
- Kosala, Raymond; Blockeel, Hendrik (2000): *Web mining research: a survey*. SIGKDD Explorations Newsletter 2 (1), S. 1–15.
- Krahmer, Emiel; Ummelen, N. (2004): *Thinking about thinking aloud: a comparison of two verbal protocols for usability testing*. IEEE Transactions on Professional Communication 47 (2), S. 105–117.
- Krotz, Friedrich (2007): *The meta-process of 'mediatization' as a conceptual frame*. Global Media and Communication 3 (3), S. 256–260.
- Krug, Steve (2006): *Don't Make Me Think!: A Common Sense Approach to Web Usability*. Berkeley, CA: New Riders.
- Kumar, B. Santhosh; Rukmani, K.V. (2010): *Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms*. International Journal of Advanced Networking and Applications 1 (6), S. 400–404.
- Kuo, Yin-hung; Wong, Man-hon (2000): *Web Document Classification based on Hyperlinks and Document Semantics*. In: Tan, Ah-Hwee; Yu, Philip S. (Hrsg.): PRICAI 2000 Workshop on Text and Web Mining. Melbourne, Australia, S. 44–51.
- Lamnek, Siegfried; Krell, Claudia (2010): *Qualitative Sozialforschung: Lehrbuch. Mit Online-Materialien*. Weinheim: Beltz.
- Lan, Bin; Bressan, Stephane; Ooi, Beng Chin; Tan, Kian-Lee (2000): *Rule-assisted prefetching in Web-server caching*. In: Proceedings of the 9th International Conference on Information and Knowledge Management. CIKM '00. New York, NY: ACM, S. 504–511.
- Langacker, Ronald W. (1986): *An Introduction to Cognitive Grammar*. Cognitive Science 10, S. 1–40.
- Larose, Daniel T. (2005): *Discovering Knowledge in Data. An Introduction in Data Mining*. Hoboken, NJ: Wiley.
- Larose, Daniel T. (2006): *Data Mining Methods and Models*. Hoboken, NJ: Wiley.
- Lassmann, Wolfgang; Schwarzer, Jens; Rogge, Rolf, Hrsg. (2006): *Wirtschaftsinformatik: Nachschlagewerk für Studium und Praxis*. Wiesbaden: Gabler.
- Laudon, Kenneth C.; Laudon, Jane P. (2015): *Management Information System*. Edinburgh Gate: Pearson Education Limited.
- Law, Chris M.; Vanderheiden, Gregg C. (2000): *Reducing Sample Sizes When User Testing with People Who Have, and Who are Simulating Disabilities - Experiences with Blindness and Public Information Kiosks*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 44 (26), S. 157–160.
- Lazar, Jonathan (2006): *Web Usability: A User-Centered Design Approach*. Boston, MA [u.a.]: Pearson.
- Lazar, Jonathan; Feng, Jinjuan Heidi; Hochheiser, Harry (2010): *Research Methods in Human-Computer Interaction*. Chichester, UK: Wiley.
- Lecerof, Andreas; Paternò, Fabio (1998): *Automatic Support for Usability Evaluation*. IEEE Transactions on Software Engineering 24 (10), S. 863–888.

- Lee, Johnny Chung; Tan, Desney S. (2006): *Using a Low-cost Electroencephalograph for Task Classification in HCI Research*. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology. UIST '06. New York, NY: ACM, S. 81–90.
- Lee, Man Wai; Chen, Sherry Y.; Chrysostomou, Kyriacos; Liu, Xiaohui (2009): *Mining students' behavior in web-based learning programs*. Expert Systems with Applications 36 (2), S. 3459–3464.
- Leußer, Wolfgang; Hippner, Hajo; Wilde, Klaus D. (2011): *CRM: Grundlagen, Konzepte und Prozesse*. Wiesbaden: Gabler.
- Levy, Paul S. (2009): *Sampling of Populations: Methods and Applications*. New York, NY: Wiley-Blackwell.
- Lewis, Clayton; Polson, Peter G.; Wharton, Cathleen; Rieman, John (1990): *Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People. CHI '90. Seattle, WA: ACM, S. 235–242.
- Lewis, Clayton; Wharton, Cathleen (1997): *Cognitive Walkthroughs*. In: Helander, Martin G.; Landauer, Thomas K.; Prabhu, Prasad V. (Hrsg.): *Handbook of Human-Computer Interaction*. Amsterdam, Netherlands: Elsevier Science B.V., S. 717–732.
- Lewis, James R. (1994): *Sample Sizes for Usability Studies: Additional Considerations*. Human Factors: The Journal of the Human Factors and Ergonomics Society 36 (2), S. 368–378.
- Lewis, James R. (2001): *Evaluation of Procedures for Adjusting Problem-Discovery Rates Estimated From Small Samples*. International Journal of Human-Computer Interaction 13 (4), S. 445–479.
- Lewis, James R. (2006): *Sample sizes for usability tests: mostly math, not magic*. interactions 13 (6), S. 29–33.
- Li, Jia; Zaïane, Osmar R. (2004): *Combining Usage, Content, and Structure Data to Improve Web Site Recommendation*. In: Bauknecht, Kurt; Bichler, Martin; Pröll, Birgit (Hrsg.): *E-Commerce and Web Technologies*. Berlin/Heidelberg: Springer, S. 305–315.
- Li, Yan; Feng, Boqin; Mao, Qinjiao (2008): *Research on Path Completion Technique in Web Usage Mining*. In: Yu, Frei; Chen, Wen; Chen, Zhigang; Yuan, Jie (Hrsg.): *Proceedings of the International Symposium on Computer Science and Computational Technology, 2008. ISCSCT '08*. Los Alamos, CA: IEEE, S. 554–559.
- Liang, Ting-Peng; Wei, Chih-Ping (2004): *Introduction to the Special Issue: Mobile Commerce Applications*. International Journal of Electronic Commerce 8 (3), S. 7–17.
- Lima, Manuel (2011): *Visual complexity: mapping patterns of information*. New York, NY: Princeton Architectural Press.
- Lin, Han X.; Choong, Yee-Yin; Salvendy, Gavriel (1997): *A proposed index of usability: A method for comparing the relative usability of different software systems*. Behaviour & Information Technology 16 (4), S. 267–277.
- Lindgaard, Gitte; Chattratichart, Jarinee (2007): *Usability testing: what have we overlooked?* In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '07. New York, NY: ACM, S. 1415–1424.
- Liu, Bing (2011): *Web data mining: exploring hyperlinks, contents, and usage data*. Berlin: Springer.
- Liu, Yanchi; Li, Zhongmou; Xiong, Hui; Gao, Xuedong; Wu, Junjie (2010): *Understanding of Internal Clustering Validation Measures*. In: IEEE 13th International Conference on Data Mining, ICDM '10. Los Alamitos, CA: IEEE, S. 911–916.
- Liu, Yanchi; Li, Zhongmou; Xiong, Hui; Gao, Xuedong; Wu, Junjie; Wu, Sen (2013): *Understanding and Enhancement of Internal Clustering Validation Measures*. IEEE Transactions on Cybernetics 43 (3), S. 982–994.
- Lohmann, Steffen; Ziegler, Jürgen; Tetzlaff, Lena (2009): *Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration*. In: Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction. INTERACT '09. Berlin/Heidelberg: Springer, S. 392–404.
- Lohr, Sharon L. (2009): *Sampling: Design and Analysis*. Boston, MA: Wadsworth Inc Fulfillment.

- Lutzky, Christian; Teichmann, Maik-Henrik (2002): *Logfiles in der Marktforschung: Gestaltungsoptionen für Analysezwecke*. Jahrbuch der Absatz- und Verbrauchsforschung 48 (3), S. 295–137.
- Mackay, Wendy E. (1998): *Triangulation within and across HCI disciplines*. Human-Computer Interaction 13 (3), S. 310–315.
- MacQueen, J. (1967): *Some methods for classification and analysis of multivariate observations*. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. Berkeley, CA: University of California Press, S. 281–297.
- Madan, Ankita; Dubey, Sanjay K. (2012): *Usability evaluation methods: a literature review*. International Journal of Engineering Science and Technology 4 (2), S. 590–599.
- Madria, Sanjay; Bhowmick, Sourav; Ng, W.-K.; Lim, E. P. (1999): *Research Issues in Web Data Mining*. In: Mohania, Mukesh; Tjoa, A. Min (Hrsg.): Data Warehousing and Knowledge Discovery. Berlin/Heidelberg: Springer, S. 805–805.
- Maisch, Josef (2006): *Wissensmanagement am Gymnasium: Anforderungen der Wissensgesellschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Malaka, Rainer; Butz, Andreas; Hußmann, Heinrich (2009): *Medieninformatik: eine Einführung*. München: Pearson Studium.
- Mandryk, Regan L.; Inkpen, Kori M. (2004): *Physiological Indicators for the Evaluation of Co-located Collaborative Play*. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work. CSCW '04. New York, NY: ACM, S. 102–111.
- Mann, Steve (1997): *Wearable computing: a first step toward personal imaging*. Computer 30 (2), S. 25–32.
- Markov, Zdravko; Larose, Daniel T. (2007): *Data mining the web: uncovering patterns in web content, structure and usage*. Hoboken, NJ: Wiley.
- Martin, Ludivine; Roladan Martinez, David; Revilla, Olga; José Aguilar, Maria; Santos, Olga C.; Boticario, Jesús G. (2008): *Usability in e-Learning Platforms: heuristics comparison between Moodle, Sakai and dotLRN*. In: OpenACS and .LRN conference 2008. International Conference and Workshops on Community based environments. Antigua, Guatemala.
- Masseglia, Florent; Poncet, Pascal; Cichetti, Rosine (2000): *An efficient algorithm for web usage mining*. Networking and Information Systems Journal, S. 571–603.
- Matera, Maristella; Costabile, Maria F.; Garzotto, Franca; Paolini, Paolo (2002): *SUE inspection: an effective method for systematic usability evaluation of hypermedia*. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 32 (1), S. 93–103.
- Mayring, Philipp (2002): *Einführung in die qualitative Sozialforschung*. Weinheim/Basel: Beltz.
- Mayring, Philipp (2010): *Qualitative Inhaltsanalyse Grundlagen und Techniken*. Weinheim: Beltz.
- Mayring, Philipp; Gläser-Zikuda, Michaela (2008): *Die Praxis der Qualitativen Inhaltsanalyse*. Weinheim/Basel: Beltz.
- McQuitty, Louis L. (1966): *Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data*. Educational and Psychological Measurement 26 (4), S. 825–831.
- Mehlenbacher, Brad; Bennett, Leslie; Bird, Tammy; Ivey, Melonie; Lucas, Jan; Morton, Janet; Whiteman, Lisa (2005): *Usable E-Learning: A Conceptual Model for Evaluation and Design*. In: Proceedings of HCI International 2005: 11th International Conference on Human-Computer Interaction, Volume 4 - Theories, Models, and Processes in HCI. Las Vegas, NV: Mira Digital, S. 1–10.
- Melton, Jay (2006): *The LMS moodle: A Usability Evaluation*. Language Issues 11/12 (1), S. 1–24.
- Merceron, Agathe; Yacef, Kalina (2007): *Revisiting interestingness of strong symmetric association rules in educational data*. In: Proceedings of the International Workshop on Applying Data Mining in e-Learning. ADML '07. Crete, Greece, S. 3–12.
- Metzger, Wolfgang; Stadler, Michael; Crabus, Heinrich (1986): *Gestalt-Psychologie: ausgewählte Werke aus den Jahren 1950 - 1982*. Frankfurt a.M.: Kramer.
- Meyen, Michael; Löblich, Maria; Pfaff-Rüdiger, Senta; Riesmeyer, Claudia, Hrsg. (2011): *Qualitative Forschung in der Kommunikationswissenschaft: eine praxisorientierte Einführung*. Studienbücher zur Kommunikations- und Medienwissenschaft. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Miles, Raymond E.; Snow, Charles C. (2003): *Organizational strategy, structure, and process*. Stanford, CA: Stanford University Press.

- Milligan, Glenn W.; Cooper, Martha C. (1985): *An examination of procedures for determining the number of clusters in a data set*. *Psychometrika* 50 (2), S. 159–179.
- Mobasher, Bamshad (2008): *Web Mining Overview*. In: Wang, John (Hrsg.): *Encyclopedia of Data Warehousing and Mining*. Hershey, PA: Information Science Reference, S. 2085–2089.
- Mobasher, Bamshad; Cooley, Robert W.; Srivastava, Jaideep (2000): *Automatic personalization based on Web usage mining*. *Communications of the ACM* 43 (8), S. 142–151.
- Mobasher, Bamshad; Dai, Honghua; Luo, Tao; Nakagawa, Miki (2002): *Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization*. *Data Mining and Knowledge Discovery* 6 (1), S. 61–82.
- Mobasher, Bamshad; Dai, Honghua; Luo, Tao; Sun, Yuqing; Zhu, Jiang (2000): *Integrating Web Usage and Content Mining for More Effective Personalization*. In: *Proceedings of the 1st International Conference on Electronic Commerce and Web Technologies. EC-WEB '00*. London, UK: Springer, S. 165–176.
- Mohamad, Siti K.; Tasir, Zaidatun (2013): *Educational Data Mining: A Review*. *Procedia - Social and Behavioral Sciences* 97, S. 320–324.
- Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012): *Foundations of Machine Learning*. Cambridge, MA: Mit Press Ltd.
- Molich, Rolf; Nielsen, Jakob (1990): *Improving a human-computer dialogue*. *Communications of the ACM* 33 (3), S. 338–348.
- Muller, Michael J.; Matheson, Lisa; Page, Colleen; Gallup, Robert (1998): *Participatory Heuristic Evaluation*. *interactions* 5 (5), S. 13–18.
- Muller, Michael J.; McClard, Anne (1995): *Validating an extension to participatory heuristic evaluation: quality of work and quality of work life*. In: *Conference Companion on Human Factors in Computing Systems. CHI '95*. New York, NY: ACM, S. 115–116.
- Nakamichi, Noboru; Shima, Kazuyuki; Sakai, Makoto; Matsumoto, Ken-ichi (2006): *Detecting low usability web pages using quantitative data of users' behavior*. In: *Proceedings of the 28th International Conference on Software Engineering. ICSE '06*. New York, NY: ACM, S. 569–576.
- Nanopoulos, Alexandros; Katsaros, Dimitrios; Manolopoulos, Yannis (2002): *Exploiting Web Log Mining for Web Cache Enhancement*. In: *Revised Papers from the Third International Workshop on Mining Web Log Data Across All Customers Touch Points. WEBKDD '01*. London, UK: Springer, S. 68–87.
- Nasraoui, Olfa; Frigui, Hichem; Joshi, Anupam; Krishnapuram, Raghu (1999): *Mining web access logs using relational competitive fuzzy clustering*. In: *Proceedings of the 8th International Fuzzy Systems Association World Congress. IFSA '99*. Taipei, Taiwan.
- Nguyen, Thao (2012): *Usability Evaluation of Web-Based Learning System*. *International Journal on E-Learning* 11 (3), S. 281–316.
- Nielsen, Jakob (1992): *Finding usability problems through heuristic evaluation*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '92*. New York, NY: ACM, S. 373–380.
- Nielsen, Jakob (1993): *Usability Engineering*. Amsterdam, Netherlands [u.a.]: Morgan Kaufmann.
- Nielsen, Jakob (1994a): *Enhancing the explanatory power of usability heuristics*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence. CHI '94*. New York, NY: ACM, S. 152–158.
- Nielsen, Jakob (1994b): *Estimating the number of subjects needed for a thinking aloud test*. *International Journal of Human-Computer Studies* 41 (3), S. 385–397.
- Nielsen, Jakob (1994c): *Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier*. In: Bias, Randolph G.; Mayhew, Deborah J. (Hrsg.): *Cost-justifying usability*. Cambridge, MA: Academic Press, S. 245–272.
- Nielsen, Jakob (1994d): *Heuristic Evaluation*. In: Nielsen, Jakob; Mack, Robert L. (Hrsg.): *Usability Inspection Methods*. New York, NY: John Wiley & Sons, S. 25–62.
- Nielsen, Jakob (1994e): *Usability inspection methods*. In: *Proceedings of the Conference Companion on Human Factors in Computing Systems. CHI '94*. New York, NY: ACM, S. 413–414.
- Nielsen, Jakob (1997): *The use and misuse of focus groups*. *IEEE Software* 14 (1), S. 94–95.

- Nielsen, Jakob; Landauer, Thomas K. (1993): *A mathematical model of the finding of usability problems*. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. CHI '93. New York, NY: ACM, S. 206–213.
- Nielsen, Jakob; Molich, Rolf (1990): *Heuristic evaluation of user interfaces*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People. CHI '90. New York, NY: ACM, S. 249–256.
- Nisbet, Robert; Elder, John; Miner, Gary (2009): *Handbook of Statistical Analysis and Data Mining Applications*. Amsterdam, Netherlands [u.a.]: Academic Press.
- Oliner, Adam; Ganapathi, Archana; Xu, Wei (2012): *Advances and challenges in log analysis*. Communications of the ACM 55 (2), S. 55–61.
- Olmsted-Hawala, Erica; Murphy, Elisabeth D.; Hawala, S.; Ashenfelter, K. T. (2010): *Think-aloud protocols: Analyzing three different think-aloud protocols with counts of verbalized frustrations in a usability study of an information-rich Web site*. In: 2010 IEEE International Professional Communication Conference. Professional Communication Conference (IPCC), 2010 IEEE International. IPCC '10. Enschede, Netherlands: IEEE, S. 60–66.
- Özgür, Arzucan; Özgür, Levent; Güngör, Tunga (2005): *Text Categorization with Class-Based and Corpus-Based Keyword Selection*. In: Yolum, Inar; Güngör, Tunga; Gürgen, Fikret; Özturan, Can (Hrsg.): Computer and Information Sciences - ISCIS 2005. Berlin/Heidelberg: Springer, S. 606–615.
- Ozok, A. Ant (2008): *Survey Design and Implementation in HCI*. In: Sears, Andrew; Jacko, Julie A. (Hrsg.): The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications. New York, NY: Lawrence Erlbaum Associates, S. 1151–1169.
- Pabarskaite, Zidrina (2002): *Implementing advanced cleaning and end-user interpretability technologies in Web log mining*. In: Proceedings of the 24th International Conference on Information Technology Interfaces. Proceedings of the 24th International Conference on Information Technology Interfaces, 2002. ITI 2002. ITI '02. Zagreb, Croatia: University of Zagreb, S. 109–113.
- Pabarskaite, Zidrina; Raudys, Aistis (2007): *A process of knowledge discovery from web log data: Systematization and critical review*. Journal of Intelligent Information Systems 28 (1), S. 79–104.
- Pal, Sankar K.; Talwar, Varun; Mitra, Pabitra (2002): *Web mining in soft computing framework: relevance, state of the art and future directions*. IEEE Transactions on Neural Networks 13 (5), S. 1163–1177.
- Paliouras, Georgios; Papatheodorou, Christos; Karkaletsis, Vangelis; Spyropoulos, Constantine D. (2000): *Clustering the Users of Large Web Sites into Communities*. In: Proceedings of the Seventeenth International Conference on Machine Learning. ICML '00. San Francisco, CA: Morgan Kaufmann, S. 719–726.
- Park, Sungjune; Suresh, Nallan C.; Jeong, Bong-Keun (2008): *Sequence-based clustering for Web usage mining: A new experimental framework and ANN-enhanced K-means algorithm*. Data & Knowledge Engineering 65 (3), S. 512–543.
- Paterno, Fabio (2009): *Model-Based Design and Evaluation of Interactive Applications*. London/Berlin/Heidelberg: Springer.
- Peacock, Ian (1998): *Showing Robots the Door*. Ariadne (15).
- Pei, Jian; Han, Jiawei; Mortazavi-Asl, Behzad; Zhu, Hua (2000): *Mining Access Patterns Efficiently from Web Logs*. In: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications. PADKK '00. London, UK: Springer, S. 396–407.
- Perlman, Gary (2015): *User Interface Usability Evaluation with Web-Based Questionnaires*. URL: <http://garyperlman.com/quest/> (aufgerufen am 27. 05. 2015).
- Perry, Mark (2003): *Distributed Cognition*. In: Carroll, John M. (Hrsg.): HCI models, theories, and frameworks. Toward a multidisciplinary science. San Francisco, CA: Morgan Kaufmann, S. 193–223.
- Piatetsky-Shapiro, Gregory (1990): *Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop*. AI Magazine 11 (4), S. 68–70.
- Pierrakos, Dimitrios; Paliouras, Georgios; Papatheodorou, Christos; Spyropoulos, Constantine D. (2003): *Web Usage Mining as a Tool for Personalization: A Survey*. User Modeling and User-Adapted Interaction 13 (4), S. 311–372.

- Pieters, Simon (2014): *HTML5 Differences from HTML4*. W3C Working Group Note 9 December 2014. URL: <http://www.w3.org/TR/html5-diff/> (aufgerufen am 01.08.2015).
- Pinelle, David; Gutwin, Carl (2000): *A Review of Groupware Evaluations*. In: Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. Washington, DC: IEEE, S. 86–91.
- Pinelle, David; Gutwin, Carl (2002): *Groupware walkthrough: adding context to groupware usability evaluation*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, changing ourselves. CHI '02. New York, NY: ACM, S. 455–462.
- Pirolli, Peter; Pitkow, James E.; Rao, Ramana (1996): *Silk from a sow's ear: extracting usable structures from the Web*. In: Bilger, Ralf; Guest, Steve; Tauber, Michael J. (Hrsg.): Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground. CHI '96. New York, NY: ACM, S. 118–125.
- Pitkow, James E. (1997): *In search of reliable usage data on the WWW*. Computer Networks and ISDN Systems 29 (8), S. 1343–1355.
- Pitt, Leyland; Berthon, Pierre; Robson, Karen (2011): *Deciding When to Use Tablets for Business Applications*. MIS Quarterly Executive 10 (3), S. 133–139.
- Polson, Peter G.; Lewis, Clayton H. (1990): *Theory-based design for easily learned interfaces*. Human-Computer Interaction 5 (2), S. 191–220.
- Polson, Peter G.; Lewis, Clayton; Rieman, John; Wharton, Cathleen (1992): *Cognitive walkthroughs: a method for theory-based evaluation of user interfaces*. International Journal of Man-Machine Studies 36 (5), S. 741–773.
- Posner, Michael I.; Petersen, Steven E.; Fox, Peter T.; Raichle, Marcus E. (1988): *Localization of cognitive operations in the human brain*. Science 240 (4859), S. 1627–1631.
- Posner, Michael I.; Snyder, Charles R. R.; Davidson, Brian J. (1980): *Attention and the detection of signals*. Journal of Experimental Psychology 109 (2), S. 160–174.
- Preece, Jenny; Rogers, Yvonne; Sharp, Helen (2002): *Interaction Design*. New York, NY: Wiley.
- Pretorius, Marco; van Biljon, Judy (2010): *Learning management systems: ICT skills, usability and learnability*. Interactive Technology and Smart Education 7 (1), S. 30–43.
- Pretschner, Alexander; Gauch, Susan (1999): *Personalization on the Web*. Technical Report ITTC-FY2000-TR-13591-01. Lawrence, KS: University of Kansas.
- Priemer, Burkhard (2004): *Logfile-Analysen: Möglichkeiten und Grenzen ihrer Nutzung bei Untersuchungen zur Mensch-Maschine-Interaktion*. MedienPädagogik 4 (1), S. 1–23.
- Pruitt, John; Aldin, Tamara (2006): *The Persona Lifecycle. Keeping People in Mind Throughout Product Design*. San Francisco, CA: Morgan Kaufmann.
- Quesenbery, Whitney (2003): *The five dimensions of usability*. In: Albers, Michael J.; Mazur, Beth (Hrsg.): Content and complexity: Information design in technical communication. Mahwah, NJ: Lawrence Erlbaum Associates, S. 81–102.
- Reeves, Thomas C.; Benson, Lisa; Elliott, Dean; Grant, Michael; Holschuh, Doug; Kim, Beaumie; Kim, Hyeonjin; Lauber, Erick; Loh, Sebastian (2002): *Usability and Instructional Design Heuristics for E-Learning Evaluation*. In: Barker, Philip; Rebelky, Samuel (Hrsg.): Proceedings of EdMedia: World Conference on Educational Media and Technology. World Conference on Educational Multimedia, Hypermedia and Telecommunications. EdMedia '02. Denver, CO: Association for the Advancement of Computing in Education (AACE), S. 1615–1621.
- Rendón, Eréndira; Abundez, Itzel; Arizmendi, Alejandra; Quiroz, Elvia M. (2011): *Internal versus External cluster validation indexes*. International Journal of computers and communications 5 (1), S. 27–34.
- Rentróia-Bonito, Maria A.; Guerreiro, Tiago; Martins, André; Fernandes, Vitor; Jorge, Joaquim (2011): *Evaluating Learning Support Systems Usability An Empirical Approach*. In: Proceedings of the 3rd E-learning conference: Computer Science Education. Coimbra, Portugal, S. 183–188.
- Roberts, Teresa L.; Moran, Thomas P. (1983): *The Evaluation of Text Editors: Methodology and Empirical Results*. Communications of the ACM 26 (4), S. 265–283.
- Rokach, Lior (2010): *A Survey of Clustering Methods*. In: Maimon, Oded; Rokach, Lior (Hrsg.): The Data Mining and Knowledge Discovery Handbook. Boston, MA: Springer, S. 269–298.

- Rokach, Lior; Maimon, Oded (2008): *Data mining with decision trees: theory and applications*. Singapore/Hackensack, NJ: World Scientific.
- Romero, Cristóbal; Ventura, Sebastián (2007): *Educational data mining: A survey from 1995 to 2005*. *Expert Systems with Applications* 33 (1), S. 135–146.
- Romero, Cristóbal; Ventura, Sebastián (2010): *Educational Data Mining: A Review of the State of the Art*. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40 (6), S. 601–618.
- Rosenbaum, Stephanie; Rohn, Janice Anne; Humburg, Judee (2000): *A toolkit for strategic usability: results from workshops, panels, and surveys*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. New York, NY: ACM, S. 337–344.
- Rousseeuw, Peter J. (1987): *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics* 20, S. 53–65.
- Royston, J. P. (1982a): *Algorithm AS 181: The W Test for Normality*. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31 (2), S. 176–180.
- Royston, J. P. (1982b): *An Extension of Shapiro and Wilk's W Test for Normality to Large Samples*. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31 (2), S. 115–124.
- Rubin, Jeffrey; Chisnell, Dana (2008): *Handbook of Usability Testing. How to Plan, Design and Conduct Effective Tests*. Indianapolis, IN: Wiley.
- Russell, Mark (2005): *Using Eye-Tracking Data to Understand First Impressions of a Website*. *Usability News* 7 (1).
- Russell, Stuart J; Norvig, Peter; Davis, Ernest (2010): *Artificial intelligence: a modern approach*. Upper Saddle River, NJ: Prentice Hall.
- Sarodnick, Florian; Brau, Henning (2011): *Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung*. Bern: Verlag Hans Huber.
- Sauro, Jeff; Kindlund, Erika (2005): *A method to standardize usability metrics into a single score*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. New York, NY: ACM, S. 401–409.
- Scapin, Dominique L.; Bastien, J. M. Christian (1997): *Ergonomic criteria for evaluating the ergonomic quality of interactive systems*. *Behaviour & Information Technology* 16 (4), S. 220–231.
- Scheirer, Jocelyn; Fernandez, Raul; Klein, Jonathan; Picard, Rosalind W. (2002): *Frustrating the user on purpose: a step toward building an affective computer*. *Interacting with Computers* 14 (2), S. 93–118.
- Schildhauer, Thomas (2004): *Business intelligence: durch E-Business Strategien und Prozesse verbessern*. Göttingen: Business Village.
- Schmidts, Hermann (2007): *Usability-Evaluation: Eine Studie zur Identifizierung von Nutzungsproblemen mittels Eye-Tracking-Parametern*. Saarbrücken: VDM-Verlag Müller.
- Schnell, Rainer; Hill, Paul B.; Esser, Elke (2008): *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Schulmeister, Rolf (2009): *Grundlagen hypermedialer Lernsysteme: Theorie - Didaktik - Design*. München: Oldenbourg.
- Schulz, Arne Hendrik; Breiter, Andreas (2012): *Was fördert Kooperation? Betrachtung von Schulinformationssystemen durch Logfileanalysen*. In: Desel, Jörg; Haake, Jörg; Spannagel, Christian (Hrsg.): *DeLFI 2012: Die 10. e-Learning Fachtagung Informatik*. Hagen: Köllen, S. 147–158.
- Schulz, Arne Hendrik; Breiter, Andreas (2013a): *Monitoring User Patterns in School Information Systems Using Logfile Analysis*. In: Passey, Don; Breiter, Andreas; Visscher, Adrie (Hrsg.): *Next generation of Information technology in Educational Management*. 10th IFIP WG 3.7 Conference, ITEM 2012. Bremen, Germany, August 2012. Revised Selected Papers. Heidelberg: Springer, S. 94–103.
- Schulz, Arne Hendrik; Breiter, Andreas (2013b): *Organization Culture Analytics auf Basis von eLearning-Systemen am Beispiel von Schulen*. In: Breiter, Andreas; Rensing, Christoph (Hrsg.): *DeLFI 2013 – Die 11. e-Learning Fachtagung Informatik*. Hagen: Köllen, S. 96–106.
- Schulz, Arne Hendrik; Jeske, Debora (2014): *Learner Differences in the Online Context: Introducing a New Method*. In: Passey, Don; Tatnall, Arthur (Hrsg.): *Key Competencies in ICT and Informatics*.

- Implications and Issues for Educational Professionals and Management. Berlin/Heidelberg: Springer, S. 306–317.
- Schwarz, Norbert (2007): *Retrospective and concurrent self-reports: The rationale for real-time data capture*. In: Stone, Arthur A.; Shiffman, Saul; Atienza, Audie A.; Nebeling, Linda (Hrsg.): *The Science of Real-Time Data Capture. Self-Reports in Health Research*. New York, NY: Oxford University Press, S. 11–26.
- Schwarz, Norbert; Oyserman, Daphna (2001): *Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction*. *American Journal of Evaluation* 22 (2), S. 127–160.
- Scime, Anthony (2005): *Web mining: applications and techniques*. Hershey, PA [u.a.]: Idea Group Publishing.
- Scriven, Michael (1967): *The methodology of evaluation*. In: Tyler, Ralph W.; Gagne, Robert M.; Scriven, Michael (Hrsg.): *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally, S. 39–83.
- Sears, Andrew (1997): *Heuristic Walkthroughs: Finding the Problems Without the Noise*. *International Journal of Human-Computer Interaction* 9 (3), S. 213–234.
- Sears, Andrew; Hess, David J. (1999): *Cognitive walkthroughs: Understanding the effect of task-description detail on evaluator performance*. *International Journal of Human-Computer Interaction* 11 (3), S. 185–200.
- Senol, Leyla; Gecili, Hakan; Onay Durdu, Pinar (2014): *Usability Evaluation of a Moodle based Learning Management System*. In: Viteli, Jarmo; Leikomaa, Marianna (Hrsg.): *Proceedings of EdMedia: World Conference on Educational Media and Technology. EdMedia: World Conference on Educational Media and Technology. EdMedia '14. Tampere, Finland: Association for the Advancement of Computing in Education (AACE)*, S. 850–858.
- Shad, Shafqat A.; Chen, Enhong; Azeem, Faisal M. (2012): *Enterprise Resource Planning - Real blessing or a Blessing in Disguise : An Exploration of the Contextual Factors in Public Sector*. *Interdisciplinary Journal of Contemporary Research in Business* 2 (10), S. 294–311.
- Shahabi, Cyrus; Kashani, Farnoush B. (2002): *A Framework for Efficient and Anonymous Web Usage Mining Based on Client-Side Tracking*. In: *Revised Papers from the 3rd International Workshop on Mining Web Log Data Across All Customers Touch Points. WEBKDD '01. London, UK: Springer*, S. 113–144.
- Shahabi, Cyrus; Zarkesh, Amri M.; Adibi, Jafar; Shah, Vishal (1997): *Knowledge discovery from users Web-page navigation*. In: *Proceedings of the 7th International Workshop on Research Issues in Data Engineering: High Performance Database Management for Large-Scale Applications. RIDE '97. Washington, DC: IEEE*, S. 20–31.
- Shehu, Visar; Besimi, Adrian; Abazi, Lejla; Shaqiri, Mehmetali (2009): *Usability issues while building a new LMS*. In: *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces. ITI '09*, S. 317–322.
- Shneiderman, Ben (1992): *Tree visualization with tree-maps: 2-d space-filling approach*. *ACM Transactions on Graphics* 11 (1), S. 92–99.
- Shneiderman, Ben; Plaisant, Catherine (2010): *Designing the User Interface. Strategies for effective Human-Computer Interaction*. Boston, MA: Addison-Wesley.
- Siegel, Sidney (2001): *Nichtparametrische statistische Methoden*. Magdeburg: Klotz.
- Simon, Herbert A.; Ericsson, K. Anders (1993): *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Singer, David; Fielding, Roy (2014): *Tracking Preference Expression (DNT)*. W3C Working Draft. W3C. URL: <http://www.w3.org/TR/2014/WD-tracking-dnt-20140128/> (aufgerufen am 14.04.2014).
- Smith, John B; Smith, Dana Kay; Kupstas, Eileen (1993): *Automated protocol analysis*. *Human-Computer Interaction* 8 (2), S. 101–145.
- Smith, Sidney L.; Mosier, Jane N. (1986): *Guidelines for designing user interface software*. ESD-TR-86-278. Toronto, Canada: Userlab Inc.
- Sneath, P. H. A. (1957): *The Application of Computers to Taxonomy*. *Journal of General Microbiology* 17 (1), S. 201–226.

- Snyder, Carolyn (2003): *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. Amsterdam, Netherlands [u.a.]: Morgan Kaufmann.
- Sokal, Robert R.; Michener, Charles D. (1958): *A Statistical Method for Evaluating Systematic Relationships*. The University of Kansas Science Bulletin 38 (6), S. 1409–1438.
- Song, Qinbao; Shepperd, Martin (2006): *Mining web browsing patterns for E-commerce*. Computers in Industry 57 (7), S. 622–630.
- Sørensen, T. (1948): *A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons*. Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter 5, S. 1–34.
- Spencer, Rick (2000): *The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company*. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '00. The Hague, Netherlands: ACM, S. 353–359.
- Spiliopoulou, Myra (1999): *Data Mining for the Web*. In: Żytkow, Jan M.; Rauch, Jan (Hrsg.): Principles of Data Mining and Knowledge Discovery. Berlin/Heidelberg: Springer, S. 588–589.
- Spiliopoulou, Myra; Faulstich, Lukas C. (1999): *WUM: A Tool for Web Utilization Analysis*. In: Atzeni, Paolo; Mendelzon, Alberto; Mecca, Giansalvatore (Hrsg.): The World Wide Web and Databases. Berlin/Heidelberg: Springer, S. 184–203.
- Spiliopoulou, Myra; Mobasher, Bamshad; Berendt, Bettina; Nakagawa, Miki (2003): *A Framework for the Evaluation of Session Reconstruction Heuristics in Web-Usage Analysis*. INFORMS Journal on Computing 15 (2), S. 171–190.
- Spool, Jared; Schroeder, Will (2001): *Testing web sites: five users is nowhere near enough*. In: CHI '01 extended abstracts on Human factors in computing systems. CHI EA '01. New York, NY, USA: ACM, S. 285–286.
- Srivastava, Jaideep; Cooley, Robert; Deshpande, Mukund; Tan, Pang-Ning (2000): *Web usage mining: discovery and applications of usage patterns from Web data*. SIGKDD Explorer Newsletter 1 (2), S. 12–23.
- Starner, Thad (1996): *Human-powered wearable computing*. IBM Systems Journal 35 (3), S. 618–629.
- Statistisches Bundesamt (2015): *Ausstattung mit Gebrauchsgütern - Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik*. URL: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/AusstattungGebrauchsguetern/Tabellen/Infotechnik_D.html (aufgerufen am 02.06.2015).
- Stegbauer, Christian (2009): *Wikipedia. Das Rätsel der Kooperation*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stegbauer, Christian (2010): *Netzwerkanalyse und Netzwerktheorie: ein neues Paradigma in den Sozialwissenschaften*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stegbauer, Christian; Hennig, Marina (2012): *Probleme der Integration von Theorie und Methode in der Netzwerkforschung*. In: Häußling, Roger; Stegbauer, Christian (Hrsg.): Die Integration von Theorie und Methode in der Netzwerkforschung. Wiesbaden: Springer, S. 7–11.
- Stern, Robert M. (2001): *Psychophysiological Recording*. New York, NY: Oxford University Press.
- Stoessel, Sabine (2002): *Methoden des Testings im Usability Engineering*. In: Beier, Markus; von Gizycki, Vittoria (Hrsg.): Usability. Nutzerfreundliches Web-Design. Berlin/Heidelberg: Springer, S. 75–96.
- Stone, Debbie; Jarrett, Caroline; Woodroffe, Mark (2005): *User Interface Design and Evaluation*. Amsterdam, Netherlands [u.a.]: Morgan Kaufmann.
- Stowasser, Sascha (2006): *Methodische Grundlagen der softwareergonomischen Evaluationsforschung*. Aachen: Shaker.
- Strybel, Thomas Z. (2011): *Task Analysis Methods and Tools for Developing Web Applications*. In: Vu, Kim-Phuong L.; Proctor, Robert W. (Hrsg.): Handbook of Human Factors in Web Design. Boca Raton, FL: CRC Press, S. 483–507.
- Suneetha, K. R.; Krishnamoorthi, R. (2009): *Identifying User Behavior by Analyzing Web Server Access Log File*. International Journal of Computer Science and Network Security 9 (4), S. 327–332.
- Svensson, Martin; Johansson, Arnold; Ereback, Anna-Lena; Höök, Kristina; Karlgren, Jussi; Bretan, Ivan (1998): *Moving Usability Testing Onto the Web*. In: Maurer, Hermann A.; Olson, Richard G. (Hrsg.): Proceedings of WebNet 98 - World Conference on the WWW and Internet & Intranet,

- Orlando, Florida, USA, November 7-12, 1998. WebNet '98. Charlottesville, VA: Association for the Advancement of Computing in Education (AACE).
- Sydow, Jörg, Hrsg. (2010): *Management von Netzwerkorganisationen: Beiträge aus der Managementforschung*. Wiesbaden: Gabler.
- Tan, Pang-Ning; Kumar, Vipin (2000): *Modeling of web robot navigational patterns*. In: WEBKDD 2000—Web Mining for ECommerce— Challenges and Opportunities, Second International Workshop, Boston, MA.
- Tan, Pang-Ning; Kumar, Vipin (2002): *Discovery of Web Robot Sessions Based on their Navigational Patterns*. Data Mining and Knowledge Discovery 6 (1), S. 9–35.
- Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin (2010): *Introduction to data mining*. Boston, MA [u.a.]: Pearson.
- Tanasa, Doru; Trousse, Brigitte (2004): *Advanced data preprocessing for intersites Web usage mining*. IEEE Intelligent Systems 19 (2), S. 59–65.
- Tauscher, Linda; Greenberg, Saul (1997): *How people revisit web pages: empirical findings and implications for the design of history systems*. International Journal of Human-Computer Studies 47 (1), S. 97–137.
- Te'eni, Dov; Carey, Jane M.; Zhang, Ping (2006): *Human-Computer Interaction: Developing Effective Organizational Information Systems*. Hoboken, NJ: John Wiley & Sons.
- Teufel, Stefanie; Sauter, Christian; Mühlherr, Thomas (1995): *Computerunterstützung für die Gruppenarbeit*. Bonn: Addison-Wesley.
- Theodoridis, Sergios; Koutroubas, Konstantinos (2009): *Pattern recognition*. Amsterdam, Netherlands [u.a.]: Elsevier Science B.V.
- Tolle, John E. (1983): *Transaction Log Analysis Online Catalogs*. In: Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '83. New York, NY: ACM, S. 147–160.
- Tukey, John W. (1977): *Exploratory Data Analysis*. Reading, MA: Pearson.
- Tullis, Thomas S.; Fleischman, Stan; McNulty, Michelle; Cianchette, Carrie; Bergel, Marguerite (2002): *An empirical comparison of lab and remote usability testing of websites*. In: Proceedings of the 11th Annual Conference of the Usability Professionals Association. Orlando, FL, S. 1–5.
- Tullis, Thomas S.; Stetson, Jacqueline N. (2004): *A comparison of questionnaires for assessing website usability*. In: Proceedings of the 11th Annual Conference of the Usability Professionals Association. Minneapolis, MN, S. 1–12.
- Tullis, Thomas; Albert, William (2008): *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies)*. Amsterdam, Netherlands [u.a.]: Morgan Kaufmann.
- Turner, Mark; Kitchenham, Barbara; Brereton, Pearl; Charters, Stuart; Budgen, David (2010): *Does the technology acceptance model predict actual use? A systematic literature review*. Information and Software Technology 52 (5), S. 463–479.
- Tyagi, Navin K.; Solanki, A. K.; Tyagi, Sanjay (2010): *An algorithmic approach to data preprocessing in web usage mining*. International Journal of Information Technology and Knowledge Management 2 (2), S. 279–283.
- Van den Haak, Maaïke; de Jong, Menno; Schellens, Peter Jan (2003): *Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue*. Behaviour & Information Technology 22 (5), S. 339–351.
- Van der Aalst, Wil M. P.; Reijers, Hajo A.; Song, Minseok (2005): *Discovering Social Networks from Event Logs*. Computer Supported Cooperative Work (CSCW) 14 (6), S. 549–593.
- Van der Meer, Debra; Dutta, Kaushik; Datta, Anindya; Ramamritham, Krithi; Navanthe, Shamkant B. (2000): *Enabling scalable online personalization on the Web*. In: Proceedings of the 2nd ACM Conference on Electronic Commerce. EC '00. New York, NY: ACM, S. 185–196.
- Van Eimeren, Birgit; Frees, Beate (2014): *79 Prozent der Deutschen online - Zuwachs bei mobiler Internetnutzung und Bewegtbild. Ergebnisse der ARD/ZDF-Onlinestudie 2014*. Media Perspektiven 8, S. 378–396.

- Van Rijn, Hedderik; Johnson, Addie; Taatgen, Niels (2011): *Cognitive User Modeling*. In: Vu, Kim-Phuong L.; Proctor, Robert W. (Hrsg.): *Handbook of Human Factors in Web Design*. Boca Raton, FL: CRC Press, S. 527–542.
- Van Rijsbergen, C. J. (1979): *Information Retrieval*. Newton, MA: Butterworth-Heinemann.
- Van Weert, Tom J.; Munro, Robert K., Hrsg. (2003): *Informatics and the digital society: social, ethical, and cognitive issues*. International Federation for Information Processing 116. Boston, MA: Kluwer Academic Publishers.
- Vanderdonckt, Jean (1999): *Development milestones towards a tool for working with guidelines*. *Interacting with Computers* 12 (2), S. 81–118.
- Vanderdonckt, Jean; Beirekdar, Abdo; Noirhomme-Fraiture, Monique (2004): *Automated Evaluation of Web Usability and Accessibility by Guideline Review*. In: Koch, Nora; Fraternali, Piero; Wirsing, Martin (Hrsg.): *Web Engineering*. Berlin/Heidelberg: Springer, S. 17–30.
- Vapnik, Vladimir N. (1998): *Statistical learning theory*. New York, NY: Wiley.
- Vapnik, Vladimir N. (2000): *The Nature of Statistical Learning Theory*. New York, NY: Springer.
- Vayssières, Marc P.; Plant, Richard E. (1998): *Identification of vegetation state and transition domains in California's hardwood rangelands*. Davis, CA: University of California. URL: http://frap.fire.ca.gov/publications/state_and_trans2.pdf (aufgerufen am 05.08.2014).
- Venkatesh, Viswanath (2000): *Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model*. *Information Systems Research* 11 (4), S. 342–365.
- Venkatesh, Viswanath; Davis, Fred D. (2000): *A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies*. *Management Science* 46 (2), S. 186–204.
- Virzi, Robert A. (1990): *Streamlining the Design Process: Running Fewer Subjects*. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 34 (4), S. 291–294.
- Virzi, Robert A. (1992): *Refining the test phase of usability evaluation: how many subjects is enough?* *Human Factors* 34 (4), S. 457–468.
- Virzi, Robert A. (1997): *Usability Inspection Methods*. In: Helander, Martin G.; Landauer, Thomas K.; Prabhu, Prasad V. (Hrsg.): *Handbook of Human-Computer Interaction*. Second, completely revised edition. Amsterdam, Netherlands: Elsevier Science B.V., S. 705–715.
- Virzi, Robert A.; Sorce, James F.; Herbert, Leslie Beth (1993): *A Comparison of Three Usability Evaluation Methods: Heuristic, Think-Aloud, and Performance Testing*. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 37 (4), S. 309–313.
- Visscher, Adrie J. (2001): *Computer-Assisted School Information Systems: The Concepts, Intended Benefits, and Stages of Development*. In: Visscher, Adrie J.; Wild, Phil; Fung, Alex C. W. (Hrsg.): *Information Technology in Educational Management*. Dordrecht, Netherlands: Springer, S. 3–18.
- Wang, Chengmin; Sun, Weiqing; Zhang, Tieyan; Zhang, Yan (2009): *Research on Transaction-Item Association Matrix Mining Algorithm in Large-scale Transaction Database*. In: *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Los Alamitos, CA: IEEE, S. 113–117.
- Ward, Joe H. (1963): *Hierarchical Grouping to Optimize an Objective Function*. *Journal of the American Statistical Association* 58 (301), S. 236–244.
- Ward, R. D.; Marsden, P. H. (2003): *Physiological responses to different WEB page designs*. *International Journal of Human-Computer Studies* 59 (1), S. 199–212.
- Wasserman, Stanley; Faust, Katherine (1994): *Social network analysis : methods and applications*. Cambridge, NY: Cambridge University Press.
- Welling, Stefan; Breiter, Andreas; Schulz, Arne Hendrik (2014): *Mediatisierte Organisationswelten in Schulen: Wie der Medienwandel die Kommunikation in den Schulen verändert*. Wiesbaden: Springer VS.
- Wentura, Dirk; Frings, Christian (2013): *Kognitive Psychologie*. Wiesbaden: Springer Fachmedien.
- Wharton, Cathleen; Bradford, Janice; Jeffries, Robin; Franzke, Marita (1992): *Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations*. In: *Proceedings*

- of the SIGCHI Conference on Human Factors in Computing Systems. CHI '92. New York, NY: ACM, S. 381–388.
- Wharton, Cathleen; Rieman, John; Lewis, Clayton; Polson, Peter G. (1994): *The Cognitive Walkthrough: A Partitioner's Guide*. In: Nielsen, Jakob; Mack, Robert L. (Hrsg.): *Usability inspection methods*. New York, NY: Wiley, S. 105–140.
- Wilson, Chauncey E. (2006): *Triangulation: The Explicit Use of Multiple Methods, Measures, and Approaches for Determining Core Issues in Product Development*. *interactions* 13 (6), S. 46–63.
- Wilson, Paul (1991): *Computer supported cooperative work: an introduction*. London, UK: Intellect Books.
- Witten, Ian H.; Frank, Eibe (2005): *Data Mining. Practical Machine Learning Tools and Techniques*. San Francisco, CA: Morgan Kaufmann.
- Wong, Cody; Shiu, Simon; Pal, Sankar (2001): *Mining fuzzy association rules for web access case adaptation*. In: *Workshop on Soft Computing in Case-Based Reasoning, International Conference on Case-Based Reasoning*. ICCBR'01.
- World Wide Web Consortium (W3C) (1995): *Common Logfile Format*. Logging Control In W3C httpd. URL: <http://www.w3.org/Daemon/User/Config/Logging.html#common-logfile-format> (aufgerufen am 26.07.2012).
- Wu, Yi-Hung; Chen, Arbee L. P. (2002): *Prediction of Web Page Accesses by Proxy Server Log*. *World Wide Web* 5 (1), S. 67–88.
- Wu, Ke-Li; Yu, Philip S.; Ballman, A. (1998): *SpeedTracer: a Web usage mining and analysis tool*. *IBM Systems Journal* 37 (1), S. 89–105.
- Xiao, Jitian; Zhang, Yanchun (2001): *Clustering of web users using session-based similarity measures*. In: *Proceedings of the 2001 International Conference on Computer Networks and Mobile Computing*. 2001 International Conference on Computer Networks and Mobile Computing, 2001. ICCNMC '01. Los Alamitos, CA: IEEE, S. 223–228.
- Xiao, Jitian; Zhang, Yanchun; Jia, Xiaohua; Li, Tianzhu (2001): *Measuring similarity of interests for clustering web-users*. In: *Proceedings of the 12th Australasian Database Conference*. ADC '01. Washington, DC: IEEE, S. 107–114.
- Xie, Yunjuan; Phoha, Vir V. (2001): *Web user clustering from access log using belief function*. In: *Proceedings of the 1st International Conference on Knowledge Capture*. K-CAP '01. New York, NY: ACM, S. 202–208.
- Xu, Beijie; Recker, Mimi (2012): *Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data*. *Educational Technology & Society* 15 (3), S. 103–115.
- Yang, Yiming; Liu, Xin (1999): *A Re-examination of Text Categorization Methods*. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '99. New York, NY: ACM, S. 42–49.
- Yau, Nathan (2011): *Visualize this: the FlowingData guide to design, visualization, and statistics*. Indianapolis, IN: Wiley.
- Yuan, Xiaoqin; MacGregor, Mike H.; Harms, Janelle (2002): *An efficient scheme to remove crawler traffic from the Internet*. In: *Proceedings of the 11th International Conference on Computer Communications and Networks*. Piscataway, NJ: IEEE, S. 90–95.
- Yuan, Yufei; Archer, Norm; Connelly, Catherine E.; Zheng, Wuping (2010): *Identifying the ideal fit between mobile work and mobile work support*. *Information & Management* 47 (3), S. 125–137.
- Zaharias, Panagiotis; Poylymenakou, Angeliki (2009): *Developing a Usability Evaluation Method for e-Learning Applications: Beyond Functional Usability*. *International Journal of Human-Computer Interaction* 25 (1), S. 75–98.
- Zhang, Jie; Ghorbani, Ali A. (2004): *The reconstruction of user sessions from a server log using improved time-oriented heuristics*. In: *Proceedings of the 2nd Annual Conference on Communication Networks and Services Research*. Second Annual Conference on Communication Networks and Services Research, 2004. *Proceedings*. CNSR '04. IEEE, S. 315–322.
- Zhang, Tian; Ramakrishnan, Raghu; Livny, Miron (1996): *BIRCH: an efficient data clustering method for very large databases*. In: *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*. SIGMOD '96. New York, NY: ACM, S. 103–114.

Zhang, Zhijun; Basili, Victor; Shneiderman, Ben (1998): *An Empirical Study of Perspective-Based Usability Inspection*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 42 (19), S. 1346–1350.

Zhou, Baoyao; Hui, Siu C.; Fong, Alvis C. M. (2006): *An Effective Approach for Periodic Web Personalization*. In: IEEE/WIC/ACM International Conference on Web Intelligence, 2006. IEEE/WIC/ACM International Conference on Web Intelligence, 2006. WI 2006. WI '06. Los Alamitos, CA: IEEE, S. 284–292.